METRON - *International Journal of Statistics*
2010, vol. LXVIII, n. 3, pp. 201-234

REINALDO B. ARELLANO-VALLE – MARC G. GENTON

Multivariate extended skew-*t* distributions and related families

Summary - A class of multivariate extended skew-t (EST) distributions is introduced and studied in detail, along with closely related families such as the subclass of extended skew-normal distributions. Besides mathematical tractability and modeling flexibility in terms of both skewness and heavier tails than the normal distribution, the most relevant properties of the EST distribution include closure under conditioning and ability to model lighter tails as well. The first part of the present paper examines probabilistic properties of the EST distribution, such as various stochastic representations, marginal and conditional distributions, linear transformations, moments and in particular Mardia's measures of multivariate skewness and kurtosis. The second part of the paper studies statistical properties of the EST distribution, such as likelihood inference, behavior of the profile log-likelihood, the score vector and the Fisher information matrix. Especially, unlike the extended skew-normal distribution, the Fisher information matrix of the univariate EST distribution is shown to be non-singular when the skewness is set to zero. Finally, a numerical application of the conditional EST distribution is presented in the context of confidential data perturbation.

Key Words - Confidential data perturbation; Elliptically contoured distributions; Fisher information; Kurtosis; Non-normal; Skewness.

1. Introduction

The quest for flexible parametric families of multivariate distributions has received sustained attention in recent years as witnessed by the book edited by Genton (2004), the review of Azzalini (2005), and references therein. Although the families under consideration can be casted in a unified framework of skewed distributions arising from certain selection mechanisms (Arellano-Valle *et al.*, 2006), the multivariate skew-t distribution has emerged as a tractable and robust

The first author was partially supported by grant FONDECYT 1085241-Chile. The second author was partially supported by NSF grants DMS-0504896, CMG ATM-0620624, and by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

Received September 2009 and revised October 2009.

model with parameters to regulate skewness and kurtosis in the data. Azzalini and Genton (2008) have advocated its use by means of various theoretical arguments and illustrative examples. Nevertheless, the skew-t distribution has some shortcomings and we propose to address them in this article.

First, the skew-t distribution is not closed under conditioning. This is an important issue for methods that require the study of the conditional distribution of a skew-t model fitted to the data. For instance, various data perturbation methods require to simulate from the conditional distribution of confidential variables given non-confidential ones, see Muralidhar and Sarathy (2003) for a recent review and references therein. Similarly, certain imputation methods for censored data necessitate to replace censored observations by values simulated from a conditional distribution, see for example Park *et al.* (2007) for the case of censored time series. Second, although the skew-t distribution has the ability to model distributional tails that are heavier than the normal, it cannot represent lighter tails. Those issues can be circumvented by the following definition of an extended version of the skew-t distribution.

Definition 1 (Extended skew-t). A continuous p-dimensional random vector Y has a multivariate extended skew-t (EST) distribution, denoted by $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$, if its density function at $y \in \mathbb{R}^p$ is

$$\frac{1}{T_1(\tau/\sqrt{1+\lambda^{\mathrm{T}}\bar{\Omega}\lambda};\nu)}\, t_p(y;\xi,\Omega,\nu)\, T_1\left\{\left(\lambda^{\mathrm{T}}z+\tau\right)\left(\frac{\nu+p}{\nu+Q(z)}\right)^{1/2};\nu+p\right\}, \quad (1)$$

where $z = \omega^{-1}(y-\xi)$, $Q(z) = z^{\mathrm{T}}\bar{\Omega}^{-1}z$, $\lambda \in \mathbb{R}^p$ is the shape parameter, $\tau \in \mathbb{R}$ is the extension parameter,

$$t_p(y;\xi,\Omega,\nu) = \frac{\Gamma((\nu+p)/2)}{|\Omega|^{1/2}(\nu\pi)^{p/2}\,\Gamma(\nu/2)} \left(1+\frac{Q(z)}{\nu}\right)^{-(\nu+p)/2}$$

denotes the density function of the usual p-dimensional Student's t distribution with location $\xi \in \mathbb{R}^p$, positive definite $p \times p$ dispersion matrix Ω, with $p \times p$ scale and correlation matrices $\omega = \mathrm{diag}(\Omega)^{1/2}$ and $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$, respectively, and degrees of freedom $\nu > 0$, and $T_1(y;\nu)$ denotes the univariate standard Student's t cumulative distribution function with degrees of freedom $\nu > 0$.

The EST distribution provides a very flexible class of statistical models. In fact, for $\lambda = \tau = 0$, we retrieve the multivariate symmetric Student's t distribution, $t_p(\xi, \Omega, \nu)$, which reduces to the multivariate normal $N_p(\xi, \Omega)$ and Cauchy $C_p(\xi, \Omega)$ distributions by letting $\nu \to \infty$ and $\nu = 1$, respectively. The same occurs when $\tau \to +\infty$, while for $\tau \to -\infty$ the density degenerates to zero. For $\tau = 0$, we have the multivariate skew-t distribution $ST_p(\xi, \Omega, \nu, \lambda)$ in the form adopted by Azzalini and Capitanio (2003). The multivariate skew-normal $SN_p(\xi, \Omega, \lambda)$ distribution of Azzalini and Dalla Valle (1996) and the skew-Cauchy $SC_p(\xi, \Omega, \lambda)$ distribution of Arnold and Beaver (2000a) arise when we further let $\nu \to \infty$ and $\nu = 1$, respectively. For $\lambda = 0$, the following new family of symmetric densities is obtained:

$$\frac{1}{T_1(\tau;\nu)} t_p(y;\xi,\Omega,\nu) T_1 \left\{ \tau \left(\frac{\nu + p}{\nu + Q(z)} \right)^{1/2} ; \nu + p \right\}, \tag{2}$$

which corresponds to the density of $\xi + CZ$, where $Z \stackrel{d}{=} (X|X_0 + \tau > 0)$ with $(X^{\mathrm{T}}, X_0)^{\mathrm{T}} \sim t_{p+1}(0, I_{p+1}, \nu)$, a Student's t distribution, and $C = \Omega^{1/2}$, *i.e.*, a square root of Ω. Finally, for $\nu \to \infty$ we obtain the extended skew-normal $ESN_p(\xi, \Omega, \lambda, \tau)$ distribution with density

$$\frac{1}{\Phi_1(\tau/\sqrt{1 + \lambda^{\mathrm{T}}\bar{\Omega}\lambda})} \phi_p(y;\xi,\Omega)\Phi_1\left(\lambda^{\mathrm{T}} z + \tau\right), \tag{3}$$

which arose with a slightly different parameterization in Azzalini and Capitanio (1999) and Arnold and Beaver (2000b). The ESN distribution has been studied in more detail by Capitanio *et al.* (2003), although without noticing that lighter tails than the normal distribution could be obtained for certain values of τ. We examine this interesting property in this paper. Finally, we note that Azzalini and Capitanio (2003, Section 4.2.4) also briefly sketched the definition of an extended skew-*t* distribution, although with a normalization based on the normal cumulative distribution function Φ_1 instead of T_1. Such an extension of the skew-*t* distribution does not seem to be as natural as (1).

In order to illustrate the shape of the extended skew-*t* distribution, we start by setting $p = 1$. In this case, a random variable $Y \sim EST_1(\xi, \omega^2, \lambda, \nu, \tau)$ has a density of the form

$$\frac{1}{T_1(\tau/\sqrt{1+\lambda^2};\nu)} t_1(y;\xi,\omega^2,\nu) T_1 \left\{ (\lambda z + \tau) \left(\frac{\nu + 1}{\nu + z^2} \right)^{1/2} ; \nu + 1 \right\}, \tag{4}$$

where $z = (y - \xi)/\omega$. Figure 1 depicts the density (4) for $\xi = 0$, $\omega = 1$, $\lambda = 2$, $\nu = 3$, and various values of τ. The solid thick curve corresponds to $\tau = 0$, that is, the classical ST density. For $\tau = 1, 3, 5$ (solid curves), the densities become more symmetric and approach the $t_1(y;\xi,\omega^2,\nu)$ density. For $\tau = -1, -3, -5$ (dashed curves), the densities become more skewed.

The moments of the EST distribution will be developed in this paper. For illustration, we present in Figure 2 the mean μ, variance σ^2, and coefficients of skewness $\sqrt{\beta_{1,1}}$ and kurtosis $\beta_{2,1}$ of the univariate standard extended skew-normal distribution, that is $ESN_1(0, 1, \lambda, \tau)$, as a function of τ for $\lambda = 1, 2, 3, 4, 5$. The thick curve is for $\lambda = 1$. It can be observed that for certain values of $\tau > 0$, the kurtosis $\beta_{2,1} < 0$, indicating lighter tails than the normal distribution.

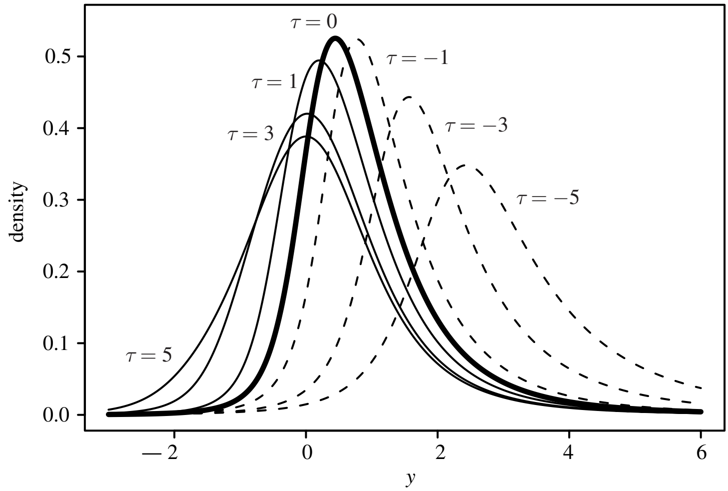

Figure 1. Densities of the $EST_1(0, 1, 2, 3, \tau)$ for $\tau = 0$ (thick solid curve), $\tau = 1, 3, 5$ (solid curves) and $\tau = -1, -3, -5$ (dashed curves).

