Journal of Multivariate Analysis 200 (2024) 105260

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A multivariate skew-normal-Tukey-h distribution

Sagnik Mondal *, Marc G. Genton

Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

AMS 2020 subject classifications:
primary 62H10
secondary 62E10

Keywords:
Heavy-tails
Lambert's-W
Non-gaussian distribution
Skew-normal
Skew-t
Tukey-h

ABSTRACT

We introduce a new family of multivariate distributions by taking the component-wise Tukey-h transformation of a random vector following a skew-normal distribution with an alternative parameterization. The proposed distribution is named the skew-normal-Tukey-h distribution and is an extension of the skew-normal distribution for handling heavy-tailed data. We compare this proposed distribution to the skew-t distribution, which is another extension of the skew-normal distribution for modeling tail-thickness, and demonstrate that when there are substantial differences in marginal kurtosis, the proposed distribution is more appropriate. Moreover, we derive many appealing stochastic properties of the proposed distribution and provide a methodology for the estimation of the parameters that can be applied to large dimensions. Using simulations, as well as a wine and a wind speed data application, we illustrate how to draw inferences based on the multivariate skew-normal-Tukey-h distribution.

1. Introduction

In recent decades, there has been a growing interest in developing parametric multivariate distributions flexible enough to handle skewness and tail-thickness for various statistical applications. In a multivariate setup, two of the most popular methods to introduce both skewness and tail-thickness are:

1. *Perturbation of symmetry* of an elliptically contoured distribution which is capable of capturing tail-thickness. Examples of such distributions include the multivariate skew-t distribution [12] and the multivariate extended skew-t distribution [6].
2. *Transformation* of a random vector following some elliptically contoured distribution for imposing skewness and tail-thickness. Examples of such transformations are the Tukey g-and-h transformation [21] and the Sinh-Arcsinh transformation [30] in the multivariate case, and the Lambert's-W transformation [26] in the univariate case.

The primary parametric model obtained by perturbing the symmetry of an elliptically contoured distribution, which instigated the research in this area, is the multivariate skew-normal distribution introduced by [14]. Many distributions such as the multivariate skew-t distribution, the multivariate extended skew-normal distribution, and the multivariate extended skew-t distribution were built upon the foundation of the skew-normal distribution. These distributions can be viewed as special cases of the multivariate unified skew-elliptical distribution studied by [7]. For more on these types of distributions, readers are referred to the books by [13,24], and to a recent review by [10]. Since the skew-normal distribution is obtained by perturbing the symmetry of the Gaussian distribution and the skew-t distribution is obtained by perturbing the symmetry of the Student's-t distribution, the skew-normal distribution is not capable of handling tail-thickness while the skew-t distribution is more apt for modeling heavy-tailed data. However, one shortcoming of the skew-t distribution is that it cannot handle different tail-thickness for different marginals, since the tail-thickness is controlled only by one parameter. There has been a proposal by [33] to introduce a multivariate Student's-t

* Corresponding author.
E-mail addresses: sagnik.mondal@kaust.edu.sa (S. Mondal), marc.genton@kaust.edu.sa (M.G. Genton).

https://doi.org/10.1016/j.jmva.2023.105260
Received 8 June 2023; Received in revised form 20 November 2023; Accepted 21 November 2023
Available online 28 November 2023

distribution with different tail-thickness parameters for different marginals. However, the probability density function (pdf) of the proposed distribution involves complicated hypergeometric functions that make inference with such a distribution very challenging.

The second approach above for introducing skewed and heavy-tailed distribution is to use some non-linear transformation on a light-tailed elliptically symmetric random variable. The Lambert's-W transformation, proposed by [26] in the univariate case, can impose both skewness and tail-thickness on a Gaussian random variable using a single parameter. However, as this transformation is not one-to-one, the pdf of its multivariate extension becomes almost impossible to track down, especially for higher dimensions. This issue was solved by [27] where he changed the Lambert's-W transformation slightly and made it one-to-one. This modified transformation is a generalized version of the Tukey-h transformation. Although [27] proposed this new distribution in the univariate setting, he only briefly mentioned how it can be extended to the multivariate setting by applying this transformation component-wise. Other examples include the Sinh-Arcsinh (SAS) transformation and the Tukey g-and-h transformation which are monotonic and control skewness and tail-thickness with separate parameters. A multivariate g-and-h distribution was presented by [21] which is based on the component-wise Tukey's g-and-h transformation of a random vector following a Gaussian distribution. As a result, it permits different kurtosis for different marginals. However, one drawback of this distribution is drawing inferences. Since the inverse of Tukey's g-and-h transformation does not have a closed form, the likelihood function cannot be readily calculated. Moreover, for parameter estimation, some definitions of multivariate quantiles are needed. This can be computationally challenging when the dimension is high because the number of directions in which the quantiles have to be computed grows exponentially with dimension. The univariate SAS distribution and its various stochastic and inferential properties were mainly discussed by [30]. The idea of the multivariate expansion of this family has also been presented by [30]. It consists in using the transformation on the marginals of a standardized but correlated multivariate Gaussian distribution. A similar approach has been taken by [37] who proposed a distribution that is capable of modeling higher skewness than the original SAS distribution by applying the two-piece transformation to the symmetric SAS distribution. The SAS distribution was used by [42] in the context of a bivariate random field for wind data and they discussed how to draw inference based on it. However, inference in the multivariate scenario is yet to be thoroughly explored.

In this article, we propose a new multivariate distribution by combining these two techniques, the perturbation of symmetry for skewness and the transformation for tail-thickness. We introduce the skew-normal-Tukey-h distribution by taking the Tukey-h transformation on the components of a skew-normal random vector to introduce tail-thickness on each component. Moreover, by changing the marginal kurtosis parameter, we can have different kurtosis for different marginals. We study some basic statistical properties of the skew-normal-Tukey-h distribution. Furthermore, we discuss how to draw inferences based on this distribution. We compare the proposed distribution with the skew-t distribution since both of them are extensions of the skew-normal distribution for handling heavy-tailed data. Finally, we justify in which scenarios the skew-normal-Tukey-h distribution is more appropriate compared to the skew-t distribution using a simulation study and two data applications.

It should be pointed out that the aforementioned two methods for constructing skewed and heavy-tailed distributions are not exhaustive. There exists a variety of proposals in the statistics literature. For example, distributions studied by [17,39] are very similar to the definition of the skew-normal distribution. A definition of generalized skew-elliptical distributions which bring such different skewed distributions defined by perturbation of symmetry under one umbrella was proposed by [25]. Another avenue for the introduction of skewness and tail-thickness was explored by [22] and further generalized by [41] under the name of location-scale mixtures of Gaussian distributions. Various other non-Gaussian distributions for modeling skewed and heavy-tailed data can also be obtained using the theory of copulas [38]. We refer interested readers to the books by [29,35], and the references therein, for more details on copulas. These are some other examples of parametric families proposed for modeling various skewed and heavy-tailed or light-tailed data.

The rest of the article is organized as follows. In Section 2, we formally define the skew-normal-Tukey-h distribution, whereas various of its stochastic properties are discussed in Section 3. In Section 4, we illustrate how to draw inferences based on the skew-normal-Tukey-h distribution. In Sections 5 and 6, we present simulation studies and two applications to wine data and to wind speed data showing when the skew-normal-Tukey-h distribution is more appropriate compared to the skew-t distribution. Finally, in Section 7, we conclude our article and discuss some avenues for future research work.

2. Multivariate skew-normal-Tukey-h distribution

In this section, we define the multivariate skew-normal-Tukey-h distribution. We start by defining an alternative parameterization of the multivariate skew-normal distribution.

2.1. Skew-normal distribution

The multivariate skew-normal distribution was introduced by [14] and later studied in [11]. A random vector $\boldsymbol{Y} \in \mathbb{R}^p$ is said to have a multivariate skew-normal distribution with location parameter $\boldsymbol{\xi} \in \mathbb{R}^p$, symmetric positive definite scale parameter $\boldsymbol{\Omega} \in \mathbb{R}^{p\times p}$, and skewness parameter $\boldsymbol{\alpha} \in \mathbb{R}^p$, if its pdf is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2\phi_p(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Omega})\, \Phi\{\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\boldsymbol{y} - \boldsymbol{\xi})\}, \quad \boldsymbol{y} \in \mathbb{R}^p, \tag{1}$$

where $\phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the pdf of a p-dimensional normal distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and positive definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and $\boldsymbol{\omega} = \mathrm{diag}(\boldsymbol{\Omega})^{1/2}$. Here, and from now on, we call this distribution with the parameterization in (1) the Azzalini skew-normal ($\mathcal{ASN}$) distribution and we denote it by $\boldsymbol{Y} \sim \mathcal{ASN}_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$.

As used in [34], the $\mathcal{ASN}_p(\xi, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ distribution can be reparameterized by means of the relations $\boldsymbol{\Omega} = \boldsymbol{\Psi} + \boldsymbol{\eta}\boldsymbol{\eta}^\top$ and $\boldsymbol{\alpha} = (1 + \boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta})^{-1/2}\boldsymbol{\omega}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}$, where $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$ is a symmetric positive definite matrix, $\boldsymbol{\eta} \in \mathbb{R}^p$ and $\boldsymbol{\omega} = \mathrm{diag}(\sqrt{\Psi_{11} + \eta_1^2}, \dots, \sqrt{\Psi_{pp} + \eta_p^2})$, with Ψ_{ii} and η_i being the ith diagonal element of $\boldsymbol{\Psi}$ and $\boldsymbol{\eta}$, respectively, for $i \in \{1, \dots, p\}$. Conversely, by letting $\boldsymbol{\omega} = \mathrm{diag}(\boldsymbol{\Omega})^{1/2}$, $\bar{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$ and $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})^{-1/2}\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$, we have $\boldsymbol{\Psi} = \boldsymbol{\omega}(\bar{\boldsymbol{\Omega}}^{-1} + \boldsymbol{\alpha}\boldsymbol{\alpha}^\top)^{-1}\boldsymbol{\omega} = \boldsymbol{\omega}(\bar{\boldsymbol{\Omega}} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)\boldsymbol{\omega}$ and $\boldsymbol{\eta} = \boldsymbol{\omega}\boldsymbol{\delta}$. With this alternative parameterization, the pdf of $\boldsymbol{Y}$ from (1) is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2\phi_p\left(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Psi} + \boldsymbol{\eta}\boldsymbol{\eta}^\top\right)\Phi\left\{\frac{\boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{y} - \boldsymbol{\xi})}{\sqrt{1 + \boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}}}\right\}, \quad \boldsymbol{y} \in \mathbb{R}^p. \tag{2}$$

[14] used this parameterization up to minor differences. Moreover, [2–4] have also used the same parameterization. With this parameterization, a p-variate random vector $\boldsymbol{Y}$ is said to have a skew-normal ($\mathcal{SN}$) distribution with location parameter $\boldsymbol{\xi} \in \mathbb{R}^p$, symmetric positive definite scale matrix $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$, and skewness parameter $\boldsymbol{\eta} \in \mathbb{R}^p$ if its pdf is given by (2). We denote it by $\boldsymbol{Y} \sim \mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$.

Many interesting properties of the $\mathcal{SN}$ distribution with the parameterization in (2) have been derived in [34]. The following results are given here as they will be useful later on, while their proofs can be found in [34]:

- *Stochastic representation of $\mathcal{SN}$ distribution:* If $\boldsymbol{Y} \sim \mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$, then $\boldsymbol{Y} = \boldsymbol{\xi} + U\boldsymbol{\eta} + \boldsymbol{W}$, where U and $\boldsymbol{W}$ are independently distributed, with half-normal U denoted by $U \sim \mathcal{HN}(0, 1)$, and $\boldsymbol{W} \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Psi})$.
- *Affine transformation of the $\mathcal{SN}$ distribution:* If $\boldsymbol{Y} \sim \mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$, then for any fixed vector $\boldsymbol{a} \in \mathbb{R}^q$ and any fixed matrix $\boldsymbol{B} \in \mathbb{R}^{q\times p}$ of full row rank and $q \le p$: $\boldsymbol{a} + \boldsymbol{B}\boldsymbol{Y} \sim \mathcal{SN}_q(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{\xi}, \boldsymbol{B}\boldsymbol{\Psi}\boldsymbol{B}^\top, \boldsymbol{B}\boldsymbol{\eta})$.
- *Marginal distributions of the $\mathcal{SN}$ distribution:* Let $\boldsymbol{Y} \sim \mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$ and consider the partition of $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top$ with $\boldsymbol{Y}_i$ of size p_i ($i \in \{1, 2\}$) and such that $p_1 + p_2 = p$, with corresponding partitions of the parameters in blocks of matching sizes. Then $\boldsymbol{Y}_i \sim \mathcal{SN}_{p_i}(\boldsymbol{\xi}_i, \boldsymbol{\Psi}_{ii}, \boldsymbol{\eta}_i)$, $i \in \{1, 2\}$.

The $\mathcal{ASN}$ and $\mathcal{SN}$ parameterizations describe the same distribution but the simplicity of the marginal distributions in the $\mathcal{SN}$ parameterization (see above) will prove useful for inferential purposes later on.

2.2. Skew-normal-Tukey-h distribution

We introduce tail-thickness in the skew-normal distribution by taking the Tukey-h transformation of each component of a random vector following a $\mathcal{SN}$ distribution. The Tukey-h transformation is

$$\tau_h(x) = x\exp(hx^2/2), \quad x \in \mathbb{R}, \quad h \ge 0. \tag{3}$$

Moreover, for $\boldsymbol{x} = (x_1, \dots, x_p)^\top \in \mathbb{R}^p$, we define

$$\boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{x}) = \{\tau_{h_1}(x_1), \dots, \tau_{h_p}(x_p)\}^\top, \quad \boldsymbol{h} = (h_1, \dots, h_p)^\top,\ h_i \ge 0,\ i \in \{1, \dots, p\}. \tag{4}$$

Definition 1 (*Skew-Normal-Tukey-h Distribution*)**.** A random vector $\boldsymbol{Y} \in \mathbb{R}^p$ with the stochastic representation $\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{Z})$, where $\boldsymbol{Z} \sim \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta})$ and $\bar{\boldsymbol{\Psi}}$ is a $p \times p$ correlation matrix, is said to have a multivariate skew-normal-Tukey-h distribution. Here $\boldsymbol{\xi} \in \mathbb{R}^p$ is the location parameter, $\boldsymbol{\omega} = \mathrm{diag}(\omega_{11}, \dots, \omega_{pp})$ is a $p \times p$ diagonal scale matrix such that $\omega_{ii} > 0$, $i \in \{1, \dots, p\}$, $\boldsymbol{\eta} \in \mathbb{R}^p$ is the skewness parameter, and $\boldsymbol{h}$ is the tail-thickness parameter vector such that $\boldsymbol{h} = (h_1, \dots, h_p)^\top \in \mathbb{R}^p$, $h_i \ge 0$, $i \in \{1, \dots, p\}$. We denote $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$.

We define the $\mathcal{SNTH}$ distribution with a correlation matrix $\bar{\boldsymbol{\Psi}}$ and a diagonal scale matrix $\boldsymbol{\omega}$. The $\bar{\boldsymbol{\Psi}}$ parameter governs the dependence structure in the model and $\boldsymbol{\omega}$ is a diagonal matrix consisting of the marginal scale parameters. To make all the parameters identifiable we restrict $\bar{\boldsymbol{\Psi}}$ to be a correlation matrix. It is immediate from the definition of the $\mathcal{SNTH}$ distribution that when $\boldsymbol{h} = \boldsymbol{0}$ the $\mathcal{SNTH}$ distribution reduces to the $\mathcal{SN}$ distribution. The Tukey-h transformation applied on the marginals of the skew-normal distribution imposes tail-thickness in the distribution. Moreover, since we can vary the components of the $\boldsymbol{h}$ parameter over the marginals, the resulting distribution can have different kurtosis for different marginals. In this way, we propose an extension of the skew-normal distribution, capable of handling different marginal tail-thickness. In that sense, the $\mathcal{SNTH}$ distribution is different from the skew-t distribution. The skew-t distribution can also be thought of as an extension of the skew-normal distribution for modeling tail-thickness in the data, but it is incapable of capturing different kurtosis for different marginals.

It should be pointed out that the proposed $\mathcal{SNTH}$ distribution belongs to the Lambert-$W \times F$ family of distributions [27], where F represents the cumulative distribution function (cdf) of the skew-normal distribution. The main difference is that [27] proposed the location-scale Lambert-$W \times F$ distribution with $\mu_X = \mathrm{E}(X)$ as the location parameter and $\sigma_X = \sqrt{\mathrm{Var}(X)}$ as the scale parameter, $X \sim F$, and the transformation was applied on $(X - \mu_X)/\sigma_X$. For defining the $\mathcal{SNTH}$ distribution, we start with a "standard" skew-normal distribution and apply the Tukey-h transformation on it, and then we use a location-scale transformation on the transformed random variable.

Since we take a monotonic marginal transformation on each component of a skew-normal random vector, the underlying copula of the $\mathcal{SNTH}$ distribution remains the same as the skew-normal copula [40]. Because the skew-normal copula can be tail-asymmetric, the copula corresponding to the $\mathcal{SNTH}$ distribution is also tail-asymmetric. This is not the case for the multivariate g-and-h distribution, for which the underlying copula is the Gaussian copula, hence tail-symmetric.

3. Properties of the $\mathcal{SNTH}$ distribution

We outline some basic probabilistic properties of the $\mathcal{SNTH}$ distribution such as its pdf, cdf, moments, marginal and conditional distributions, and canonical form. Due to the $\mathcal{SNTH}$ definition using the $\mathcal{SN}$ distribution, many of the $\mathcal{SN}$ appealing properties get transferred to the $\mathcal{SNTH}$ distribution. This is one of the reasons we defined the $\mathcal{SNTH}$ with the $\mathcal{SN}$ distribution parameterized in (2).

3.1. Probability density function of $\mathcal{SNTH}$

In the next proposition we present the pdf of the $\mathcal{SNTH}$ distribution. The univariate $\mathcal{SNTH}$ pdf can be found using Theorem 1 of [27] using F as the skew-normal distribution. We extend Theorem 1 of [27] with F as the skew-normal distribution to the multivariate setup in the next proposition.

Proposition 1. *The pdf of $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$ is, for $\boldsymbol{y} \in \mathbb{R}^p$:*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 2\phi_p\{\boldsymbol{g}(\boldsymbol{y}); \boldsymbol{0}, \bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top\}\Phi\left\{\frac{\boldsymbol{\eta}^\top\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{g}(\boldsymbol{y})}{\sqrt{1+\boldsymbol{\eta}^\top\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{\eta}}}\right\}\prod_{i=1}^{p}\left\{\frac{1}{\omega_{ii}}\left(\frac{\exp[\frac{1}{2}W_0\{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2\}]}{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2+\exp[W_0\{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2\}]}\right)\right\}, \tag{5}$$

where $\boldsymbol{g}(\boldsymbol{y}) = \{g_1(y_1), \dots, g_p(y_p)\}^\top$, $g_i(y_i) = (\frac{y_i-\xi_i}{\omega_{ii}})\exp[-\frac{1}{2}W_0\{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2\}]$, $i \in \{1, \dots, p\}$, and $W_0(\cdot)$ is the principal branch of the Lambert's-W function.

