

Article

A Multivariate Flexible Skew-Symmetric-Normal Distribution: Scale-Shape Mixtures and Parameter Estimation via Selection Representation

Abbas Mahdavi ¹, Vahid Amirzadeh ¹, Ahad Jamalizadeh ¹ and Tsung-I Lin ^{2,3,*} 

¹ Department of Statistics, Faculty of Mathematics & Computer, Shahid Bahonar University of Kerman, Kerman 7616914111, Iran; a.mahdavi@vru.ac.ir (A.M.); v.amirzadeh@uk.ac.ir (V.A.); a.jamalizadeh@uk.ac.ir (A.J.)

² Institute of Statistics, National Chung Hsing University, Taichung 402, Taiwan

³ Department of Public Health, China Medical University, Taichung 404, Taiwan

* Correspondence: tilin@nchu.edu.tw

Abstract: Multivariate skew-symmetric-normal (MSSN) distributions have been recognized as an appealing tool for modeling data with non-normal features such as asymmetry and heavy tails, rendering them suitable for applications in diverse areas. We introduce a richer class of MSSN distributions based on a scale-shape mixture of (multivariate) flexible skew-symmetric normal distributions, called the SSMFSSN distributions. This very general class of SSMFSSN distributions can capture various shapes of multimodality, skewness, and leptokurtic behavior in the data. We investigate some of its probabilistic characterizations and distributional properties which are useful for further methodological developments. An efficient EM-type algorithm designed under the selection mechanism is advocated to compute the maximum likelihood (ML) estimates of parameters. Simulation studies as well as applications to a real dataset are employed to illustrate the usefulness of the presented methods. Numerical results show the superiority of our proposed model in comparison to several existing competitors.

Keywords: ECME algorithm; Hadamard product; scale-shape mixtures; skew-contaminated-normal distribution; skew-symmetric distribution; truncated normal distribution



Citation: Mahdavi, A.; Amirzadeh, V.; Jamalizadeh, A.; Lin, T.I. A Multivariate Flexible Skew-Symmetric-Normal Distribution: Scale-Shape Mixtures and Parameter Estimation via Selection Representation. *Symmetry* **2021**, *13*, 1343. <https://doi.org/10.3390/sym13081343>

Academic Editor: Basil Papadopoulos

Received: 2 July 2021

Accepted: 22 July 2021

Published: 25 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of multivariate normal (MN) distribution plays a central role in statistical modeling. However, there are some situations where the data are not in agreement with the MN distribution. Departure from normality can take place in different ways, such as multimodality, lack in central symmetry, and positive or negative excess of kurtosis. The class of scale mixtures of skew-normal distributions (SMSN) whose general form was first introduced by Branco and Dey [1] includes many multivariate skew symmetric (MSS) distributions with only one mode.

More formally, a scale mixture distribution can be constituted by mixing a base density over a scaling distribution. Its density can be expressed in the form of the integral given by

$$f(\mathbf{y}) = \int_0^\infty g(\mathbf{y}|\kappa(\tau))dH(\tau;\boldsymbol{\nu}), \quad (1)$$

where $g(\mathbf{y}|\kappa(\tau))$ is the conditional density of a $p \times 1$ random vector \mathbf{Y} given $\kappa(\tau)$. Herein, $\kappa(\cdot)$ is a positive function of a scaling variable τ with cumulative distribution function (cdf) $H(\tau;\boldsymbol{\nu})$, indexed by the parameter vector $\boldsymbol{\nu}$.

Using (1), the family of SMSN distributions can be generated by assuming a multivariate skew-normal (MSN) distribution [2] with location $\boldsymbol{\xi}$, scale covariance matrix $\kappa(\tau)\boldsymbol{\Sigma}$,

and shape parameter λ for $g(\mathbf{y}|\kappa(\tau))$. The marginal probability density function (pdf) of \mathbf{Y} can be obtained as follows:

$$f(\mathbf{y}) = 2 \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\xi}, \kappa(\tau)\boldsymbol{\Sigma}) \Phi(\kappa(\tau)^{-1/2} \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\xi})) dH(\tau; \nu), \quad (2)$$

where $\phi_p(\cdot; \boldsymbol{\xi}, \boldsymbol{\Sigma})$ denotes the pdf of the p -variate MN distribution with mean vector $\boldsymbol{\xi}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\Phi(\cdot)$ is the cdf of the standard normal distribution. Some simple scaling functions such as $\kappa(\tau) = \tau$ and $\kappa(\tau) = 1/\tau$ lead to well-known distributions. As described by Branco and Dey [1], some remarkable examples are the multivariate skew- t (MST), multivariate skew-slash (MSSL), and multivariate skew-contaminated-normal (MSCN) distributions, to name just a few. The SMSN family collapses to the class of scale mixture of normal distributions when the skewness parameter vanishes. The MSN distribution [2] with the parameterization given by Arellano-Valle and Genton [3] can be recovered when $H(\tau; \nu)$ is degenerated by imposing $\tau = 1$.

The multivariate skew-scale mixtures of normal (MSSMN) distributions [4] is established when $g(\mathbf{y}|\kappa(\tau))$ defined in (1) is given by

$$g(\mathbf{y}|\kappa(\tau) = \tau^{-1}) = 2\phi_p(\mathbf{y}; \boldsymbol{\xi}, \tau^{-1}\boldsymbol{\Sigma}) \Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\xi})). \quad (3)$$

Therefore, the pdf of MSSMN distribution is obtained as follows:

$$f(\mathbf{y}) = 2\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\xi})) \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\xi}, \tau^{-1}\boldsymbol{\Sigma}) dH(\tau; \nu). \quad (4)$$

Recently, Arellano-Valle et al. [5] proposed a multivariate class of scale-shape mixtures of skew-normal (MSSMSN) distributions which provides alternative candidates for modeling asymmetric data. A convenient hierarchical representation of the MSSMSN distribution is given by

$$\mathbf{Y}|\boldsymbol{\tau} = (\tau_1, \tau_2)^\top \sim SN_p(\boldsymbol{\xi}, \kappa(\tau_1)\boldsymbol{\Sigma}, \eta(\tau_1, \tau_2)\boldsymbol{\lambda}), \quad (5)$$

where $\boldsymbol{\tau} = (\tau_1, \tau_2)^\top$ is a mixing vector with a joint cdf $H(\tau_1, \tau_2; \nu)$, and $\eta(\tau_1, \tau_2) : (\mathbb{R}^+, \mathbb{R}^+) \rightarrow \mathbb{R}$ is a real-valued shape mixing function which is not necessarily symmetric about zero. The family of MSSMSN distributions encapsulates several renowned unimodal asymmetric distributions generated by varying the scale and shape functions, $\kappa(\tau_1)$ and $\eta(\tau_1, \tau_2)$, for a given distribution of $\boldsymbol{\tau}$, or alternatively by fixing $\kappa(\tau_1)$ and $\eta(\tau_1, \tau_2)$ but varying the distribution of $\boldsymbol{\tau}$. A convenient setup for the mixing functions is to choose $\kappa(\tau_1) = 1/\tau_1$ and $\eta(\tau_1, \tau_2) = \sqrt{\tau_2}$, leading to a particular form of the shape mixture of SMSN distributions. If we choose $\eta(\tau_1, \tau_2) = (\tau_1/\tau_2)^{-1/2}$, it produces another form of shape mixtures of MSSMN distributions. See [5] for a more comprehensive discussion and detailed investigation.

As shown by Azzalini and Capitanio [6], the pdf of the MSS distribution can be written as follows:

$$f(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Sigma}) = 2|\boldsymbol{\Sigma}|^{-1/2} f_0(\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\xi})) G_0(w\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\xi})\}), \quad (6)$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$ are, respectively, the location vector and scale covariance matrix; $f_0 : \mathbb{R}^p \rightarrow \mathbb{R}^+$ is a p -variate centrally symmetric pdf with respect to the origin, i.e., $f_0(\mathbf{x}) = f_0(-\mathbf{x})$; $G_0 : \mathbb{R} \rightarrow [0, 1]$ is a univariate cdf satisfying $G_0(-x) = 1 - G_0(x)$; and $w : \mathbb{R}^p \rightarrow \mathbb{R}$ is an odd real-valued function, namely $w\{-\mathbf{x}\} = w\{\mathbf{x}\}$. The MSS class is equivalent to the one studied by Wang et al. [7] for which $G_0(w\{\mathbf{x}\})$ is replaced by $\pi : \mathbb{R}^p \rightarrow [0, 1]$ satisfying $\pi(-\mathbf{x}) = 1 - \pi(\mathbf{x})$.

In light of (6), it is possible to generate a wide range of asymmetric families of unimodal and multimodal skew distributions depending on the specification of the function $w\{\mathbf{x}\}$. This essential property induces greater flexibility in the available shapes. For example, Ma and Genton [8] proposed a flexible class of skew-symmetric distributions by choosing $w\{\mathbf{x}\} = P_K(\mathbf{x})$, where $P_K(\mathbf{x})$ is an odd polynomial function of order K .

The Hadamard product, also known as the Schur product, is a type of matrix multiplication that is commutative and simpler than the matrix product. See [9] for a comprehensive review and its applications to multivariate statistical analysis. The Hadamard product is advantageous in computations and algebraic manipulations because the products are entry-wise, the multiplication is commutative, and, particularly, the inverse is very easy to obtain and the computation of power matrices is straightforward. By convention, we use “ \odot ” to denote the Hadamard operations (product and power). Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $p \times q$ matrices of the same dimension but not necessarily square. Then, the Hadamard product between these two matrices, denoted by $A \odot B$, is the element-wise product of A and B , that is, a $p \times q$ matrix whose (i, j) entry is $a_{ij}b_{ij}$. Accordingly, the n th Hadamard power of matrix A is defined as $A^{\odot n} = [a_{ij}^n]$. More key properties concerning the multiplication and partial derivatives of the Hadamard product are deferred to Appendix A.

In the multivariate setup, the odd polynomial of order K has various combinations of the coefficients. For example, the bivariate case under an odd polynomial of order $K = 3$ is given by $P_K(x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1^3 + \alpha_4 x_2^3 + \alpha_5 x_1^2 x_2 + \alpha_6 x_1 x_2^2$. As an alternative to that of Ma and Genton [8] in more a general setting, we introduce the multivariate flexible skew-symmetric-normal (MFSSN) distribution, denoted by $MFSSN_p(\xi, \Sigma, \alpha)$, which has the following pdf:

$$f_{MFSSN}(\mathbf{y}; \xi, \Sigma, \alpha) = 2\phi_p(\mathbf{y}; \xi, \Sigma)\Phi(\lambda_1^\top \eta_1 + \lambda_2^\top \eta_3 + \cdots + \lambda_m^\top \eta_{2m-1}), \quad (7)$$

where $\alpha = (\lambda_1^\top, \dots, \lambda_m^\top)^\top$ is a $pm \times 1$ multiple-scaled vector of shape parameters and $\eta_{2k-1} = [\Sigma^{-1/2}(\mathbf{y} - \xi)]^{\odot 2k-1}$ remains a $p \times 1$ vector through an odd Hadamard power transformation of order $2K - 1$, for $K = 1, \dots, m$. Notably, the MFSSN distribution encompasses the flexible generalized skew-normal (FGSN) distribution introduced by Ma and Genton [8] as the univariate case. Figure 1 illustrates the scatter-contour plots coupled with their marginal histograms of the bivariate MFSSN distribution under $\xi = \mathbf{0}$, $\Sigma = I_2$ and various specifications of shape parameters arisen from two setups of m . As can be seen, many different non-elliptically distributional shapes with multiple modes and asymmetric patterns can be produced. Additional flexibility can be gained by expanding the thickness of tails. This motivates us to propose a more general class of scale-shape mixtures of MFSSN (SSMFSSN) distributions.

The class of SSMFSSN distributions can be hierarchically represented by

$$\mathbf{Y} \mid \boldsymbol{\tau} = (\tau_1, \tau_2)^\top \sim MFSSN_p(\xi, \tau_1^{-1}\Sigma, \boldsymbol{\vartheta}), \quad (8)$$

where $\boldsymbol{\vartheta} = (\lambda_1^\top \tau_2^{1/2} \tau_1^{-1/2}, \dots, \lambda_m^\top \tau_2^{1/2} \tau_1^{-(2m-1)/2})^\top$. Further, we denote the pdf of $\boldsymbol{\tau} = (\tau_1, \tau_2)^\top$ by $h_\tau(\tau_1, \tau_2; \nu)$. From (8), the marginal pdf of \mathbf{Y} is given by

$$f_{SSMFSSN}(\mathbf{y}; \xi, \Sigma, \zeta, \nu) = \int \phi_p(\mathbf{y}; \xi, \tau_1^{-1}\Sigma)\Phi(\tau_2^{1/2}\{\lambda_1^\top \eta_1 + \lambda_2^\top \eta_3 + \cdots + \lambda_m^\top \eta_{2m-1}\})dH(\boldsymbol{\tau}; \nu). \quad (9)$$

Obviously, the MFSSN model is obtained by setting $\tau_1 = \tau_2 = 1$ in (9).