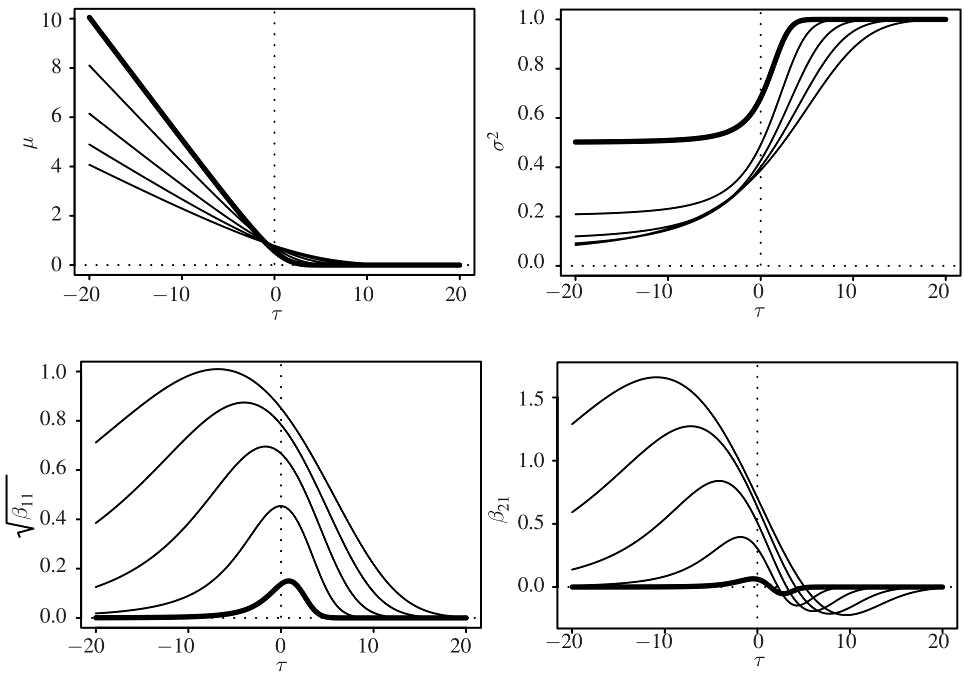

Figure 2. Mean μ, variance σ^2, and coefficients of skewness $\sqrt{\beta_{1,1}}$ and kurtosis $\beta_{2,1}$ of the univariate standard extended skew-normal distribution, that is $ESN_1(0, 1, \lambda, \tau)$, as a function of τ for $\lambda = 1, 2, 3, 4, 5$. The thick curve is for $\lambda = 1$.

Although the extended skew-t distribution of Definition 1 is appealing for practical applications, there are two natural generalizations that are of interest from a theoretical point of view. First, a continuous p-dimensional random vector Y has a multivariate unified skew-t (SUT) distribution, denoted by $Y \sim SUT_{p,q}(\xi, \Omega, \Lambda, \nu, \tau, \Gamma)$, if its density function at $y \in \mathbb{R}^p$ is

$$\frac{1}{T_q(\tau; \Gamma+\Lambda\bar{\Omega}\Lambda^{\mathrm{T}}, \nu)} t_p(y; \xi, \Omega, \nu) T_q \left\{ (\Lambda z + \tau) \left(\frac{\nu+p}{\nu+Q(z)} \right)^{1/2} ; \Gamma, \nu+p \right\}, \quad (5)$$

where now Λ is a $q \times p$ real matrix controlling shape, $\tau \in \mathbb{R}^q$ is the extension parameter, Γ is a $q \times q$ positive definite correlation matrix, and $T_r(y; \Sigma, \upsilon)$ denotes the r-dimensional centered Student's t cumulative distribution function with $r \times r$ positive definite dispersion matrix Σ and degrees of freedom $\upsilon > 0$. For $q = 1$, we have that $\Gamma = 1$, $\Lambda = \lambda^{\mathrm{T}}$ and $T_q(x; a, \nu) = T_1(x/\sqrt{a}; \nu)$, hence (5) reduces to (1), that is, $SUT_{p,1} = EST_p$. When $\nu \to \infty$ it reduces to the following unified skew-normal (SUN) density

$$\frac{1}{\Phi_q(\tau; \Gamma + \Lambda\bar{\Omega}\Lambda^{\mathrm{T}})} \phi_p(y; \xi, \Omega)\, \Phi_q\left(\Lambda z + \tau; \Gamma\right),$$

which is denoted by $SUN_{p,q}(\xi, \Omega, \Lambda, \tau, \Gamma)$ and was studied in Arellano-Valle and Azzalini (2006) with a different but equivalent parameterization. Although numerical algorithms exist to evaluate the multivariate cumulative distribution functions Φ_q and T_q for $q > 1$, see Genz (1992) and Genz and Bretz (2002), this aspect is less appealing from a practical point of view than using $q = 1$. However, when $p = q$ and the matrices $\bar{\Omega}$, Λ, $\Gamma = I_p$ are diagonal, the $SUN_{p,p}$ distribution reduces to the product of p ESN_1 distributions, hence it includes the particular case of independent and identically distributed $ESN_1(\xi, \omega^2, \lambda, \tau)$ random variables. The SUT and the SUN are only two important subclass of a more general family originating from elliptically contoured (EC) distributions (see, *e.g.*, Fang *et al.*, 1990) that we describe next. A continuous p-dimensional random vector Y has a multivariate unified skew-elliptical (SUE) distribution, denoted by $Y \sim SUE_{p,q}(\xi, \Omega, \Lambda, h^{(p+q)}, \tau, \Gamma)$, if its density function at $y \in \mathbb{R}^p$ is

$$\frac{1}{F_q(\tau; \Gamma + \Lambda\bar{\Omega}\Lambda^T, h^{(q)})} f_p(y; \xi, \Omega, h^{(p)})\, F_q\left(\Lambda z + \tau; \Gamma, h^{(q)}_{Q(z)}\right),$$

where $f_p(y; \xi, \Omega, h^{(p)}) = |\Omega|^{-1/2} h^{(p)}(Q(z))$ denotes the density function of an elliptically contoured distribution with location $\xi \in \mathbb{R}^p$, positive definite $p \times p$ dispersion matrix Ω, and density generator $h^{(p)}$, $F_r(y; \Sigma, h^{(r)})$ denotes the r-dimensional centered elliptical cumulative distribution function with $r \times r$ dispersion matrix Σ and density generator $h^{(q)}$, and $h^{(q)}_{Q(z)}(u) = h^{(p+q)}(u +$

$Q(z))/h^{(p)}(Q(z))$. The SUE distribution was also considered in Arellano-Valle and Azzalini (2006) with a different but equivalent parameterization, see also Arellano-Valle *et al.* (2006), Arellano-Valle and Genton (2005), and González-Farías *et al.* (2004). However, these authors did not study systematically the main properties of the SUE distributions, for instance such as marginal and conditional distributions, and moments. Except for the factorization of the dispersion matrix Ω, we are using a similar parametrization as González-Farías *et al.* (2004) for the closed skew-normal distribution. The SUE distribution reduces to the SUT when $h^{(p)}$ is the p-dimensional Student's t density generator function with ν degrees of freedom, that is,

$$h^{(p)}(u) = c(\nu, p)\,\{1 + (u/\nu)\}^{-(\nu+p)/2}, \quad \text{where} \quad c(a,b) = \frac{\Gamma((a+b)/2)}{\Gamma(a/2)(\pi a)^{b/2}}. \tag{6}$$

Moreover, when $q = 1$ the SUE distributions will be called extended skew-elliptical (ESE) distributions by analogy with Definition 1. They reduce directly to the EST distributions when adopting the Student's t density generator function (6).

The paper is now organized as follows. We present the probabilistic properties of the EST distribution in Section 2. Their proofs are given in the Appendix. Those theoretical properties include various stochastic representations, marginal and conditional distributions, linear transformations, moments and in particular Mardia's measures of multivariate skewness and kurtosis. In Section 3, we study statistical properties of the EST distribution, such as likelihood inference, behavior of the profile log-likelihood, the score vector and the Fisher information matrix. Especially, the Fisher information matrix of the univariate EST distribution (hence also the ST) is shown to be non-singular when the skewness is set to zero, unlike the case of the ESN distribution (hence also the SN). Finally, we present a numerical application of the conditional EST distribution in the context of confidential data perturbation.

2. Probabilistic properties

2.1. *Stochastic representations*

We present three stochastic representations of the EST distribution. They are useful for the generation of random numbers and for deriving moments and other formal properties.

Proposition 1 (Selection representation of EST). *Let $Y = \xi + \omega Z$, where*

$$Z \overset{d}{=} (X | X_0 < \lambda^{\mathrm{T}} X + \tau) \tag{7}$$

and

$$\begin{pmatrix} X \\ X_0 \end{pmatrix} \sim t_{p+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \bar{\Omega} & 0 \\ 0 & 1 \end{pmatrix}, \nu\right). \tag{8}$$

Then $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$.

The stochastic representation (7) is a natural extension of a stochastic representation considered by Azzalini (1985) to introduce the univariate skew-normal distribution. There are however two other equivalent and convenient stochastic representations for Z, which yield the well-known δ-parameterization of the shape parameter. These stochastic representations are given next.

First, let $\tilde{X}_0 = (1 + \lambda^{\mathrm{T}}\bar{\Omega}\lambda)^{-1/2}(\lambda^{\mathrm{T}} X - X_0)$ and $\bar{\tau} = (1 + \lambda^{\mathrm{T}}\bar{\Omega}\lambda)^{-1/2}\tau$. Then

$$Z \overset{d}{=} (X | \tilde{X}_0 + \bar{\tau} > 0), \tag{9}$$

where

$$\begin{pmatrix} X \\ \tilde{X}_0 \end{pmatrix} \sim t_{p+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \bar{\Omega} & \delta \\ \delta^{\mathrm{T}} & 1 \end{pmatrix}, \nu\right), \quad \text{and} \quad \delta = \frac{\bar{\Omega}\lambda}{\sqrt{1 + \lambda^{\mathrm{T}}\bar{\Omega}\lambda}}. \tag{10}$$

The representation (9), sometimes called a conditioning method, was used by Azzalini and Dalla Valle (1996) to introduce the multivariate skew-normal distribution and also by Branco and Dey (2001) to extend this model to skew-elliptical distributions. This stochastic representation yields the δ-parameterization on $(-1, 1)^p$ of the shape parameter. Therefore, we need to make a reparameterization from δ to λ in order to express the model in term of the λ-parameterization in $\mathbb{R}^p$. Another advantage of this stochastic representation is that it allows to obtain the distribution of functions $\psi(Z)$, for example such as linear functions, in a simple way, because $\psi(Z) \overset{d}{=} (\psi(X) | \tilde{X}_0 + \bar{\tau} > 0)$.