Proof. Consider the transformation $z = x\exp(hx^2/2)$. Then $hz^2 = hx^2\exp(hx^2) \Rightarrow hx^2 = W_0(hz^2) \Rightarrow x = z\exp\{-W_0(hz^2)/2\}$, where $W_0(\cdot)$ is the principal branch of the Lambert's-W function [20]. This essentially means that $W_0(\cdot)$ is the inverse function of the function $f(x) = x\exp(x)$, $x \in \mathbb{R}$. Although the inverse of $f(x)$ is not unique when $x < 0$, it is unique when $x > 0$. For us the argument of $W_0(\cdot)$ is $hz^2 \geq 0$, which makes the inverse of the Tukey-h transformation unique (see also Lemma 5 in [27]). Hence, the inverse of the Tukey-h transformation (3) is

$$\tau_h^{-1}(z) = z\exp\{-W_0(hz^2)/2\}, \tag{6}$$

and it is a one-to-one function as it should be since $\tau_h(z)$ is one-to-one for $h \geq 0$. Moreover,

$$\frac{\partial}{\partial z}\tau_h^{-1}(z) = \frac{\exp\{W_0(hz^2)/2\}}{hz^2+\exp\{W_0(hz^2)\}},$$

and is obtained using the fact that $W_0'(z) = 1/[z+\exp\{W_0(z)\}]$. With the form of $\tau_h^{-1}(z)$ and $\partial\tau_h^{-1}(z)/\partial z$ it is straightforward to conclude that the pdf of $\boldsymbol{Y}$ reduces to (5). □

The pdf of the $\mathcal{SNTH}$ distribution is given in closed form in Proposition 1 and it involves the principal branch $W_0(\cdot)$ of the Lambert's-W function. Although $W_0(\cdot)$ does not have a closed form, it is a well studied function and the function has been already implemented in many softwares, including in R [36] in the `LambertW` package by [26]. This is an advantage of the $\mathcal{SNTH}$ distribution over the multivariate g-and-h distribution in the sense that the inverse of the Tukey g-and-h transformation is not in a closed form. As a result, the computation of the pdf and the log-likelihood function of the $\mathcal{SNTH}$ distribution is somewhat simpler compared to that of the multivariate g-and-h distribution.

To illustrate the effects of the skewness and the tail-thickness parameters of the $\mathcal{SNTH}$ distribution, we present the contour plots of $\mathcal{SNTH}_2(\boldsymbol{0}, \mathrm{diag}(1,1), \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$ probability densities with $\bar{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$ for four different pairs of $\boldsymbol{\eta}$ and $\boldsymbol{h}$: $\boldsymbol{\eta} = (0,0)^\top$ and $\boldsymbol{h} = (0,0)^\top$ corresponding to a normal density; $\boldsymbol{\eta} = (0,0)^\top$ and $\boldsymbol{h} = (0.05, 0.1)^\top$ corresponding to a normal-Tukey-h density; $\boldsymbol{\eta} = (-1,2)^\top$ and $\boldsymbol{h} = (0,0)^\top$ corresponding to a $\mathcal{SN}$ density; and $\boldsymbol{\eta} = (-1,2)^\top$ and $\boldsymbol{h} = (0.05, 0.1)^\top$ corresponding to a $\mathcal{SNTH}$ density. For comparison we also plot the density contours of a skew-t distribution with $\boldsymbol{\xi} = (0,0)^\top$, $\boldsymbol{\Omega} = \begin{pmatrix} 2 & -1.6 \\ -1.6 & 5 \end{pmatrix}$, $\boldsymbol{\alpha} = (-1.02, 2.15)^\top$, and $\nu = 5$, and a Student's-t distribution with these same parameters (i.e., the same skew-t with $\boldsymbol{\alpha} = \boldsymbol{0}$). The $\boldsymbol{\Omega}$ and $\boldsymbol{\alpha}$ parameters are obtained so that they correspond to $\bar{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$ and $\boldsymbol{\eta} = (-1,2)^\top$ using the relationship between the parameters of the $\mathcal{ASN}$ and the $\mathcal{SN}$ parameterizations. All the density contours are plotted in Fig. 1. The contours are drawn for the levels with approximate coverage probabilities 0.05, 0.25, 0.5, 0.75, and 0.95. The density contour plots in the first row correspond to the density contours of the second row when the corresponding skewness parameters are set to zero. Although the contours in the first row are all symmetric, their symmetry differs from each other. More precisely, in Fig. 1, the normal and the Student's-t probability contours are centrally symmetric whereas the normal-Tukey-h probability contours are sign-invariant symmetric, which is a special case of central symmetry. It can be concluded from Fig. 1 that the shapes of the Student's-t and skew-t density contours are similar to that of the normal and the skew-normal densities, respectively, with more spacing in-between the different levels for the formers due to thicker tails. The contours of the normal-Tukey-h density and the $\mathcal{SNTH}$ density look similar to the normal and the skew-normal density contours, respectively, but the former have been stretched along the two axes. Since the extent of this stretching can be different along the two axes, the $\mathcal{SNTH}$ density contours can represent a variety of shapes with changes in the skewness and the tail-thickness parameters.

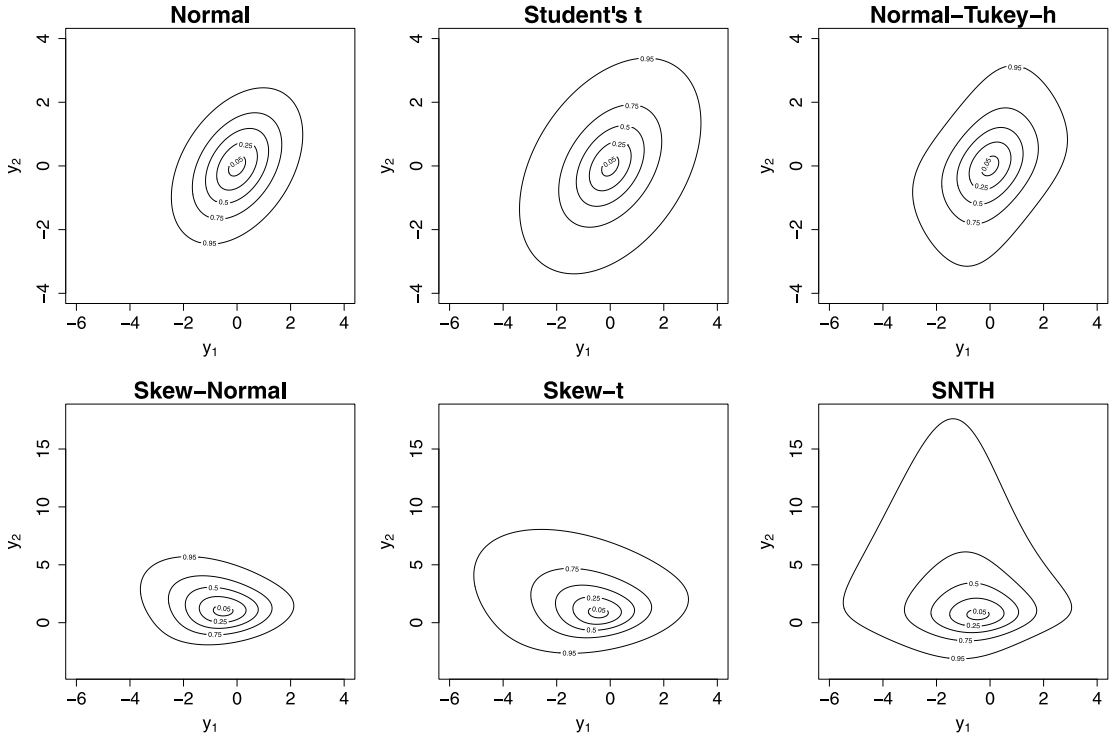


Fig. 1. Bivariate probability density contours of various distributions. Contours are given so that their coverage probabilities are approximately 0.05, 0.25, 0.5, 0.75, and 0.95.

3.2. Cumulative distribution function of $\mathcal{SNTH}$

The cdf of the $\mathcal{SNTH}$ distribution can be obtained in closed form involving the principal branch $W_0(\cdot)$ of the Lambert's-W function as shown next.

Proposition 2. *The cdf of $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$ is $F_{\boldsymbol{Y}}(\boldsymbol{y}) = 2\Phi_{p+1}(\boldsymbol{y}_{**}; \boldsymbol{0}, \boldsymbol{\Omega}_{**})$ where Φ_{p+1} is the multivariate Gaussian cdf of dimension $p+1$, $\boldsymbol{y}_{**} = \left\{ \tau_{h_1}^{-1}\left(\frac{y_1-\xi_1}{\omega_{11}}\right), \dots, \tau_{h_p}^{-1}\left(\frac{y_p-\xi_p}{\omega_{pp}}\right), 0 \right\}^\top$ and $\boldsymbol{\Omega}_{**} = \begin{pmatrix} \bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top & -\boldsymbol{\eta} \\ -\boldsymbol{\eta}^\top & 1 \end{pmatrix}$.*

Proof. Let $\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{Z})$, where $\boldsymbol{Z} \sim \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta})$. Then the cdf of $\boldsymbol{Y}$ is

$$\begin{aligned} F_{\boldsymbol{Y}}(\boldsymbol{y}) &= \Pr(Y_1 \le y_1, \dots, Y_p \le y_p) = \Pr\left[Z_1 \le \tau_{h_1}^{-1}\left(\frac{y_1-\xi_1}{\omega_{11}}\right), \dots, Z_p \le \tau_{h_p}^{-1}\left(\frac{y_p-\xi_p}{\omega_{pp}}\right) \right] \\ &= F_{\boldsymbol{Z}}\left\{ \tau_{h_1}^{-1}\left(\frac{y_1-\xi_1}{\omega_{11}}\right), \dots, \tau_{h_p}^{-1}\left(\frac{y_p-\xi_p}{\omega_{pp}}\right) \right\} = 2\Phi_{p+1}(\boldsymbol{y}_{**}; \boldsymbol{0}, \boldsymbol{\Omega}_{**}), \quad \boldsymbol{y} \in \mathbb{R}^p, \end{aligned}$$

where $\tau_h^{-1}(z)$ is given in (6). The cdf of $\boldsymbol{Z}$, $F_{\boldsymbol{Z}}(\cdot)$, is obtained using Proposition 12 of [34]. □

3.3. Marginal distributions of $\mathcal{SNTH}$

Similar to the $\mathcal{SN}$ distribution, the marginals of the $\mathcal{SNTH}$ distribution are also from the same family, as shown in the next proposition.

Proposition 3. *Let $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$ and consider the partition $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top$ with $\boldsymbol{Y}_i$ of size p_i ($i \in \{1,2\}$) and such that $p_1 + p_2 = p$, with corresponding partitions of the parameters in blocks of matching sizes, as follows:*

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\omega}_{22} \end{pmatrix}, \bar{\boldsymbol{\Psi}} = \begin{pmatrix} \bar{\boldsymbol{\Psi}}_{11} & \bar{\boldsymbol{\Psi}}_{12} \\ \bar{\boldsymbol{\Psi}}_{21} & \bar{\boldsymbol{\Psi}}_{22} \end{pmatrix}, \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix}, \boldsymbol{h} = \begin{pmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{pmatrix}.$$

Then $\boldsymbol{Y}_i \sim \mathcal{SNTH}_{p_i}(\boldsymbol{\xi}_i, \boldsymbol{\omega}_{ii}, \bar{\boldsymbol{\Psi}}_{ii}, \boldsymbol{\eta}_i, \boldsymbol{h}_i)$, $i \in \{1,2\}$.

Proof. Since, $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$, then by definition there exists a random vector $\boldsymbol{Z} \sim \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta})$ such that $\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{Z})$. Consider the partition $\boldsymbol{Z} = (\boldsymbol{Z}_1^\top, \boldsymbol{Z}_2^\top)^\top$, similar to $\boldsymbol{Y}$. Then, $\boldsymbol{Z}_i \sim \mathcal{SN}_{p_i}(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}_{ii}, \boldsymbol{\eta}_i)$, $i \in \{1,2\}$, and $\boldsymbol{Y}_i = \boldsymbol{\xi}_i + \boldsymbol{\omega}_{ii}\boldsymbol{\tau}_{\boldsymbol{h}_i}(\boldsymbol{Z}_i)$. Hence, $\boldsymbol{Y}_i \sim \mathcal{SNTH}_{p_i}(\boldsymbol{\xi}_i, \boldsymbol{\omega}_{ii}, \bar{\boldsymbol{\Psi}}_{ii}, \boldsymbol{\eta}_i, \boldsymbol{h}_i)$, $i \in \{1,2\}$. □

Although the marginals of the $\mathcal{SNTH}$ remain in the same family, the same cannot be said for any general affine transformation of the $\mathcal{SNTH}$ distribution. The distribution of an arbitrary affine transformation of a $\mathcal{SNTH}$ random vector is not of a known type.

3.4. Mean and variance–covariance of $\mathcal{SNTH}$

The mean vector and the variance–covariance matrix of the $\mathcal{SNTH}$ distribution can be obtained in closed form. The next proposition presents these results.

Proposition 4. *Let $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$. The mean vector $\boldsymbol{\mu} = E(\boldsymbol{Y})$ and variance–covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij}) = Var(\boldsymbol{Y})$ are defined by:*

$$\mu_i = \xi_i + \omega_{ii}\sqrt{\frac{2}{\pi}}\frac{\eta_i}{\sqrt{1-h_i}\{1-h_i(1+\eta_i^2)\}}, \quad \text{if } h_i < \frac{1}{1+\eta_i^2},$$

$$\sigma_{ii} = \omega_{ii}^2\left[\frac{1+\eta_i^2}{\{1-2h_i(1+\eta_i^2)\}^{3/2}} - \frac{2}{\pi}\frac{\eta_i^2}{(1-h_i)\{1-h_i(1+\eta_i^2)\}^2}\right], \quad \text{if } h_i < \frac{1}{2(1+\eta_i^2)},$$

$$\sigma_{ij} = \omega_i\omega_j\left[\frac{\sqrt{\det(\boldsymbol{A}^{(ij)})}}{\sqrt{\det(\bar{\boldsymbol{\Psi}}_{i,j} + \boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top)}}a_{12}^{(ij)} - \frac{2}{\pi}\frac{\eta_i\eta_j}{\sqrt{(1-h_i)(1-h_j)}\{1-h_i(1+\eta_i^2)\}\{1-h_j(1+\eta_j^2)\}}\right],$$

$$\text{if } \boldsymbol{A}^{(ij)} \text{ is positive definite},$$

where $\boldsymbol{\eta}_{i,j} = (\eta_i, \eta_j)^\top$, $\bar{\boldsymbol{\Psi}}_{i,j} = \begin{pmatrix} 1 & \bar{\Psi}_{ij} \\ \bar{\Psi}_{ij} & 1 \end{pmatrix}$, $\boldsymbol{A}^{(ij)} = \{(\bar{\boldsymbol{\Psi}}_{i,j} + \boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top)^{-1} - \boldsymbol{H}_{i,j}\}^{-1} = \begin{pmatrix} a_{11}^{(ij)} & a_{12}^{(ij)} \\ a_{12}^{(ij)} & a_{22}^{(ij)} \end{pmatrix}$, $\boldsymbol{H}_{i,j} = diag(h_i, h_j)$, $i \neq j$, and $i, j \in \{1, \dots, p\}$.

Proof. Since $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$, then $\boldsymbol{Y}$ can be written as $\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{Z})$, where $\boldsymbol{Z} \sim \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta})$. Then using the fact $Z_i \sim \mathcal{SN}(0, 1, \eta_i)$:

$$\begin{aligned}
\mathrm{E}\{\tau_{h_i}(Z_i)\} &= \int_{\mathbb{R}} x\exp(h_i x^2/2)2\phi(x; 0, 1+\eta_i^2)\Phi\left(\frac{\eta_i x}{\sqrt{1+\eta_i^2}}\right)\mathrm{d}x \\
&= \int_{\mathbb{R}} \frac{\sqrt{1+\eta_i^2}}{\sqrt{1-h_i(1+\eta_i^2)}}t\frac{2}{\sqrt{2\pi}\sqrt{1+\eta_i^2}}\exp(-t^2/2)\Phi\left(\frac{\eta_i t}{\sqrt{1-h_i(1+\eta_i^2)}}\right)\frac{\sqrt{1+\eta_i^2}}{\sqrt{1-h_i(1+\eta_i^2)}}\mathrm{d}t \\
&\quad \left(\text{using the change of variable } t = \left\{\sqrt{1-h_i(1+\eta_i^2)}/\sqrt{1+\eta_i^2}\right\}x\right) \\
&= \frac{\sqrt{1+\eta_i^2}}{1-h_i(1+\eta_i^2)}\mathrm{E}(X_i) \quad \text{with } X_i \sim \mathcal{ASN}\left(0, 1, \eta_i/\sqrt{1-h_i(1+\eta_i^2)}\right) \\
&= \sqrt{\frac{2}{\pi}}\frac{\eta_i}{\sqrt{1-h_i}\{1-h_i(1+\eta_i^2)\}}, \quad h_i(1+\eta_i^2) < 1,\ i \in \{1, \dots, p\};
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[\{\tau_{h_i}(Z_i)\}^2] &= \int_{\mathbb{R}} x^2\exp(h_i x^2)2\phi(x; 0, 1+\eta_i^2)\Phi\left(\frac{\eta_i x}{\sqrt{1+\eta_i^2}}\right)\mathrm{d}x \\
&= \int_{\mathbb{R}} \frac{1+\eta_i^2}{1-2h_i(1+\eta_i^2)}t^2\frac{2}{\sqrt{2\pi}\sqrt{1+\eta_i^2}}\exp(-t^2/2)\Phi\left(\frac{\eta_i t}{\sqrt{1-2h_i(1+\eta_i^2)}}\right)\frac{\sqrt{1+\eta_i^2}}{\sqrt{1-2h_i(1+\eta_i^2)}}\mathrm{d}t \\
&\quad \left(\text{using the change of variable } t = \left\{\sqrt{1-2h_i(1+\eta_i^2)}/\sqrt{1+\eta_i^2}\right\}x\right) \\
&= \frac{1+\eta_i^2}{\{1-2h_i(1+\eta_i^2)\}^{3/2}}\mathrm{E}(X_i^2) \quad \text{with } X_i \sim \mathcal{ASN}\left(0, 1, \eta_i/\sqrt{1-2h_i(1+\eta_i^2)}\right) \\
&= \frac{1+\eta_i^2}{\{1-2h_i(1+\eta_i^2)\}^{3/2}}, \quad 2h_i(1+\eta_i^2) < 1,\ i \in \{1, \dots, p\}.
\end{aligned}$$

Hence:

$$\mathrm{Var}\{\tau_{h_i}(Z_i)\} = \frac{1+\eta_i^2}{\{1-2h_i(1+\eta_i^2)\}^{3/2}} - \frac{2}{\pi}\frac{\eta_i^2}{(1-h_i)\{1-h_i(1+\eta_i^2)\}^2},\ h_i < \frac{1}{2(1+\eta_i^2)},\ i \in \{1,\dots,p\};$$

$$\mathrm{E}\{\tau_{h_i}(Z_i)\tau_{h_j}(Z_j)\} = \int_{\mathbb{R}^2} x_1x_2 \exp\{(h_ix_1^2+h_jx_2^2)/2\}2\phi_2(\boldsymbol{x};\boldsymbol{0},\bar{\boldsymbol{\Psi}}_{i,j}+\boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top)\boldsymbol{\Phi}\left(\frac{\boldsymbol{\eta}_{i,j}^\top\bar{\boldsymbol{\Psi}}_{i,j}^{-1}\boldsymbol{x}}{\sqrt{1+\boldsymbol{\eta}_{i,j}^\top\bar{\boldsymbol{\Psi}}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}}\right)\mathrm{d}\boldsymbol{x}$$

$$= \int_{\mathbb{R}^2} x_1x_2 \frac{\sqrt{\det(\boldsymbol{A}^{(ij)})}}{\sqrt{\det(\bar{\boldsymbol{\Psi}}_{i,j}+\boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top)}}2\phi_2(\boldsymbol{x};\boldsymbol{0},\boldsymbol{A}^{(ij)})\boldsymbol{\Phi}\left(\frac{\boldsymbol{\eta}_{i,j}^\top\bar{\boldsymbol{\Psi}}_{i,j}^{-1}\boldsymbol{\omega}_{\boldsymbol{A}^{(ij)}}\boldsymbol{\omega}_{\boldsymbol{A}^{(ij)}}^{-1}\boldsymbol{x}}{\sqrt{1+\boldsymbol{\eta}_{i,j}^\top\bar{\boldsymbol{\Psi}}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}}\right)\mathrm{d}\boldsymbol{x}$$

$$= \frac{\sqrt{\det(\boldsymbol{A}^{(ij)})}}{\sqrt{\det(\bar{\boldsymbol{\Psi}}_{i,j}+\boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top)}}\mathrm{E}(X_iX_j) = \frac{\sqrt{\det(\boldsymbol{A}^{(ij)})}}{\sqrt{\det(\bar{\boldsymbol{\Psi}}_{i,j}+\boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top)}}\mathrm{E}(X_iX_j)a_{12}^{(ij)},\quad \boldsymbol{A}^{(ij)} \text{ is positive definite},\ i\neq j,\ i,j\in\{1,\dots,p\},$$

where $(X_i,X_j)^\top \sim \mathcal{ASN}_2\left(\boldsymbol{0},\boldsymbol{A}^{(ij)},\boldsymbol{\omega}_{\boldsymbol{A}^{(ij)}}\bar{\boldsymbol{\Psi}}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}/\sqrt{1+\boldsymbol{\eta}_{i,j}^\top\bar{\boldsymbol{\Psi}}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}\right)$ and $\boldsymbol{\omega}_{\boldsymbol{A}^{(ij)}} = \{\mathrm{diag}(\boldsymbol{A}^{(ij)})\}^{1/2}$. The moments related to the $\mathcal{ASN}$ distribution are obtained from Chapter 2 (univariate) and Chapter 5 (multivariate) of [13]. The rest of the proof is straightforward and hence omitted. □

To this point, we have closed-form expressions of the mean vector and the variance–covariance matrix for the $\mathcal{SNTH}$ distribution. However, we cannot have a closed-form expression for its moment generating function or characteristic function. This is because the distribution of any general affine transformation of the $\mathcal{SNTH}$ distribution is not known.