The family of SSMFSSN distributions introduced in (8) is quite vast, containing several subfamilies of asymmetric and multimodal distributions which have never been considered in the literature. Notice that the MSSMSN distribution described in (5) is a simple case of SSMFSSN by taking $\kappa(\tau_1) = 1/\tau_1$, $\eta(\tau_1, \tau_2) = (\tau_1/\tau_2)^{-1/2}$, and $\lambda_j = \mathbf{0}$, for $j = 2, \dots, m$. More importantly, the scale-shape mixtures of flexible generalized skew-normal (SSMFGSN) distributions proposed very recently by Mahdavi et al. [10] can be thought of as a univariate case of SSMFSSN when the dimension $p = 1$.

The EM algorithm [11] and some of its extraordinary variants such as the expectation conditional maximization (ECM) algorithm [12] and the expectation conditional maximization either (ECME) algorithm [13] are broadly applicable methods to carry out

ML estimation for multivariate skew distributions in a complete-data framework. To the best of our knowledge, previous developments of the EM-type algorithm are based on the convolution-type representations, see, e.g., [14] for the MSN distribution, [15] for the MST distribution, [1] for the SMSN distribution, and [5] for the MSSMSN distribution. Since our proposed SSMFSSN model cannot be explicitly expressed by a convolution-type representation, we develop a novel EM-based procedure designed under the selection mechanism to compute the ML estimates.

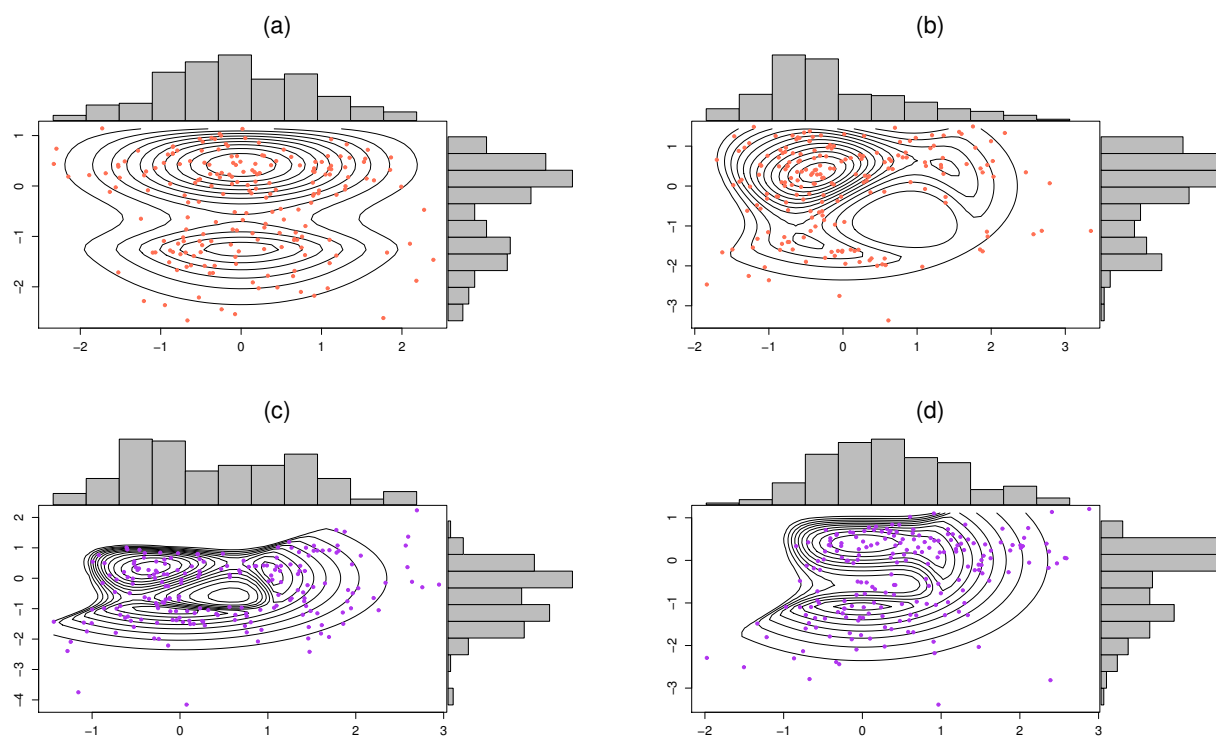


Figure 1. Four scatter-contour plots coupled with their marginal histograms of the bivariate MFSSN distribution: (a) $m = 2$, $\lambda_1 = (0, 2)^\top$ and $\lambda_2 = (0, -2)^\top$ (bimodal); (b) $m = 2$, $\lambda_1 = (-2, 2)^\top$ and $\lambda_2 = (1, -1)^\top$ (trimodal); (c) $m = 3$, $\lambda_1 = (-2, 2)^\top$, $\lambda_2 = (1, -1)^\top$ and $\lambda_3 = (2, -2)^\top$ (trimodal); (d) $m = 3$, $\lambda_1 = (0, 2)^\top$, $\lambda_2 = (0, -2)^\top$ and $\lambda_3 = (2, -1)^\top$ (trimodal).

The rest of the paper is organized as follows. Section 2 presents the formulation of the general SSMFSSN model and discusses how to deploy the ECME algorithm for ML estimation based on the selection-type mechanism. Section 3 exemplifies some particular cases of SSMFSSN distributions constructed by setting different mixing distributions for τ . The proposed techniques are illustrated by conducting two simulation studies in Section 4 and analyzing a real data example in Section 5. We conclude in Section 6 with a few remarks and offer directions for future research.

2. Methodology

2.1. The Family of SSMFSSN Distributions

A p -variate random vector $Y \sim \text{SSMFSSN}_p(\xi, \Sigma, \alpha, \nu)$ is asserted to follow the SSMFSSN distribution with location vector ξ , scale covariance matrix Σ , shape parameters $\alpha = (\lambda_1^\top, \dots, \lambda_m^\top)^\top$, and flatness parameters ν if it has the following selection-type representation:

$$Y \stackrel{d}{=} V \mid (U > 0), \quad (10)$$

where $V = \xi + \Sigma^{1/2} \tau_1^{-1/2} Z_1$, $U = \lambda_1^\top (\tau_1^{-1/2} Z_1) + \dots + \lambda_m^\top (\tau_1^{-(2m-1)/2} Z_1^{\odot 2m-1}) - \tau_2^{-1/2} Z_2$ and $(Z_1^\top, Z_2^\top)^\top \sim N_{p+1}(\mathbf{0}, I_{p+1})$. Herein, ' $\stackrel{d}{=}$ ' stands for equality in distribution, and U is obviously a continuous random variable symmetric about zero. Using this character-

ization, the random samples for $SSMFSSN_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\alpha}, \nu)$ can be simulated through the following scheme

$$\mathbf{X} = \begin{cases} \tau_1^{-1/2} \mathbf{Z}_1, & \text{if } U > 0, \\ -\tau_1^{-1/2} \mathbf{Z}_1, & \text{otherwise.} \end{cases} \quad (11)$$

As a result, the random samples of the general $SSMFSSN_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu)$ can be obtained by the affine transformation $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\Sigma}^{1/2} \mathbf{X}$.

For fitting the SSMFSSN model (10) within the complete-data framework via the EM-type algorithm, we introduce two latent variables $W \stackrel{d}{=} U \mid (U > 0)$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top \stackrel{d}{=} (\tau_1, \tau_2)^\top \mid (U > 0)$. Then, $(\mathbf{Y}^\top, W, \boldsymbol{\gamma}^\top)^\top \stackrel{d}{=} (\mathbf{V}^\top, U, \boldsymbol{\tau}^\top)^\top \mid (U > 0)$ has the following joint pdf:

$$\begin{aligned} f_{\mathbf{Y}, W, \boldsymbol{\gamma}}(\mathbf{y}, W, \boldsymbol{\gamma}) &= \frac{1}{\Pr(U > 0)} f_{\mathbf{V}, U, \boldsymbol{\tau}}(\mathbf{y}, W, \boldsymbol{\gamma}) \\ &= 2f_{\boldsymbol{\tau}}(\boldsymbol{\gamma}) f_{\mathbf{V}|\boldsymbol{\tau}}(\mathbf{y}) f_{U|\mathbf{V}, \boldsymbol{\tau}}(W) \\ &= 2\gamma_2^{1/2} h_{\boldsymbol{\tau}}(\gamma_1, \gamma_2; \nu) \phi_p(\mathbf{y}; \boldsymbol{\xi}, \gamma_1^{-1} \boldsymbol{\Sigma}) \phi(\gamma_2^{1/2} \{W - \zeta\}), \end{aligned} \quad (12)$$

where $\zeta = \boldsymbol{\lambda}_1^\top \boldsymbol{\eta}_1 + \cdots + \boldsymbol{\lambda}_m^\top \boldsymbol{\eta}_{2m-1}$, $h_{\boldsymbol{\tau}}(\gamma_1, \gamma_2; \nu)$ is the pdf of $\boldsymbol{\tau} = (\tau_1, \tau_2)^\top$ evaluated at point $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top$.

Integrating out W from (12) gives the following joint pdf

$$f_{\mathbf{Y}, \boldsymbol{\gamma}}(\mathbf{y}, \gamma_1, \gamma_2) = 2h_{\boldsymbol{\tau}}(\gamma_1, \gamma_2; \nu) \phi_p(\mathbf{y}; \boldsymbol{\xi}, \gamma_1^{-1} \boldsymbol{\Sigma}) \Phi(\gamma_2^{1/2} \zeta). \quad (13)$$

Therefore, the marginal pdf of \mathbf{Y} is given by

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \int_0^\infty \int_0^\infty f_{\mathbf{Y}, \boldsymbol{\gamma}}(\mathbf{y}, \gamma_1, \gamma_2) d\gamma_1 d\gamma_2 \\ &\stackrel{\tau_1 \perp \tau_2}{=} 2 \int \phi_p(\mathbf{y}; \boldsymbol{\xi}, \gamma_1^{-1} \boldsymbol{\Sigma}) dH_{\tau_1}(\gamma_1; \nu_1) \int \Phi(\gamma_2^{1/2} \zeta) dH_{\tau_2}(\tau_2; \nu_2), \end{aligned} \quad (14)$$

where the second equality holds if we further assume τ_1 and τ_2 are mutually independent. It is noteworthy that the shape mixtures of MSSMN distribution established by Arellano-Valle et al. [5] also belongs to the family of our proposed SSMFSSN model by imposing $\lambda_2 = \cdots = \lambda_m = \mathbf{0}$.

Dividing (12) by (13) yields the following relation

$$f(W \mid \mathbf{y}, \gamma_1, \gamma_2) = \frac{\gamma_2^{1/2} \phi(\gamma_2^{1/2} (W - \zeta))}{\Phi(\gamma_2^{1/2} \zeta)} \equiv f(W \mid \mathbf{y}, \gamma_2), \quad (15)$$

for which ' \equiv ' in (15) means that W and γ_1 are conditionally independent. Moreover, it is straightforward to show that

$$W \mid (\mathbf{y}^\top, \gamma_2)^\top \sim TN(\zeta, \gamma_2^{-1}) I_{(0, \infty)}, \quad (16)$$

where $TN(\mu, \sigma^2) I_{\mathbb{A}}$ represents a doubly truncated normal distribution confined within the interval $\mathbb{A} = \{a_1 < x < a_2\}$, and $I_{\mathbb{A}}$ is an indicator function of set \mathbb{A} . Using Lemma 2 of Lin et al. [16], we have the following conditional expectation:

$$E(W \mid \mathbf{y}, \gamma_2) = \zeta + \gamma_2^{-1/2} \frac{\phi(\gamma_2^{1/2} \zeta)}{\Phi(\gamma_2^{1/2} \zeta)}. \quad (17)$$

By Bayes' rule, the conditional pdf of $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top$ given $\mathbf{Y} = \mathbf{y}$ is

$$f(\gamma_1, \gamma_2 \mid \mathbf{y}) = \frac{2h_{\boldsymbol{\tau}}(\gamma_1, \gamma_2; \nu) \phi_p(\mathbf{y}; \boldsymbol{\xi}, \gamma_1^{-1} \boldsymbol{\Sigma}) \Phi(\gamma_2^{1/2} \zeta)}{f_{\mathbf{Y}}(\mathbf{y})}. \quad (18)$$

2.2. Parameter Estimation via the ECME Algorithm

The EM algorithm [11] is a widely used iterative technique to deal with ML estimation in models that involve incomplete data or latent variables. One primary virtue of EM lies in the fact of attractive monotone convergence properties and the preservation of simplicity and stability. In practice, a major limitation of EM is often that some estimators in the M-step cannot be solved in terms of closed-form expressions. To overcome this obstacle, the ECM algorithm proposed by Meng and Rubin [12] recommends replacing the E-step of EM with a sequence of simpler conditional maximization (CM) steps, yet it also enjoys the same convergence properties as EM. However, in certain problems, some of the CM-steps of ECM may become analytically intractable or suffer from slow convergence. As a further flexible extension, the ECME algorithm [13] divides the CM-steps of ECM to maximize either the Q -function, called the CMQ-step, or the corresponding constrained actual log-likelihood function, named as the CML-step. In what follows, we describe in greater detail how the proposed SSMFSSN model can be fitted by using the ECME algorithm.