Second, let $\tilde{X} = X - \delta\tilde{X}_0$. Then

$$Z \overset{d}{=} [(\tilde{X} + \delta\tilde{X}_0) | \tilde{X}_0 + \bar{\tau} > 0], \tag{11}$$

where

$$\begin{pmatrix} \tilde{X} \\ \tilde{X}_0 \end{pmatrix} \sim t_{p+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \bar{\Omega} - \delta\delta^{\mathrm{T}} & 0 \\ 0 & 1 \end{pmatrix}, \nu\right).$$

The representation (11) reduces to $Z \overset{d}{=} \tilde{X} + \delta\tilde{X}_{(-\bar{\tau},\infty)}$, where $\tilde{X}_{(-\bar{\tau},\infty)} \overset{d}{=} (\tilde{X}_0 | \tilde{X}_0 + \bar{\tau} > 0)$, when $\tilde{X}$ and $\tilde{X}_0$ are independent, which holds for the normal distribution (*i.e.*, $\nu \to \infty$) only. In that case, it is sometimes called a convolution method and is useful to compute moments, to perform simulations, and also to implement both the EM algorithm and MCMC procedures. For additional comments about the different but equivalent representations (7), (9) and (11), see Arellano-Valle and Azzalini (2006) and Arellano-Valle *et al.* (2006). For an alternative definition of the EST distribution based on the representation (11), see Adcock (2010). We give in Proposition 2 below a novel and very convenient convolution type of representation for the EST random vector Z, which is based on (11).

Proposition 2 (Convolution type stochastic representation of EST). *Let Z be a $p \times 1$ random vector defined by*

$$Z = \sqrt{\frac{\nu + T^2_{(-\bar{\tau},\infty)}}{\nu + 1}} X_1 + \delta T_{(-\bar{\tau},\infty)}, \tag{12}$$

where $X_1 \sim t_p(0, \bar{\Omega} - \delta\delta^{\mathrm{T}}, \nu + 1)$ and is independent of $T_{(-\bar{\tau},\infty)} \stackrel{d}{=} (X_0 | X_0 + \bar{\tau} > 0)$, with $X_0 \sim t(0, 1, \nu)$. Then $Z \sim EST_p(0, \bar{\Omega}, \nu, \lambda, \bar{\tau})$.

The cumulative distribution function of the EST distribution can also easily be computed from (7) or (9). Indeed, if $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$ then

$$P(Y \leq y) = \frac{1}{T_1(\bar{\tau}; \nu)} T_{p+1}\left(\begin{pmatrix} z \\ \bar{\tau} \end{pmatrix}; \begin{pmatrix} \bar{\Omega} & -\delta \\ -\delta^{\mathrm{T}} & 1 \end{pmatrix}, \nu + 1\right),$$

where $z = \omega^{-1}(y - \xi)$ and $\bar{\tau}$ and δ are defined above.

Let $Z = V^{-1/2} Z_0$, where the random scale factor V is defined on $(0, \infty)$, $Z_0 \sim ESN_p(0, \bar{\Omega}, \lambda, \tau)$ and they are independent. By conditioning on $V = v$, we have $(Z | V = v) \sim ESN_p(0, v^{-1}\bar{\Omega}, \lambda, \tau)$, that is, a conditional ESN density given by

$$f_{Z|V=v}(z) = \frac{1}{\Phi_1(\bar{\tau})} \phi_p(z; v^{-1}\bar{\Omega}) \Phi_1(v^{1/2}\lambda^{\mathrm{T}} z + \tau). \tag{13}$$

Thus, if for example we assume in (13) that $V \sim Gamma(\nu/2, \nu/2)$, then we can conclude that the ST distribution (but not the EST distribution) can be expressed as a scale mixture of the skew-normal distribution. In other words, from (13) and $E\{f_{Z|V}(z)\} = f_Z(z)$ we have for $\tau = 0$ that:

$$E\left\{V^{p/2}\phi_p(V^{1/2}z; \bar{\Omega})\Phi_1(V^{1/2}\lambda^{\mathrm{T}}z)\right\} = t_p(z; \bar{\Omega}, \nu) T_1\left\{\lambda^{\mathrm{T}} z \left(\frac{\nu + p}{\nu + Q(z)}\right)^{1/2}; \nu + p\right\}.$$

Hence, if $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$, then only for $\tau = 0$ we obtain the following well-known stochastic representation of the ST distribution as a scale mixture of the SN distribution:

$$Y \stackrel{d}{=} \xi + \omega V^{-1/2} \tilde{Z},$$

where $V \sim Gamma(\nu/2, \nu/2)$ and is independent of $\tilde{Z} \sim SN_p(0, \bar{\Omega}, \lambda)$.

2.2. Distribution theory

We start by describing the marginal distribution of the EST.

Proposition 3 (Marginal distribution of EST). *Let $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$. Consider the partition $Y^{\mathrm{T}} = (Y_1^{\mathrm{T}}, Y_2^{\mathrm{T}})$ with $\dim(Y_1) = p_1$, $\dim(Y_2) = p_2$, $p_1 + p_2 = p$, and the corresponding partition of the parameters (ξ, Ω, λ). Then $Y_i \sim EST_{p_i}(\xi_i, \Omega_{ii}, \lambda_{i(j)}, \nu, \tau_{i(j)})$ for $i = 1, 2$, where*

$$\lambda_{i(j)} = \frac{\lambda_i + \bar{\Omega}_{ii}^{-1}\bar{\Omega}_{ij}\lambda_j}{\sqrt{1 + \lambda_j^T \tilde{\Omega}_{jj\cdot i}\lambda_j}}, \quad \tau_{i(j)} = \frac{\tau}{\sqrt{1 + \lambda_j^{\mathrm{T}} \tilde{\Omega}_{jj\cdot i}\lambda_j}}, \quad \tilde{\Omega}_{jj\cdot i} = \bar{\Omega}_{jj} - \bar{\Omega}_{ji}\bar{\Omega}_{ii}^{-1}\bar{\Omega}_{ij},$$

for $i = 1, 2$ and $j = 2, 1$.

Note that $\lambda_2 = 0$ with $\lambda_1 \neq 0$ does not imply symmetry in the marginal distribution of Y_2 unlike $\delta_2 = 0$ in the parameterizarion defined by (10). In fact, a necessary and sufficient condition to obtain $\lambda_{2(1)} = 0$ is that $\lambda_2 = -\bar{\Omega}_{11}^{-1}\bar{\Omega}_{12}\lambda_1$. Next, we present the conditional distribution of the EST and emphasize that the resulting distribution remains in the original family.

Proposition 4 (Conditional distribution of EST). *Let $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$. Consider the partition $Y^{\mathrm{T}} = (Y_1^{\mathrm{T}}, Y_2^{\mathrm{T}})$ with $\dim(Y_1) = p_1$, $\dim(Y_2) = p_2$, $p_1 + p_2 = p$, and the corresponding partition of the parameters. Then*

$$(Y_2|Y_1 = y_1) \sim EST_{p_2}\left(\xi_{2\cdot 1}, \alpha_{Q_1}\Omega_{22\cdot 1}, \lambda_{2\cdot 1}, \nu_{2\cdot 1}, \tau_{2\cdot 1}^*\right),$$

where $\xi_{2\cdot 1} = \xi_2 + \Omega_{21}\Omega_{11}^{-1}(y_1 - \xi_1)$, $Q_1(z_1) = z_1^{\mathrm{T}}\bar{\Omega}_{11}^{-1}z_1$, $z_1 = \omega_1^{-1}(y_1 - \xi_1)$, $\Omega_{22\cdot 1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$, $\lambda_{2\cdot 1} = \omega_{2\cdot 1}\omega_2^{-1}\lambda_2$, $\omega_2 = \mathrm{diag}(\Omega_{22})^{1/2}$, $\omega_{2\cdot 1} = \mathrm{diag}(\Omega_{22\cdot 1})^{1/2}$, $\nu_{2\cdot 1} = \nu + p_1$, $\tau_{2\cdot 1} = (\lambda_2^{\mathrm{T}}\bar{\Omega}_{21}\bar{\Omega}_{11}^{-1} + \lambda_1^{\mathrm{T}})z_1 + \tau$, $\tau_{2\cdot 1}^* = \alpha_{Q_1}^{-1/2}\tau_{2\cdot 1}$ and $\alpha_{Q_1} = (\nu + Q_1(z_1))/(\nu + p_1)$.

Unlike the marginal distribution of Y_2, we note for the conditional distribution of Y_2 given $Y_1 = y_1$ that $\lambda_2 = 0$ implies symmetry. Another interesting consequence of the result in Proposition 4 is that even if one starts with an ST distribution, that is an EST with $\tau = 0$, then the resulting conditional distribution has an extension parameter $\tau_{2\cdot 1}^* \neq 0$, unless the condition $\lambda_1 = -\bar{\Omega}_{11}^{-1}\bar{\Omega}_{12}\lambda_2$ holds or $z_1 = 0$. In order to illustrate this effect, consider the bivariate random vector $Y \sim EST_2(0, I_2, \lambda, \nu, 0) \equiv ST_2(0, I_2, \lambda, \nu)$. Then the conditional distribution is

$$(Y_2|Y_1 = y_1) \sim EST_1\left(0, \frac{\nu + y_1^2}{\nu + 1}, \lambda_2, \nu + 1, \sqrt{\frac{\nu + 1}{\nu + y_1^2}}\lambda_1 y_1\right).$$

Therefore, we have $\lim_{y_1 \to \pm\infty} \tau_{2\cdot 1}^* = \pm\lambda_1\sqrt{\nu + 1}$ and the new conditional distribution can become more symmetric or more skewed depending on the sign of λ_1. Larger values of $y_1 > 0$ tend to produce more symmetric conditional densities in this setting.