3.5. Marginal skewness and kurtosis of $\mathcal{SNTH}$

Here we discuss some results related to the skewness and kurtosis of the $\mathcal{SNTH}$ distribution. The Mardia's measures of multivariate skewness and kurtosis [32] for the $\mathcal{SNTH}$ distribution cannot be derived in closed form. However, their univariate counterparts can be derived. Similar to the skew-t distribution, the Pearson's measures of skewness and excess-kurtosis are also unbounded for the univariate $\mathcal{SNTH}$ distribution, suggesting that it is also the case in the multivariate setting.

Proposition 5. *The Pearson's measures of skewness and excess-kurtosis of $Y \sim \mathcal{SNTH}_1(0,1,1,\eta,h)$ are $\gamma_1 = \mu_3/\mu_2^{3/2}$ and $\gamma_2 = \mu_4/\mu_2^2 - 3$, where $\mu_2 = Var(Y)$, $\mu_3 = E\{Y-E(Y)\}^3 = E(Y^3) - 3E(Y^2)E(Y) + 2E(Y)^2$, $\mu_4 = E\{Y-E(Y)\}^4 = E(Y^4) - 4E(Y^3)E(Y) + 6E(Y^2)E(Y)^2 - 3E(Y)^4$ with:*

$$E(Y^3) = \sqrt{\frac{2}{\pi}}\frac{(1+\eta^2)^{3/2}}{\{1-3h(1+\eta^2)\}^2}\left[\frac{2\eta^3+3\eta\{1-3h(1+\eta^2)\}}{\{(1+\eta^2)(1-3h)\}^{3/2}}\right],\quad h<\frac{1}{3(1+\eta^2)},$$

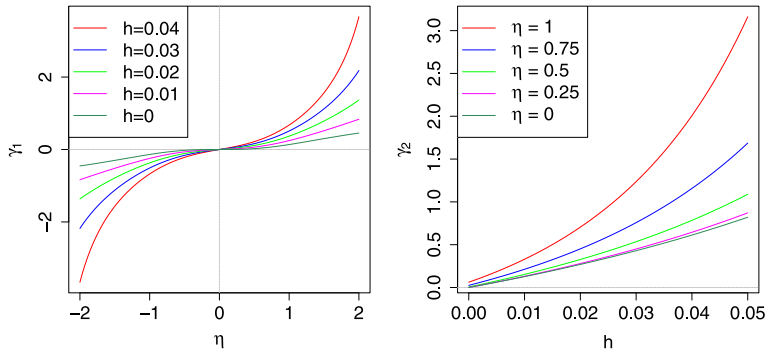
$$E(Y^4) = \frac{3(1+\eta^2)}{\{1-4h(1+\eta^2)\}^{5/2}},\quad h<\frac{1}{4(1+\eta^2)}.$$

Proof. The expressions of $\mathrm{E}(Y)$, $\mathrm{E}(Y^2)$, and $\mathrm{Var}(Y)$ are given in Proposition 4. Since $Y \sim \mathcal{SNTH}(0,1,1,\eta,h)$, we have $Y = \tau_h(Z)$, where $Z \sim \mathcal{SN}(0,1,\eta)$. Hence,

$$\begin{aligned}
\mathrm{E}(Y^3) &= \int_{\mathbb{R}} x^3 \exp(3hx^2/2)2\phi(x;0,1+\eta^2)\Phi\left(\frac{\eta x}{\sqrt{1+\eta^2}}\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{1-3h(1+\eta^2)}}\int_{\mathbb{R}} x^3 2\phi\left(x;0,\frac{(1+\eta^2)}{1-3h(1+\eta^2)}\right)\Phi\left(\frac{\eta}{\{1-3h(1+\eta^2)\}^{1/2}}\frac{(1+\eta^2)^{-1/2}}{\{1-3h(1+\eta^2)\}^{-1/2}}x\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{1-3h(1+\eta^2)}}\mathrm{E}(X^3)\quad \text{with } X \sim \mathcal{ASN}\left(0,\frac{(1+\eta^2)}{1-3h(1+\eta^2)},\frac{\eta}{\{1-3h(1+\eta^2)\}^{1/2}}\right) \\
&= \sqrt{\frac{2}{\pi}}\frac{(1+\eta^2)^{3/2}}{\{1-3h(1+\eta^2)\}^2}\left[\frac{2\eta^3+3\eta\{1-3h(1+\eta^2)\}}{\{(1+\eta^2)(1-3h)\}^{3/2}}\right],\quad h<\frac{1}{3(1+\eta^2)},
\end{aligned}$$

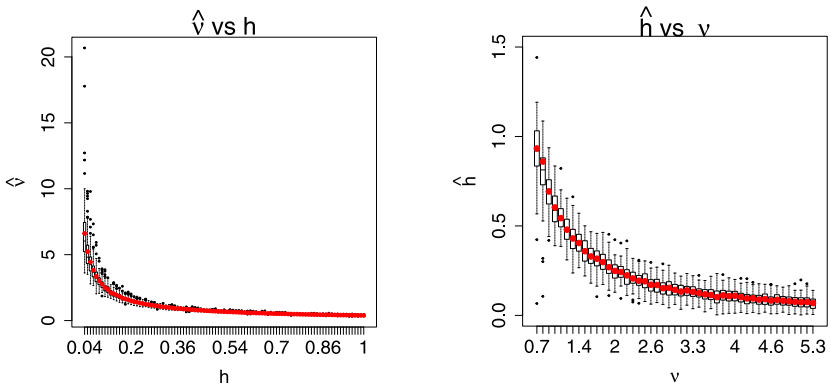
and

$$\begin{aligned}
\mathrm{E}(Y^4) &= \int_{\mathbb{R}} x^4 \exp(2hx^2)2\phi(x;0,1+\eta^2)\Phi\left(\frac{\eta x}{\sqrt{1+\eta^2}}\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{1-4h(1+\eta^2)}}\int_{\mathbb{R}} x^4 2\phi\left(x;0,\frac{(1+\eta^2)}{1-4h(1+\eta^2)}\right)\Phi\left\{\frac{\eta}{\{1-4h(1+\eta^2)\}^{1/2}}\frac{(1+\eta^2)^{-1/2}}{\{1-4h(1+\eta^2)\}^{-1/2}}x\right\}\mathrm{d}x \\
&= \frac{1}{\sqrt{1-4h(1+\eta^2)}}\mathrm{E}(X^4)\quad \text{with } X \sim \mathcal{ASN}\left(0,\frac{(1+\eta^2)}{1-4h(1+\eta^2)},\frac{\eta}{\{1-3h(1+\eta^2)\}^{1/2}}\right)
\end{aligned}$$



(a) Plots of γ_1 against η for different fixed h **(b)** Plots of γ_2 against h for different fixed η

Fig. 2. Plots of the measures of skewness and kurtosis for the $\mathcal{SNTH}_1(0,1,1,\eta,h)$ distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Boxplots of $\hat{\nu}$ against the true h parameter for data simulated from $\mathcal{SNTH}$ **(b)** Boxplots of $\hat{h}$ against the true ν parameter for data simulated from skew-t

Fig. 3. Boxplots of estimated ν parameter against the true h parameter in (a) and estimated h against the true ν parameter in (b). The red dots in each plot correspond to the means of the estimates based on 100 replicates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$= \frac{3(1+\eta^2)}{\{1-4h(1+\eta^2)\}^{5/2}}, \quad h < \frac{1}{4(1+\eta^2)}.$$

The 3rd and 4th order moments of the $\mathcal{ASN}$ distribution are obtained from Chapter 2 of [13]. □

We provide plots of the γ_1 and γ_2 measures for the $\mathcal{SNTH}_1(0,1,1,\eta,h)$ distribution against η and h for different fixed h and η, respectively, in Fig. 2. From the plots, it is clear that the parameter η dictates the extent of skewness in the distribution. Moreover, for a fixed η, the extent of skewness increases with increase in h and vice-versa. Similarly, the extent of the tail-thickness is dictated by the parameter h and for a fixed h, the tail-thickness increases with increase in η and vice-versa. Here we only plot γ_2 against h for positive η as γ_2 is only a function of η^2. The plots show how the effect of η and h on skewness and kurtosis are intertwined. Nevertheless, we associate the parameter η with the skewness and the parameter h with the tail-thickness of the $\mathcal{SNTH}$ distribution. It is also worth pointing out from the plots that the γ_2 measure cannot be less than zero for the $\mathcal{SNTH}$ distribution. Hence, the $\mathcal{SNTH}$ distribution is not suitable for scenarios when tail-thickness of the data is less than that of the Gaussian distribution.

The h parameter of the $\mathcal{SNTH}$ distribution is the counterpart of the ν parameter of the skew-t distribution since these two parameters primarily control the tail-thickness in their respective distributions. The relationship between h and ν is studied here using two simulation experiments. In the first experiment we simulate 500 realizations from $\mathcal{SNTH}_1(0,1,1,1.5,h)$, where h varies in the interval $[0.02,1]$. We fit the skew-t distribution to the simulated $\mathcal{SNTH}$ data for varying h with the R [36] package sn [9] and note the estimate of ν. For each h, we repeat this experiment 100 times and present the boxplots of ν estimates as a function of h in Fig. 3(a). Moreover, the estimates' means are indicated by the red dots. Similar experiment results are provided in Fig. 3(b), where we present the boxplots of the 100 estimates of h obtained by fitting the $\mathcal{SNTH}$ distribution to 100 replicates of size 500 from the skew-t distribution with location, scale, and skewness parameter as 0, 1, and 1.5, with varying degrees of freedom $\nu \in [0.7, 5.3]$. From the two boxplots in Fig. 3, we can see how the two tail-thickness parameters of the $\mathcal{SNTH}$ and the skew-t are related. As ν

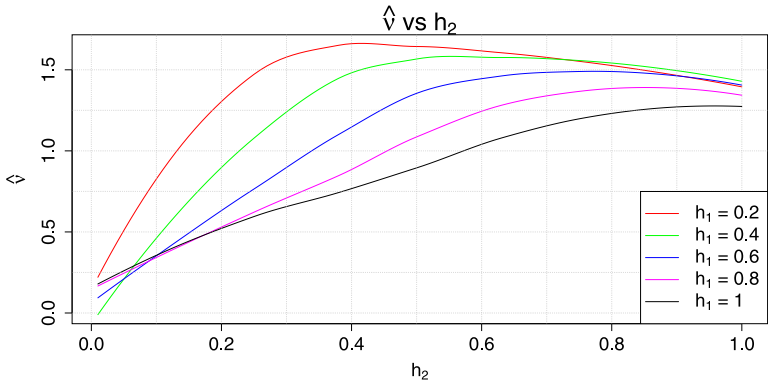


Fig. 4. Curves obtained by smoothing the median of $\hat{\nu}$s from fitted bivariate skew-t based on 100 replicates from $\mathcal{SNTH}_2$ as a function of true h_2 for different values of h_1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in the skew-t distribution increases, the kurtosis decreases and that corresponds to the decrease in h in the $\mathcal{SNTH}$ distribution and vice-versa.

A similar experiment in the bivariate case yields some interesting results. In this experiment, we simulate 500 realizations from $\mathcal{SNTH}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \mathbf{I}_2, \begin{pmatrix}1 & 0.3\\0.3 & 1\end{pmatrix}, \begin{pmatrix}-1.5\\2\end{pmatrix}, \begin{pmatrix}h_1\\h_2\end{pmatrix}\right)$ with varying $h_2 \in [0.01, 1]$, for fixed $h_1 \in \{0.2, 0.4, 0.6, 0.8, 1\}$. We fit a bivariate skew-t distribution to the $\mathcal{SNTH}$ data and note the estimate of ν. Based on 100 replicates, we plot the median of $\hat{\nu}$s against h_2 for different h_1 in Fig. 4. Moreover, we smooth the curve using local polynomial fitting. From this plot we see that a particular $\hat{\nu}$ can be obtained for different pairs of h_1 and h_2. For instance, the line $\hat{\nu} = 1$ cuts all the curves in the plot. From here we conclude that the skew-t distribution is not suitable for scenarios when there is a great disparity between marginal kurtosis. When h_2 is very small and h_1 is large, the skew-t model puts more emphasis on h_1 and the overall estimate of ν in that case becomes small, which corresponds to heavier tail in the fitted distribution. As h_2 increases, the true distribution becomes more heavy-tailed but the fitted distribution becomes less heavy-tailed.

3.6. Conditional distribution of $\mathcal{SNTH}$

Before deriving the conditional distribution of the $\mathcal{SNTH}$ family, we first discuss the result about the conditional distribution of the $\mathcal{SN}$ distribution. To do that, we need to revisit the family of the extended skew-normal distribution [4,6,8,19] but with the $\boldsymbol{\Psi}$-$\boldsymbol{\eta}$ parameterization, similar to the definition of the $\mathcal{SN}$ distribution in Section 2.1. A p-variate random vector $\boldsymbol{Y}$ has an extended skew-normal distribution if its pdf is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{\Phi(\tau)}\phi_p(\boldsymbol{y}; \boldsymbol{\xi} + \tau\boldsymbol{\eta}, \boldsymbol{\Psi} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)\Phi\left\{\frac{\tau + \boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{y} - \boldsymbol{\xi})}{\sqrt{1 + \boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}}}\right\}, \quad \boldsymbol{y} \in \mathbb{R}^p, \tag{7}$$

where $\boldsymbol{\xi} \in \mathbb{R}^p$ is the location parameter, $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$ is the symmetric positive definite scale matrix, $\boldsymbol{\eta} \in \mathbb{R}^p$ is the skewness parameter, and $\tau \in \mathbb{R}$ is the extension parameter. We denote $\boldsymbol{Y} \sim \mathcal{ESN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta}, \tau)$. From the pdf of the $\mathcal{ESN}$ distribution in (7) we have, when this extension parameter $\tau = 0$, that the $\mathcal{ESN}$ distribution reduces to the $\mathcal{SN}$ distribution. Like the $\mathcal{SN}$ distribution, a random vector $\boldsymbol{Y} \sim \mathcal{ESN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta}, \tau)$ also has a concise stochastic representation

$$\boldsymbol{Y} = \boldsymbol{\xi} + \tau\boldsymbol{\eta} + \boldsymbol{\eta}U + \boldsymbol{W}, \tag{8}$$

where $U \overset{d}{=} (Z|Z + \tau > 0)$, $Z \sim \mathcal{N}(0, 1)$, $\boldsymbol{W} \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Psi})$, and Z and $\boldsymbol{W}$ are independently distributed. The last statement is directly obtained from Proposition 1 of [6] (see (10) there with $\nu \to \infty$). As a consequence of this stochastic representation, the marginals of the $\mathcal{ESN}$ distribution also remain in the same family and the parameters of the marginal distribution are just the corresponding marginal parameters, similar to the $\mathcal{SN}$ distribution. We need this definition of the $\mathcal{ESN}$ distribution because the conditionals of the $\mathcal{SN}$ family belong to the $\mathcal{ESN}$ family.

Let $\boldsymbol{Y} \sim \mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$, and consider the partition of $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top$ with $\boldsymbol{Y}_i$ of size p_i ($i \in \{1, 2\}$) and such that $p_1 + p_2 = p$, with corresponding partitions of the parameters in blocks of matching sizes. Then the conditional distribution of $\boldsymbol{Y}_1$ given $\boldsymbol{Y}_2 = \boldsymbol{y}_2$, $\boldsymbol{y}_2 \in \mathbb{R}^{p_2}$, is

$$(\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2) \sim \mathcal{ESN}_{p_1}(\boldsymbol{\xi}_{1.2}, \bar{\boldsymbol{\Psi}}_{11.2}, \bar{\boldsymbol{\eta}}_{1.2}, \bar{\tau}_{1.2}), \tag{9}$$

where $\boldsymbol{\xi}_{1.2} = \boldsymbol{\xi}_1 + \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\xi}_2)$, $\boldsymbol{\Psi}_{11.2} = \boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}$, $\boldsymbol{\eta}_{1.2} = \boldsymbol{\eta}_1 - \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\eta}_2$, $\bar{\boldsymbol{\eta}}_{1.2} = \boldsymbol{\eta}_{1.2}/\sqrt{1 + \boldsymbol{\eta}_2^\top\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\eta}_2}$, and $\bar{\tau}_{1.2} = \boldsymbol{\eta}_2^\top\boldsymbol{\Psi}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\xi}_2)/\sqrt{1 + \boldsymbol{\eta}_2^\top\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\eta}_2}$. This result can be verified by the fact that the conditional distribution of the $\mathcal{ASN}$ family belongs to the extended skew-normal distribution proposed by [6] (see Section 5.3.2 in [13]) and by reparameterizing to the $\boldsymbol{\Psi}$-$\boldsymbol{\eta}$ parameterization.

In the next proposition we derive the conditional distribution of the $\mathcal{SNTH}$ family. We show that the conditional distributions of the $\mathcal{SN}$ family and the $\mathcal{SNTH}$ family are related.