Suppose that $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ constitutes a set of p -dimensional observed samples of size n arising from the SSMFSSN model. Under the EM framework, the latent variables $\mathbf{W} = (W_1, \dots, W_n)^\top$ and $\boldsymbol{\gamma} = (\gamma_1^\top, \dots, \gamma_n^\top)^\top$ introduced in the preceding subsection are treated as missing data. Then, the complete data are given by $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{W}^\top, \boldsymbol{\gamma}^\top)^\top$. Further, we let $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \text{vech}(\boldsymbol{\Sigma})^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\nu}^\top)^\top$ denote the entire unknown parameters to be estimated in the SSMFSSN model, where $\text{vech}(\cdot)$ is the half-vectorization operator that stacks the lower triangular elements of a $p \times p$ symmetric matrix into a single $p(p+1)/2$ vector.

According to (12), the log-likelihood function of $\boldsymbol{\theta}$ corresponding to the complete-data \mathbf{y}_c , excluding additive constants and terms that do not involve parameters of the model, is given by

$$\begin{aligned} \ell_c(\boldsymbol{\theta} | \mathbf{y}_c) = & -\frac{n}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n \left\{ \gamma_{1i} (\mathbf{y}_i - \boldsymbol{\xi})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\xi}) + \gamma_{2i} (W_i - \zeta_i)^2 \right. \\ & \left. - 2 \ln h(\gamma_{1i}, \gamma_{2i}; \boldsymbol{\nu}) \right\}, \end{aligned} \quad (19)$$

with

$$\begin{aligned} \zeta_i &= \lambda_1^\top \boldsymbol{\eta}_{i,1} + \lambda_2^\top \boldsymbol{\eta}_{i,3} + \dots + \lambda_m^\top \boldsymbol{\eta}_{i,2m-1}, \\ &= \lambda_1^\top [\boldsymbol{\Sigma}^{-1/2} (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 1} + \lambda_2^\top [\boldsymbol{\Sigma}^{-1/2} (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 3} + \dots + \lambda_m^\top [\boldsymbol{\Sigma}^{-1/2} (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 2m-1} \\ &= \mathbf{1}_p^\top \boldsymbol{\Lambda}_1 [\boldsymbol{\Sigma}^{-1/2} (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 1} + \mathbf{1}_p^\top \boldsymbol{\Lambda}_2 [\boldsymbol{\Sigma}^{-1/2} (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 3} + \dots + \mathbf{1}_p^\top \boldsymbol{\Lambda}_m [\boldsymbol{\Sigma}^{-1/2} (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 2m-1} \\ &= \mathbf{1}_p^\top [\boldsymbol{\Delta}_1 (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 1} + \mathbf{1}_p^\top [\boldsymbol{\Delta}_2 (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 3} + \dots + \mathbf{1}_p^\top [\boldsymbol{\Delta}_m (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 2m-1} \\ &= \mathbf{1}_p^\top \sum_{j=1}^m [\boldsymbol{\Delta}_j (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 2j-1}, \end{aligned}$$

where $\mathbf{1}_p$ is a $p \times 1$ vector of ones, $\boldsymbol{\eta}_{i,2j-1} = [\boldsymbol{\Sigma}^{-1/2} (\mathbf{y}_i - \boldsymbol{\xi})]^{\odot 2j-1}$, $\boldsymbol{\Lambda}_j = \text{Diag}\{\lambda_j\}$ is a $p \times p$ diagonal matrix containing the elements of λ_j on the main diagonal, and $\boldsymbol{\Delta}_j = \boldsymbol{\Lambda}_j^{1/(2j-1)} \boldsymbol{\Sigma}^{-1/2}$ is a $p \times p$ reparameterized matrix.

On the k th iteration, the E-step requires the calculation of the so-called Q -function, which is the conditional expectation of (19) given the observed data \mathbf{y} and the current estimate $\hat{\boldsymbol{\theta}}^{(k)}$, where the superscript (k) denote the updated estimates at iteration k . To evaluate the Q -function, we require the following conditional expectations:

$$\hat{s}_{1i}^{(k)} = E(\gamma_{1i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)}), \quad \hat{s}_{2i}^{(k)} = E(\gamma_{2i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)}), \quad \hat{s}_{3i}^{(k)} = E(W_i \gamma_{2i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)}), \quad (20)$$

which have explicit expressions that are discussed in detail in a subsequent section along with

$$\hat{s}_{4i}^{(k)}(\boldsymbol{\nu}) = E(\ln h(\gamma_{1i}, \gamma_{2i}; \boldsymbol{\nu}) | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)}), \quad (21)$$

which may not have standard form for some subfamilies. Substituting (20) and (21) into (19) yields the following Q-function:

$$Q(\theta | \hat{\theta}^{(k)}) = -\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n \left\{ \hat{s}_{1i}^{(k)} (\mathbf{y}_i - \boldsymbol{\zeta})^\top \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\zeta}) + \hat{s}_{2i}^{(k)} \zeta_i^2 - 2\hat{s}_{3i}^{(k)} \zeta_i - 2\hat{s}_{4i}^{(k)}(\nu) \right\}. \quad (22)$$

The CM-steps are implemented to update estimates of θ in the order of $\boldsymbol{\zeta}$, Σ , α and ν by maximizing, one by one, the Q-function obtained in the E-step. After some algebraic manipulations, they are summarized by the following CMQ and CML steps:

CMQ-Step 1: Fixing $\Sigma = \hat{\Sigma}^{(k)}$ and $\alpha = \hat{\alpha}^{(k)}$, we update $\hat{\boldsymbol{\zeta}}^{(k)}$ via Proposition A2 by taking the partial derivative of (22) with respect to $\boldsymbol{\zeta}$. Since the derivation cannot get a closed-form expression for its maximizer, the solution of $\hat{\boldsymbol{\zeta}}^{(k+1)}$ is validated by numerically solving the root of the following equation:

$$\sum_{i=1}^n \left\{ \hat{s}_{1i}^{(k)} \hat{\Sigma}^{(k)-1} (\mathbf{y}_i - \boldsymbol{\zeta}) + \hat{s}_{2i}^{(k)} \zeta_i^{(k)} \hat{\mathbf{a}}_i - \hat{s}_{3i}^{(k)} \hat{\mathbf{a}}_i \right\} = \mathbf{0}, \quad (23)$$

where the two terms $\zeta_i^{(k)} = \mathbf{1}_p^\top \sum_{j=1}^m [\hat{\Delta}_j^{(k)} (\mathbf{y}_i - \boldsymbol{\zeta})]^{\odot 2j-1}$ and $\hat{\mathbf{a}}_i = \sum_{j=1}^m (2j-1) \hat{\Delta}_j^{(k)\top} [\hat{\Delta}_j^{(k)} (\mathbf{y}_i - \boldsymbol{\zeta})]^{\odot 2j-2}$ are nonlinear functions of $\boldsymbol{\zeta}$ with $\hat{\Delta}_j^{(k)} = \text{Diag}\{\hat{\lambda}_j^{(k)}\}^{1/(2j-1)} \hat{\Sigma}^{(k)-1/2}$.

CMQ-Step 2: Fixing $\boldsymbol{\zeta} = \hat{\boldsymbol{\zeta}}^{(k+1)}$ and then updating $\hat{\Sigma}^{(k)}$ by maximizing (22) over Σ gives

$$\hat{\Sigma}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \hat{s}_{1i}^{(k)} (\mathbf{y}_i - \hat{\boldsymbol{\zeta}}^{(k+1)}) (\mathbf{y}_i - \hat{\boldsymbol{\zeta}}^{(k+1)})^\top. \quad (24)$$

CMQ-Step 3: Fixing $\boldsymbol{\zeta} = \hat{\boldsymbol{\zeta}}^{(k+1)}$, we update $\hat{\Delta}_j^{(k)}$ via Proposition A3 by taking the partial derivative of (22) with respect to Δ_j each, $j = 1, \dots, m$. Since their solutions cannot be isolated and set equal to zeros, we have the following equation for finding the nonlinear roots of Δ_j :

$$\sum_{i=1}^n \left(\hat{s}_{3i}^{(k)} - \hat{s}_{2i}^{(k)} \zeta_i^{(k+1)} (\Delta_j) \right) (\Delta_j (\mathbf{y}_i - \hat{\boldsymbol{\zeta}}^{(k+1)}))^{\odot 2j-2} (\mathbf{y}_i - \hat{\boldsymbol{\zeta}}^{(k+1)})^\top = \mathbf{0}, \quad (25)$$

where $\zeta_i^{(k+1)} (\Delta_j) = \mathbf{1}_p^\top \sum_{j=1}^m [\Delta_j (\mathbf{y}_i - \hat{\boldsymbol{\zeta}}^{(k+1)})]^{\odot 2j-1}$ is a nonlinear function of Δ_j . After simplification, we can transform $\hat{\Delta}_j^{(k+1)}$ back to

$$\hat{\lambda}_j^{(k+1)} = (\Delta_j^{(k+1)} \Sigma^{1/2(k+1)})^{\odot 2j-1} \mathbf{1}_p, \quad j = 1, \dots, m. \quad (26)$$

Collecting the above solutions turn out to be $\hat{\alpha}^{(k+1)} = (\hat{\lambda}_1^{(k+1)\top}, \dots, \hat{\lambda}_m^{(k+1)\top})^\top$.

For some members of SSMFSSN, the calculation of $\hat{s}_{4i}^{(k)}(\nu)$ is not straightforward. An update of $\hat{\nu}^{(k)}$ can be achieved by directly maximizing the constrained actual log-likelihood function. This gives rise to the following CML-step:

CML-Step: $\hat{\nu}^{(k)}$ is updated by optimizing the following constrained log-likelihood function:

$$\hat{\nu}^{(k+1)} = \arg \max_{\nu} \sum_{i=1}^n \ln f_{SSMFSSN}(\mathbf{y}_i, \hat{\boldsymbol{\zeta}}^{(k+1)}, \hat{\Sigma}^{(k+1)}, \hat{\boldsymbol{\zeta}}^{(k+1)}, \nu). \quad (27)$$

Note that the maximization in the above CML-step requires p -dimensional search of the objective function (constrained log-likelihood), which can be easily accomplished by using, for example, the `optim` routine in R Development Core Team [17]. The iterations of

the above algorithms are alternately repeated until a suitable convergence rule is satisfied, e.g., the relative difference $|\ell(\hat{\theta}^{(k+1)})/\ell(\hat{\theta}^{(k)}) - 1|$ is sufficiently small, e.g. less than 10^{-6} , which we consider in the numerical experiments, where $\ell(\theta) = \sum_{i=1}^n \ln f_{SSMFSSN}(y_i; \theta)$. To prevent infinite loop from adopting this criterion, the maximum number of iterations is set to 5000.

On the initialization of parameters for starting the algorithm, the location vector $\hat{\xi}^{(0)}$ and the scale covariance matrix $\hat{\Sigma}^{(0)}$ are specified as the sample mean vector and sample covariance matrix, respectively. The initial values for the shape parameters $\hat{\lambda}_1^{(0)}$ are taken as the sample skewness of p variables, while the remaining $\hat{\lambda}_2^{(0)}, \dots, \hat{\lambda}_m^{(0)}$ are fixed around 0. As for $\hat{\nu}^{(0)} = (\hat{\nu}_1^{(0)}, \hat{\nu}_2^{(0)})^\top$, their initial values are chosen as relatively small values depending on the settings of parameter domain. To avoid getting stuck in one of the many local maxima of the likelihood function, a convenient method is to try a variety different of initial values with perturbations or using the bootstrap resampling method [14]. The solution with the highest log-likelihood value is treated as the ML estimates, denoted by $\hat{\theta} = (\hat{\xi}^\top, \text{vech}(\hat{\Sigma})^\top, \hat{\alpha}^\top, \hat{\nu}^\top)^\top$.

3. Examples of SSMFSSN Distributions

We present some special cases of SSMFSSN distributions which are induced by setting different mixing distributions for τ . For each case, additional conditional expectations are also derived for the implementation of ECME.