Next, we describe the distribution of linear transformations of random vectors having an EST distribution.

Proposition 5 (Linear transformation of EST). *Let $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$. Then $AY + b \sim EST_r(\xi_A, \Omega_A, \lambda_A, \nu, \tau_A)$ for any $r \times p$ matrix A of rank $r \leq p$ and $r \times 1$ vector b, where $\xi_A = A\xi + b$, $\Omega_A = A\Omega A^{\mathrm{T}}$, with scale and correlation matrices $\omega_A = \mathrm{diag}(\Omega_A)^{1/2}$ and $\bar{\Omega}_A = \omega_A^{-1}\Omega_A\omega_A^{-1}$, respectively, and*

$$\lambda_A = \gamma_A^{-1}\bar{\lambda}_A, \quad \bar{\lambda}_A = \bar{\Omega}_A^{-1}\omega_A^{-1}A\omega\bar{\Omega}\lambda, \quad \tau_A = \gamma_A^{-1}\tau, \quad \gamma_A = (1 + \lambda^{\mathrm{T}}\bar{\Omega}\lambda - \bar{\lambda}_A^{\mathrm{T}}\bar{\Omega}_A\bar{\lambda}_A)^{1/2}.$$

In particular, if $r = p$, i.e., A is a $p \times p$ non-singular matrix, then $AY + b \sim EST_p(A\xi + b, A\Omega A^{\mathrm{T}}, \omega_A(A^{-1})^{\mathrm{T}}\omega^{-1}\lambda, \nu, \tau)$.

Considering the special case $AY = Y_1 + Y_2$ with $Y = (Y_1^{\mathrm{T}}, Y_2^{\mathrm{T}})^{\mathrm{T}} \sim EST_{p_1+p_2}(\xi = (\xi_1^{\mathrm{T}}, \xi_2^{\mathrm{T}})^{\mathrm{T}}, \Omega = \mathrm{diag}(\Omega_1, \Omega_2), \lambda = (\lambda_2^{\mathrm{T}}, \lambda_2^{\mathrm{T}})^{\mathrm{T}}, \nu, \tau)$ and $p_1 = p_2 = r$, we obtain by Proposition 3 that $Y_i \sim EST_r(\xi_i, \Omega_i, (1+\lambda_j^{\mathrm{T}}\bar{\Omega}_j\lambda_j)^{-1/2}\lambda_i, \nu, (1+\lambda_j^{\mathrm{T}}\bar{\Omega}_j\lambda_j)^{-1/2}\tau)$, $i, j = 1, 2$, and by Proposition 5 that

$$Y_1 + Y_2 \sim EST_r(\xi_1 + \xi_2, \Omega_1 + \Omega_2, \gamma_+^{-1}\lambda_+, \nu, \gamma_+^{-1}\tau),$$

where $\lambda_+ = (\bar{\Omega}_1 + \bar{\Omega}_2)^{-1}(\omega_1^2 + \omega_2^2)^{-1}(\omega_1\bar{\Omega}_1\lambda_1 + \omega_2\bar{\Omega}_2\lambda_2)$, and $\gamma_+ = \{1 + \lambda_1^{\mathrm{T}}\bar{\Omega}_1\lambda_1 + \lambda_2^{\mathrm{T}}\bar{\Omega}_2\lambda_2 - \lambda_+^{\mathrm{T}}(\bar{\Omega}_1 + \bar{\Omega}_2)\lambda_+\}^{1/2}$. However, if Y_1 and Y_2 are independent EST random vectors, then the distribution of $Y_1 + Y_2$ does neither belong to the EST nor to the SUT classes anymore, unless $\nu = +\infty$. In fact, as was shown in González-Farías *et al.* (2004), the distribution of the sum of independent SUN random vectors remains in the SUN class. Another relevant example is given by the standardized EST version $Z = \Omega^{-1/2}(Y - \xi) \sim EST_p(0, I_p, \bar{\lambda}, \nu, \tau)$, where $\bar{\lambda} = \bar{\Omega}^{1/2}\lambda$, for which $A = \Omega^{-1/2}$ and $b = -\Omega^{-1/2}\xi$. In addition, if we consider a $p \times p$ orthogonal matrix Γ such that $\Gamma\bar{\lambda} = \|\bar{\lambda}\|e_1$, with e_1 being the first unitary p-dimensional vector, then we obtain the canonical EST representation given by $\Gamma Z \sim EST_p(0, I_p, \|\bar{\lambda}\|e_1, \nu, \tau)$.

2.3. *Moments*

To compute the moments of the EST distribution, we can use directly the new stochastic representations (12). We start by computing the moments for the univariate case defined by $p = 1$. To this end, we use the results given in the following lemma.

Lemma 1. *Let $T = V^{-1/2}N$, where $V \sim Gamma(\nu/2, \nu/2)$ and $N \sim N(0,1)$ and they are independent. Let also $T_{(a,b)} \overset{d}{=} (T|a < T < b)$ and $N_{(a,b)} \overset{d}{=} (N|a < N < b)$, where $T \sim t(0,1,\nu)$ and $N \sim N(0,1)$. Then*

$$E(V^{-k/2}\phi_1(\sqrt{V}\,a)) = \left(\frac{\nu}{2}\right)^{k/2} \frac{\Gamma[(\nu-k)/2](1+(a^2/\nu))^{-(\nu-k)/2}}{(2\pi)^{1/2}\Gamma(\nu/2)}, \quad \nu > k \geq 0, \quad (14)$$

$$E(V^{-k/2}\Phi_1(\sqrt{V}\,a)) = \left(\frac{\nu}{2}\right)^{k/2} \frac{\Gamma[(\nu-k)/2]}{\Gamma(\nu/2)} T_1\left(\sqrt{\frac{\nu-k}{\nu}}a; \nu-k\right), \quad \nu > k \geq 0. \quad (15)$$

Moreover, for any integrable function h,

$$E(h(T_{(a,b)})) = \frac{E\{\left[\Phi_1(V^{1/2}b) - \Phi_1(V^{1/2}a)\right] E[h\left(V^{-1/2}N_{(V^{1/2}a, V^{1/2}b)}\right)|V]\}}{T_1(b;\nu) - T_1(a;\nu)}.$$

In particular, for $h(x) = x^k$, we have

$$E(T^k_{(a,b)}) = \frac{E\{V^{-k/2}\left[\Phi_1(V^{1/2}b) - \Phi_1(V^{1/2}a)\right] E[N^k_{(V^{1/2}a, V^{1/2}b)}|V]\}}{T_1(b;\nu) - T_1(a;\nu)}, \quad \nu > k.$$

Proposition 6 (Moments of univariate ESN). *Let $Z \sim ESN_1(0,1,\lambda,\tau)$. Define $N_{(-a,\infty)} \overset{d}{=} (N_0|N_0 + a > 0)$ where $N_0 \sim N(0,1)$. The moments $\mu_k = E(Z^k)$, $k = 1,2,3,4$, are given by*:

$$\begin{aligned}
\mu_1 &= \delta E(N_{(-\bar{\tau},\infty)}),\\
\mu_2 &= (1-\delta^2) + \delta^2 E(N^2_{(-\bar{\tau},\infty)}),\\
\mu_3 &= 3\delta(1-\delta^2)E(N_{(-\bar{\tau},\infty)}) + \delta^3 E(N^3_{(-\bar{\tau},\infty)}),\\
\mu_4 &= 3(1-\delta^2)^2 + 6\delta^2(1-\delta^2)E(N^2_{(-\bar{\tau},\infty)}) + \delta^4 E(N^4_{(-\bar{\tau},\infty)}),
\end{aligned}$$

where:

$$\begin{aligned}
&E(N_{(-a,\infty)}) = \zeta_1(a) \equiv \frac{\phi_1(a)}{\Phi_1(a)}, \quad E(N^2_{(-a,\infty)}) = 1 - a\zeta_1(a),\\
&E(N^3_{(-a,\infty)}) = (2+a^2)\zeta_1(a), \quad E(N^4_{(-a,\infty)}) = 3 - (3a + a^3)\zeta_1(a).
\end{aligned}$$

Proposition 7 (Moments of univariate EST). *Let $Z \sim EST_1(0,1,\lambda,\nu,\tau)$. Define $T_{(-a,\infty)} \overset{d}{=} (T_0|T_0 + a > 0)$ where $T_0 \sim t(0,1,\nu)$. The moments $\mu_k =$*

$E(Z^k)$, $k = 1, 2, 3, 4$, *are given by*:

$$\mu_1 = \delta E(T_{(-\bar{\tau},\infty)}), \quad \nu > 1,$$
$$\mu_2 = \frac{\nu}{\nu - 1}(1 - \delta^2)\left(1 + \frac{1}{\nu}E(T^2_{(-\bar{\tau},\infty)})\right) + \delta^2 E(T^2_{(-\bar{\tau},\infty)}), \quad \nu > 2,$$
$$\mu_3 = \frac{3\nu}{\nu - 1}\delta(1 - \delta^2)\left(E(T_{(-\bar{\tau},\infty)}) + \frac{1}{\nu}E(T^3_{(-\bar{\tau},\infty)})\right) + \delta^3 E(T^3_{(-\bar{\tau},\infty)}), \quad \nu > 3,$$
$$\mu_4 = \frac{3\nu^2}{(\nu - 1)(\nu - 3)}(1 - \delta^2)^2\left(1 + \frac{2}{\nu}E(T^2_{(-\bar{\tau},\infty)}) + \frac{1}{\nu^2}E(T^4_{(-\bar{\tau},\infty)})\right)$$
$$+ \frac{6\nu}{\nu - 1}\delta^2(1 - \delta^2)\left(E(T^2_{(-\bar{\tau},\infty)}) + \frac{1}{\nu}E(T^4_{(-\bar{\tau},\infty)})\right) + \delta^4 E(T^4_{(-\bar{\tau},\infty)}), \quad \nu > 4,$$

where:

$$E(T_{(-\bar{\tau},\infty)}) = \frac{\nu}{\nu - 1}\left(1 + \frac{\bar{\tau}^2}{\nu}\right)\frac{t_1(\bar{\tau};\nu)}{T_1(\bar{\tau};\nu)}, \quad \nu > 1,$$
$$E(T^2_{(-\bar{\tau},\infty)}) = \frac{\nu}{\nu - 2}\frac{T_1(\bar{\tau}_{-2};\nu - 2)}{T_1(\bar{\tau};\nu)} - \bar{\tau}\, E(T_{(-\bar{\tau},\infty)}), \quad \nu > 2,$$
$$E(T^3_{(-\bar{\tau},\infty)}) = \frac{2\nu^2}{(\nu - 1)(\nu - 3)}\left(1 + \frac{\bar{\tau}^2}{\nu}\right)^2\frac{t_1(\bar{\tau};\nu)}{T_1(\bar{\tau};\nu)} + \bar{\tau}^2\, E(T_{(-\bar{\tau},\infty)}), \quad \nu > 3,$$
$$E(T^4_{(-\bar{\tau},\infty)}) = \frac{3\nu^2}{(\nu - 2)(\nu - 4)}\frac{T_1(\bar{\tau}_{-4};\nu - 4)}{T_1(\bar{\tau};\nu)} - \frac{3}{2}\bar{\tau}\, E(T^3_{(-\bar{\tau},\infty)})$$
$$+ \frac{1}{2}\bar{\tau}^3 E(T_{(-\bar{\tau},\infty)}), \quad \nu > 4,$$

with $\bar{\tau}_r = \sqrt{\frac{\nu+r}{\nu}}\,\bar{\tau}$, $\nu + r > 0$.

For $\tau = 0$, the moments μ_k in Proposition 7 reduce to those of the classical $ST_1(0, 1, \lambda, \nu)$ distribution, obtained by Azzalini and Capitanio (2003, Section 4.2.2). If instead $\lambda = 0$, the moments μ_k reduce to those of the symmetric univariate distribution that arises from (2) when $\xi = 0$ and $\Omega = 1$, *i.e.*, to the moments of $Z \overset{d}{=} (X|X_0 + \bar{\tau} > 0)$, where $(X, X_0)^{\mathrm{T}} \sim t_2(0, I_2, \nu)$.

The mean μ, variance σ^2, and coefficients of skewness $\sqrt{\beta_{1,1}}$ and kurtosis $\beta_{2,1}$ of the univariate ESN and EST distributions can then be computed by means of Propositions 6 and 7 and their well-known relations to the moments μ_k, see, *e.g.*, Stuart and Ord (1987, p. 73). Also, in terms of the truncated moments $\mu_{*k} = E(T^k_{(-\bar{\tau},\infty)})$, $k = 1, 2, 3, 4$, we have the following relations for the EST distribution:

$$\mu = \delta\mu_*, \quad \nu > 1, \tag{16}$$
$$\sigma^2 = \delta^2\sigma_*^2 + (1-\delta^2)a_*^2, \qquad \nu > 2, \tag{17}$$
$$\sqrt{\beta_{1,1}} = \delta^3\left(\frac{\sigma_*}{\sigma}\right)^3\sqrt{\beta_{1,1}^*} + \frac{3}{\nu-1}\delta(1-\delta^2)\,\frac{\mu_{*3}-\mu_{*2}\mu_{*1}}{\sigma^3}, \quad \nu > 3, \tag{18}$$
$$\begin{aligned}\beta_{2,1} &= \delta^4\left(\frac{\sigma_*}{\sigma}\right)^4(\beta_{2,1}^*+3)\\ &\quad+\frac{6\nu}{\nu-1}\delta^2(1-\delta^2)\frac{1}{\sigma^4}\left[\sigma_*^2+\frac{1}{\nu}\left(\mu_{*4}-2\mu_{*3}\mu_{*1}+\mu_{*2}\mu_{*1}^2\right)\right]\\ &\quad+\frac{3\nu^2}{(\nu-1)(\nu-3)}(1-\delta^2)^2\frac{1}{\sigma^4}\left(1+\frac{2}{\nu}\mu_{*2}+\frac{1}{\nu^2}\mu_{*4}\right), \quad \nu > 4,\end{aligned} \tag{19}$$

where $a_*^2 = (\nu/(\nu-1))(1+\mu_{*2}/\nu)$ and $\mu_* = \mu_{*1}$, $\sigma_*^2 = \mu_{*2} - \mu_{*1}^2$, $\sqrt{\beta_{1,1}^*} = (\mu_{*3} - 3\mu_{*2}\mu_{*1} + 2\mu_{*1}^3)/\sigma_*^3$ and $\beta_{2,1}^* = (\mu_{*4} - 4\mu_{*3}\mu_{*1} + 6\mu_{*2}\mu_{*1}^2 - 3\mu_{*1}^4)/\sigma_*^4$ are, respectively, the mean, variance, skewness and kurtosis of $T_{(-\bar{\tau},\infty)}$. The quantities in (16)-(19) reduce to the corresponding ones for the symmetric distribution (2) when $\delta = 0$, while for $\delta = \pm 1$ (or $\lambda \to \pm\infty$) they converge to the corresponding ones for the truncated random variable $T_{(-\bar{\tau},\infty)}$. Also, when $\nu \to \infty$, these quantities converges to the corresponding ESN ones. That is, as $\nu \to \infty$,

$$\begin{aligned}\mu &\to \delta\zeta_1(\bar{\tau}),\\ \sigma^2 &\to 1-\delta^2(\bar{\tau}+\zeta_1(\bar{\tau}))\zeta_1(\bar{\tau}),\\ \sqrt{\beta_{1,1}} &\to \frac{\delta^3\{(\bar{\tau}^2-1)\zeta_1(\bar{\tau})+3\bar{\tau}[\zeta_1(\bar{\tau})]^2+2[\zeta_1(\bar{\tau})]^3\}}{[1-\delta^2(\bar{\tau}+\zeta_1(\bar{\tau}))\zeta_1(\bar{\tau})]^{3/2}},\\ \beta_{2,1} &\to \frac{\delta^4\{3-(\bar{\tau}^3+3\bar{\tau})\zeta_1(\bar{\tau})-(4\bar{\tau}^2+2)[\zeta_1(\bar{\tau})]^2-6\bar{\tau}[\zeta_1(\bar{\tau})]^3-3[\zeta_1(\bar{\tau})]^4\}}{[1-\delta^2(\bar{\tau}+\zeta_1(\bar{\tau}))\zeta_1(\bar{\tau})]^2}\\ &\quad+\frac{6\delta^2(1-\delta^2)[1-(\bar{\tau}+\zeta_1(\bar{\tau}))\zeta_1(\bar{\tau})]}{[1-\delta^2(\bar{\tau}+\zeta_1(\bar{\tau}))\zeta_1(\bar{\tau})]^2}+\frac{3(1-\delta^2)^2}{[1-\delta^2(\bar{\tau}+\zeta_1(\bar{\tau}))\zeta_1(\bar{\tau})]^2}.\end{aligned}$$

As mentioned in the introduction, the ESN distribution can have lighter tails than the normal distribution. This is the case when $\beta_{2,1} - 3 < 0$ and Figure 3 depicts such regions as a function of λ and τ. A numerical analysis shows that $-0.243 < \beta_{2,1} - 3 < 6.182$ for an $ESN_1(0, 1, \lambda, \tau)$ distribution, whereas $0 < \beta_{2,1} - 3 < 0.869$ for a classical $SN_1(0, 1, \lambda)$ distribution. For fixed δ and τ, the kurtosis $\beta_{2,1}$ in (24) of the EST is minimized when $\nu \to \infty$, hence at the ESN distribution. Similarly, $-1.995 < \sqrt{\beta_{1,1}} < 1.995$ for an $ESN_1(0, 1, \lambda, \tau)$

and $-0.995 < \sqrt{\beta_{1,1}} < 0.995$ for an $SN_1(0, 1, \lambda)$. Therefore, the ESN has more flexibility than the SN to model skewness and kurtosis.

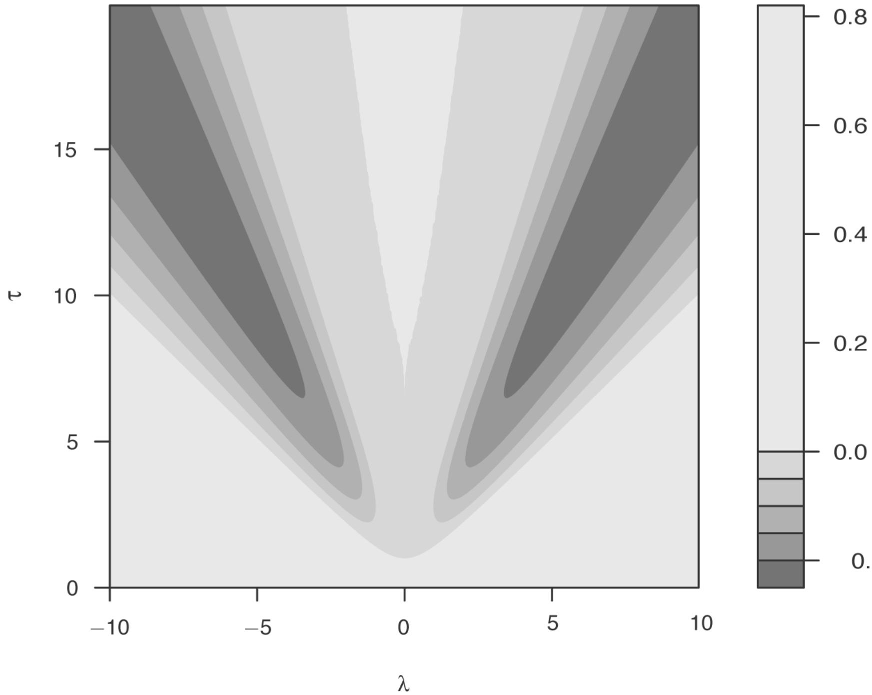

Figure 3. Contours of the kurtosis $\beta_{2,1} - 3$ as a function of λ and τ for an $ESN_1(0, 1, \lambda, \tau)$. Negative values correspond to lighter tails than the normal distribution.