Proposition 6. *Let $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\xi, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$, and consider the partition of $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top$ with $\boldsymbol{Y}_i$ of size p_i ($i \in \{1,2\}$) and such that $p_1 + p_2 = p$, with corresponding partitions of the parameters in blocks of matching sizes. Then the conditional distribution of $\boldsymbol{Y}_1$ given $\boldsymbol{Y}_2 = \boldsymbol{y}_2$ is*

$$(\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2) \overset{d}{=} \boldsymbol{\tau}_{\boldsymbol{h}_1}(\boldsymbol{Y}_0), \quad \boldsymbol{Y}_0 \sim \mathcal{ESN}_{p_1}(\boldsymbol{\xi}_{1.2}, \bar{\boldsymbol{\Psi}}_{11.2}, \bar{\boldsymbol{\eta}}_{1.2}, \bar{\tau}_{1.2}),$$

where $\boldsymbol{\xi}_{1.2} = \bar{\boldsymbol{\Psi}}_{12}\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{g}_2(\boldsymbol{y}_2)$, $\bar{\boldsymbol{\Psi}}_{11.2} = \bar{\boldsymbol{\Psi}}_{11} - \bar{\boldsymbol{\Psi}}_{12}\bar{\boldsymbol{\Psi}}_{22}^{-1}\bar{\boldsymbol{\Psi}}_{21}$, $\boldsymbol{\tau}_{\boldsymbol{h}_1}(\cdot)$ is the same as in (4), $\boldsymbol{g}(\boldsymbol{y})$ is the same as in (5), $\boldsymbol{g}(\boldsymbol{y}) = \{\boldsymbol{g}_1(\boldsymbol{y}_1), \boldsymbol{g}_2(\boldsymbol{y}_2)\}^\top$ with $\boldsymbol{g}_1(\boldsymbol{y}_1) = \{g_1(y_1), \ldots, g_{p_1}(y_{p_1})\}^\top$ and $\boldsymbol{g}_2(\boldsymbol{y}_2) = \{g_{p_1+1}(y_{p_1+1}), \ldots, g_p(y_p)\}^\top$, $\bar{\boldsymbol{\eta}}_{1.2} = (\boldsymbol{\eta}_1 - \bar{\boldsymbol{\Psi}}_{12}\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{\eta}_2)/\sqrt{1 + \boldsymbol{\eta}_2^\top\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{\eta}_2}$, and $\bar{\tau}_{1.2} = \boldsymbol{\eta}_2^\top\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{g}_2(\boldsymbol{y}_2)/\sqrt{1 + \boldsymbol{\eta}_2^\top\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{\eta}_2}$.

Proof. From Proposition 3, the marginal pdf of $\boldsymbol{Y}_2$ is

$$f_{\boldsymbol{Y}_2}(\boldsymbol{y}_2) = 2\phi_{p_2}\{\boldsymbol{g}_2(\boldsymbol{y}_2); \boldsymbol{0}, \bar{\boldsymbol{\Psi}}_{22} + \boldsymbol{\eta}_2\boldsymbol{\eta}_2^\top\}\Phi\left\{\frac{\boldsymbol{\eta}_2^\top\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{g}_2(\boldsymbol{y}_2)}{\sqrt{1 + \boldsymbol{\eta}_2^\top\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{\eta}_2}}\right\}\prod_{i=p_1+1}^{p}\left\{\frac{1}{\omega_{ii}}\left(\frac{\exp[\frac{1}{2}W_0\{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2\}]}{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2 + \exp[W_0\{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2\}]}\right)\right\}, \quad \boldsymbol{y}_2 \in \mathbb{R}^{p_2}.$$

Hence, the conditional pdf of $\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2$ is, for $y_1 \in \mathbb{R}_1^p$:

$$f_{\boldsymbol{Y}_1|\boldsymbol{Y}_2=\boldsymbol{y}_2}(\boldsymbol{y}_1) = \frac{f_{\boldsymbol{Y}}(\boldsymbol{y})}{f_{\boldsymbol{Y}_2}(\boldsymbol{y}_2)} = \frac{\phi_p\{\boldsymbol{g}(\boldsymbol{y}); \boldsymbol{0}, \bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top\}\Phi\left\{\frac{\boldsymbol{\eta}^\top\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{g}(\boldsymbol{y})}{\sqrt{1 + \boldsymbol{\eta}^\top\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{\eta}}}\right\}}{\phi_{p_2}\{\boldsymbol{g}_2(\boldsymbol{y}_2); \boldsymbol{0}, \bar{\boldsymbol{\Psi}}_{22} + \boldsymbol{\eta}_2\boldsymbol{\eta}_2^\top\}\Phi\left\{\frac{\boldsymbol{\eta}_2^\top\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{g}_2(\boldsymbol{y}_2)}{\sqrt{1 + \boldsymbol{\eta}_2^\top\bar{\boldsymbol{\Psi}}_{22}^{-1}\boldsymbol{\eta}_2}}\right\}}\prod_{i=1}^{p_1}\left\{\frac{1}{\omega_{ii}}\left(\frac{\exp[\frac{1}{2}W_0\{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2\}]}{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2 + \exp[W_0\{h_i(\frac{y_i-\xi_i}{\omega_{ii}})^2\}]}\right)\right\}.$$

From the pdf given above, we can see that it is the density function of $\boldsymbol{\tau}_{\boldsymbol{h}_1}(\boldsymbol{Y}_0)$, where $\boldsymbol{Y}_0 \overset{\mathrm{d}}{=} [\boldsymbol{Z}_1|\{\boldsymbol{Z}_2 = \boldsymbol{g}_2(\boldsymbol{y}_2)\}]$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1^\top, \boldsymbol{Z}_2^\top)^\top \sim \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta})$. Hence, from (9), we have $\boldsymbol{Y}_0 \sim \mathcal{ESN}_{p_1}(\boldsymbol{\xi}_{1.2}, \bar{\boldsymbol{\Psi}}_{11.2}, \bar{\boldsymbol{\eta}}_{1.2}, \bar{\tau}_{1.2})$. □

Since the conditional distribution of the $\mathcal{SNTH}$ family can be viewed as a component-wise Tukey-h transformation on the $\mathcal{ESN}$, closed-form expressions of its mean vector and variance–covariance matrix can be derived. The conditional mean and the variance–covariance matrix will be helpful for using the $\mathcal{SNTH}$ model for various formal statistical purposes such as regression modeling, time-series analysis, and spatial modeling. In the next three propositions we provide the mathematical expressions of the elements of the conditional mean vector and the conditional variance–covariance matrix.

Proposition 7. *Let $\boldsymbol{Y}_0$ be defined as in Proposition 6. The mean vector $\boldsymbol{\mu} = E\{\boldsymbol{\tau}_{\boldsymbol{h}_1}(\boldsymbol{Y}_0)\}$ is:*

$$\mu_i = \frac{1}{\sqrt{1 - (\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i}}\exp\left\{\frac{(\xi_{1.2_i} + \bar{\tau}_{1.2}\bar{\eta}_{1.2_i})^2 h_i}{2(1 - (\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i)}\right\}\frac{\Phi(\tilde{\tau}_i)}{\Phi(\bar{\tau}_{1.2})}\left\{\tilde{\xi}_i + \tilde{\omega}_i\tilde{\delta}_i\frac{\phi(\tilde{\tau}_i)}{\Phi(\tilde{\tau}_i)}\right\},$$

where $\boldsymbol{\xi}_{1.2} = (\xi_{1.2_1}, \ldots, \xi_{1.2_{p_1}})^\top$, $diag(\bar{\boldsymbol{\Psi}}_{11.2}) = (\bar{\Psi}_{11.2_{11}}, \ldots, \bar{\Psi}_{11.2_{p_1p_1}})^\top$, $\bar{\boldsymbol{\eta}}_{1.2} = (\bar{\eta}_{1.2_1}, \ldots, \bar{\eta}_{1.2_{p_1}})^\top$, $\tilde{\xi}_i = \frac{\xi_{1.2_i} + \bar{\tau}_{1.2}\bar{\eta}_{1.2_i}}{1 - (\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i}$, $\tilde{\omega}_i = \sqrt{\frac{\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2}{1 - (\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i}}$, $\tilde{\alpha}_i = \frac{\bar{\eta}_{1.2_i}}{\sqrt{\bar{\Psi}_{11.2_{ii}}}}\frac{1}{\sqrt{1 - (\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i}}$, $\tilde{\alpha}_{0_i} = \frac{\bar{\tau}_{1.2}\sqrt{\bar{\Psi}_{11.2_{ii}}} + \frac{\bar{\eta}_{1.2_i}}{\sqrt{\bar{\Psi}_{11.2_{ii}}}}\left\{\frac{\bar{\tau}_{1.2}\bar{\eta}_{1.2_i} + \xi_{1.2_i}(\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i}{1 - (\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i}\right\}}{\sqrt{\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2}}$, $\tilde{\delta}_i = \frac{\tilde{\alpha}_i}{\sqrt{1 + \tilde{\alpha}_i^2}}$, $\tilde{\tau}_i = \frac{\tilde{\alpha}_{0_i}}{\sqrt{1 + \tilde{\alpha}_i^2}}$, $h_i < \frac{1}{\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2}$, $i \in \{1, \ldots, p_1\}$.

Proof. From (8) it can be established that $Y_{0_i} \sim \mathcal{ESN}_1(\xi_{1.2_i}, \bar{\Psi}_{11.2_{ii}}, \bar{\eta}_{1.2_i}, \bar{\tau}_{1.2})$, $i \in \{1, \ldots, p_1\}$. Then:

$$\begin{aligned}\mu_i = \mathrm{E}(Y_{0_i}) &= \int_{\mathbb{R}} x\exp(h_i x^2/2)\frac{1}{\Phi(\bar{\tau}_{1.2})}\phi(x; \xi_{1.2_i} + \bar{\tau}_{1.2}\bar{\eta}_{1.2_i}, \bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)\Phi\left\{\frac{\bar{\tau}_{1.2} + \bar{\eta}_{1.2_i}(x - \xi_{1.2_i})/\bar{\Psi}_{11.2_{ii}}}{\sqrt{1 + \bar{\eta}_{1.2_i}^2/\bar{\Psi}_{11.2_{ii}}}}\right\}\mathrm{d}x \\ &= \exp\left[\frac{(\xi_{1.2_i} + \bar{\tau}_{1.2}\bar{\eta}_{1.2_i})^2 h_i}{2\{1 - (\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)h_i\}}\right]\frac{1}{\Phi(\bar{\tau}_{1.2})}\frac{1}{\sqrt{2\pi}\sqrt{\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2}}\end{aligned}$$

$$\times \int_{\mathbb{R}} x \exp\left[-\frac{1}{2}\frac{\left\{x-\frac{\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i}}{1-(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}^2}{\left\{\frac{\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i}}{1-(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}}\right]\Phi\left\{\frac{\bar{\tau}_{1.2}+\bar{\eta}_{1.2_i}(x-\xi_{1.2_i})/\bar{\Psi}_{11.2_{ii}}}{\sqrt{1+\bar{\eta}^2_{1.2_i}/\bar{\Psi}_{11.2_{ii}}}}\right\}\mathrm{d}x$$

$$= \exp\left\{\frac{(\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i})^2 h_i}{2(1-(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i)}\right\}\frac{1}{\sqrt{1-(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}\frac{\Phi(\tilde{\tau}_i)}{\Phi(\bar{\tau}_{1.2})}\int_{\mathbb{R}} x\frac{1}{\Phi(\tilde{\tau}_i)}\phi(x;\tilde{\xi}_i,\tilde{\omega}_i^2)\Phi\{\tilde{\alpha}_{0_i}+\tilde{\alpha}_i\tilde{\omega}_i^{-1}(x-\tilde{\xi}_i)\}\mathrm{d}x$$

$$= \frac{1}{\sqrt{1-(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}\exp\left\{\frac{(\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i})^2 h_i}{2(1-(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i)}\right\}\frac{\Phi(\tilde{\tau}_i)}{\Phi(\bar{\tau}_{1.2})}\left\{\tilde{\xi}_i+\tilde{\omega}_i\tilde{\delta}_i\frac{\phi(\tilde{\tau}_i)}{\Phi(\tilde{\tau}_i)}\right\}.$$

The last step is obtained from the moments of the extended skew-normal distribution from [13] (see Section 5.3.4). □

Proposition 8. *Let $\boldsymbol{Y}_0$ be defined as in Proposition 6, and let $\boldsymbol{\Sigma}=(\sigma_{ij})=Var\{\boldsymbol{\tau}_{\boldsymbol{h}_1}(\boldsymbol{Y}_0)\}$. Then:*

$$\sigma_{ii} = \frac{1}{\sqrt{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}\exp\left\{\frac{(\xi_{1.2_i}+\tau\bar{\eta}_{1.2_i})^2 h_i}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}\frac{\Phi(\tilde{\tau}_i)}{\Phi(\bar{\tau}_{1.2})}\left\{\tilde{\xi}_i^2+\tilde{\omega}_i^2-\tilde{\tau}_i\frac{\phi(\tilde{\tau}_i)}{\Phi(\tilde{\tau}_i)}\tilde{\omega}_i^2\tilde{\delta}_i^2+2\frac{\phi(\tilde{\tau}_i)}{\Phi(\tilde{\tau}_i)}\tilde{\xi}_i\tilde{\omega}_i\tilde{\delta}_i\right\}-\mu_i^2,$$

where $\boldsymbol{\xi}_{1.2} = (\xi_{1.2_1},\dots,\xi_{1.2_{p_1}})^\top$, $diag(\bar{\boldsymbol{\Psi}}_{11.2}) = (\bar{\Psi}_{11.2_{11}},\dots,\bar{\Psi}_{11.2_{p_1p_1}})^\top$, $\bar{\boldsymbol{\eta}}_{1.2} = (\bar{\eta}_{1.2_1},\dots,\bar{\eta}_{1.2_{p_1}})^\top$, $\tilde{\xi}_i = \frac{\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i}}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}$, $\tilde{\omega}_i = \sqrt{\frac{\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i}}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}$, $\tilde{\alpha}_i = \frac{\bar{\eta}_{1.2_i}}{\sqrt{\bar{\Psi}_{11.2_{ii}}}}\frac{1}{\sqrt{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}$, $\tilde{\alpha}_{0_i} = \frac{\bar{\tau}_{1.2}\sqrt{\bar{\Psi}_{11.2_{ii}}}+\frac{\bar{\eta}_{1.2_i}}{\sqrt{\bar{\Psi}_{11.2_{ii}}}}\left\{\frac{\bar{\tau}_{1.2}\bar{\eta}_{1.2_i}+2\xi_{1.2_i}(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}}{\sqrt{\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i}}}$, $\tilde{\delta}_i = \frac{\tilde{\alpha}_i}{\sqrt{1+\tilde{\alpha}_i^2}}$, $\tilde{\tau}_i = \frac{\tilde{\alpha}_{0_i}}{\sqrt{1+\tilde{\alpha}_i^2}}$, *$\mu_i$ is the same as in Proposition 7, and $h_i < \frac{1}{2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})}$, $i\in\{1,\dots,p_1\}$.*

Proof. We have:

$$\mathrm{E}(Y_{0_i}^2) = \int_{\mathbb{R}} x^2\exp(h_i x^2)\frac{1}{\Phi(\bar{\tau}_{1.2})}\phi(x;\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i},\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})\Phi\left\{\frac{\bar{\tau}_{1.2}+\bar{\eta}_{1.2_i}(x-\xi_{1.2_i})/\bar{\Psi}_{11.2_{ii}}}{\sqrt{1+\bar{\eta}^2_{1.2_i}/\bar{\Psi}_{11.2_{ii}}}}\right\}\mathrm{d}x$$

$$= \exp\left\{\frac{(\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i})^2 h_i}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}\frac{1}{\Phi(\bar{\tau}_{1.2})}\frac{1}{\sqrt{2\pi}\sqrt{\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i}}}$$

$$\times \int_{\mathbb{R}} x^2\exp\left[-\frac{1}{2}\frac{\left\{x-\frac{\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i}}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}^2}{\frac{\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i}}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}\right]\Phi\left\{\frac{\bar{\tau}_{1.2}+\bar{\eta}_{1.2_i}(x-\xi_{1.2_i})/\bar{\Psi}_{11.2_{ii}}}{\sqrt{1+\bar{\eta}^2_{1.2_i}/\bar{\Psi}_{11.2_{ii}}}}\right\}\mathrm{d}x$$

$$= \exp\left\{\frac{(\xi_{1.2_i}+\bar{\tau}_{1.2}\bar{\eta}_{1.2_i})^2 h_i}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}\frac{1}{\sqrt{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}\frac{\Phi(\tilde{\tau}_i)}{\Phi(\bar{\tau}_{1.2})}\int_{\mathbb{R}} x^2\frac{1}{\Phi(\tilde{\tau}_i)}\phi(x;\tilde{\xi}_i,\tilde{\omega}_i^2)\Phi\{\tilde{\alpha}_{0_i}+\tilde{\alpha}_i\tilde{\omega}_i^{-1}(x-\tilde{\xi}_i)\}\mathrm{d}x$$

$$= \frac{1}{\sqrt{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}}\exp\left\{\frac{(\xi_{1.2_i}+\tau\bar{\eta}_{1.2_i})^2 h_i}{1-2(\bar{\Psi}_{11.2_{ii}}+\bar{\eta}^2_{1.2_i})h_i}\right\}\frac{\Phi(\tilde{\tau}_i)}{\Phi(\bar{\tau}_{1.2})}\left\{\tilde{\xi}_i^2+\tilde{\omega}_i^2-\tilde{\tau}_i\frac{\phi(\tilde{\tau}_i)}{\Phi(\tilde{\tau}_i)}\tilde{\omega}_i^2\tilde{\delta}_i^2+2\frac{\phi(\tilde{\tau}_i)}{\Phi(\tilde{\tau}_i)}\tilde{\xi}_i\tilde{\omega}_i\tilde{\delta}_i\right\},$$

The last step is obtained from the moments of the extended skew-normal distribution from [13] (see Section 5.3.4). □

Proposition 9. *Let $\boldsymbol{Y}_0$ be defined as in Proposition 6, and let $\boldsymbol{\Sigma}=(\sigma_{ij})=Var\{\boldsymbol{\tau}_{\boldsymbol{h}_1}(\boldsymbol{Y}_0)\}$. Then:*

$$\sigma_{ij} = \frac{\sqrt{\det\{(\boldsymbol{\Omega}_{i,j}^{-1}-\boldsymbol{H}_{i,j})^{-1}\}}}{\sqrt{\det(\boldsymbol{\Omega}_{i,j})}}\exp\left[-\frac{1}{2}\{\tilde{\boldsymbol{\mu}}_{i,j}^\top\boldsymbol{\Omega}_{i,j}^{-1}\tilde{\boldsymbol{\mu}}_{i,j}-\tilde{\boldsymbol{\mu}}_{i,j}^\top(\boldsymbol{\Omega}_{i,j}-\boldsymbol{\Omega}_{i,j}\boldsymbol{H}_{i,j}\boldsymbol{\Omega}_{i,j})^{-1}\tilde{\boldsymbol{\mu}}_{i,j}\}\right]$$

$$\times\frac{\Phi(\tilde{\tau}_{i,j})}{\Phi(\bar{\tau}_{1.2})}\left\{(\tilde{\boldsymbol{\Omega}}_{i,j})_{12}-\tilde{\tau}_{i,j}\frac{\phi(\tilde{\tau}_{i,j})}{\Phi(\tilde{\tau}_{i,j})}(\tilde{\boldsymbol{\omega}}_{i,j})_{11}(\tilde{\boldsymbol{\omega}}_{i,j})_{22}(\tilde{\boldsymbol{\delta}}_{i,j})_1(\tilde{\boldsymbol{\delta}}_{i,j})_2+\xi_{1.2_i}\xi_{1.2_j}\right.$$

$$\left.+\frac{\phi(\tilde{\tau}_{i,j})}{\Phi(\tilde{\tau}_{i,j})}\xi_{1.2_i}(\tilde{\boldsymbol{\omega}}_{i,j})_{22}(\tilde{\boldsymbol{\delta}}_{i,j})_2+\frac{\phi(\tilde{\tau}_{i,j})}{\Phi(\tilde{\tau}_{i,j})}\xi_{1.2_j}(\tilde{\boldsymbol{\omega}}_{i,j})_{11}(\tilde{\boldsymbol{\delta}}_{i,j})_1\right\}-\mu_i\mu_j,$$

where $\boldsymbol{\xi}_{i,j} = (\xi_{1.2_i}, \xi_{1.2_j})^\top$, $\boldsymbol{\Psi}_{i,j} = \begin{pmatrix} \bar{\Psi}_{11.2_{ii}} & \bar{\Psi}_{11.2_{ij}} \\ \bar{\Psi}_{11.2_{ij}} & \bar{\Psi}_{11.2_{jj}} \end{pmatrix}$, $\boldsymbol{\eta}_{i,j} = (\bar{\eta}_{1.2_i}, \bar{\eta}_{1.2_j})^\top$, $\boldsymbol{\Omega}_{i,j} = \boldsymbol{\Psi}_{i,j} + \boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top$, $\tilde{\boldsymbol{\mu}}_{i,j} = \boldsymbol{\xi}_{i,j} + \bar{\tau}_{1.2}\boldsymbol{\eta}_{i,j}$, $\boldsymbol{H}_{i,j} = \begin{pmatrix} h_i & 0 \\ 0 & h_j \end{pmatrix}$, $\tilde{\boldsymbol{\xi}}_{i,j} = (\mathbf{I}_2 - \boldsymbol{\Omega}_{i,j}\boldsymbol{H}_{i,j})^{-1}\tilde{\boldsymbol{\mu}}_{i,j}$, $\tilde{\boldsymbol{\Omega}}_{i,j} = (\boldsymbol{\Omega}_{i,j}^{-1} - \boldsymbol{H}_{i,j})^{-1}$, $\tilde{\alpha}_{0_{i,j}} = \frac{\bar{\tau}_{1.2} + \boldsymbol{\eta}_{i,j}^\top \boldsymbol{\Psi}_{i,j}^{-1}(\tilde{\xi}_{i,j} - \xi_{i,j})}{\sqrt{1 + \boldsymbol{\eta}_{i,j}^\top \boldsymbol{\Psi}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}}$, $\tilde{\boldsymbol{\alpha}}_{i,j} = \frac{\tilde{\boldsymbol{\omega}}_{i,j}\boldsymbol{\Psi}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}{\sqrt{1 + \boldsymbol{\eta}_{i,j}^\top \boldsymbol{\Psi}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}}$, $\tilde{\boldsymbol{\omega}}_{i,j} = \{diag(\tilde{\boldsymbol{\Omega}}_{i,j})\}^{1/2}$, $\bar{\tilde{\boldsymbol{\Omega}}}_{i,j} = \tilde{\boldsymbol{\omega}}_{i,j}^{-1}\tilde{\boldsymbol{\Omega}}_{i,j}\tilde{\boldsymbol{\omega}}_{i,j}^{-1}$, $\boldsymbol{\delta}_{i,j} = (1 + \tilde{\boldsymbol{\alpha}}_{i,j}^\top \bar{\tilde{\boldsymbol{\Omega}}}_{i,j}\tilde{\boldsymbol{\alpha}}_{i,j})^{-1/2}\bar{\tilde{\boldsymbol{\Omega}}}_{i,j}\tilde{\boldsymbol{\alpha}}_{i,j}$, μ_i, μ_j *are the same as in Proposition 7, and* $h_i < \frac{1}{2(\bar{\Psi}_{11.2_{ii}} + \bar{\eta}_{1.2_i}^2)}$, $h_j < \frac{1}{2(\bar{\Psi}_{11.2_{jj}} + \bar{\eta}_{1.2_j}^2)}$, $i, j \in \{1, \dots, p_1\}$, $i \neq j$.