3.1. The Multivariate Flexible Skew-Symmetric- t -Normal Distribution

The multivariate flexible skew-symmetric- t -normal (MFSSTN) distribution, denoted by $Y \sim \text{MFSSTN}_p(\xi, \Sigma, \alpha, \nu)$, is produced by taking $\tau_1 \sim \Gamma(\nu/2, \nu/2)$ and $\tau_2 = 1$ in (10). In this case, $h_\tau(\gamma_1, \gamma_2; \nu) = g(\gamma_1; \nu/2, \nu/2)$, where $g(\cdot; \alpha, \beta)$ represents the pdf of the gamma distribution with mean α/β . The pdf of the MFSSTN distribution can be expressed as

$$f_{\text{MFSSTN}}(y; \xi, \Sigma, \alpha, \nu) = 2t_p(y; \xi, \Sigma, \nu)\Phi(\zeta), \quad (28)$$

where $t_p(\cdot; \xi, \Sigma, \nu)$ stands for the pdf of p -dimensional multivariate t distribution with location vector ξ , scale covariance matrix Σ , and degree of freedom (DOF) equal to ν . One special case of the MFSSTN distribution is the multivariate skew- t -normal (MSTN) distribution of Lin et al. [18], obtained by letting $\lambda_j = 0$ for $j = 2, \dots, m$. In addition, the MFSSTN distribution reduces to MFSSN as $\nu \rightarrow \infty$.

According to (18), it is easy to observe that $\gamma_1 | Y = y \sim \Gamma((\nu + p)/2, (\nu + \delta^2)/2)$, where $\delta^2 = (y - \xi)^\top \Sigma^{-1}(y - \xi)$ denotes the squared Mahalanobis distance. Therefore,

$$E(\gamma_1 | Y = y) = \frac{\nu + p}{\nu + \delta^2} \text{ and } E(\ln \gamma_1 | Y = y) = \psi\left(\frac{\nu + p}{2}\right) - \ln\left(\frac{\nu + \delta^2}{2}\right), \quad (29)$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function.

From (20) and (21), the E-step involves the calculation of $\hat{s}_{1i}^{(k)} = E(\gamma_{1i} | y_i, \hat{\theta}^{(k)})$, $\hat{s}_{2i}^{(k)} = 1$, $\hat{s}_{3i}^{(k)} = E(W_i | y_i, \hat{\theta}^{(k)})$ and $\hat{s}_{4i}^{(k)} = E(\ln \gamma_{1i} | y_i, \hat{\theta}^{(k)})$, which can be easily evaluated via the results of (17) and (29). As an alternative way of estimating ν , the CML-Step for the MFSSTN distribution can be altered to the following CMQ-Step.

CMQ-Step 4: $\hat{\nu}^{(k+1)}$ is obtained by solving the root of the following equation:

$$1 + \ln\left(\frac{\nu}{2}\right) - \psi\left(\frac{\nu}{2}\right) + \frac{1}{n} \sum_{i=1}^n (\hat{s}_{4i}^{(k)} - \hat{s}_{1i}^{(k)}) = 0. \quad (30)$$

3.2. The Multivariate Flexible Skew-Symmetric-Slash-Normal Distribution

Let Y be a p -dimensional random vector with the following representation

$$Y = \xi + \tau^{-1/2} \Sigma^{1/2} Z, \quad (31)$$

where $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ and is independent of $\tau \stackrel{d}{=} U^{1/\nu} \sim \text{Beta}(\nu, 1)$, where U is the uniform distribution on the interval $(0, 1)$. From (31), the conditional distribution \mathbf{Y} given τ is $N_p(\xi, \tau^{-1}\Sigma)$. Then, \mathbf{Y} has a multivariate slash distribution with pdf given by

$$f_{SL}(\mathbf{y}; \xi, \Sigma, \nu) = \nu \int_0^1 \tau^{\nu-1} \phi_p(\mathbf{y}; \xi, \tau^{-1}\Sigma) d\tau = \begin{cases} \frac{2^\nu \nu |\Sigma|^{-1/2} \Gamma(\nu+p/2) G(\delta^2/2; \nu+p/2)}{\pi^{p/2} \delta^{2\nu+p}}, & \mathbf{y} \neq \xi \\ \frac{\nu}{(\nu+p/2)(2\pi)^{p/2}} |\Sigma|^{-1/2}, & \mathbf{y} = \xi \end{cases} \quad (32)$$

where $G(\cdot; r)$ denotes the cdf of the gamma distribution with scale parameter 1 and shape parameter r . Using the law of iterated expectations, the mean vector and variance-covariance matrix of \mathbf{Y} are

$$E(\mathbf{Y}) = \xi \text{ and } \text{cov}(\mathbf{Y}) = \frac{\nu}{\nu-1} \Sigma. \quad (33)$$

If we select $\tau_1 \sim \text{Beta}(\nu, 1)$ and $\tau_2 = 1$ for (10), this generates the multivariate flexible skew-symmetric-slash-normal (MFSSSLN) distribution, denoted by $\mathbf{Y} \sim \text{MFSSSLN}_p(\xi, \Sigma, \alpha, \nu)$, with the pdf taking the form of

$$f_{\text{MFSSSLN}}(\mathbf{y}; \xi, \Sigma, \alpha, \nu) = 2f_{SL}(\mathbf{y}; \xi, \Sigma, \nu) \Phi(\zeta). \quad (34)$$

Note that the MFSSSLN distribution contains the MFSSN distribution as a limiting case for $\nu \rightarrow \infty$ and encloses the multivariate skew-slash distribution considered by Wang and Genton [19] as a reduced case when $\lambda_2 = \dots = \lambda_m = \mathbf{0}$.

According to (18), the conditional distribution $\gamma_1 \mid \mathbf{y}$ is given by

$$f(\gamma_1 \mid \mathbf{y}) = \begin{cases} \frac{|\delta|^{2\nu+p} \gamma_1^{\nu+p/2-1} \exp(-\gamma_1 \delta^2/2)}{2^{\nu+p/2} \Gamma(\nu+p/2) G(\delta^2/2; \nu+p/2)}, & \mathbf{y} \neq \xi \\ (\nu + \frac{p}{2}) \gamma_1^{\nu+p/2-1}, & \mathbf{y} = \xi. \end{cases} \quad (35)$$

In addition, some algebraic manipulations yield

$$E(\gamma_1 \mid \mathbf{Y} = \mathbf{y}) = \begin{cases} \left(\frac{2\nu+p}{\delta^2}\right) \frac{G(\delta^2/2; \nu+p/2+1)}{G(\delta^2/2; \nu+p/2)}, & \mathbf{y} \neq \xi \\ \frac{2\nu+p}{2\nu+p+2}, & \mathbf{y} = \xi. \end{cases} \quad (36)$$

and

$$E(\ln \gamma_1 \mid \mathbf{Y} = \mathbf{y}) = \begin{cases} \ln\left(\frac{2}{\delta^2}\right) + \frac{\int_0^{\delta^2/2} \ln(x) x^{\nu+p/2-1} e^{-x} dx}{\Gamma(\nu+p/2) G(\delta^2/2; \nu+p/2)}, & \mathbf{y} \neq \xi \\ \frac{-2}{2\nu+p}, & \mathbf{y} = \xi. \end{cases} \quad (37)$$

To implement the ECME procedure for fitting MFSSSLN, the conditional expectations involved in (20) and (21) can be easily evaluated via the results of (17), (36), and (37). Besides, the DOF $\hat{\nu}^{(k)}$ can be alternatively updated by maximizing the Q -function over ν , leading to the following CMQ-Step:

CMQ-Step 4:

$$\hat{\nu}^{(k+1)} = -\frac{n}{\sum_{i=1}^n \hat{s}_{4i}^{(k)}}. \quad (38)$$

3.3. The Multivariate Flexible Skew-Symmetric-Contaminated-Normal Distribution

The multivariate flexible skew-symmetric-contaminated-normal (MFSSCN) distribution, denoted by $Y \sim MFSSCN_p(\xi, \Sigma, \alpha, \nu_1, \nu_2)$, arises when $\tau_2 = 1$, while τ_1 has a discrete distribution taking value $\nu_2 \in (0, 1)$ with probability ν_1 and value 1 with probability $1 - \nu_1$. More precisely,

$$h(\tau_1, \nu) = \nu_1 \mathbb{I}_{(\tau_1=\nu_2)} + (1 - \nu_1) \mathbb{I}_{(\tau_1=1)}, 0 < \nu_1 < 1 \text{ and } 0 < \nu_2 < 1, \quad (39)$$

where ν_1 is the proportion of outliers and ν_2 is an inflation parameter denoting the degree of contamination.

Using (14), the pdf of Y is given by

$$f_{MFSSCN}(y; \xi, \Sigma, \nu_1, \nu_2) = 2 \left\{ \nu_1 \phi_p(y; \xi, \nu_2^{-1} \Sigma) + (1 - \nu_1) \phi_p(y; \xi, \Sigma) \right\} \Phi(\zeta). \quad (40)$$

Obviously, the MFSSCN distribution reduces to the MFSSN distribution when $\nu_2 = 1$, to the multivariate skew-contaminated-normal distribution [4] when $\lambda_2 = \dots = \lambda_m = \mathbf{0}$, and is said to follow the "MFSSCNe" distribution if ν_1 and ν_2 are restricted to be equal, namely $\nu_1 = \nu_2 = \nu$.

To obtain $\hat{s}_{4i}^{(k)}$, we require the following conditional expectation

$$E(\gamma_1 | Y = y) = \frac{1 - \nu_1 + \nu_1 \nu_2^{1+p/2} \exp\{(1 - \nu_2)\delta^2/2\}}{1 - \nu_1 + \nu_1 \nu_2^{p/2} \exp\{(1 - \nu_2)\delta^2/2\}}. \quad (41)$$

The resulting Q -function can be easily evaluated through (17) and (41) since $\hat{s}_{2i}^{(k)} = 1$. To estimate ν_1 and ν_2 , we perform the CML-Step, so the calculation of $\hat{s}_{4i}^{(k)}$ can be omitted.

3.4. The Multivariate Flexible Skew-Symmetric- t Distribution

The multivariate flexible skew-symmetric- t (MFSST) distribution, denoted by $Y \sim MFSST_p(\xi, \Sigma, \alpha, \nu)$, is created by setting $\tau_1 = \tau_2 = \tau$ with $\tau \sim \Gamma(\nu/2, \nu/2)$. Utilizing (8), the hierarchical representation for Y can be simplified as

$$Y | \tau \sim MFSSN_p(\xi, \tau_1^{-1} \Sigma, \vartheta), \quad (42)$$

where $\vartheta = (\lambda_1^\top, \tau^{-1} \lambda_2^\top, \dots, \tau^{-m+1} \lambda_m^\top)^\top$. Therefore, it can be verified that the MFSST distribution has the following pdf

$$f_{MFSST}(y; \xi, \Sigma, \nu) = 2t_p(y; \xi, \Sigma, \nu) T\left(\zeta \sqrt{\frac{\nu+p}{\nu+\delta^2}}; \nu+p\right), \quad (43)$$

where $T(\cdot, \nu)$ denotes the cdf of the t distribution with DOF ν . The detailed proof of (43) is sketched in Appendix B. It is notable that the MFSST distribution includes the multivariate skew- t distribution of Azzalini and Capitanio [6] as a particular member by letting $\lambda_2 = \dots = \lambda_m = \mathbf{0}$ and the MFSSN distribution as a limiting case when ν grows to infinity.

Using (18) subject to $\gamma_1 = \gamma_2 = \gamma$, it suffices to show that

$$f(\gamma | y) = \frac{1}{T(M; \nu+p)} g\left(\gamma; \frac{\nu+p}{2}, \frac{\nu+\delta^2}{2}\right) \Phi(\gamma^{1/2} \zeta), \quad (44)$$

where $M = \zeta \sqrt{\frac{\nu+p}{\delta^2+\nu}}$.