In the multivariate case $p > 1$, we obtain from the stochastic representation (12) the following expressions for the mean vector and covariance matrix of $Z \sim EST_p(0, \bar{\Omega}, \lambda, \nu, \tau)$:

$$E(Z) = \delta\mu_*, \quad \nu > 1,$$
$$Var(Z) = a_*^2(\bar{\Omega} - \delta\delta^{\mathrm{T}}) + \sigma_*^2\delta\delta^{\mathrm{T}}, \quad \nu > 2,$$

where μ_*, σ_*^2 and a_*^2 are defined above. The measures of multivariate skewness and kurtosis proposed by Mardia (1970) are $\beta_{1,p} = E\{[(Y - \mu_Y)^{\mathrm{T}}\Sigma_Y^{-1}(Y' - \mu_Y)]^3\}$ and $\beta_{2,p} = E\{[(Y - \mu_Y)^{\mathrm{T}}\Sigma_Y^{-1}(Y - \mu_Y)]^2\}$, respectively, where $Y \sim EST_p(\xi, \Omega, \nu, \lambda, \tau)$, $\mu_Y = E(Y)$, $\Sigma_Y = Var(Y)$, and Y' is an independent replicate of Y.

Proposition 8 (Mardia's measures of multivariate skewness and kurtosis for the EST). *Let $Y \sim EST_p(\xi, \Omega, \nu, \lambda, \tau)$. Then,*

$$a_*^6\beta_{1,p} = \frac{3(p-1)}{(\nu-1)^2}\|\bar{\delta}\|^2(1+\alpha_*\|\bar{\delta}\|^2)(\mu_{*3}-\mu_{*2}\mu_{*1})^2+(1+\alpha_*\|\bar{\delta}\|^2)^3\bar{\sigma}^3\left(\sqrt{\bar{\beta}_{1,1}}\right)^3$$

and

$$
\begin{aligned}
&a_*^4 \beta_{2,p} \\
&= \left\{ \frac{(p-1)(p+1)\nu^2}{(\nu-1)(\nu-3)} + \frac{2(p-1)\nu^2}{(\nu-1)^2}(1+\alpha_*\|\bar{\delta}\|^2)(1-\|\bar{\delta}\|^2) \right\} \left(1+\frac{2}{\nu}\mu_{*2}+\frac{1}{\nu^2}\mu_{*4}\right) \\
&\quad + \frac{2(p-1)}{\nu-1}\|\bar{\delta}\|^2(1+\alpha_*\|\bar{\delta}\|^2)(\mu_{*3}-\mu_{*2}\mu_{*1}) + (1+\alpha_*\|\bar{\delta}\|^2)^2\bar{\sigma}^4\bar{\beta}_{2,1},
\end{aligned}
$$

where $a_*^2 = (\nu/(\nu-1))(1+\mu_{*2}/\nu)$, $\mu_{*k} = E(T^k_{(-\bar{\tau},\infty)})$, $k = 1, 2, 3, 4$,

$$
\alpha_* = \frac{a_*^2 - \sigma_*^2}{a_*^2 - (a_*^2 - \sigma_*^2)\|\bar{\delta}\|^2},
$$

with $\sigma_*^2 = Var(T_{(-\bar{\tau},\infty)})$, *and where* $\bar{\sigma}^2$, $\sqrt{\bar{\beta}_{1,1}}$ *and* $\bar{\beta}_{2,1}$ *are defined as in* (17)-(19) *but with the* μ_i*'s computed by replacing* δ *by* $\|\bar{\delta}\| = \sqrt{\delta^{\mathrm{T}}\bar{\Omega}^{-1}\delta}$.

3. Statistical properties

3.1. Likelihood inference

Consider n independent observations $y_1, \ldots, y_n$ from $Y_i \sim EST_p(\xi_i, \Omega, \lambda, \nu, \tau)$, $i = 1, \ldots, n$, with a regression structure $\xi_i = B^{\mathrm{T}} x_i$, where $x_i \in \mathbb{R}^d$ is a vector of covariates and B is a $d \times p$ matrix of parameters. The log-likelihood function $\ell(\theta) = \log L(\theta)$ to estimate the parameters $\theta = (B, \Omega, \lambda, \nu, \tau)$, based on the density (1), is given by

$$
\begin{aligned}
\ell(\theta) = \sum_{i=1}^{n} \Bigg[& \log t_p(y_i; \xi_i, \Omega, \nu) + \log T_1 \left\{ (\lambda^{\mathrm{T}} z_i + \tau) \left(\frac{\nu+p}{\nu+Q(z_i)} \right)^{1/2}; \nu+p \right\} \\
& - \log T_1(\tau/\sqrt{1+\lambda^{\mathrm{T}}\bar{\Omega}\lambda}; \nu) \Bigg],
\end{aligned}
\tag{20}
$$

where $z_i = \omega^{-1}(y_i - \xi_i)$, $i = 1, \ldots, n$. This function cannot be maximized in closed form and numerical methods are necessary. We propose to take advantage of existing procedures for the case $\tau = 0$ and suggest the following approach. First estimate θ for fixed $\tau = 0$ using, for instance, the command `mst.mle` of the `R` package `sn` developed by Azzalini (2006). Then, for a fixed increasing sequence of $\tau > 0$, maximize (20) using the previous parameter estimates as starting values in the optimization. Proceed similarly for a fixed decreasing sequence of $\tau < 0$. This leads to a profile log-likelihood function for τ which is numerically stable.

An example of the outcome of this scheme is depicted in Figure 4 for $p = 1$ and $\xi_i = \xi$. The data consists of $n = 278$ wind speeds recorded at midnight at the Vansycle meteorological tower between February 25 and November 30, 2003. This is part of a wind power production study as reported by Azzalini and Genton (2008). The EST profile log-likelihood for τ is maximized at $\tau = -2.1$, identified by the vertical dashed line in Figure 4. The likelihood ratio test of the hypothesis $H_0 : \tau = 0$ yields a p-value of 0.519, hence the classical ST distribution is preferred over the EST in this example. Nevertheless, we believe that the strength of the EST model lies in applications that require the use of the conditional distribution of the classical ST, see Section 3.3.

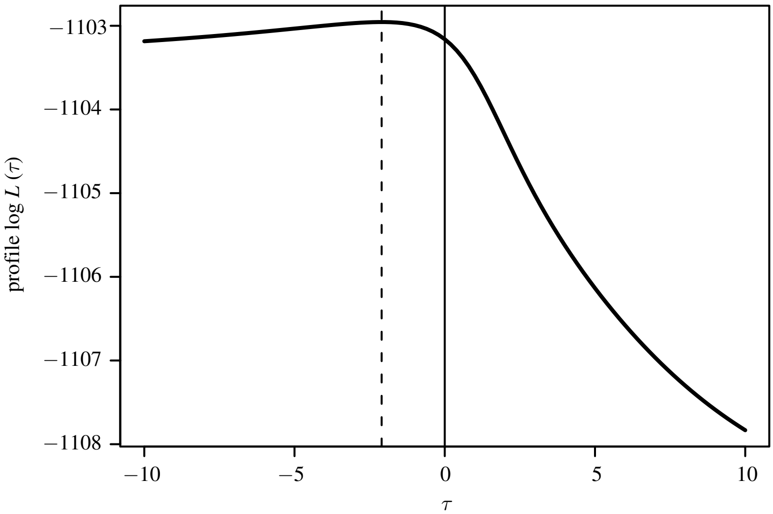

Figure 4. Profile log-likelihood for τ for the wind speed data at Vansycle. The maximum is identified by the vertical dashed line at $\tau = -2.1$.

We further investigate the fit of the ESN distribution to the wind speed data. The likelihood ratio test of the hypothesis $H_0 : \tau = 0$ yields a p-value of 0.031, indicating that the ESN distribution is preferred over the SN in this example. This is not surprising as the sample skewness and kurtosis of this dataset are $\sqrt{b_{1,1}} = -0.849$ and $b_{2,1} = 1.376$, respectively. Hence the sample kurtosis is outside the possible range offered by the SN model, whereas it is within the range for the ESN distribution, see the range values given in Section 2.3. Figure 5 depicts a histogram of the wind speed data with fitted densities: SN (thin dashed curve), ESN (thick dashed curve) with maximum likelihood estimate $\hat{\tau} = -19.3$, ST (thin solid curve), and EST (thick solid curve) with maximum likelihood estimate $\hat{\tau} = -2.1$. Clearly, the ESN model captures the central peak of the distribution and its left tail better than the SN model. Moreover, as expected from the discussion above, the EST model is not significantly different from the ST model, except for a small difference near the mode of the distribution.

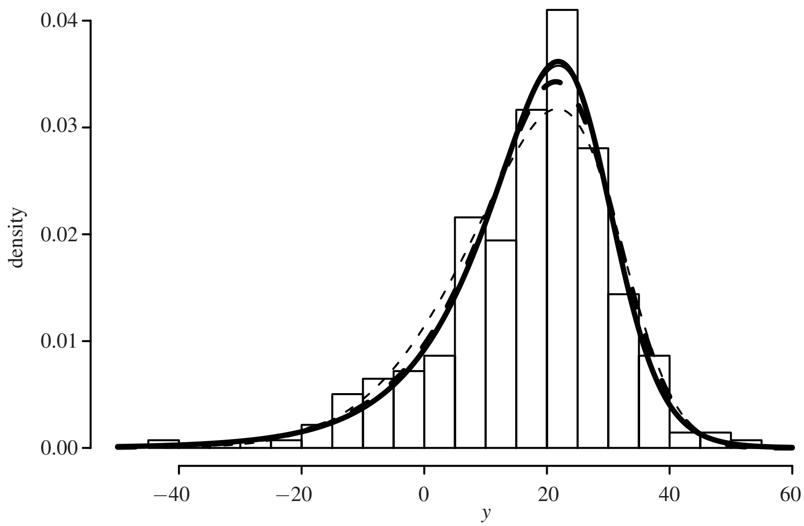

Figure 5. Histogram of wind speed data at Vansycle with fitted densities: SN (thin dashed curve), ESN (thick dashed curve), ST (thin solid curve), and EST (thick solid curve).