Proof. We have, for $\boldsymbol{x}_{i,j} = (x_i, x_j)^\top$:

$$
\begin{aligned}
\mathrm{E}(Y_iY_j) &= \int_{\mathbb{R}^2} x_ix_j \exp(h_ix_i^2/2)\exp(h_jx_j^2/2)\frac{\phi_2(\boldsymbol{x}_{i,j}; \boldsymbol{\xi}_{i,j} + \bar{\tau}_{1.2}\boldsymbol{\eta}_{i,j}, \boldsymbol{\Psi}_{i,j} + \boldsymbol{\eta}_{i,j}\boldsymbol{\eta}_{i,j}^\top)}{\Phi(\bar{\tau}_{1.2})}\Phi\left\{\frac{\bar{\tau}_{1.2} + \boldsymbol{\eta}_{i,j}^\top\boldsymbol{\Psi}_{i,j}^{-1}(\boldsymbol{x}_{i,j} - \boldsymbol{\xi}_{i,j})}{\sqrt{1 + \boldsymbol{\eta}_{i,j}^\top\boldsymbol{\Psi}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}}\right\}\mathrm{d}\boldsymbol{x}_{i,j} \\
&= \frac{\sqrt{\det\{(\boldsymbol{\Omega}_{i,j}^{-1} - \boldsymbol{H}_{i,j})^{-1}\}}}{\sqrt{\det(\boldsymbol{\Omega}_{i,j})}}\exp\left[-\frac{1}{2}\{\tilde{\boldsymbol{\mu}}_{i,j}^\top\boldsymbol{\Omega}_{i,j}^{-1}\tilde{\boldsymbol{\mu}}_{i,j} - \tilde{\boldsymbol{\mu}}_{i,j}^\top(\boldsymbol{\Omega}_{i,j} - \boldsymbol{\Omega}_{i,j}\boldsymbol{H}_{i,j}\boldsymbol{\Omega}_{i,j})^{-1}\tilde{\boldsymbol{\mu}}_{i,j}\}\right] \\
&\quad \times \frac{1}{\Phi(\bar{\tau}_{1.2})}\int_{\mathbb{R}^2} x_ix_j\phi_2\left(\boldsymbol{x}_{i,j}; (\mathbf{I}_2 - \boldsymbol{\Omega}_{i,j}\boldsymbol{H}_{i,j})^{-1}\tilde{\boldsymbol{\mu}}_{i,j}, (\boldsymbol{\Omega}_{i,j}^{-1} - \boldsymbol{H}_{i,j})^{-1}\right)\Phi\left\{\frac{\bar{\tau}_{1.2} + \boldsymbol{\eta}_{i,j}^\top\boldsymbol{\Psi}_{i,j}^{-1}(\boldsymbol{x}_{i,j} - \boldsymbol{\xi}_{i,j})}{\sqrt{1 + \boldsymbol{\eta}_{i,j}^\top\boldsymbol{\Psi}_{i,j}^{-1}\boldsymbol{\eta}_{i,j}}}\right\}\mathrm{d}\boldsymbol{x}_{i,j} \\
&= \frac{\sqrt{\det\{(\boldsymbol{\Omega}_{i,j}^{-1} - \boldsymbol{H}_{i,j})^{-1}\}}}{\sqrt{\det(\boldsymbol{\Omega}_{i,j})}}\exp\left[-\frac{1}{2}\{\tilde{\boldsymbol{\mu}}_{i,j}^\top\boldsymbol{\Omega}_{i,j}^{-1}\tilde{\boldsymbol{\mu}}_{i,j} - \tilde{\boldsymbol{\mu}}_{i,j}^\top(\boldsymbol{\Omega}_{i,j} - \boldsymbol{\Omega}_{i,j}\boldsymbol{H}_{i,j}\boldsymbol{\Omega}_{i,j})^{-1}\tilde{\boldsymbol{\mu}}_{i,j}\}\right] \\
&\quad \times \frac{1}{\Phi(\bar{\tau}_{1.2})}\int_{\mathbb{R}^2} x_ix_j\phi_2(\boldsymbol{x}_{i,j}; \tilde{\boldsymbol{\xi}}_{i,j}, \bar{\tilde{\boldsymbol{\Omega}}}_{i,j})\Phi\{\tilde{\alpha}_{0_{i,j}} + \tilde{\boldsymbol{\alpha}}_{i,j}^\top\tilde{\boldsymbol{\omega}}_{i,j}^{-1}(\boldsymbol{x}_{i,j} - \tilde{\boldsymbol{\xi}}_{i,j})\}\mathrm{d}\boldsymbol{x}_{i,j} \\
&= \frac{\sqrt{\det\{(\boldsymbol{\Omega}_{i,j}^{-1} - \boldsymbol{H}_{i,j})^{-1}\}}}{\sqrt{\det(\boldsymbol{\Omega}_{i,j})}}\exp\left[-\frac{1}{2}\{\tilde{\boldsymbol{\mu}}_{i,j}^\top\boldsymbol{\Omega}_{i,j}^{-1}\tilde{\boldsymbol{\mu}}_{i,j} - \tilde{\boldsymbol{\mu}}_{i,j}^\top(\boldsymbol{\Omega}_{i,j} - \boldsymbol{\Omega}_{i,j}\boldsymbol{H}_{i,j}\boldsymbol{\Omega}_{i,j})^{-1}\tilde{\boldsymbol{\mu}}_{i,j}\}\right] \\
&\quad \times \frac{\Phi(\tilde{\tau}_{i,j})}{\Phi(\bar{\tau}_{1.2})}\left\{(\bar{\tilde{\boldsymbol{\Omega}}}_{i,j})_{12} - \tilde{\tau}_{i,j}\frac{\phi(\tilde{\tau}_{i,j})}{\Phi(\tilde{\tau}_{i,j})}(\tilde{\boldsymbol{\omega}}_{i,j})_{11}(\tilde{\boldsymbol{\omega}}_{i,j})_{22}(\tilde{\boldsymbol{\delta}}_{i,j})_1(\tilde{\boldsymbol{\delta}}_{i,j})_2 + \xi_{1.2_i}\xi_{1.2_j}\right. \\
&\qquad \left. + \frac{\phi(\tilde{\tau}_{i,j})}{\Phi(\tilde{\tau}_{i,j})}\xi_{1.2_i}(\tilde{\boldsymbol{\omega}}_{i,j})_{22}(\tilde{\boldsymbol{\delta}}_{i,j})_2 + \frac{\phi(\tilde{\tau}_{i,j})}{\Phi(\tilde{\tau}_{i,j})}\xi_{1.2_j}(\tilde{\boldsymbol{\omega}}_{i,j})_{11}(\tilde{\boldsymbol{\delta}}_{i,j})_1\right\}.
\end{aligned}
$$

The last step is obtained from the moments of the extended skew-normal distribution from [13] (see Section 5.3.4). □

3.7. Canonical form of the $\mathcal{SNTH}$ distribution

Consider a p-variate random vector $\boldsymbol{X} \sim \mathcal{ASN}_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. It can be shown that there exists a matrix $\boldsymbol{H} \in \mathbb{R}^{p\times p}$ such that $\boldsymbol{H}(\boldsymbol{X} - \boldsymbol{\xi}) \sim \mathcal{ASN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\alpha}^*)$, where $\boldsymbol{\alpha}^* = (\alpha^*, 0, \dots, 0)^\top$, $\alpha^* = \sqrt{\boldsymbol{\alpha}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}}$, and $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$. [18] showed that the matrix $\boldsymbol{H}$ is of the form $\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}$, where $\boldsymbol{Q}$ is obtained from the spectral decomposition of $\boldsymbol{Q}^\top\boldsymbol{\Lambda}\boldsymbol{Q} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Omega}^{-1/2}$, $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Omega} - \frac{2}{\pi}\boldsymbol{\omega}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\omega}$, and $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})^{-1/2}\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}$. The distribution of $\boldsymbol{H}(\boldsymbol{X} - \boldsymbol{\xi})$ is defined as the canonical form of the $\mathcal{ASN}$ distribution.

Similarly, we can define the canonical form of the $\mathcal{SN}$ distribution. Consider a random vector $\boldsymbol{X} \sim \mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$. Using $\boldsymbol{\Omega} = \boldsymbol{\Psi} + \boldsymbol{\eta}\boldsymbol{\eta}^\top$ and $\boldsymbol{\eta} = \boldsymbol{\omega}\boldsymbol{\delta}$, the relations between the parameterizations of the $\mathcal{ASN}$ and the $\mathcal{SN}$, the distribution of $\boldsymbol{H}(\boldsymbol{X} - \boldsymbol{\xi})$ is obtained as $\mathcal{SN}_p(\mathbf{0}, \mathbf{I}_p - \frac{\boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\top}}{1 + \boldsymbol{\alpha}^{*\top}\boldsymbol{\alpha}^*}, \frac{\boldsymbol{\alpha}^*}{\sqrt{1 + \boldsymbol{\alpha}^{*\top}\boldsymbol{\alpha}^*}})$. Hence, the canonical form of the $\mathcal{SN}$ distribution is defined by the distribution of $\boldsymbol{H}^*(\boldsymbol{X} - \boldsymbol{\xi}) \sim \mathcal{SN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\eta}^*)$, where $\boldsymbol{\eta}^* = (\eta^*, 0, \dots, 0)^\top$, $\eta^* = \sqrt{\boldsymbol{\alpha}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}}$, and $\boldsymbol{H}^* = \begin{pmatrix} \sqrt{1 + \boldsymbol{\alpha}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}} & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{I}_{p-1} \end{pmatrix}\boldsymbol{H}$.

The canonical form of the $\mathcal{ASN}$ or the $\mathcal{SN}$ distribution is useful for deriving Mardia's measures of multivariate skewness and kurtosis [32] and the measures of multivariate skewness and kurtosis introduced by [31] since they are invariant under affine transformations of the variable. Moreover, using the canonical form, the unique mode of the $\mathcal{ASN}$ distribution can be derived; see Proposition 5.14 in [13]. Hence, the canonical form is used mainly to reduce the dimensionality of various problems when applicable.

For the $\mathcal{SNTH}$ distribution, we define the canonical form by taking the component-wise Tukey-h transformation of the canonical form of the latent $\mathcal{SN}$ random vector.

Proposition 10. *Suppose* $\boldsymbol{Y} \sim \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$. *We define the canonical form of the* $\mathcal{SNTH}$ *by the distribution of*

$$\boldsymbol{\omega}^{-1}(\boldsymbol{Y}^* - \boldsymbol{\xi}) = \boldsymbol{\tau}_{\boldsymbol{h}}[\boldsymbol{H}^*\boldsymbol{\tau}_{\boldsymbol{h}}^{-1}\{\boldsymbol{\omega}^{-1}(\boldsymbol{Y} - \boldsymbol{\xi})\}] \sim \mathcal{SNTH}_p(\mathbf{0}, \mathbf{I}_p, \mathbf{I}_p, \boldsymbol{\eta}^*, \boldsymbol{h}),$$

where $\boldsymbol{\tau}_{\boldsymbol{h}}^{-1}(\boldsymbol{z}) = \{\tau_{h_1}^{-1}(z_1), \dots, \tau_{h_p}^{-1}(z_p)\}^\top$, $\tau_h^{-1}(z)$ *is same as in* (6), $\boldsymbol{\eta}^* = (\eta^*, 0, \dots, 0)^\top$, $\eta^* = \sqrt{\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}}$, $\boldsymbol{\Omega} = \bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top$, $\boldsymbol{\alpha} = (1 + \boldsymbol{\eta}^\top \bar{\boldsymbol{\Psi}}^{-1} \boldsymbol{\eta})^{-1/2} \{diag(\boldsymbol{\Omega})\}^{1/2} \bar{\boldsymbol{\Psi}}^{-1} \boldsymbol{\eta}$, $\bar{\boldsymbol{\Omega}} = \{diag(\boldsymbol{\Omega})\}^{-1/2} \boldsymbol{\Omega} \{diag(\boldsymbol{\Omega})\}^{-1/2}$, $\boldsymbol{H}^* = \begin{pmatrix} \sqrt{1 + \boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}} & \boldsymbol{0}^\top \\ \boldsymbol{0} & \mathbf{I}_{p-1} \end{pmatrix} \boldsymbol{H}$, $\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}$, $\boldsymbol{Q}$ *is obtained from the spectral decomposition of* $\boldsymbol{Q}^\top \boldsymbol{\Lambda} \boldsymbol{Q} = \boldsymbol{\Omega}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1/2}$, *and* $\boldsymbol{\Sigma} = \bar{\boldsymbol{\Psi}} + \left(1 - \frac{2}{\pi}\right) \boldsymbol{\eta}\boldsymbol{\eta}^\top$.

Proof. We have $\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{Z})$, where $\boldsymbol{Z} \sim \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta})$. Moreover, let $\boldsymbol{Z}^*$ be the canonical transform of $\boldsymbol{Z}$, and $\boldsymbol{Z}^* = \boldsymbol{H}^* \boldsymbol{Z} \sim \mathcal{SN}_p(\boldsymbol{0}, \mathbf{I}_p, \boldsymbol{\eta}^*)$. Here, $\boldsymbol{H}^* = \begin{pmatrix} \sqrt{1 + \boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}} & \boldsymbol{0}^\top \\ \boldsymbol{0} & \mathbf{I}_{p-1} \end{pmatrix} \boldsymbol{H}$, $\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{\Omega}^{-1/2}$, $\boldsymbol{Q}$ is obtained from the spectral decomposition of $\boldsymbol{Q}^\top \boldsymbol{\Lambda} \boldsymbol{Q} = \boldsymbol{\Omega}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1/2}$, and $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{Z}) = \bar{\boldsymbol{\Psi}} + \left(1 - \frac{2}{\pi}\right) \boldsymbol{\eta}\boldsymbol{\eta}^\top$. Hence, $\boldsymbol{\omega}^{-1}(\boldsymbol{Y} - \boldsymbol{\xi}) = \boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{Z}^*) \sim \mathcal{SNTH}_p(\boldsymbol{0}, \mathbf{I}_p, \mathbf{I}_p, \boldsymbol{\eta}^*, \boldsymbol{h})$. □

Since the canonical form of the $\mathcal{SNTH}$ distribution is not exactly an affine transformation, it cannot be used for deriving the measures of multivariate skewness and kurtosis introduced by [31,32]. However, it can be used for reducing the dimensionality of the problem, when applicable, such as simulating observations from the $\mathcal{SNTH}$ distribution.

4. Inference for the $\mathcal{SNTH}$ distribution

In this section, we discuss how to estimate parameters and perform tests for the $\mathcal{SNTH}$ distribution.

4.1. Parameter estimation for the $\mathcal{SNTH}$ distribution

To estimate the parameters of the $\mathcal{SNTH}$ distribution, we use the method of maximizing the likelihood function. Suppose $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ is a random sample of size n from the $\mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$ distribution with $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{ip})^\top$, $i \in \{1, \dots, n\}$. For an observed sample $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$, with $\boldsymbol{y}_i = (y_{i1}, \dots, y_{ip})^\top$, $i \in \{1, \dots, n\}$, the log-likelihood function based on (5) is

$$\begin{aligned} \ell(\boldsymbol{\theta}) = \ln(2) - \frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\{\det(\bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)\} - \frac{1}{2}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{y}_i)^\top (\bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)^{-1} \boldsymbol{g}(\boldsymbol{y}_i) + \sum_{i=1}^n \Phi\left\{\frac{\boldsymbol{\eta}^\top \bar{\boldsymbol{\Psi}}^{-1} \boldsymbol{g}(\boldsymbol{y}_i)}{\sqrt{1 + \boldsymbol{\eta}^\top \bar{\boldsymbol{\Psi}} \boldsymbol{\eta}}}\right\} \\ - n\sum_{j=1}^p \ln(\omega_{jj}) + \sum_{i=1}^n \sum_{j=1}^p \frac{1}{2} W_0\left\{h_j\left(\frac{y_{ij} - \xi_j}{\omega_{jj}}\right)^2\right\} - \sum_{i=1}^n \sum_{j=1}^p \ln\left(h_j\left(\frac{y_{ij} - \xi_j}{\omega_{jj}}\right)^2 + \exp\left[W_0\left\{h_j\left(\frac{y_{ij} - \xi_j}{\omega_{jj}}\right)^2\right\}\right]\right), \end{aligned} \tag{10}$$

where $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \mathrm{diag}(\boldsymbol{\omega})^\top, \mathrm{vech}(\bar{\boldsymbol{\Psi}})^\top, \boldsymbol{\eta}^\top, \boldsymbol{h}^\top)^\top$, where $\mathrm{vech}(\bar{\boldsymbol{\Psi}})^\top$ is the vector of all the upper-off-diagonal elements of $\bar{\boldsymbol{\Psi}}$. We estimate the parameters in $\boldsymbol{\theta}$ by maximizing $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. This maximization cannot be done analytically and has to be performed numerically. Hence, for a p-dimensional problem, we need to perform a $\{4p + p(p-1)/2\}$-dimensional numerical optimization, which becomes difficult when p is large. We can tackle this problem in a different way.