With arguments similar to those of Lin et al. [20], it is straightforward to derive

$$E(\gamma | Y = y) = \left(\frac{\nu + p}{\delta^2 + \nu} \right) \frac{T(M\sqrt{\frac{\nu+p+2}{\nu+p}}; \nu + p + 2)}{T(M; \nu + p)} \quad (45)$$

and

$$\begin{aligned} E(\ln \gamma | Y = y) &= \psi\left(\frac{\nu + p}{2}\right) - \ln\left(\frac{\delta^2 + \nu}{2}\right) \\ &+ \frac{\nu + p}{\delta^2 + \nu} \left\{ \frac{T(M\sqrt{\frac{\nu+p+2}{\nu+p}}; \nu + p + 2)}{T(M; \nu + p)} - 1 \right\} \\ &+ \frac{\zeta(\delta^2 - 1)}{\sqrt{(\nu + p)(\nu + \delta^2)^3}} \frac{t(M; \nu + p)}{T(M; \nu + p)} \\ &+ \frac{1}{T(M; \nu + p)} \int_{-\infty}^M \kappa_\nu(x) t(x; \nu + p) dx, \end{aligned} \quad (46)$$

where

$$\kappa_\nu(x) = \psi\left(\frac{\nu + p + 1}{2}\right) - \psi\left(\frac{\nu + p}{2}\right) - \ln\left(1 + \frac{x^2}{\nu + p}\right) + \frac{(\nu + p)x^2 - \nu - p}{(\nu + p)(\nu + p + x^2)}.$$

Additionally, using the law of iterated expectations, we can deduce that

$$\begin{aligned} E(W\gamma | Y = y) &= \frac{1}{T(M; \nu + p)} \left\{ M\sqrt{\frac{\nu + p}{\delta^2 + \nu}} T\left(M\sqrt{\frac{\nu + p + 2}{\nu + p}}; \nu + p + 2\right) \right. \\ &\left. + (\delta^2 + \nu)^{-1/2} \frac{\Gamma((\nu + p + 1)/2)}{\sqrt{\pi}\Gamma((\nu + p)/2)} \left(1 + \frac{\zeta^2}{\delta^2 + \nu}\right)^{-\frac{\nu + p + 1}{2}} \right\}. \end{aligned} \quad (47)$$

As a consequence, the E-step in (20) and (21) requires the calculation of $\hat{s}_{1i}^{(k)} = \hat{s}_{2i}^{(k)} = E(\gamma_i | y_i, \theta^{(k)})$, $\hat{s}_{3i}^{(k)} = E(W_i \gamma_i | y_i, \hat{\theta}^{(k)})$, and $\hat{s}_{4i}^{(k)} = E(\ln \gamma_i | y_i, \hat{\theta}^{(k)})$, which can be directly evaluated via (45)–(47). Moreover, the procedure of updating $\hat{\nu}^{(k)}$ is the same as (30).

3.5. The Multivariate Flexible Skew-Symmetric- t - t Distribution

Consider two independent random variables $\tau_1 \sim \Gamma(\nu_1/2, \nu_1/2)$ and $\tau_2 \sim \Gamma(\nu_2/2, \nu_2/2)$ with joint pdf given by

$$h_\tau(\gamma_1, \gamma_2; \nu_1, \nu_2) = g(\gamma_1; \nu_1/2, \nu_1/2)g(\gamma_2; \nu_2/2, \nu_2/2). \quad (48)$$

Using (14), we thus generate the multivariate flexible skew-symmetric- t - t (MFSSTT) distribution, denoted by $Y \sim \text{MFSSTT}_p(\xi, \Sigma, \alpha, \nu_1, \nu_2)$, whose pdf is of the form

$$f_{\text{MFSSTT}}(y; \xi, \Sigma, \alpha, \nu) = 2t_p(y; \xi, \Sigma, \nu_1)T(\zeta; \nu_2). \quad (49)$$

When the two DOFs are restricted to be equal, namely $\nu_1 = \nu_2 = \nu$, Y is said to follow the ‘MFSSTTe’ distribution. One thing worth noting is that the MFSSTT distribution reduces to the MFSSN distribution when $(\nu_1, \nu_2) \rightarrow (\infty, \infty)$, to the MFSSTN distribution by letting $\nu_1 = \nu$ and $\nu_2 = \infty$, and embeds the multivariate skew- t - t (MSTT) distribution introduced by Wang et al. [21] as a special case under the setting of $\lambda_2 = \dots = \lambda_m = \mathbf{0}$.

According to (18), it is easy to see $\gamma_1 | Y = y \sim \Gamma((\nu_1 + p)/2, (\nu_1 + \delta^2)/2)$. The conditional pdf of γ_2 given $Y = y$ is

$$f(\gamma_2 | y) = \frac{g(\gamma_2; \nu_2/2, \nu_2/2)\Phi(\gamma_2^{1/2}\zeta)}{T(\zeta; \nu_2)}. \quad (50)$$

Further, we have the following conditional expectations:

$$E(\gamma_2 | \mathbf{Y} = \mathbf{y}) = \frac{T(\zeta \sqrt{\frac{\nu_2+2}{\nu_2}}; \nu_2 + 2)}{T(\zeta; \nu_2)} \quad (51)$$

and

$$\begin{aligned} E(\ln \gamma_2 | \mathbf{Y} = \mathbf{y}) &= \psi\left(\frac{\nu_2+1}{2}\right) - \ln\left(\frac{\nu_2}{2}\right) + \frac{T(\zeta \sqrt{\frac{\nu_2+2}{\nu_2}}; \nu_2 + 2)}{T(\zeta; \nu_2)} - 1 \\ &\quad - \left(\frac{\zeta}{\nu_2}\right) \frac{t(\zeta; \nu_2)}{T(\zeta; \nu_2)} - \frac{1}{T(\zeta; \nu_2)} \int_{-\infty}^{\zeta} \ln\left(1 + \frac{x^2}{\nu_2}\right) t(x; \nu_2) dx. \end{aligned} \quad (52)$$

Applying the law of iterated expectations to (15) and (50) gives

$$E(W\gamma_2 | \mathbf{Y} = \mathbf{y}) = \frac{1}{T(\zeta; \nu_2)} \left\{ \zeta T\left(\zeta \sqrt{\frac{\nu_2+2}{\nu_2}}; \nu_2 + 2\right) + t(\zeta; \nu_2) \right\}. \quad (53)$$

With slight modifications as defined in (20) and (21), the necessary conditional expectations in the E-step include $\hat{s}_{1i}^{(k)} = E(\gamma_{1i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)})$, $\hat{s}_{2i}^{(k)} = E(\gamma_{2i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)})$, $\hat{s}_{3i}^{(k)} = E(W_i \gamma_{2i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)})$, $\hat{s}_{4i}^{(k)} = E(\ln \gamma_{1i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)})$, and $\hat{s}_{5i}^{(k)} = E(\ln \gamma_{2i} | \mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(k)})$, which can be easily evaluated via the results given in (29) and (51)–(53), respectively. To numerically estimate ν_1 and ν_2 for the MFSSTT distribution, we resort to the following two root-finding equations:

CMQ-Step 4: $\hat{\nu}_1^{(k+1)}$ and $\hat{\nu}_2^{(k+1)}$ are obtained by solving the roots of the following two equations:

$$1 + \ln\left(\frac{\nu_1}{2}\right) - \psi\left(\frac{\nu_1}{2}\right) + \frac{1}{n} \sum_{i=1}^n \left(\hat{s}_{4i}^{(k)} - \hat{s}_{1i}^{(k)}\right) = 0 \quad (54)$$

and

$$1 + \ln\left(\frac{\nu_2}{2}\right) - \psi\left(\frac{\nu_2}{2}\right) + \frac{1}{n} \sum_{i=1}^n \left(\hat{s}_{5i}^{(k)} - \hat{s}_{2i}^{(k)}\right) = 0. \quad (55)$$

4. Simulation Studies

4.1. Recovery of the True Underlying Parameters

The first experiment intends to investigate the ability of the proposed ECME algorithm to recover the true underlying parameters. Monte Carlo samples of different sample sizes $n = 100, 250, 500$, and 1000 were generated from the MFSSN distributions specified in (7) and five examples of SSMFSSN distributions studied in Section 3. For ease of exposition, we considered the Hadamard power transformation of order three ($K = 3$) that allows two shape parameters in the skewing function $\Phi(\cdot)$. Moreover, the flatness parameters for MFSSCN and MFSSTT were assumed to be equal, say $\nu_1 = \nu_2 = \nu$, referred to as the MFSSCNe and MFSSTTe distributions. The presumed true parameters were $\boldsymbol{\xi} = (1, -1)^\top$, $\boldsymbol{\sigma} = \text{vech}(\boldsymbol{\Sigma}) = (1, 0.5, 4)^\top$, $\boldsymbol{\lambda}_1 = (-2, 2)^\top$, and $\boldsymbol{\lambda}_2 = (1, -1)^\top$. Furthermore, $\nu = 3$ was taken for the MFSSTN, MFSSSLN, MFSST, and MFSSTTe distributions, while $\nu = 0.7$ was adopted for the MFSSCNe distribution since its support lies within the interval $(0, 1)$. As an illustration, Figure 2 displays the scatter plots superimposed on the fitted contours for each type of data simulated from one trail.

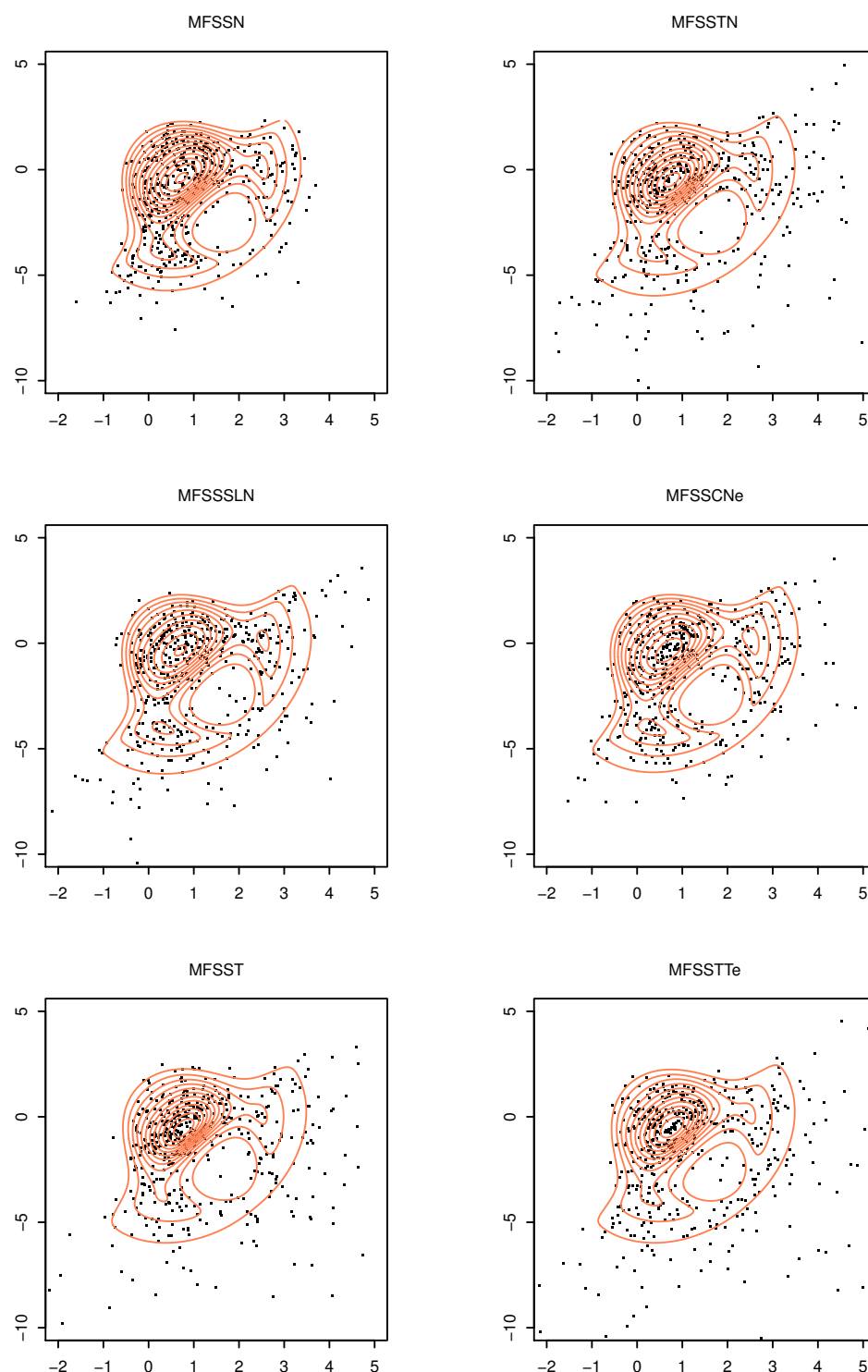


Figure 2. Scatter-contour plots of one simulation case with 500 random samples generated from six subfamilies of the proposed model.

For all scenarios, the accuracies of the parameter estimates are assessed by computing the mean absolute bias (MAB) and the root mean square error (RMSE) over $R = 100$ replications. For a vector of parameters $\theta = (\theta_1, \dots, \theta_p)^\top$, these measures are, respectively, defined as

$$\text{MAB} = \frac{1}{pR} \sum_{k=1}^p \sum_{r=1}^R |\hat{\theta}_{kr} - \theta_k^A| \text{ and } \text{RMSE} = \sqrt{\frac{1}{pR} \sum_{k=1}^p \sum_{r=1}^R (\hat{\theta}_{kr} - \theta_k^A)^2}, \quad (56)$$

where $\hat{\theta}_{kr}$ denotes the ML estimate of the k th parameter at the r th replication and θ_k^A represents the actual value of θ_k .

The experimental results are summarized in Table 1. It is readily seen both MAB and RMSE values tend to approach zero with increasing the sample size. While this study is limited to the simplest case ($p = 2; m = 2$), our developed ECME algorithm shows favorable ability to recover the true parameter values with data generated exactly according to model assumptions. Similar experiments have also been undertaken on more complex scenarios ($p = 3; m = 3$). The extensive results would not necessarily be excessively reported since the conclusions are in accordance with those already presented.