3.2. *Profile log-likelihood, likelihood score, and fisher information*

Azzalini (1985) and Azzalini and Capitanio (1999) have shown that the profile log-likelihood for the shape parameter of a univariate SN distribution always has a stationary point at $\lambda = 0$ and that the Fisher information matrix is singular at this point. Pewsey (2006) extended this result to generalized forms of the univariate SN distribution. This problematic feature carries over to the multivariate SN distribution as demonstrated by Azzalini and Genton (2008). However, the behavior of the profile log-likelihood for the ST distribution turns out to be much more regular; see Azzalini and Capitanio (2003) and Azzalini and Genton (2008), although a formal proof has not been given. We investigate next these issues for the ESN and EST distributions.

As an illustration, we plot in Figure 6 the profile log-likelihood for λ for the wind speed data at Vansycle described in Section 3.1 for the distributions: SN (thin dashed curve), ESN (thick dashed curve), ST (thin solid curve), and EST (thick solid curve). The vertical dotted line represents $\lambda = 0$. The triangles identify the maximum of each profile log-likelihood. The stationary point at $\lambda = 0$ seems to remain for the ESN distribution, whereas it does not occur for the EST distribution. This suggests that the singularity of the Fisher information matrix at $\lambda = 0$ will happen for the ESN distribution as well. We give a formal result next.

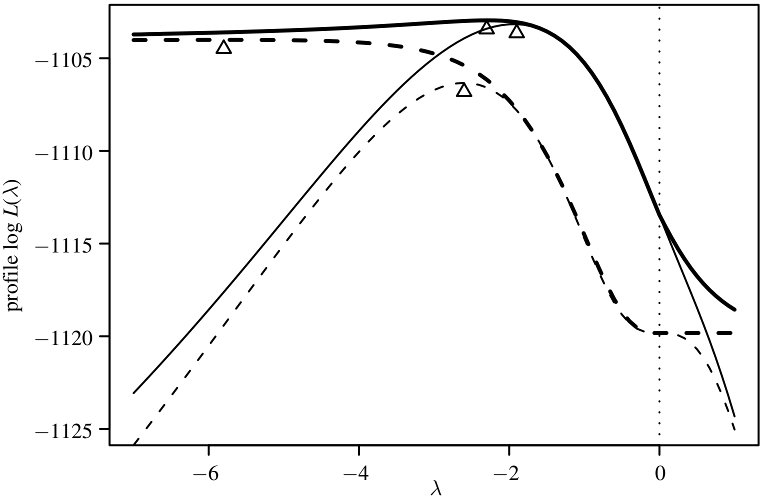

Figure 6. Profile log-likelihood for λ for the wind data at Vansycle for the distributions: SN (thin dashed curve), ESN (thick dashed curve), ST (thin solid curve), and EST (thick solid curve). The vertical dotted line represents $\lambda = 0$. The triangles identify the maximum of each profile log-likelihood.

For a random sample from an $ESN_1(\xi, \omega^2, \lambda, \tau)$ distribution, the log-likelihood (20) for $\theta = (\xi, \omega, \lambda, \tau)^{\mathrm{T}}$ reduces to:

$$\begin{aligned} \ell(\theta) = \text{constant} - n\log\omega - \frac{1}{2}\sum_{i=1}^{n} z_i^2 + \sum_{i=1}^{n} \log\Phi_1(\lambda z_i + \tau) \\ - n\log\Phi_1(\tau/\sqrt{1+\lambda^2}). \end{aligned} \tag{21}$$

We now have the following result, the proof of which is given in the Appendix.

Proposition 9 (Stationary point of profile log-likelihood and singularity of observed information of univariate ESN at $\lambda = 0$). *Denote by $y_1, \ldots, y_n$ a random sample of size $n \geq 3$ from an $ESN_1(\xi, \omega^2, \lambda, \tau)$ distribution with density given by* (3) *when $p = 1$. If we denote the sample mean by $\bar{y}$ and the sample variance by s^2, then*

(a) *$\xi = \bar{y}$, $\omega = s$, $\lambda = 0$, $\tau \in \mathbb{R}$ is a solution to the score equations for* (21)*;*
(b) *the observed information matrix is singular when $\lambda = 0$.*

Similarly, the multivariate ESN distribution with $p > 1$ also has a stationarity point of its profile log-likelihood at $\lambda = 0 \in \mathbb{R}^p$. The proof follows the steps in Azzalini and Genton (2008, Section 2.1) given for the multivariate SN distribution.

Although there is plenty of numerical evidence suggesting that the EST distribution (as well as the ST) does not have a stationary point of its profile log-likelihood at $\lambda = 0$, a formal proof is needed. In order to investigate this issue, we consider the case with $p = 1$ and derive the score vector for the $EST_1(\xi, \omega^2, \lambda, \nu, \tau)$ model based on a sample of size one. The log-likelihood function for $\theta = (\xi, \omega, \lambda, \nu, \tau)^{\mathrm{T}}$ is then

$$\begin{aligned}\ell(\theta) = c_\nu - \log\omega - (1/2)(\nu+1)\log\left(1 + z^2/\nu\right) \\ + \log T_1(q; \nu+1) - \log T_1(\bar{\tau}; \nu),\end{aligned} \tag{22}$$

where $z = (y - \xi)/\omega$, $\bar{\tau} = (1+\lambda^2)^{-1/2}\tau$, $\delta = (1+\lambda^2)^{-1/2}\lambda$, and

$$\begin{aligned}c_\nu &= \log t_1(0; \nu) = \log\,\Gamma\,((\nu+1)/2) - \log\Gamma\,(\nu/2) - (1/2)\log(\pi\nu),\\ q &= q(z; \theta) = (1 + 1/\nu)^{1/2}\left(1 + z^2/\nu\right)^{-1/2}(\lambda z + \tau).\end{aligned}$$

Denote by $r_1(x; \nu) = \partial \log\, T_1(x; \nu)/\partial x = t_1(x; \nu)/T_1(x; \nu)$. The behavior of the score vector $S_\theta = (S_\xi, S_\omega, S_\lambda, S_\nu, S_\tau)^{\mathrm{T}}$, where $S_\alpha = \partial l(\theta)/\partial\alpha$, and of the Fisher information matrix are given next.

Proposition 10 (Behavior of score vector and Fisher information of univariate EST). *Denote by y an observation from $Y \sim EST_1(\xi, \omega^2, \lambda, \nu, \tau)$. Then*:

(a) *The score vector S_θ is given by*:

$$\begin{aligned}S_\xi &= (1/\omega)(1 + 1/\nu)\left(1 + z^2/\nu\right)^{-1} z\\ &\quad + r_1(q; \nu+1)(1/\omega)\,(1 + 1/\nu)^{1/2}\left\{\left(1 + z^2/\nu\right)^{-3/2}(z/\nu)(\lambda z + \tau)\right.\\ &\quad \left. - \left(1 + z^2/\nu\right)^{-1/2}\lambda\right\},\\ S_\omega &= -(1/\omega) + (1/\omega)(1 + 1/\nu)\left(1 + z^2/\nu\right)^{-1} z^2\\ &\quad + r_1(q; \nu+1)(1/\omega)\,(1 + 1/\nu)^{1/2}\left\{\left(1 + z^2/\nu\right)^{-3/2}(z^2/\nu)(\lambda z + \tau)\right.\\ &\quad \left. - \left(1 + z^2/\nu\right)^{-1/2}\lambda z\right\},\end{aligned}$$

$$S_\lambda = r_1(q;\nu+1)\,(1+1/\nu)^{1/2}\left(1+z^2/\nu\right)^{-1/2} z + r_1(\bar\tau;\nu)(1+\lambda^2)^{-1}\lambda\bar\tau,$$

$$\begin{aligned}S_\nu = {} & c'_\nu - (1/2)\log\left(1+z^2/\nu\right) + (1/2)(1+1/\nu)\left(1+z^2/\nu\right)^{-1}(z^2/\nu)\\ & + (1/T_1(q;\nu+1))(\partial T_1(q;\nu+1)/\partial\nu) - (1/T_1(\bar\tau;\nu))(\partial T_1(\bar\tau;\nu)/\partial\nu),\end{aligned}$$

$$S_\tau = r_1(q;\nu+1)\,(1+1/\nu)^{1/2}\left(1+z^2/\nu\right)^{-1/2} - r_1(\bar\tau;\nu)(1+\lambda^2)^{-1/2},$$

where

$$\begin{aligned}c'_\nu &= \partial c_\nu/\partial\nu = (1/2)\left\{\psi\left((\nu+1)/2\right) - \psi\left(\nu/2\right) - 1/\nu\right\},\\ \partial T_1(q;\nu+1)/\partial\nu &= c'_{\nu+1}T_1(q;\nu+1) - (1/2)J\left(q;\nu+1\right)\\ &\quad + (1/2)(1/\nu)(1+1/\nu)^{-1}\left[T_1(q;\nu+1) - q t_1(q;\nu+1)\right]\\ &\quad - (1/2)(1/\nu^2)t_1(q;\nu+1)\,(1+1/\nu)^{-1/2}\left(1+z^2/\nu\right)^{-1/2}(\lambda z+\tau),\\ \partial T_1(\bar\tau;\nu)/\partial\nu &= c'_\nu T_1(\bar\tau;\nu) - (1/2)J(\bar\tau;\nu) + (1/2)(1/\nu)[T_1(\bar\tau;\nu) - \bar\tau t_1(\bar\tau;\nu)],\end{aligned}$$

$\psi(x)$ is the digamma function, and $J(x;\nu) = \int_{-\infty}^{x} t_1(u;\nu)\log\left(1+u^2/\nu\right)du$.