Since $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n \overset{\text{i.i.d.}}{\sim} \mathcal{SNTH}_p(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$, we also have that $Y_{1j}, \dots, Y_{nj} \overset{\text{i.i.d.}}{\sim} \mathcal{SNTH}_1(\xi_j, \omega_{jj}, 1, \eta_j, h_j)$, $j \in \{1, \dots, p\}$, from Proposition 3. Based on the jth marginal data, the marginal log-likelihood function is

$$\begin{aligned} \ell_j(\xi_j, \omega_{jj}, \eta_j, h_j) = \ln(2) - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(1 + \eta_j^2) - \frac{1}{2}\sum_{i=1}^n \frac{g_j(y_{ij})^2}{1 + \eta_j^2} + \sum_{i=1}^n \Phi\left\{\frac{\eta_j g_j(y_{ij})}{\sqrt{1 + \eta_j^2}}\right\} - n\ln(\omega_{jj}) \\ + \sum_{i=1}^n \frac{1}{2} W_0\left\{h_j\left(\frac{y_{ij} - \xi_j}{\omega_{jj}}\right)^2\right\} - \sum_{i=1}^n \ln\left(h_j\left(\frac{y_{ij} - \xi_j}{\omega_{jj}}\right)^2 + \exp\left[W_0\left\{h_j\left(\frac{y_{ij} - \xi_j}{\omega_{jj}}\right)^2\right\}\right]\right), \end{aligned} \tag{11}$$

$j \in \{1, \dots, p\}$. We estimate ξ_j, ω_{jj}, η_j, and h_j, by maximizing the log-likelihood function for the jth marginal $\ell_j(\xi_j, \omega_{jj}, \eta_j, h_j)$, $j \in \{1, \dots, p\}$. Therefore, by performing four-dimensional numerical optimization p times, we obtain the marginal maximum likelihood estimates (MLEs) for $\boldsymbol{\xi}$, $\boldsymbol{\omega}$, $\boldsymbol{\eta}$, and $\boldsymbol{h}$.

At this point, we are yet to obtain the estimate for $\bar{\boldsymbol{\Psi}}$. From the definition of the $\mathcal{SNTH}$ distribution, we have $\boldsymbol{Y}_i \overset{\text{d}}{=} \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\tau}_{\boldsymbol{h}}(\boldsymbol{Z}_i)$, $i \in \{1, \dots, n\}$ and $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n \overset{\text{i.i.d.}}{\sim} \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta})$. With the marginal MLEs $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\omega}}$, $\hat{\boldsymbol{\eta}}$, and $\hat{\boldsymbol{h}}$ of $\boldsymbol{\xi}$, $\boldsymbol{\omega}$, $\boldsymbol{\eta}$, and $\boldsymbol{h}$, we can compute an estimate for the latent $\mathcal{SN}$ observations. Then, $\hat{\boldsymbol{Z}}_i = \boldsymbol{\tau}_{\hat{\boldsymbol{h}}}^{-1}\{\hat{\boldsymbol{\omega}}^{-1}(\boldsymbol{Y}_i - \hat{\boldsymbol{\xi}})\}$, $i \in \{1, \dots, n\}$ are the estimates for $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$. Assuming that, $\hat{\boldsymbol{Z}}_1, \dots, \hat{\boldsymbol{Z}}_n \overset{\text{i.i.d.}}{\sim} \mathcal{SN}_p(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}, \hat{\boldsymbol{\eta}})$ we can estimate $\bar{\boldsymbol{\Psi}}$.

We use the EM algorithm for the $\mathcal{SN}$ distribution for estimating $\bar{\boldsymbol{\Psi}}$, keeping the location and the skewness parameters fixed at $\boldsymbol{0}$ and $\hat{\boldsymbol{\eta}}$. The EM algorithm does not ensure that the estimate of $\bar{\boldsymbol{\Psi}}$ will be a correlation matrix, but the estimate is a covariance matrix, which can be easily converted to its corresponding correlation matrix. We use this correlation matrix as an estimate for $\bar{\boldsymbol{\Psi}}$. In the next section, we will justify the effectiveness of the described method for estimating parameters using a simulation study. Moreover, if we use the marginal MLEs of $\boldsymbol{\xi}$, $\boldsymbol{\omega}$, $\boldsymbol{\eta}$ and $\boldsymbol{h}$ and the estimate of $\bar{\boldsymbol{\Psi}}$ obtained from the EM algorithm as the initial value for the numerical maximization of $\ell(\boldsymbol{\theta})$ in (10), we can converge to the joint MLEs of $\boldsymbol{\theta}$ in very few iterations. Although it does not completely tackle the problem of high-dimensional numerical maximization, this specific selection of initial values reduces the run-time of the numerical maximization greatly. Moreover, we will show in our simulation study that the initial parameter values obtained in the aforementioned way are close to the joint MLEs and can be directly used for high-dimensional problems as the

computation required for estimating the initial estimates is lower than that for estimating the MLEs numerically when p is large. In the next subsection, we describe the EM algorithm for the $\mathcal{SN}$ distribution in details. Note that instead of computing the marginal MLEs of the parameters one can use the iterative generalized method of moments (IGMM) estimators proposed by [26]. IGMM is also based on the estimates of the latent observations and from there estimating the parameters corresponding to the latent random vector. While using the IGMM estimators for the $\mathcal{SNTH}$ distribution one has to keep in mind that the location and the scale parameters used in its definition are not the mean and the marginal standard deviation of the latent random vectors, unlike the proposal of [26]. The IGMM has to be adapted accordingly for getting the correct estimates of the parameters.

4.2. EM algorithm for the $\mathcal{SN}$ distribution

The EM algorithm for the skew-normal distribution is a well-researched topic. Interested readers are directed to the recent paper by [1] and the references therein for more on this topic. In this section, we put forward an EM algorithm for the skew-normal distribution with $\boldsymbol{\Psi}$-$\boldsymbol{\eta}$ parameterization (see (2)), which is new in the literature. Moreover, we are only concerned with the scenario when we need to estimate the scale parameter $\boldsymbol{\Psi}$ while the location $\boldsymbol{\xi} = \mathbf{0}$ and the skewness parameter $\boldsymbol{\eta}$ is known.

Consider a random sample $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n \overset{\text{i.i.d.}}{\sim} \mathcal{SN}_p(\mathbf{0}, \boldsymbol{\Psi}, \boldsymbol{\eta}_0)$, where $\boldsymbol{\eta}_0$ is given. The log-likelihood of an observed sample $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ is

$$\ell(\boldsymbol{\Psi}) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\{\det(\boldsymbol{\Psi} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top)\} - \frac{1}{2}\sum_{i=1}^n \boldsymbol{z}_i^\top(\boldsymbol{\Psi} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top)^{-1}\boldsymbol{z}_i + \sum_{i=1}^n \ln\left\{2\Phi\left(\frac{\boldsymbol{\eta}_0^\top\boldsymbol{\Psi}^{-1}\boldsymbol{z}_i}{\sqrt{1+\boldsymbol{\eta}_0^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_0}}\right)\right\}.$$

Using the stochastic representation of the $\mathcal{SN}$ distribution we can represent $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ as $(\boldsymbol{Z}_i|U_i = u_i) \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(u_i\boldsymbol{\eta}_0, \boldsymbol{\Psi})$, $U_i \overset{\text{i.i.d.}}{\sim} \mathcal{HN}(0,1)$, $i \in \{1, \dots, n\}$ and obtain the conditional pdf of $(U_i|\boldsymbol{Z}_i = \boldsymbol{z}_i)$ as

$$\begin{aligned} f_{U_i|\boldsymbol{Z}_i=\boldsymbol{z}_i}(u) &\propto \phi_p(\boldsymbol{z}_i; u_i\boldsymbol{\eta}_0, \boldsymbol{\Psi})\phi(u;0,1) = \phi_p(\boldsymbol{z}_i; \mathbf{0}, \boldsymbol{\Psi} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top)\phi\{u; \boldsymbol{\eta}_0^\top(\boldsymbol{\Psi} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top)^{-1}\boldsymbol{z}_i, 1 - \boldsymbol{\eta}_0^\top(\boldsymbol{\Psi} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top)^{-1}\boldsymbol{\eta}_0\} \\ &= \phi_p(\boldsymbol{z}_i; \mathbf{0}, \boldsymbol{\Psi} + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top)\phi\left(u; \tau_i, 1/(1+\alpha^2)\right), \quad u > 0, \quad i \in \{1, \dots, n\}, \end{aligned}$$

where $\alpha^2 = \boldsymbol{\eta}_0^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_0$ and $\tau_i = (\boldsymbol{\eta}_0^\top\boldsymbol{\Psi}^{-1}\boldsymbol{z}_i)/(1+\alpha^2)$. Hence, the conditional distribution of the latent variables U_i given the observable $\boldsymbol{Z}_i$ is truncated normal:

$$(U_i|\boldsymbol{Z}_i = \boldsymbol{z}_i) \overset{\text{i.i.d}}{\sim} \mathcal{TN}\left(0; \tau_i, 1/(1+\alpha^2)\right), \quad i \in \{1, \dots, n\}.$$

Moreover, the first and second order raw moments of $(U_i|\boldsymbol{Z}_i = \boldsymbol{z}_i)$ are

$$\upsilon_{1i} = \mathrm{E}(U_i|\boldsymbol{Z}_i = \boldsymbol{z}_i) = \frac{\bar{\tau}_i + \frac{\phi(\bar{\tau}_i)}{\Phi(\bar{\tau}_i)}}{\sqrt{1+\alpha^2}}, \quad \upsilon_{2i} = \mathrm{E}(U_i^2|\boldsymbol{Z}_i = \boldsymbol{z}_i) = \frac{1 + \bar{\tau}_i^2 + \bar{\tau}_i\frac{\phi(\bar{\tau}_i)}{\Phi(\bar{\tau}_i)}}{1+\alpha^2}, \quad i \in \{1, \dots, n\},$$

where $\bar{\tau}_i = \sqrt{1+\alpha^2}\tau_i = (\boldsymbol{\eta}_0^\top\boldsymbol{\Psi}^{-1}\boldsymbol{z}_i)/(\sqrt{1+\alpha^2})$.

From the hierarchical representation above, the complete log-likelihood for $\boldsymbol{\Psi}$ based on the observed data $\boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_n)^\top$ and the missing data $\boldsymbol{u} = (u_1, \dots, u_n)^\top$ is

$$\ell_c(\boldsymbol{\Psi}|\boldsymbol{z}, \boldsymbol{u}) = -\frac{np}{2}\ln(2\pi) + \frac{n}{2}\ln\{\det(\boldsymbol{\Lambda})\} - \frac{1}{2}\sum_{i=1}^n \boldsymbol{z}_i^\top\boldsymbol{\Lambda}\boldsymbol{z}_i + \boldsymbol{\eta}_0^\top\boldsymbol{\Lambda}\sum_{i=1}^n u_i\boldsymbol{z}_i - \frac{1}{2}\boldsymbol{\eta}_0^\top\boldsymbol{\Lambda}\boldsymbol{\eta}_0\sum_{i=1}^n u_i^2 + \frac{n}{2}\ln\left(\frac{2}{\pi}\right) - \frac{1}{2}\sum_{i=1}^n u_i^2,$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Psi}^{-1}$.

Let $\boldsymbol{Z} = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n)^\top$ be the observable random sample and $\boldsymbol{U} = (U_1, \dots, U_n)^\top$ be the latent random sample. Then the E-Step at the $(k+1)$th iteration of the EM algorithm is

$$\begin{aligned} Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(k)}) &= \mathrm{E}_{\boldsymbol{\Psi}^{(k)}}\{\ell_c(\boldsymbol{\Psi}|\boldsymbol{Z}, \boldsymbol{U})|\boldsymbol{Z} = \boldsymbol{z}\} \\ &= -\frac{np}{2}\ln(2\pi) + \frac{n}{2}\ln\{\det(\boldsymbol{\Lambda})\} - \frac{1}{2}\sum_{i=1}^n \boldsymbol{z}_i^\top\boldsymbol{\Lambda}\boldsymbol{z}_i + \boldsymbol{\eta}_0^\top\boldsymbol{\Lambda}\sum_{i=1}^n \upsilon_{1i}^{(k)}\boldsymbol{z}_i - \frac{1}{2}\boldsymbol{\eta}_0^\top\boldsymbol{\Lambda}\boldsymbol{\eta}_0\sum_{i=1}^n \upsilon_{2i}^{(k)} + \frac{n}{2}\ln\left(\frac{2}{\pi}\right) - \frac{1}{2}\sum_{i=1}^n \upsilon_{2i}^{(k)}, \end{aligned}$$

where $\boldsymbol{\Psi}^{(k)}$ is the estimated value of $\boldsymbol{\Psi}$ in the kth step, $\boldsymbol{\Lambda}^{(k)} = \{\boldsymbol{\Psi}^{(k)}\}^{-1}$,

$$\upsilon_{1i}^{(k)} = \frac{\bar{\tau}_i^{(k)} + \frac{\phi(\bar{\tau}_i^{(k)})}{\Phi(\bar{\tau}_i^{(k)})}}{\sqrt{1+\{\alpha^{(k)}\}^2}}, \quad \upsilon_{2i}^{(k)} = \frac{1 + \{\bar{\tau}_i^{(k)}\}^2 + \bar{\tau}_i^{(k)}\frac{\phi(\bar{\tau}_i^{(k)})}{\Phi(\bar{\tau}_i^{(k)})}}{1+\{\alpha^{(k)}\}^2},$$

$\bar{\tau}_i^{(k)} = [1 + \{\alpha^{(k)}\}^2]^{-1/2}\boldsymbol{\eta}_0^\top\boldsymbol{\Lambda}^{(k)}\boldsymbol{z}_i$, $\alpha^{(k)} = \sqrt{\boldsymbol{\eta}_0^\top\boldsymbol{\Lambda}^{(k)}\boldsymbol{\eta}_0}$. To get the $(k+1)$th estimate of $\boldsymbol{\Psi}$, we maximize $Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(k)})$ with respect to $\boldsymbol{\Psi}$ and update $\boldsymbol{\Psi}^{(k+1)} = \mathrm{argmax}\{Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(k)})\}$.

Since $\boldsymbol{\Psi}$ is a symmetric positive definite matrix according to our definition of the $\mathcal{SN}$ distribution, we can write $\boldsymbol{\Psi}^{-1} = \boldsymbol{\Lambda} = \boldsymbol{C}^\top\boldsymbol{C}$, where $\boldsymbol{C} \in \mathbb{R}^{p\times p}$ is a nonsingular matrix. Hence,

$$Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(k)}) \propto \frac{n}{2}\ln\{\det(\boldsymbol{C}^\top\boldsymbol{C})\} - \frac{1}{2}\sum_{i=1}^n \boldsymbol{z}_i^\top\boldsymbol{C}^\top\boldsymbol{C}\boldsymbol{z}_i + \boldsymbol{\eta}_0^\top\boldsymbol{C}^\top\boldsymbol{C}\sum_{i=1}^n \upsilon_{1i}^{(k)}\boldsymbol{z}_i - \frac{1}{2}\boldsymbol{\eta}_0^\top\boldsymbol{C}^\top\boldsymbol{C}\boldsymbol{\eta}_0\sum_{i=1}^n \upsilon_{2i}^{(k)}$$

$$\Rightarrow \frac{\partial Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(k)})}{\partial \boldsymbol{C}} = n(\boldsymbol{C}^\top)^{-1} - \boldsymbol{C}\sum_{i=1}^{n} \boldsymbol{z}_i\boldsymbol{z}_i^\top + \boldsymbol{C}\sum_{i=1}^{n}\left(\boldsymbol{\eta}_0 \upsilon_{1i}^{(k)}\boldsymbol{z}_i^\top + \upsilon_{1i}^{(k)}\boldsymbol{z}_i\boldsymbol{\eta}_0^\top\right) - \boldsymbol{C}\boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top \sum_{i=1}^{n}\upsilon_{2i}^{(k)} = \boldsymbol{0}$$

$$\Rightarrow (\boldsymbol{C}^\top\boldsymbol{C})^{-1} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{z}_i^\top + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top\left(\frac{1}{n}\sum_{i=1}^{n}\upsilon_{2i}^{(k)}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\eta}_0\upsilon_{1i}^{(k)}\boldsymbol{z}_i^\top + \upsilon_{1i}^{(k)}\boldsymbol{z}_i\boldsymbol{\eta}_0^\top\right).$$

Therefore, we update

$$\boldsymbol{\Psi}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{z}_i^\top + \boldsymbol{\eta}_0\boldsymbol{\eta}_0^\top\left(\frac{1}{n}\sum_{i=1}^{n}\upsilon_{2i}^{(k)}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\eta}_0\upsilon_{1i}^{(k)}\boldsymbol{z}_i^\top + \upsilon_{1i}^{(k)}\boldsymbol{z}_i\boldsymbol{\eta}_0^\top\right).$$

We stop the algorithm when $\{\ell(\boldsymbol{\Psi}^{(k+1)})/\ell(\boldsymbol{\Psi}^{(k)}) - 1\}$ is sufficiently close to zero.

4.3. Tests based on the $\mathcal{SNTH}$ distribution

It is a well-known fact [28] that the Fisher information matrix of the $\mathcal{ASN}$ and the $\mathcal{SN}$ distributions is singular when the skewness parameter, $\boldsymbol{\alpha}$ or $\boldsymbol{\eta}$, is set to zero. As a result, we cannot use the Wald type test or the likelihood ratio test (LRT) for testing the null hypothesis that the skewness parameter is zero based on the $\mathcal{ASN}$ or the $\mathcal{SN}$ distribution. Although the asymptotic distribution of the LRT statistic is χ^2_p for the univariate $\mathcal{ASN}$ or the univariate $\mathcal{SN}$ distribution, i.e., for $p = 1$, the same is not true for $p > 1$; see [34]. The explanation of why the asymptotic distribution of the LRT statistic is χ^2_1 for the univariate $\mathcal{ASN}$ or the univariate $\mathcal{SN}$ is still an open problem.

For the skew-t distribution, this singularity of the Fisher information matrix does not occur when the skewness parameter is set to zero. Hence, we can perform the test of the null hypothesis that the skewness parameter is zero based on the skew-t distribution using the Wald type test or the LRT. Next, we show that the Fisher information matrix of the $\mathcal{SNTH}_2$ distribution, when the skewness parameter is set to zero, remains nonsingular.