Table 1. Simulation results based on 100 replications with different sample sizes.

Model	Parameter	$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
		MAB	RMSE	MAB	RMSE	MAB	RMSE	MAB	RMSE
MFSSN	ξ	0.090	0.122	0.057	0.074	0.037	0.048	0.027	0.033
	σ	0.257	0.382	0.143	0.209	0.115	0.168	0.081	0.123
	λ_1	0.476	0.625	0.203	0.270	0.111	0.155	0.079	0.112
	λ_2	0.339	0.489	0.160	0.209	0.104	0.131	0.064	0.082
MFSSTN	ξ	0.095	0.127	0.062	0.086	0.040	0.053	0.028	0.036
	σ	0.419	0.673	0.244	0.383	0.158	0.241	0.121	0.179
	λ_1	0.382	0.469	0.349	0.426	0.317	0.382	0.306	0.355
	λ_2	0.335	0.439	0.194	0.244	0.177	0.216	0.151	0.181
	ν	2.495	10.745	0.458	0.651	0.319	0.413	0.203	0.258
MFSSSLN	ξ	0.091	0.121	0.052	0.069	0.036	0.047	0.025	0.035
	σ	0.424	0.641	0.306	0.489	0.190	0.314	0.127	0.200
	λ_1	0.458	0.588	0.267	0.347	0.216	0.269	0.174	0.206
	λ_2	0.470	0.614	0.259	0.378	0.174	0.227	0.114	0.138
	ν	6.530	11.849	3.425	8.206	1.191	3.251	0.481	0.721
MFSSCNe	ξ	0.087	0.117	0.048	0.064	0.036	0.047	0.026	0.034
	σ	0.363	0.546	0.285	0.440	0.223	0.339	0.133	0.202
	λ_1	0.446	0.595	0.288	0.353	0.202	0.247	0.138	0.174
	λ_2	0.426	0.586	0.265	0.347	0.178	0.219	0.131	0.160
	ν	0.199	0.216	0.162	0.176	0.116	0.137	0.083	0.098
MFSST	ξ	0.114	0.152	0.069	0.091	0.042	0.055	0.032	0.040
	σ	0.393	0.586	0.250	0.386	0.160	0.238	0.132	0.216
	λ_1	0.453	0.559	0.265	0.329	0.207	0.253	0.194	0.228
	λ_2	0.355	0.468	0.215	0.273	0.137	0.175	0.110	0.139
	ν	1.152	2.210	0.430	0.621	0.266	0.366	0.207	0.265
MFSSTTe	ξ	0.114	0.146	0.070	0.094	0.047	0.061	0.029	0.038
	σ	0.400	0.654	0.213	0.323	0.165	0.255	0.115	0.188
	λ_1	0.437	0.546	0.411	0.476	0.406	0.453	0.414	0.438
	λ_2	0.313	0.409	0.216	0.260	0.183	0.217	0.167	0.193
	ν	1.332	2.733	0.424	0.573	0.300	0.408	0.210	0.279

4.2. Comparing the Proposed Procedure with Convolution-Type EM Algorithms

The second experiment aims to compare the performance of the proposed selection-type ECME procedure outlined in Section 2.2 with the traditional EM-based algorithms derived based on convolution-type representations. As an illustration, we consider the fitting of MSN, MST, and MSCN distributions arisen from the multivariate skew-normal independent (SNI) family studied by Cabral et al. [22]. As discussed above, they are special cases of our proposed MFSSN, MFSST, and MFSSCN distributions by setting $\lambda_2 = \dots = \lambda_m = \mathbf{0}$. Accordingly, the CMQ-Step 1 in Section 2 can be simplified as follows:

CMQ-Step 1: Fixing $\Sigma = \hat{\Sigma}^{(k)}$ and $\lambda_1 = \hat{\lambda}_1^{(k)}$, we obtain $\hat{\xi}^{(k+1)}$ by

$$\begin{aligned} \hat{\xi}^{(k+1)} = & \left[I_p \sum_{i=1}^n \hat{s}_{1i}^{(k)} + \hat{\Sigma}^{(k)} \hat{\Delta}_1^{\top(k)} \mathbf{1}_p \mathbf{1}_p^{\top} \hat{\Delta}_1^{(k)} \sum_{i=1}^n \hat{s}_{2i}^{(k)} \right]^{-1} \left\{ \sum_{i=1}^n \hat{s}_{1i}^{(k)} \mathbf{y}_i \right. \\ & \left. + \hat{\Sigma}^{(k)} \hat{\Delta}_1^{\top(k)} \mathbf{1}_p \mathbf{1}_p^{\top} \hat{\Delta}_1^{(k)} \sum_{i=1}^n \hat{s}_{2i}^{(k)} \mathbf{y}_i - \hat{\Sigma}^{(k)} \hat{\Delta}_1^{\top(k)} \mathbf{1}_p \sum_{i=1}^n \hat{s}_{3i}^{(k)} \right\}. \end{aligned} \quad (57)$$

In total, 100 Monte Carlo (MC) samples of sizes $n = 100, 500$, and 1000 were generated from each of the three distributions. The true parameters were the same as those given in the previous experiment except for $\lambda_2 = 0$. Each simulated sample was fitted twice with the proposed selection-type ECME procedure and the EM-type algorithm based on convolution-type representations, as implemented by the `mixsmsn` R package [23]. For a fair comparison, we started the two algorithms using the same initial values as described at the end of Section 2. All computations were carried out by Microsoft R package 3.5.1 in win64 environment of a desktop computer with 2.80-GHz/Intel Core(TM) i7-7700HQ CPU Processor and 16.0 GB RAM. Performance evaluation was assessed by the execution CPU time and the converged log-likelihood maxima.

The box plots depicted in Figure 3 reveal the selection-type algorithm demands much lower computational cost than those required for the convolution-type algorithm. The phenomenon is more apparent for the MST and MSCN distributions, particularly for larger n . The high efficiency of the selection-type algorithm can be ascribed to the fact that its E-step is designed by virtue of simplification. Finally, it is worth mentioning that both algorithms can achieve the same final log-likelihood, as demonstrated by violin plots in Figure 4.

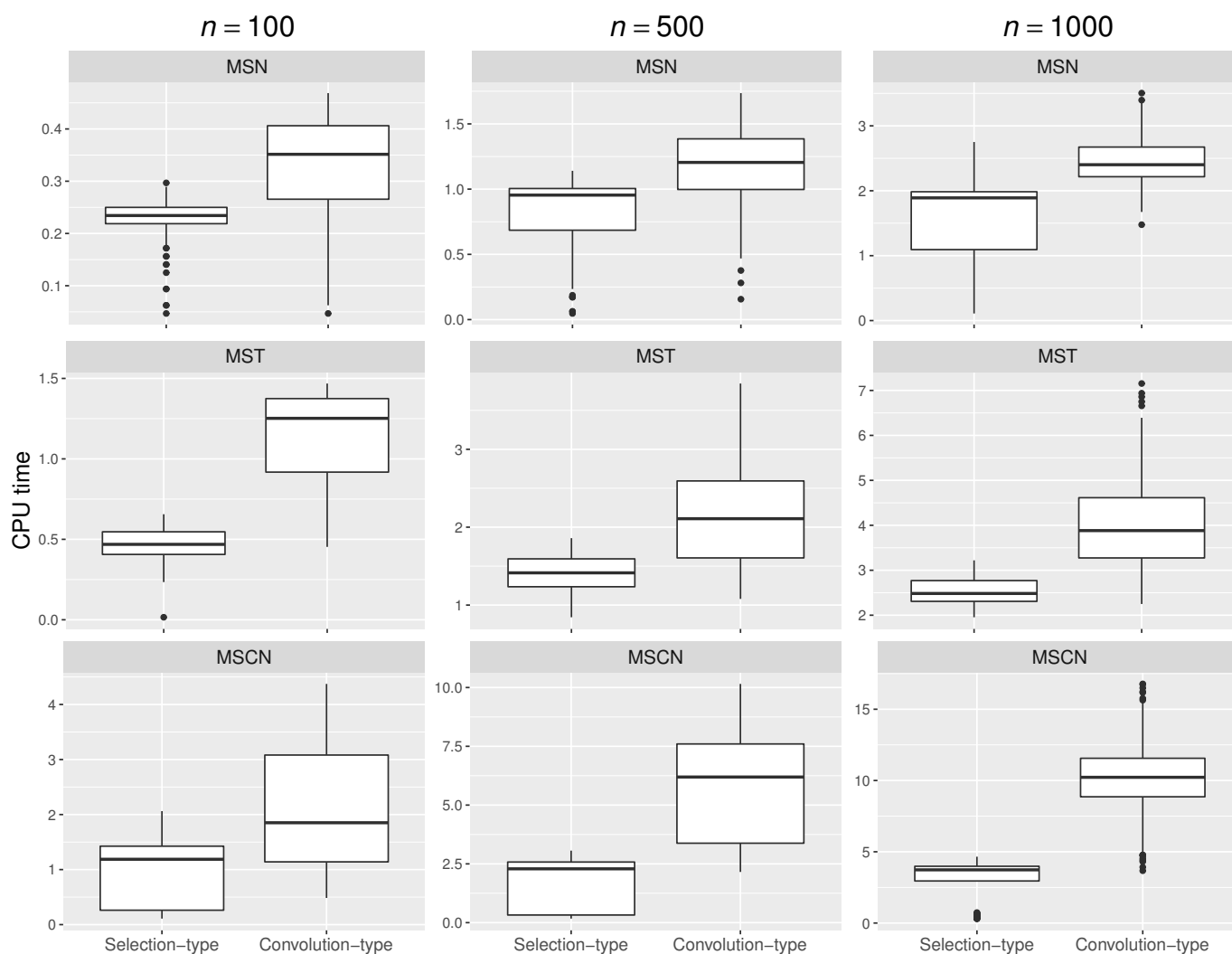


Figure 3. Box plots of CPU time for convergence of selection-type and convolution-type algorithms for fitting MSN, MST, and MSCN distributions under various sample sizes.

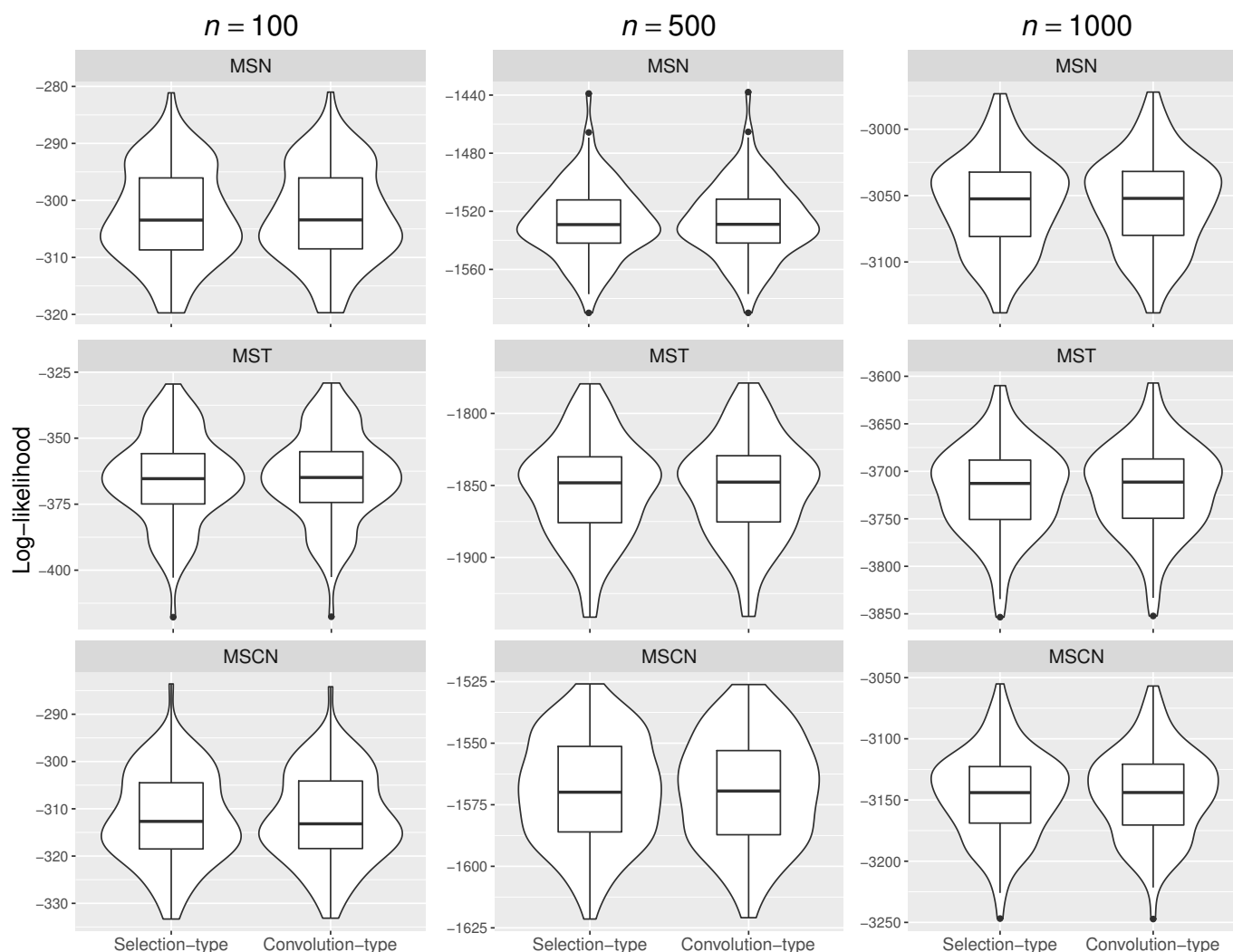


Figure 4. Violin plots of converged log-likelihood obtained by selection and convolution EM-type algorithms for fitting MSN, MST, and MSCN distributions under various sample sizes.