Proposition 11. *The Fisher information matrix for a bivariate random vector $\boldsymbol{Y} \sim \mathcal{SNTH}_2(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$ is nonsingular when $\boldsymbol{\eta} = \boldsymbol{0}$.*

Proof. From (10), the log-likelihood function for $\boldsymbol{Y} = \boldsymbol{y} = (y_1, y_2)^\top$ is

$$\ell(\boldsymbol{\theta}) = -\ln(\pi) - \frac{1}{2}\ln\{\det(\bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)\} - \frac{1}{2}\boldsymbol{g}(\boldsymbol{y})^\top(\bar{\boldsymbol{\Psi}} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)^{-1}\boldsymbol{g}(\boldsymbol{y}) + \ln\left[\Phi\left\{\frac{\boldsymbol{\eta}^\top\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{g}(\boldsymbol{y})}{\sqrt{1 + \boldsymbol{\eta}^\top\bar{\boldsymbol{\Psi}}^{-1}\boldsymbol{\eta}}}\right\}\right]$$
$$+ \sum_{i=1}^{2}\left(-\ln(\omega_{ii}) + \frac{1}{2}W_0\left(h_i x_i^2\right) - \ln\left[h_i x_i^2 + \exp\left\{W_0\left(h_i x_i^2\right)\right\}\right]\right),$$

where $x_i = (y_i - \xi_i)/\omega_{ii}$, $i \in \{1, 2\}$. The score functions of all the parameters are obtained by differentiating the log-likelihood with respect to the parameters. Assuming that $\bar{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, the score functions of all the parameters, when $\boldsymbol{\eta} = \boldsymbol{0}$, are listed below for $i = 1, j = 2$ or $i = 2, j = 1$:

$$S_{\xi_i} = \frac{1}{\omega_{ii}}\left(\frac{x_i - \rho x_j \exp\left\{\frac{1}{2}W_0(h_i x_i^2) - \frac{1}{2}W_0(h_j x_j^2)\right\}}{(1-\rho^2)[h_i x_i^2 + \exp\{W_0(h_i x_i^2)\}]} + \frac{h_i x_i[h_i x_i^2 + 3\exp\{W_0(h_i x_i^2)\}]}{[h_i x_i^2 + \exp\{W_0(h_i x_i^2)\}]^2}\right),$$

$$S_{\omega_{ii}} = \frac{1}{\omega_{ii}}\left(\frac{x_i^2 - \rho x_i x_j \exp\left\{\frac{1}{2}W_0(h_i x_i^2) - \frac{1}{2}W_0(h_j x_j^2)\right\}}{(1-\rho^2)[h_i x_i^2 + \exp\{W_0(h_i x_i^2)\}]} + \frac{\exp\{W_0(h_i x_i^2)\}[h_i x_i^2 - \exp\{W_0(h_i x_i^2)\}]}{[h_i x_i^2 + \exp\{W_0(h_i x_i^2)\}]^2}\right),$$

$$S_{\eta_i} = \sqrt{\frac{2}{\pi}}\left\{\frac{g_i(y_i) - \rho g_j(y_j)}{(1-\rho^2)}\right\},$$

$$S_{h_i} = \frac{1}{2}\left(\frac{x_i^4\exp\left\{-W_0(h_i x_i^2)\right\} - \rho x_i^3 x_j \exp\left\{-\frac{1}{2}W_0(h_i x_i^2) - \frac{1}{2}W_0(h_j x_j^2)\right\}}{(1-\rho^2)[h_i x_i^2 + \exp\{W_0(h_i x_i^2)\}]} - \frac{h_i x_i^4 + 3x_i^2\exp\{W_0(h_i x_i^2)\}}{[h_i x_i^2 + \exp\{W_0(h_i x_i^2)\}]^2}\right),$$

$$S_\rho = \frac{g_1(y_1)g_2(y_2)}{(1-\rho^2)} - \frac{\rho}{(1-\rho^2)^2}\{g_1^2(y_1) + g_2^2(y_2) - 2\rho g_1(y_1)g_2(y_2)\} + \frac{\rho}{(1-\rho^2)}.$$

From the form of the score functions we can observe that they are not linearly dependent when $\boldsymbol{\eta} = \boldsymbol{0}$ and hence the Fisher information matrix, which is the variance–covariance matrix of the score vector, is nonsingular when $\boldsymbol{\eta} = \boldsymbol{0}$. □

Proposition 11 demonstrates that the Fisher information matrix of the $\mathcal{SNTH}$ distribution is nonsingular when $\boldsymbol{\eta} = \boldsymbol{0}$ for $p = 2$. Our conjecture is that this statement remains true for $p > 2$. We justify this by plotting, in Fig. 5, the histogram of the LRT statistic for testing $H_0 : \boldsymbol{\eta} = \boldsymbol{0}$ vs $H_1 : \boldsymbol{\eta} \neq \boldsymbol{0}$ for $p \in \{2, 3, 4\}$, based on samples of size 5000 and 1000 replicates. Along with the histograms,

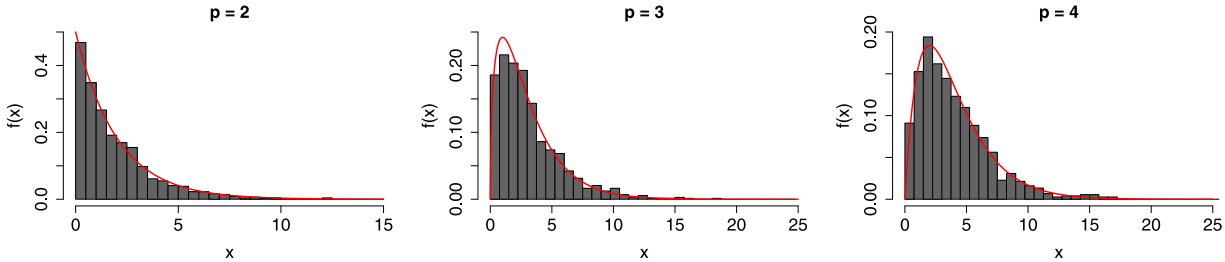


Fig. 5. Histograms of the LRT statistic for testing $H_0 : \boldsymbol{\eta} = \mathbf{0}$ vs $H_1 : \boldsymbol{\eta} \neq \mathbf{0}$ for $\mathcal{SNTH}_p$ when $p \in \{2, 3, 4\}$ based on samples of size 5000 and 1000 replicates. The red curves indicate the pdf of the χ^2_p distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

we also plot the χ^2_p pdf. The plots indicate that the asymptotic distribution of the LRT statistic indeed follows χ^2_p, for $p \in \{2, 3, 4\}$. This would not have been the case if the Fisher information matrix was singular for $\boldsymbol{\eta} = \mathbf{0}$.

Although we have justified the nonsingularity of the Fisher information matrix for the $\mathcal{SNTH}$ distribution when $\boldsymbol{\eta} = \mathbf{0}$, we do not have the mathematical form of the Fisher information matrix. As a result, we cannot use the Wald type test for testing $\boldsymbol{\eta} = \mathbf{0}$. We have to rely on the LRT for that:

- *Testing* $H_0 : \boldsymbol{\eta} = \mathbf{0}$ *vs* $H_1 : \boldsymbol{\eta} \neq \mathbf{0}$*, given that* $\boldsymbol{h} \neq \mathbf{0}$: Since the Fisher information matrix of the $\mathcal{SNTH}$ distribution when $\boldsymbol{\eta} = \mathbf{0}$ is nonsingular, given that $\boldsymbol{h} \neq \mathbf{0}$, we use the asymptotic distribution of the LRT statistic for conducting the test.
- *Testing* $H_0 : \boldsymbol{h} = \mathbf{0}$ *vs* $H_1 : \boldsymbol{h} \neq \mathbf{0}$*, given that* $\boldsymbol{\eta} \neq \mathbf{0}$: Under the null hypothesis the $\mathcal{SNTH}$ distribution becomes the $\mathcal{SN}$ distribution. The Fisher information matrix of the $\mathcal{SN}$ distribution is nonsingular when $\boldsymbol{\eta} \neq \mathbf{0}$. Hence, under the null hypothesis we can use the asymptotic distribution of the LRT statistic for conducting the test.
- *Testing* $H_0 : \boldsymbol{\eta} = \mathbf{0}$ *and* $\boldsymbol{h} = \mathbf{0}$ *vs* $H_1 : \boldsymbol{\eta} \neq \mathbf{0}$ *or* $\boldsymbol{h} \neq \mathbf{0}$: Under the null hypothesis, the Fisher information matrix is singular. Hence, we cannot use the LRT anymore for this testing problem. However, since the asymptotic distribution of the LRT statistic for testing $\eta = 0$ vs $\eta \neq 0$ based on the univariate $\mathcal{SN}$ is χ^2_1, we can use the LRT for testing $H_{i0} : \eta_i = 0, h_i = 0$ vs $H_{i1} : \eta_i \neq 0$ or $h_i \neq 0$, $i \in \{1, \dots, p\}$. We reject H_0 if any of the H_{i0} gets rejected. Note here that the rejection region for testing H_{i0} vs H_{i1}, $i \in \{1, \dots, p\}$, has to be computed subject to Bonferroni's correction.

5. Simulation study

We conduct two simulation studies in this section: one to demonstrate the effectiveness of the parameter estimation method described in Sections 4.1 and 4.2, and another to show in which scenarios the $\mathcal{SNTH}$ distribution is more suitable compared to the skew-t distribution.

5.1. $\mathcal{SNTH}$ parameter estimation

We test the methodology for $\mathcal{SNTH}$ parameter estimation in a simulation study. We simulate observations of size $n = 50$, 100, 200, 500, and 1000 from a $\mathcal{SNTH}_3(\boldsymbol{\xi}, \boldsymbol{\omega}, \bar{\boldsymbol{\Psi}}, \boldsymbol{\eta}, \boldsymbol{h})$, with $\boldsymbol{\xi} = (0.8, -0.6, 1.3)^\top$, $\boldsymbol{\omega} = \text{diag}(3, 5, 2)$, $\bar{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & -0.5 & 0.3 \\ -0.5 & 1 & -0.2 \\ 0.3 & -0.2 & 1 \end{pmatrix}$, $\boldsymbol{\eta} = (-1.5, 2, 0.5)^\top$ and $\boldsymbol{h} = (0.02, 0.08, 0.03)^\top$. Based on the simulated data, we estimate the parameters by the methodology described in Sections 4.1 and 4.2. We repeat the process 100 times and summarize the estimated parameter in boxplots in Fig. 6. Alongside the estimates obtained from the methodology described in Section 4.1 (indicated as mMLE (short for marginal MLE) for $\boldsymbol{\xi}$, $\boldsymbol{\omega}$, $\boldsymbol{\eta}$, $\boldsymbol{h}$ and as EM for $\bar{\boldsymbol{\Psi}}$ in Fig. 6) we also report the MLEs of all the parameters. The boxplots indicate that the methodology is working reasonably well for estimating the parameters from the $\mathcal{SNTH}$ model. Moreover, as the sample size increases, the variance of the estimates decreases, as it should. Hence, we can say that the parameter estimation methodology described in Sections 4.1 and 4.2 is justified. The boxplots also show that the estimates of the parameters obtained from the EM algorithm are not very different from the MLEs, although they have more variability. The variability difference between the two estimation methods also decreases as the sample size increases. For problems with high dimensions where the computation of the exact MLEs are infeasible, one can use the methodology described in Sections 4.1 and 4.2 as an alternative. Moreover, these estimates are an excellent choice for the starting values of the parameters when optimizing the exact log-likelihood for computing the MLEs.

5.2. Comparison between the $\mathcal{SNTH}$ and the skew-t distributions

In this simulation study we show that when there is a great disparity between the marginal kurtosis values in a multivariate dataset, the $\mathcal{SNTH}$ distribution is more appropriate than the skew-t distribution. We generate 500 random samples from a three-dimensional vine copula to create a trivariate dataset in Uniform$(0, 1)$ scale. In this vine copula model, variables 1 and 2 are related

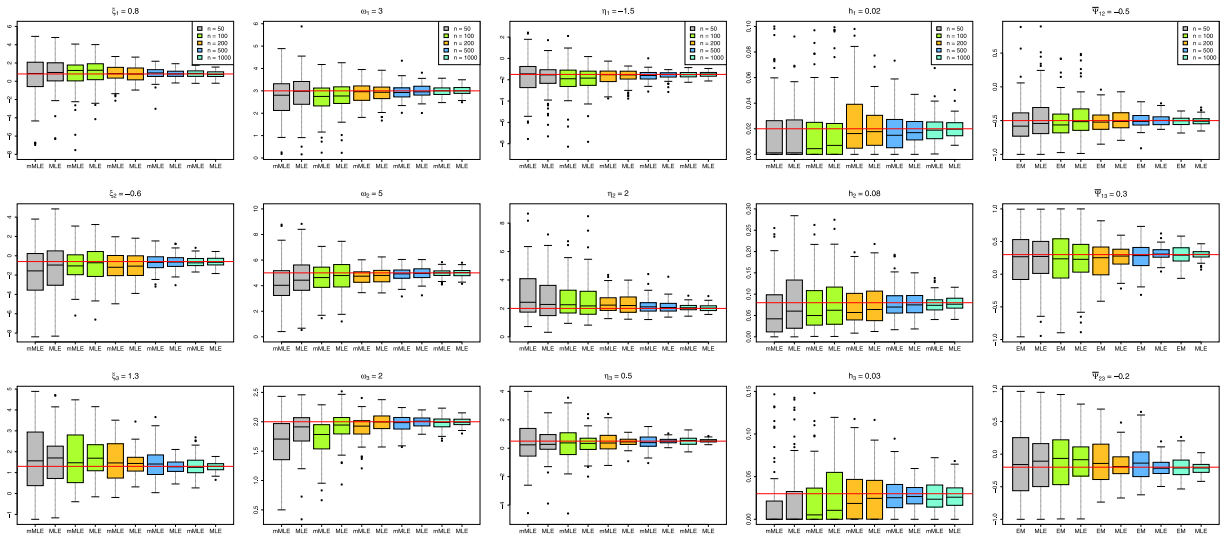


Fig. 6. Boxplots of the parameter estimates (100 replicates) of a $\mathcal{SNTH}_3$ distribution obtained from the methodology in Sections 4.1 and 4.2 for different sample sizes n, given as mMLE (marginal MLE) for $\boldsymbol{\xi}$, $\boldsymbol{\omega}$, $\boldsymbol{\eta}$, $\boldsymbol{h}$ and as EM for $\bar{\boldsymbol{\Psi}}$ along with the MLE boxplots. The red line in each plot indicates the true parameter value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with a Gaussian copula with $\rho = 0.5$, variables 1 and 3 are related with a Clayton copula with parameter 4.8 and variables 2 and 3 given variable 1 are related with a Gumbel copula with parameter 1.9. On the trivariate simulated data, we transform the 1st component to the standard normal scale, the 2nd component to the Cauchy t_1 scale, and the 3rd component to the Student's-t_{10} scale. We fit both the $\mathcal{SNTH}$ and the skew-t distribution to this simulated data. The Akaike information criterion (AIC) computed for the $\mathcal{SNTH}$ and the skew-t are 4393 and 4848, respectively, suggesting the $\mathcal{SNTH}$ distribution is more suitable for this simulated dataset, compared to the skew-t distribution.

We perform similar experiments where we generate 500 observations from a three-dimensional multiple-scaled generalized hyperbolic ($\mathcal{MSGH}$) distribution [41] and from a three-dimensional t-SAS distribution [16]. For the $\mathcal{MSGH}$ distribution we use the following parameters: $\boldsymbol{\mu} = (0,0,0)^\top$, $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.3 & -0.2 \\ 0.3 & 1 & -0.4 \\ -0.2 & -0.4 & 1 \end{pmatrix}$, $\boldsymbol{\beta} = (3, 0.5, -0.2)^\top$, $\boldsymbol{\lambda} = (2,1,4)^\top$, $\boldsymbol{\gamma} = (\sqrt{3}, \sqrt{0.2}, \sqrt{0.25})^\top$, and $\delta = 1$. For the t-SAS distribution, we use a three-dimensional t-copula with correlation matrix $\begin{pmatrix} 1 & 0.3 & -0.2 \\ 0.3 & 1 & -0.4 \\ -0.2 & -0.4 & 1 \end{pmatrix}$ to generate observations on the uniform scale. For the Sinh-Arcsinh (SAS) transformation, we use $(-0.7, 1)$, $(0.2, 0.6)$, and $(0.5, 0.8)$ as our (g, h) (for skewness and tail-thickness, as used in [16]) parameters for the three marginals, respectively. Finally, we scale the marginals by 1, 1.2, and 1.8, respectively. When the $\mathcal{SNTH}$ and the skew-t models are fitted to the $\mathcal{MSGH}$ dataset the obtained AICs are 9982 and 10110, and for the t-SAS dataset, the AICs are 6606 and 6634. The AICs for both studies suggest that the $\mathcal{SNTH}$ is a better fit to these two datasets compared to the skew-t model.

We provide the contour plots of the bivariate marginal pdfs of the $\mathcal{SNTH}$ and the skew-t distributions fitted to the three simulated datasets in Fig. 7. The bivariate marginal pdfs for the $\mathcal{SNTH}$ distribution are obtained based on the MLEs and also based on the estimates from the EM algorithm. The contours are plotted for the 0.25, 0.5, 0.75, and 0.95 approximate probability regions. The plots show that, as expected, the skew-t distribution cannot handle different tail-thickness for different marginals, and instead tries to find the best compromise with a single parameter, ν. In scenarios like this, the $\mathcal{SNTH}$ distribution is more appropriate. Moreover, in the first row of Fig. 7 we see from the contour plots that the difference between the bivariate marginal pdfs obtained based on the MLE and the EM algorithm is small for the vine copula dataset. However, in the second and third rows of Fig. 7 the dissimilarity between the two $\mathcal{SNTH}$ parameter estimation methods is more prominent, especially for the (Y_1, Y_3) pair. Finally, it is clear from the plots that the marginal bivariate $\mathcal{SNTH}$ pdf contours obtained from the MLEs are more suitable for all three datasets compared to the skew-t counterparts.

6. Data applications

We use two data applications to illustrate the effectiveness of the $\mathcal{SNTH}$ distribution over the skew-t in certain situations. The parameter estimates and standard errors for the two data applications, as well as log-likelihood and AIC values along with computing times, are given in Sections S1 and S2 of the supplementary material.

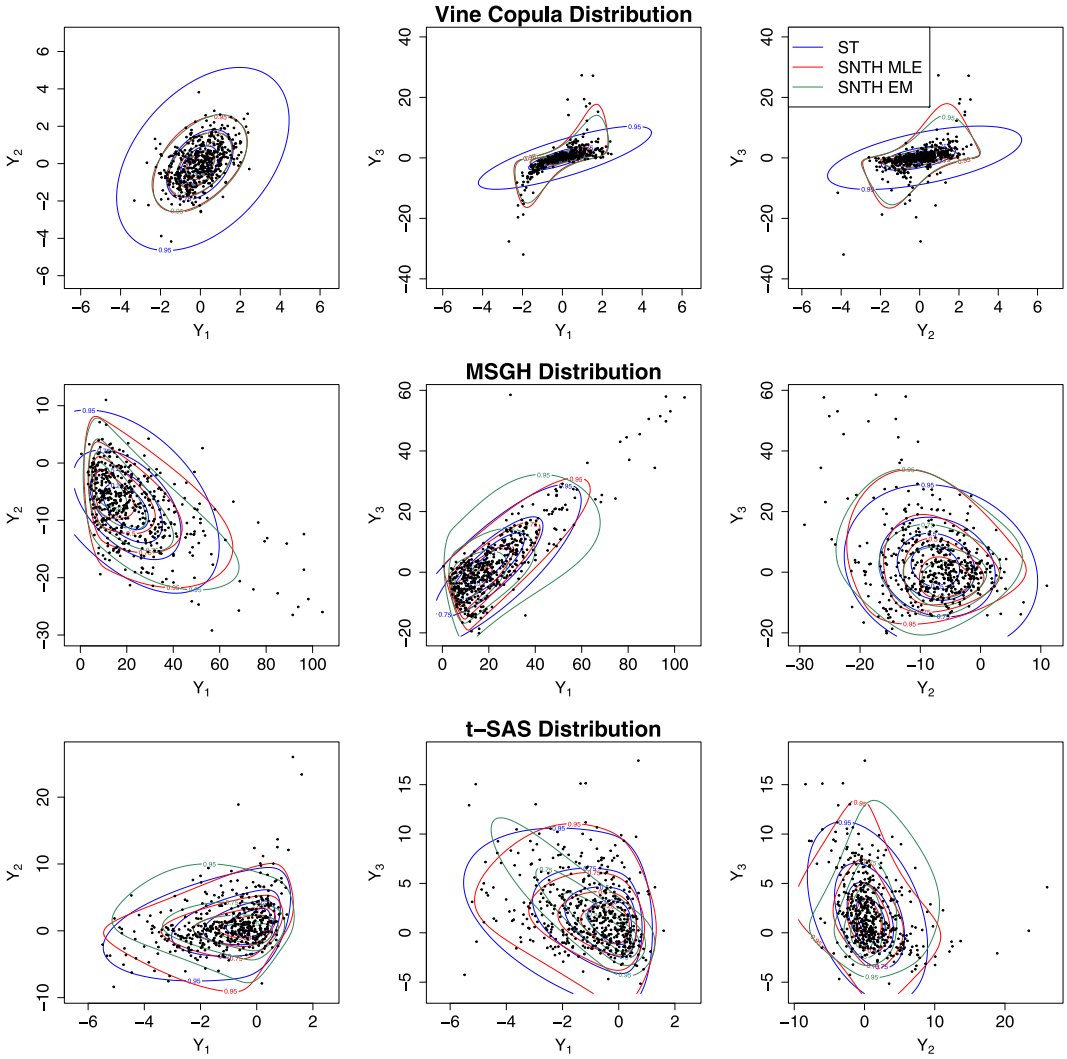


Fig. 7. Bivariate contours of the marginal bivariate pdfs obtained from the fitted $\mathcal{SNTH}$ using Sections 4.1 and 4.2 methodology (green), from the fitted $\mathcal{SNTH}$ using MLEs (red) and from the fitted skew-t (blue) distributions to trivariate vine copula data (first row), $\mathcal{MSGH}$ data (second row), and t-SAS data (third row). The contours correspond to 0.25, 0.5, 0.75, and 0.95 approximate probability regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.1. Italian wine dataset

We consider a trivariate dataset consisting of the amount of chloride, glycerol and magnesium in a particular type of wine. The data were obtained from [23] and originally consist of measurements on 28 chemicals from 178 samples of Italian wines. Among these 178 samples, 48 originated from the Barbera region, 59 from the Barolo region, and 71 from the Grignolino region. Here we use the variables chloride, glycerol and magnesium for the Grignolino region as previously analyzed by Azzalini and Capitanio [13] with a skew-t distribution, hence $p = 3$ variables and $n = 71$ observations.

The sample estimate of the marginal Pearson's measure of kurtosis for this dataset are 7.7, 21.1, and 7.9, which suggest that the $\mathcal{SNTH}$ distribution might be more suitable for this dataset compared to the skew-t distribution. We fit both the $\mathcal{SNTH}$ and the skew-t distribution to this dataset. The contour plots of the bivariate marginal pdfs obtained from the two fitted distributions are presented in Fig. 8. For the $\mathcal{SNTH}$ model we have produced the contours of the bivariate marginal pdfs using MLEs (in red) and the EM algorithm estimates (in green) along with the skew-t bivariate marginal pdfs (in blue). One can see visually that the $\mathcal{SNTH}$ distribution fits the data better than the skew-t. Moreover, the contour plots indicate that there are some discrepancies between the two estimation methodologies based on the $\mathcal{SNTH}$ distribution, specifically for the magnesium-chloride pair, but much less in the other two pairs. The difference is likely due to a relatively small sample size ($n = 71$). The AIC corresponding to the $\mathcal{SNTH}$ distribution and the skew-t distribution are 1474 and 1492, respectively. Hence, for this dataset, the $\mathcal{SNTH}$ distribution is a better model than the skew-t distribution. Moreover, assuming that $\boldsymbol{\eta} \neq \mathbf{0}$, the p-value for testing $H_0 : \boldsymbol{h} = \mathbf{0}$ vs $H_1 : \boldsymbol{h} \neq \mathbf{0}$ is 2.53×10^{-14},

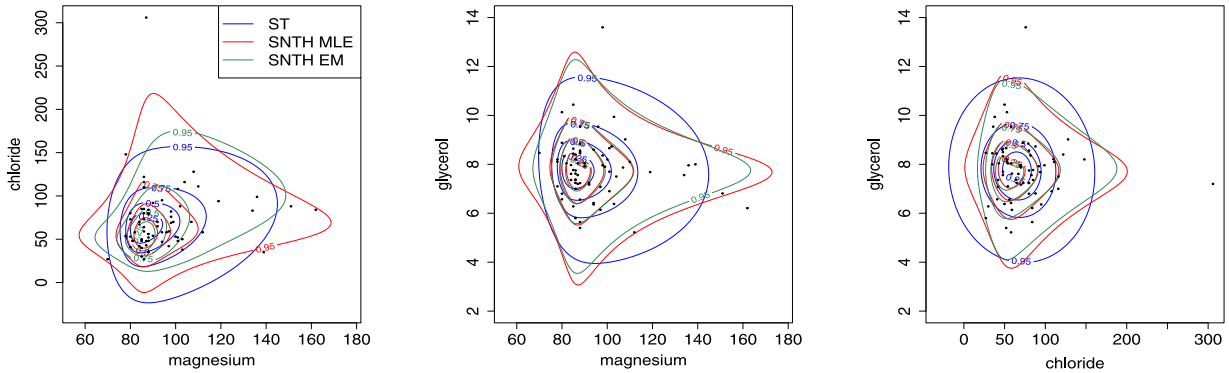


Fig. 8. Bivariate contours of the marginal bivariate pdfs obtained from the fitted $\mathcal{SNTH}$ using Sections 4.1 and 4.2 methodology (green), from the fitted $\mathcal{SNTH}$ using MLE (red) and the skew-t (blue) distributions to the wine data. The contours correspond to 0.25, 0.5, 0.75, and 0.95 approximate probability regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

using the LRT based on the $\mathcal{SNTH}$ distribution. This suggests that $\boldsymbol{h} \neq \boldsymbol{0}$ for this dataset. Using the LRT for testing $H_0 : \boldsymbol{\eta} = \boldsymbol{0}$ vs $H_1 : \boldsymbol{\eta} \neq \boldsymbol{0}$ when $\boldsymbol{h} \neq \boldsymbol{0}$ is 1.6×10^{-5}, hence confirming the apparent skewness in the data.

Although we have considered this classical dataset, originally used to motivate the skew-t distribution in the book [13], for the purpose of comparing the differences between the $\mathcal{SNTH}$ and the skew-t distributions, a further analysis shows that the variables in this dataset are not much dependent. In fact, the independent marginal $\mathcal{SNTH}$ and independent marginal skew-t distributions yield better AIC (1466 and 1467, respectively) compared to the multivariate $\mathcal{SNTH}$ distribution. The AIC suggests for this dataset that the independent marginal $\mathcal{SNTH}$ distribution is the most suitable. It is worth pointing out here that both the independent marginal $\mathcal{SNTH}$ and independent marginal skew-t distributions are not special cases of the multivariate $\mathcal{SNTH}$ and multivariate skew-t distributions, respectively. The next data application shows a clearer comparison between the multivariate skew-t and the multivariate $\mathcal{SNTH}$ distribution as we use a multivariate dataset with stronger dependence.

6.2. Saudi Arabian wind speed dataset

We analyze the dependence structure of the daily average, minimum, and maximum wind speed in the city of Sharurah in southern Saudi Arabia, at 100 meters in height (a typical hub height for wind turbines), in the year 2015. Understanding the dependence and distribution of these variables is important for setting up wind farms for harvesting wind energy. We remove a quadratic trend from all three variables and fit an AR(1) time series model to the detrended data marginally to obtain residuals. A Ljung–Box test shows that there is no significant serial correlation left in all three residuals. Hence, the residuals can be treated as a random sample of size $n = 365$ from a trivariate distribution.

The sample estimates of the marginal Pearson's measure of kurtosis for the three variables are 3.0, 7.5, and 4.4, which means that the residuals corresponding to the average windspeed have a Gaussian-like tail and the other two residuals have heavier tails than the Gaussian distribution. This indicates that the $\mathcal{SNTH}$ distribution may be more apt for this dataset compared to the skew-t distribution. We fit both the $\mathcal{SNTH}$ and the skew-t distribution to the residuals. Similar to the previous contour plots, we have produced in Fig. 9 the contours of the bivariate marginal pdfs using MLEs (in red) and the EM algorithm estimates (in green) along with the skew-t bivariate marginal pdfs (in blue). The plots indicate that the $\mathcal{SNTH}$ distribution is more suitable here for capturing different tail-thickness for different marginals, compared to the skew-t distribution. This conclusion is further validated by the AIC which is 3274 for the $\mathcal{SNTH}$ distribution and is 3432 for the skew-t distribution. Moreover, the difference between the contours obtained from the MLEs and from the EM algorithm estimates for the $\mathcal{SNTH}$ distribution are very close to each other. Similar to the wine dataset, we can perform the following tests: assuming that $\boldsymbol{\eta} \neq \boldsymbol{0}$, the p-value for testing $H_0 : \boldsymbol{h} = \boldsymbol{0}$ vs $H_1 : \boldsymbol{h} \neq \boldsymbol{0}$ is 9.7×10^{-35}, using the LRT based on the $\mathcal{SNTH}$ distribution, which confirms that the data here are not from a skew-normal distribution; $H_0 : \boldsymbol{\eta} = \boldsymbol{0}$ vs $H_1 : \boldsymbol{\eta} \neq \boldsymbol{0}$ when $\boldsymbol{h} \neq \boldsymbol{0}$ is 2.56×10^{-13}, which confirms the presence of skewness in the data. The independent marginal $\mathcal{SNTH}$ and independent marginal skew-t distributions yield much worse AIC (3551 and 3562, respectively) compared to the multivariate $\mathcal{SNTH}$ distribution.

7. Discussion

In this article, we have introduced the multivariate $\mathcal{SNTH}$ distribution, a new extension of the multivariate skew-normal distribution for modeling heavy-tailed data. We have compared our proposed distribution with the skew-t distribution, another extension of the skew-normal distribution for adapting tail-thickness. Unlike the skew-t distribution, our proposal is capable of handling data with different kurtosis for different marginals. As a consequence, the $\mathcal{SNTH}$ model can be used as a robust model, as suggested by [15] for the skew-t, for modeling outliers. Moreover, the $\mathcal{SNTH}$ distribution can capture outliers in some marginals while having Gaussian-like distributions in other marginals. We have discussed various appealing stochastic and

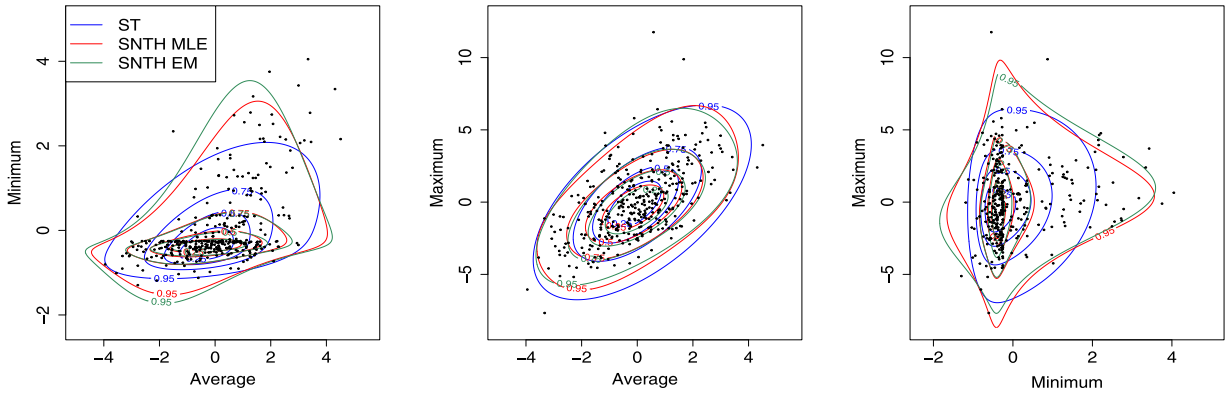


Fig. 9. Bivariate contours of the marginal bivariate pdfs obtained from the fitted $\mathcal{SNTH}$ using the EM algorithm (green), from the fitted $\mathcal{SNTH}$ using MLE (red) and the skew-t (blue) distributions to the wind speed residuals. The contours correspond to 0.25, 0.5, 0.75, and 0.95 approximate probability regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

inferential properties of the $\mathcal{SNTH}$ distribution in detail. A methodology for parameter estimation of the $\mathcal{SNTH}$ distribution was also provided.

There are other proposals in the multivariate setup for modeling varying marginal tail-thickness, such as the $\mathcal{MSGH}$ distribution by [41] and the t-SAS distribution by [16]. However, they lack appealing stochastic properties, such as a tractable conditional distribution and an explicit form of conditional mean and variance, unlike the $\mathcal{SNTH}$ model. How the $\mathcal{SNTH}$ model performs compared to these other multivariate models for modeling varying marginal tail-thickness is left as a future research direction.

The $\mathcal{SNTH}$ distribution can be further generalized by extending the idea of using transformation to induce tail-thickness in the distribution to the extended skew-normal ($\mathcal{ESN}$) family and the unified skew-normal ($\mathcal{SUN}$) family [5]. In Section 3.1, we have discussed how the $\mathcal{SNTH}$ distribution induces tail-thickness in the $\mathcal{SN}$ distribution by stretching the distribution along different axes, and this stretching can be different for different marginals. This idea could be further generalized where the stretching occurs along arbitrary directions.

The EM algorithm in Section 4.2 discussed how we can estimate the scale matrix $\boldsymbol{\Psi}$ of an $\mathcal{SN}_p(\mathbf{0}, \boldsymbol{\Psi}, \boldsymbol{\eta}_0)$ distribution, given that $\boldsymbol{\eta}_0$ is fixed. However, we need this $\boldsymbol{\Psi}$ to be a correlation matrix, not a covariance matrix. This is achieved by transforming the final estimate of $\boldsymbol{\Psi}$ from covariance to a correlation matrix. The EM algorithm for the scenario when $\boldsymbol{\Psi}$ is a correlation matrix is an open problem.

The R-codes and real data for Sections 5 and 6 are available on a GitHub repository: https://github.com/sagnikind/Skew-normal-Tukey-h.

CRediT authorship contribution statement

Sagnik Mondal: Conceptualization, Methodology Writing. **Marc G. Genton:** Conceptualization, Methodology, Writing.

Acknowledgments

The authors thank the Editor, Associate Editor and referees for comments that improved this paper, as well as the King Abdullah City for Atomic and Renewable Energy (K.A.CARE) for providing the wind speed observational data. This research was supported by the King Abdullah University of Science and Technology (KAUST) in Saudi Arabia.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2023.105260.

References

[1] T. Abe, H. Fujisawa, T. Kawashima, C. Ley, EM algorithm using overparameterization for the multivariate skew-normal distribution, Econom. Stat. 19 (2021) 151–168.
[2] C.J. Adcock, Capital asset pricing for UK stocks under the multivariate skew-normal distribution, in: M.G. Genton (Ed.), Skew-Elliptical Distributions and their Applications: A Journey beyond Normality, Chapman and Hall, 2004, pp. 191–204 (Chapter 11).
[3] C.J. Adcock, Exploiting skewness to build an optimal hedge fund with a currency overlay, Eur. J. Finance 11 (5) (2005) 445–462.
[4] C.J. Adcock, K. Shutes, Portfolio selection based on the multivariate skew normal distribution, Financ. Model. (2001) 167–177.
[5] R.B. Arellano-Valle, A. Azzalini, On the unification of families of skew-normal distributions, Scand. J. Stat. 33 (3) (2006) 561–574.
[6] R.B. Arellano-Valle, M.G. Genton, Multivariate extended skew-t distributions and related families, Metron 68 (3) (2010) 201–234.
[7] R.B. Arellano-Valle, M.G. Genton, Multivariate unified skew-elliptical distributions, Chil. J. Stat. 1 (1) (2010) 17–33.

[8] B.C. Arnold, R.J. Beaver, Hidden truncation models, Sankhyā 62 (2000) 23–35.
[9] A. Azzalini, Package 'sn', 2015, R package version 2.1.1.
[10] A. Azzalini, An overview on the progeny of the skew-normal family—A personal perspective, J. Multivariate Anal. 188 (2022) 104851.
[11] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (3) (1999) 579–602.
[12] A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution, J. R. Stat. Soc. Ser. B Stat. Methodol. 65 (2) (2003) 367–389.
[13] A. Azzalini, A. Capitanio, The Skew-Normal and Related Families, Vol. 3, Cambridge University Press, 2014.
[14] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (4) (1996) 715–726.
[15] A. Azzalini, M.G. Genton, Robust likelihood methods based on the skew-t and related distributions, Internat. Statist. Rev. 76 (1) (2008) 106–129.
[16] S. Babić, C. Ley, D. Veredas, Comparison and classification of flexible distributions for multivariate skew and heavy-tailed data, Symmetry 11 (10) (2019) 1216.
[17] M.D. Branco, D.K. Dey, A general class of multivariate skew-elliptical distributions, J. Multivariate Anal. 79 (1) (2001) 99–113.
[18] A. Capitanio, On the canonical form of scale mixtures of skew-normal distributions, Statistica 80 (2) (2020) 145–160.
[19] A. Capitanio, A. Azzalini, E. Stanghellini, Graphical models for skew-normal variates, Scand. J. Stat. 30 (1) (2003) 129–144.
[20] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth, On the Lambert W function, Adv. Comput. Math. 5 (4) (1996) 329–359.
[21] C. Field, M.G. Genton, The multivariate g-and-h distribution, Technometrics 48 (1) (2006) 104–111.
[22] F. Forbes, D. Wraith, A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: Application to robust clustering, Stat. Comput. 24 (6) (2014) 971–984.
[23] M. Forina, C. Armanino, M. Castino, M. Ubigli, Multivariate data analysis as a discriminating method of the origin of wines, Vitis 25 (59) (1986) 189–201.
[24] M.G. Genton, Skew-Elliptical Distributions and their Applications: A Journey beyond Normality, CRC Press, 2004, Edited Volume.
[25] M.G. Genton, N.M. Loperfido, Generalized skew-elliptical distributions and their quadratic forms, Ann. Inst. Statist. Math. 57 (2) (2005) 389–401.
[26] G.M. Goerg, Lambert W random variables—A new family of generalized skewed distributions with applications to risk estimation, Ann. Appl. Stat. 5 (3) (2011) 2197–2230.
[27] G.M. Goerg, The lambert way to Gaussianize heavy-tailed data with the inverse of Tukey's h transformation as a special case, Sci. World J. 2015 (2015).
[28] M. Hallin, C. Ley, Skew-symmetric distributions and Fisher information–a tale of two densities, Bernoulli 18 (3) (2012) 747–763.
[29] H. Joe, Multivariate Models and Multivariate Dependence Concepts, CRC Press, 1997.
[30] M.C. Jones, A. Pewsey, Sinh-arcsinh distributions, Biometrika 96 (4) (2009) 761–780.
[31] J.F. Malkovich, A. Afifi, On tests for multivariate normality, J. Amer. Statist. Assoc. 68 (341) (1973) 176–179.
[32] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (3) (1970) 519–530.
[33] K. Miller, Some multivariate t-distributions, Ann. Math. Stat. 39 (5) (1968) 1605–1609.
[34] S. Mondal, R.B. Arellano-Valle, M.G. Genton, A multivariate modified skew-normal distribution, Statist. Papers (2023) 1–45.
[35] R.B. Nelsen, An Introduction to Copulas, Springer Science & Business Media, 2007.
[36] R. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022.
[37] F.J. Rubio, E.O. Ogundimu, J.L. Hutton, On modelling asymmetric data using two-piece sinh–arcsinh distributions, Braz. J. Probab. Stat. (2016) 485–501.
[38] M. Sklar, Fonctions de Répartition à n Dimensions et Leurs Marges, Publ. Inst. Stat. Univ. Paris 8 (1959) 229–231.
[39] J. Wang, J. Boyer, M.G. Genton, A skew-symmetric representation of multivariate distributions, Statist. Sinica 14 (4) (2004) 1259–1270.
[40] Z. Wei, S. Kim, D. Kim, Multivariate skew normal copula for non-exchangeable dependence, Procedia Comput. Sci. 91 (2016) 141–150.
[41] D. Wraith, F. Forbes, Location and scale mixtures of Gaussians with flexible tail behaviour: Properties, inference and application to multivariate clustering, Comput. Statist. Data Anal. 90 (2015) 61–73.
[42] Y. Yan, J. Jeong, M.G. Genton, Multivariate transformed Gaussian processes, Jpn. J. Stat. Data Sci. 3 (1) (2020) 129–152.