5. An Illustrative Example: The Wind Speed Data

We considered a trivariate dataset analyzed by Azzalini and Genton [24] for the study of spatial distribution of wind speed by means of the MST and various MSSMSN distributions proposed by Azzalini and Capitanio [6] and Arellano-Valle et al. [5], respectively. This dataset contains 278 hourly average speed assembled at three meteorological towers: Goodnoe Hills (gh), Kennewick (kw), and Vansycle (vs) from 23 February to 30 November 2003 recorded at midnight when wind speeds tend to peak. The positive and negative signs of wind speed measurements represent a westerly wind direction and an easterly wind direction, respectively. The Ljung–Box test indicates weak serial correlation for observations measured at the three stations. For modeling these data, we followed Azzalini and Genton [24] to treat the observations as being independent and identically distributed. Figure 5 presents histograms overlaid with kernel density curves obtained by using R `density()` function for measurements collected at each tower.

Six SSMFSSN models of order 3 were considered to fit the wind speed data. For the sake of comparison, the MST, MSTN, and MSTC distributions belonging to the MSSMSN family [5] were also fitted as sub-models of SSMFSSN subject to the constraint of $\lambda_2 = 0$. To select an appropriate model from the candidates, we adopted the Akaike information criterion (AIC) [25] and the Bayesian information criterion (BIC) [26], which are the two

most widely used model selection indices based on penalized likelihood and applicable for both nested and non-nested models. The two criteria are defined as

$$\text{AIC} = 2d - 2\ell_{\max} \quad \text{and} \quad \text{BIC} = d \log n - 2\ell_{\max}, \quad (58)$$

where d is the number of free parameters in the model and ℓ_{\max} is the maximized log-likelihood value. A lower AIC or BIC value indicates that a closer fit of the model to the data.

Table 2 compares the ML estimation results for nine candidate models. As can be seen, our proposed SSMFSSN models perform favorably as compared to three MSSMSN analogs because they suffer from a lack of ability to capture the possibly bimodal behavior of the wind speed data (Figure 5). Accordingly, the MFSSTT distribution provides the best fit in terms of the lowest value of AIC as well as BIC, followed by the MFSST distribution. The MSTC and MST are the top two MSSMSN models with smaller AIC and BIC values.

Table 2. Summary results from fitting various models to the wind speed data. The model with the smallest value of AIC and BIC is displayed in bold.

Family	Model	ℓ_{\max}	d	AIC	BIC
MSSMSN	MSTC	−3178.7	13	6383.4	6430.5
	MST	−3180.7	13	6387.5	6434.6
	MSTN	−3180.9	13	6387.8	6434.9
SSMFSSN	MFSSN	−3171.7	15	6373.4	6427.8
	MFSSN	−3145.6	16	6323.1	6381.2
	MFSSSLN	−3147.1	16	6326.2	6384.2
	MFSSCN	−3145.6	16	6323.3	6381.3
	MFSST	−3143.0	16	6318.1	6376.1
	MFSSTT	−3138.2	17	6310.4	6372.1

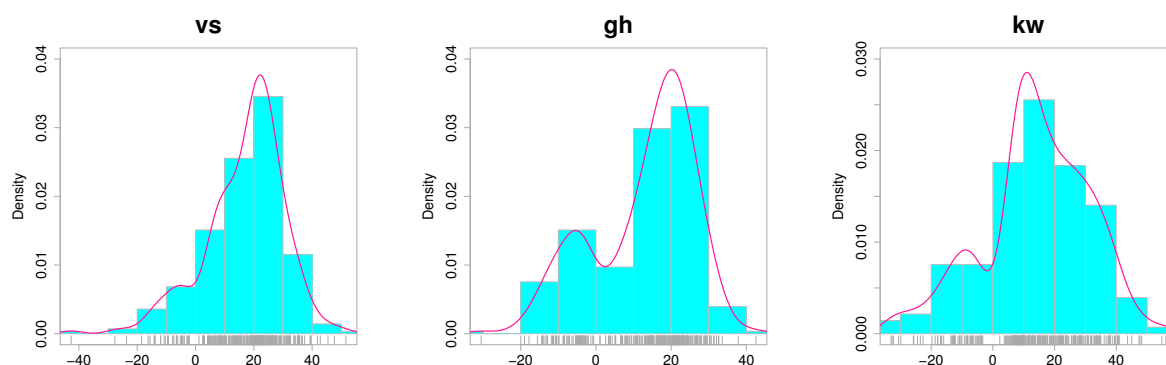
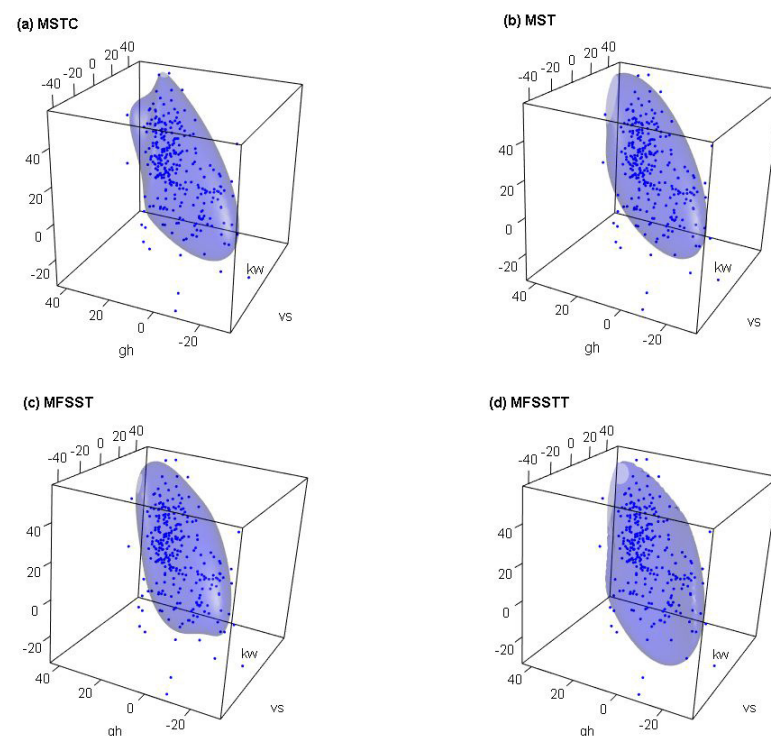


Figure 5. Histograms of univariate measurements overlaid with kernel density curves for the hourly average wind speed collected at three meteorological towers.

Table 3 summarizes the resulting ML estimates of 6 considered SSMFSSN models together with their asymptotic standard errors obtained by performing the parametric bootstrap method [27]. Notably, the estimates of the shape and flat parameters indicate the presence of skewed and leptokurtic characteristics toward different directions among the three variables. As an illustration, the fitted 3D contour densities of MSTC, MST, MFSST, and MFSSTT distributions are depicted in Figure 6. It is interesting to see that the MFSSTT model having the smallest AIC and BIC can adapt the shape of the wind speed data more closely than the other three competitors.

Table 3. ML estimates of key parameters for six SSMFSSN models. The associated standard errors are shown in parentheses.

Parameter	MFSSN	MFSSTN	MFSSSLN	MFSSCN	MFSST	MFSSTT
ξ_1	23.0(0.05)	21.3 (0.09)	21.1 (0.04)	21.5 (0.07)	19.9 (0.06)	18.6 (0.08)
ξ_2	14.8 (0.04)	15.6 (0.08)	15.2 (0.04)	15.0 (0.07)	15.1 (0.04)	15.0 (0.06)
ξ_3	14.6 (0.04)	14.9 (0.06)	14.8 (0.02)	15.4 (0.04)	13.4 (0.08)	12.6 (0.09)
σ_{11}	221.2 (0.26)	122.9 (0.49)	80.6 (0.24)	115.8 (0.31)	115.6 (0.21)	115.6 (0.25)
σ_{21}	138.8 (0.16)	96.4 (0.36)	64.6 (0.18)	91.5 (0.24)	92.8 (0.15)	93.2 (0.17)
σ_{31}	151.0 (0.13)	104.8 (0.26)	69.5 (0.13)	99.1 (0.16)	102.9 (0.18)	106.0 (0.19)
σ_{22}	181.3 (0.10)	134.5 (0.27)	90.3 (0.16)	128.0 (0.18)	131.2 (0.18)	132.1 (0.20)
σ_{32}	111.4 (0.07)	80.8 (0.18)	54.4 (0.09)	79.6 (0.12)	78.8 (0.15)	79.1 (0.15)
σ_{33}	296.4 (0.07)	203.4 (0.19)	134.0 (0.13)	187.2 (0.09)	204.4 (0.24)	206.7 (0.27)
λ_{11}	−1.7 (0.01)	−0.4 (0.04)	−0.2 (0.01)	−0.3 (0.02)	0.1 (0.02)	1.0 (0.02)
λ_{12}	1.6 (0.04)	1.1 (0.04)	1.0 (0.03)	1.3 (0.04)	1.3 (0.01)	3.2 (0.01)
λ_{13}	1.9 (0.04)	1.4 (0.04)	1.1 (0.02)	1.3 (0.02)	1.4 (0.01)	2.8 (0.02)
λ_{21}	−0.7 (1.99)	−0.1 (2.36)	0.0 (1.03)	−0.1 (1.53)	−0.2 (0.04)	1.2 (0.32)
λ_{22}	−0.7 (0.05)	−0.4 (0.25)	−0.2 (0.05)	−0.4 (0.19)	−0.5 (0.01)	−1.6 (0.21)
λ_{23}	−0.1 (1.52)	−0.2 (1.98)	−0.1 (0.78)	−0.2 (1.18)	−0.1 (0.02)	−2.3 (0.17)
ν_1	—	5.0 (0.18)	1.5 (0.03)	0.2 (0.04)	4.7 (0.06)	4.7 (0.08)
ν_2	—	—	—	0.2 (0.54)	—	1.0 (0.16)

**Figure 6.** The 3-D contour densities for MSTC, MST, MFSST, and MFSSTT distributions fitted to the wind speed data.

6. Conclusions

We introduce a novel family of SSMFSSN distributions as a generalization of the work of Mahdavi et al. [10] that can capture simultaneously the dependency among multivariate responses, skewness, heavy-tailedness, and, in particular, multimodal density shapes without resorting to the use of finite mixtures [14,15,21]. Since the SSMFSSN model cannot be represented by a convolution-type form, this stimulates us to devise a feasible ECME algorithm for ML estimation under the selection-type mechanism. The effectiveness and efficiency of the algorithm are evaluated by conducting two simulation studies. Numerical analysis of a real dataset highlights the potential and capability of our proposed approach as a promising alternative tool for modeling multimodal multivariate data with asymmetrical behavior. Computer programs for implementation of our methods can be installed as a R package from Github devtools::install_github("a-mahdavi/SSMFSSN.EM"). Further

developments of the current approach could be exploited for powerful extensions of the factor analysis model or finite mixtures thereof with censored or possibly missing values that were considered recently by the authors of [28–35]. One limitation of the SSMFSSN model is that it may not be suited to the data with modes having too far distances. Another worthwhile extension of this work is to pursue a mixture modeling framework of the current approach that would be an effective way to resolve this problem.

Author Contributions: A.M. and T.-I.L. conceived the project, developed the statistical methods, designed the approach, and analyzed the results. All authors contributed to the development of the methodology and to writing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: TIL was partially supported by the Ministry of Science and Technology of Taiwan (Grant No. MOST 109-2118-M-005-005-MY3).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge three anonymous referees for their comments and suggestions that greatly improved this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Hadamard Product

Below, we present three propositions concerning the multiplication and partial derivatives of the Hadamard product which are useful for our methodological developments.

Proposition A1. Let A be a $p \times q$ matrix and D be a $p \times p$ diagonal matrix. Then, $(DA)^{\odot n} = D^{\odot n} A^{\odot n}$, where $n \in \mathbb{N}$.

Proof. Multiplication of a Hadamard product by diagonal matrix D satisfies the following equation:

$$D(A \odot B) = (DA) \odot B = A \odot (DB). \quad (A1)$$

By taking $B = A$ in (A1), we obtain

$$D(A \odot A) = (DA) \odot A = A \odot (DA). \quad (A2)$$

Now, consider $n = 2$; using (A2), we have

$$D^{\odot 2} A^{\odot 2} = D[(DA) \odot A] = (DA) \odot (DA) = (DA)^{\odot 2}.$$

Similarly, for $n = 3$, we have

$$D^{\odot 3} A^{\odot 3} = D[(DA)^{\odot 2} \odot A] = (DA)^{\odot 2} \odot (DA) = (DA)^{\odot 3}.$$

Clearly, the desired statement can be established by mathematical induction on n . \square

Proposition A2. Let A be a $p \times q$ matrix and $\mathbf{b} \in \mathbb{R}^p$ and $\mathbf{x} \in \mathbb{R}^q$ be two column vector. Then,

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{b}^\top (\mathbf{A}\mathbf{x})^{\odot n} = n \mathbf{A}^\top ((\mathbf{A}\mathbf{x})^{\odot n-1} \odot \mathbf{b}),$$

Proof. Let $f = \mathbf{b}^\top (\mathbf{A}\mathbf{x})^{\odot n}$. It is easy to show

$$f = \text{Trace}(\mathbf{b}^\top (\mathbf{A}\mathbf{x})^{\odot n}) = \mathbf{b} : (\mathbf{A}\mathbf{x})^{\odot n},$$

where “:” denotes Frobenius products defined as $\mathbf{A} : \mathbf{B} = \text{Trace}(\mathbf{A}^\top \mathbf{B})$.

Making use of the following facts

$$\begin{aligned}
(A \odot B) : C &= A : (B \odot C), \\
A : BC &= B^\top A : C = AC^\top : B, \\
\frac{\partial}{\partial \mathbf{x}}(A \odot B) &= \left(\frac{\partial A}{\partial \mathbf{x}}\right) \odot B + A \odot \left(\frac{\partial B}{\partial \mathbf{x}}\right),
\end{aligned}$$

the differential of the scalar function f yields

$$\begin{aligned}
df &= \mathbf{b} : n(\mathbf{Ax})^{\odot n-1} \odot A(d\mathbf{x}) = n((\mathbf{Ax})^{\odot n-1} \odot \mathbf{b}) : A(d\mathbf{x}) \\
&= nA^\top((\mathbf{Ax})^{\odot n-1} \odot \mathbf{b}) : d\mathbf{x}.
\end{aligned}$$

This completes the proof. \square

Proposition A3. Let \mathbf{X} be a $p \times q$ matrix and $\mathbf{a} \in \mathbb{R}^p$ and $\mathbf{b} \in \mathbb{R}^q$ be two column vector then,

$$\frac{\partial}{\partial \mathbf{X}} \mathbf{a}^\top (\mathbf{Xb})^{\odot n} = n \left((\mathbf{Xb})^{\odot n-1} \odot \mathbf{a} \right) \mathbf{b}^\top.$$

Proof. Similar to the proof of Proposition A2, we define $f = \mathbf{a} : (\mathbf{Xb})^{\odot n}$ and obtain its differential as below

$$\begin{aligned}
df &= \mathbf{a} : n(\mathbf{Xb})^{\odot n-1} \odot (d\mathbf{X})\mathbf{b} = n((\mathbf{Xb})^{\odot n-1} \odot \mathbf{a}) : (d\mathbf{X})\mathbf{b} \\
&= n((\mathbf{Xb})^{\odot n-1} \odot \mathbf{a}) \mathbf{b}^\top : d\mathbf{X},
\end{aligned}$$

which completes the proof. \square

Appendix B. Proof of Equation (43)

Proof. Without loss of generality, we assume that $\boldsymbol{\zeta} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$. Let $\mathbf{Y} \stackrel{d}{=} (\tau^{-1/2})\mathbf{Z}_1 | (\tau^{-1/2}\mathbf{Z}_2 < \lambda_1^\top (\tau^{-1/2}\mathbf{Z}_1) + \dots + \lambda_m^\top (\tau^{-1/2}\mathbf{Z}_1)^{\odot 2m-1})$, where $\mathbf{Z}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$ and $\mathbf{Z}_2 \sim N(0, 1)$ are two independent random variables and $\tau \sim \Gamma(\nu/2, \nu/2)$. Clearly, $(\mathbf{X}_1, \mathbf{X}_2)^\top \stackrel{d}{=} \tau^{-1/2}(\mathbf{Z}_1, \mathbf{Z}_2)^\top \sim t_{p+1}(\mathbf{0}, \mathbf{I}_{p+1}, \nu)$. Therefore, it is easy to verify that $\mathbf{X}_1 \sim t_p(\mathbf{0}, \mathbf{I}_p, \nu)$, $\mathbf{X}_2 \sim t(0, 1, \nu)$ and

$$\sqrt{\frac{\nu+p}{\nu+\mathbf{x}_1^\top \mathbf{x}_1}} \mathbf{X}_2 | (\mathbf{X}_1 = \mathbf{x}_1) \sim t(0, 1, \nu+p).$$

By Bayes' theorem, the pdf of $\mathbf{Y} \stackrel{d}{=} \mathbf{X}_1 | (\mathbf{X}_2 < \lambda_1^\top \mathbf{X}_1 + \dots + \lambda_m^\top \mathbf{X}_1^{\odot 2m-1})$ is

$$\begin{aligned}
f_Y(\mathbf{y}) &= \frac{f_{\mathbf{X}_1}(\mathbf{y}) \Pr(\mathbf{X}_2 < \lambda_1^\top \mathbf{X}_1 + \dots + \lambda_m^\top \mathbf{X}_1^{\odot 2m-1} | \mathbf{X}_1 = \mathbf{y})}{\Pr(\mathbf{X}_2 < \lambda_1^\top \mathbf{X}_1 + \dots + \lambda_m^\top \mathbf{X}_1^{\odot 2m-1})} \\
&= 2f_{\mathbf{X}_1}(\mathbf{y}) \Pr\left(\sqrt{\frac{\nu+p}{\nu+\mathbf{y}^\top \mathbf{y}}} \mathbf{X}_2 < \sqrt{\frac{\nu+p}{\nu+\mathbf{y}^\top \mathbf{y}}} (\lambda_1^\top \mathbf{y} + \dots + \lambda_m^\top \mathbf{y}^{\odot 2m-1}) | \mathbf{X}_1 = \mathbf{y}\right) \\
&= 2t_p(\mathbf{y}; \mathbf{0}, \mathbf{I}_p, \nu) T\left(\sqrt{\frac{\nu+p}{\nu+\mathbf{y}^\top \mathbf{y}}} (\lambda_1^\top \mathbf{y} + \dots + \lambda_m^\top \mathbf{y}^{\odot 2m-1}); \nu+p\right).
\end{aligned}$$

This completes the proof. \square

References

1. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [\[CrossRef\]](#)
2. Azzalini, A.; Valle, A.D. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [\[CrossRef\]](#)
3. Arellano-Valle, R.B.; Genton, M.G. On fundamental skew distributions. *J. Multivar. Anal.* **2005**, *96*, 93–116. [\[CrossRef\]](#)
4. Ferreira, C.S.; Lachos, V.H.; Bolfarine, H. Likelihood-based inference for multivariate skew scale mixtures of normal distributions. *ASTA Adv. Stat. Anal.* **2016**, *100*, 421–441. [\[CrossRef\]](#)

5. Arellano-Valle, R.B.; Ferreira, C.S.; Genton, M.G. Scale and shape mixtures of multivariate skew-normal distributions. *J. Multivar. Anal.* **2018**, *166*, 98–110. [CrossRef]
6. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution. *J. R. Stat. Soc. Ser. B* **2003**, *65*, 367–389. [CrossRef]
7. Wang, J.; Boyer, J.; Genton, M.G. A skew-symmetric representation of multivariate distributions. *Stat. Sin.* **2004**, 1259–1270. Available online: <https://www.jstor.org/stable/24307231> (accessed on 5 January 2005).
8. Ma, Y.; Genton, M.G. Flexible class of skew-symmetric distributions. *Scand. J. Stat.* **2004**, *31*, 459–468. [CrossRef]
9. Styan, G.P. Hadamard products and multivariate statistical analysis. *Linear Algebra Its Appl.* **1973**, *6*, 217–240. [CrossRef]
10. Mahdavi, A.; Amirzadeh, V.; Jamalizadeh, A.; Lin, T.I. Maximum likelihood estimation for scale-shape mixtures of flexible generalized skew normal distributions via selection representation. *Comput. Stat.* **2021**, *36*, 2201–2230. [CrossRef]
11. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–22. [CrossRef]
12. Meng, X.L.; Rubin, D.B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **1993**, *80*, 267–278. [CrossRef]
13. Liu, C.; Rubin, D.B. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **1994**, *81*, 633–648. [CrossRef]
14. Lin, T.I. Maximum likelihood estimation for multivariate skew normal mixture models. *J. Multivar. Anal.* **2009**, *100*, 257–265. [CrossRef]
15. Lin, T.I. Robust mixture modeling using multivariate skew t distributions. *Stat. Comp.* **2010**, *20*, 343–356. [CrossRef]
16. Lin, T.I.; Lee, J.C.; Yen, S.Y. Finite mixture modelling using the skew normal distribution. *Stat. Sin.* **2007**, 909–927. Available online: <https://www.jstor.org/stable/24307705> (accessed on 5 December 2007).
17. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: <http://www.R-project.org> (accessed on 10 January 2021).
18. Lin, T.I.; Ho, H.J.; Lee, C.-R. Flexible mixture modelling using the multivariate skew- t -normal distribution. *Stat. Comp.* **2014**, *24*, 531–546. [CrossRef]
19. Wang, J.; Genton, M.G. The multivariate skew-slash distribution. *J. Stat. Plan. Inference* **2006**, *136*, 209–220. [CrossRef]
20. Lin, T.I.; Lee, J.C.; Hsieh, W.J. Robust mixture modeling using the skew t distribution. *Stat. Comp.* **2007**, *17*, 81–92. [CrossRef]
21. Wang, W.L.; Jamalizadeh, A.; Lin, T.I. Finite mixtures of multivariate scale-shape mixtures of skew-normal distributions. *Stat. Pap.* **2020**, *61*, 2643–2670. [CrossRef]
22. Cabral, C.R.B.; Lachos, V.H.; Prates, M.O. Multivariate mixture modeling using skew-normal independent distributions. *Comput. Stat. Data Anal.* **2012**, *56*, 126–142. [CrossRef]
23. Prates, M.O.; Lachos, V.H.; Cabral, C.R.B. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *J. Stat. Softw.* **2013**, *54*, 1–20. [CrossRef]
24. Azzalini, A.; Genton, M.G. Robust likelihood methods based on the skew- t and related distributions. *Int. Stat. Rev.* **2008**, *76*, 106–129. [CrossRef]
25. Akaike, H. A new look at the statistical model identification. *IEEE. Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
26. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
27. Efron, B.; Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1986**, 54–75. Available online: <https://www.jstor.org/stable/2245500> (accessed on 22 May 2003). [CrossRef]
28. Lin, T.I.; Wu, P.H.; McLachlan, G.J.; Lee, S.X. A robust factor analysis model using the restricted skew- t distribution. *Test* **2015**, *24*, 510–531. [CrossRef]
29. Liu, M.; Lin, T. Skew-normal factor analysis models with incomplete data. *J. Appl. Stat.* **2015**, *42*, 789–805. [CrossRef]
30. Wang, W.L.; Liu, M.; Lin, T.I. Robust skew- t factor analysis models for handling missing data. *Stat. Methods Appl.* **2017**, *26*, 649–672. [CrossRef]
31. Lin, T.I.; Wang, W.L.; McLachlan, G.J.; Lee, S.X. Robust mixtures of factor analysis models using the restricted multivariate skew- t distribution. *Stat. Model.* **2018**, *28*, 50–72. [CrossRef]
32. Wang, W.L.; Castro, L.M.; Lachos, V.H.; Lin, T.I. Model-based clustering of censored data via mixtures of factor analyzers. *Comput. Stat. Data Anal.* **2019**, *140*, 104–121. [CrossRef]
33. Wang, W.L.; Castro, L.M.; Hsieh, W.C.; Lin, T.I. Mixtures of factor analyzers with covariates for modeling multiply censored dependent variables. *Stat. Pap.* **2021**. [CrossRef]
34. Wang, W.L.; Lin, T.I. Robust clustering of multiply censored data via mixtures of t factor analyzers. *TEST* **2021**. [CrossRef]
35. Wang, W.L.; Lin, T.I. Robust clustering via mixtures of t factor analyzers with incomplete data. *Adv. Data Anal. Classif.* **2021**. [CrossRef]