(b) *When $\nu \to \infty$ we have*:

$$\begin{aligned}S_\xi &\to \tilde S_\xi = (z/\omega) - (\lambda/\omega)\zeta_1(\lambda z+\tau),\\ S_\omega &\to \tilde S_\omega = (z^2/\omega) - (\lambda z/\omega)\zeta_1(\lambda z+\tau) - (1/\omega)\\ &\qquad\;\; = z\tilde S_\xi - (1/\omega),\\ S_\lambda &\to \tilde S_\lambda = z\zeta_1(\lambda z+\tau) + c_\lambda, \quad c_\lambda = \lambda\bar\tau\zeta_1(\bar\tau)/(1+\lambda^2),\\ S_\nu &\to \tilde S_\nu = 0,\\ S_\tau &\to \tilde S_\tau = \zeta_1(\lambda z+\tau) - c_\tau, \quad c_\tau = \zeta_1(\bar\tau)/\sqrt{1+\lambda^2},\end{aligned}$$

where $\tilde S_\xi$, $\tilde S_\omega$, $\tilde S_\lambda$ and $\tilde S_\tau$ are the score components corresponding to the log-likelihood function of $\tilde Y \sim EST_1(\xi,\omega^2,\lambda,\infty,\tau) \equiv ESN_1(\xi,\omega^2,\lambda,\tau)$, which are linearly dependent at $\lambda = 0$. Therefore, the Fisher information matrix associated with the EST_1 is always singular when $\nu \to \infty$. Moreover, its rank is 3 when $\lambda = 0$ for any τ.

(c) *For finite values of ν, the Fisher information matrix of the EST_1 distribution is non-singular at $\tau = \lambda = 0$ and it is equal to*:

$$
\begin{pmatrix}
\frac{1}{\omega^2}\frac{\nu+1}{\nu+3} & 0 & \frac{2}{\omega}\frac{t_1(0;\,\nu)(\nu+1)}{\nu+2} & 0 & 0 \\
0 & \frac{2}{\omega^2}\frac{\nu}{\nu+3} & 0 & -\frac{2}{\omega}\frac{1}{(\nu+1)(\nu+3)} & \frac{2}{\omega}\left\{\frac{t_1(0;\,\nu)(\nu+1)}{\nu+2} - t_1(0;\,\nu)\right\} \\
\frac{2}{\omega}\frac{t_1(0;\,\nu)(\nu+1)}{\nu+2} & 0 & 4[t_1(0;\,\nu+1)]^2 & 0 & 0 \\
0 & -\frac{2}{\omega}\frac{1}{(\nu+1)(\nu+3)} & 0 & -J'(0;\,\nu) - \frac{1}{2}\frac{\nu+5}{\nu(\nu+1)(\nu+3)} & 0 \\
0 & \frac{2}{\omega}\left\{\frac{t_1(0;\,\nu)(\nu+1)}{\nu+2} - t_1(0;\,\nu)\right\} & 0 & 0 & 4[t_1(0;\,\nu+1)]^2 - 4[t_1(0;\,\nu)]^2
\end{pmatrix}
$$

where $-J'(0;\,\nu) = \frac{1}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right)\right]$.

For $\tau \neq 0$, the structure of the EST information matrix at $\lambda = 0$ is similar to $\tau = \lambda = 0$, but some additional numerical procedures are necessary to compute the non-null elements. Thus, the Fisher information matrix of the univariate EST and ST distributions is non-singular at $\lambda = 0$, and consequently their profile log-likelihoods do not have a stationary point at $\lambda = 0$. The case of the multivariate EST distributions is still an open problem, but we conjecture that the non-singularity will hold in that setting as well. The full information matrix when $\tau = 0$, *i.e.*, for the univariate ST distribution, is studied by DiCiccio and Monti (2009), whereas the case of the multivariate ST distribution is provided by Arellano-Valle (2010).

3.3. Numerical example: confidential data perturbation

Data perturbation methods aim at protecting confidential variables in commercial, governmental and medical databases. One possible approach to maintain confidentiality is to replace confidential variables by new perturbed variables. By doing so, distributional properties of the perturbed variables should remain as close as possible to those of the original confidential variables while preserving confidentiality. A popular method consists in fitting a multivariate normal distribution to the database and simulating perturbed variables from the conditional distribution of confidential variables given non-confidential ones, see the review by Muralidhar and Sarathy (2003). Because the distribution of databases is most often unlikely to be multivariate normal, it is tempting to use a classical ST distribution instead. Proposition 4 tells us that the conditional distribution must be of EST type, from which simulations can be performed easily by means of Proposition 1. A detailed analysis of such a procedure is beyond the scope of this article and has been pursued by Lee *et al.* (2010). Nevertheless, we illustrate some of these ideas next and refer the interested reader to the aforementioned article.

We consider a dataset from Cook and Weisberg (1994) on characteristics of $n = 202$ Australian athletes collected by the Australian Institute of Sport (AIS). The variables of interest are height (Ht), weight (Wt) and body mass index (Bmi) of these athletes. Because Ht and Wt would possibly allow to identify the athletes, and then to infer their Bmi, the dataset cannot be released to the public in its original form if confidentiality issues are of concern. Therefore, we regard Ht and Wt as confidential variables. We fit a trivariate ST distribution to the AIS dataset. The estimates of the shape and degrees of freedom are $\hat{\lambda} = (-56.5,\ 113.9,\ -62.5)^{\mathrm{T}}$ and $\hat{\nu} = 2.6$, respectively. The likelihood ratio test of the hypothesis of normality yields a p-value which is essentially zero, so we strongly reject this hypothesis. The EST form of the conditional distribution of $(\mathrm{Ht}, \mathrm{Wt})$ given Bmi is described in Proposition 4. We investigate its departure from the ST distribution by means of its extension parameter $\tau^*_{1.2}$ in Figure 7. The vertical dashed lines represent the minimum and maximum of Bmi values in the AIS dataset, corresponding to an extension parameter of -14.8 and 17.8, respectively. The vertical dotted line indicates the Bmi value, 21.5, corresponding to an extension parameter of zero. Hence, simulations from the EST conditional distribution will have a positive extension parameter (*i.e.*, more symmetric distribution) for Bmi values above 21.5, and negative extension parameter (*i.e.*, more skewed distribution) for Bmi values below 21.5. Overall, Figure 7 confirms that the extension parameter of the EST distribution arising from this example can be quite different from zero, hence the properties of the EST distribution developed in this article are relevant for such applications.

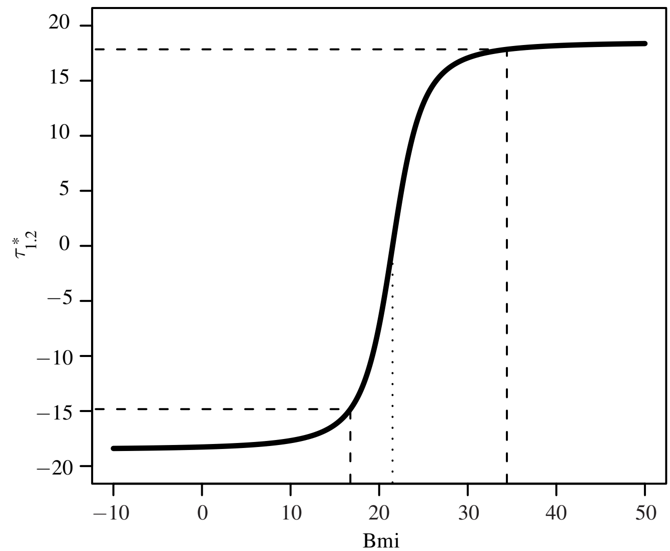

Figure 7. Extension parameter $\tau^*_{1.2}$ of the conditional EST distribution of (Ht, Wt) given Bmi based on the AIS dataset. The vertical dashed lines represent the minimum and maximum of Bmi values in the AIS dataset. The vertical dotted line indicates the Bmi value corresponding to an extension parameter of zero.

Appendix: Proofs

Proof of Proposition 1. Note first that $f_Y(y) = |\omega|^{-1} f_Z(z)$, where $z = \omega^{-1}(y - \xi)$. Now, the density of $Z \stackrel{d}{=} (X|X_0 < \lambda^{\mathrm{T}} X + \tau)$ is given by (see, *e.g.*, Arellano-Valle *et al.*, 2002)

$$f_Z(z) = \frac{1}{P(X_0 - \lambda^{\mathrm{T}} X < \tau)} f_X(z) P(X_0 < \lambda^{\mathrm{T}} z + \tau | X = z). \tag{23}$$

Thus the proof follows from $(X_0|X = z) \sim t_1\left(0, \alpha_{Q(z)}, \nu + p\right)$, where $\alpha_{Q(z)} = (\nu + Q(z))/(\nu + p)$, $Q(z) = z^{\mathrm{T}}\bar{\Omega}^{-1}z$, $X \sim t_p(0, \bar{\Omega}, \nu)$ and $X_0 - \lambda^{\mathrm{T}}X \sim t_1(0, 1 + \lambda^{\mathrm{T}}\bar{\Omega}\lambda, \nu)$. □

Proof of Proposition 2. Let $X_* = T_{(-\bar{\tau},\infty)}$, and note that $Z|X_* = u \sim t(\delta u, [(\nu + u^2)/(\nu + 1)](\bar{\Omega} - \delta\delta^{\mathrm{T}}), \nu + 1)$, where $f_{X_*}(u) = t_1(u; \nu)/T_1(\bar{\tau}; \nu)$ for $u > -\bar{\tau}$. So $f_{Z|X_*=u}(z) = f_{X|X_0=u}(z)$, where $(X^{\mathrm{T}}, X_0)^{\mathrm{T}}$ has the distribution (10). Thus, from the symmetry of f_{X_0} and the identity $f_{X|X_0=u}(z)f_{X_0}(u) = f_X(z)f_{X_0|X=z}(u)$ we obtain that

$$
\begin{aligned}
f_Z(z) &= \int_{-\bar{\tau}}^{\infty} f_{X|X_0=u}(z)f_{X_0}(u)du/T_1(\bar{\tau}; \nu) = f_X(z)\int_{-\infty}^{\bar{\tau}} f_{X_0|X=z}(u)du/T_1(\bar{\tau}; \nu) \\
&= t_p(z; \bar{\Omega}, \nu)T_1\left(\sqrt{(\nu + p)/(\nu + \|z\|^2)}\,(\lambda^{\mathrm{T}}z + \tau); \nu + p\right)/T_1(\bar{\tau}; \nu),
\end{aligned}
$$

since $(X_0|X=z)\sim t_1(\lambda^{\mathrm{T}}z,(\nu+\|z\|^2)/(\nu+p),\nu+p)$, with $\lambda=\bar{\Omega}^{-1}\delta/\sqrt{1-\delta^{\mathrm{T}}\bar{\Omega}^{-1}\delta}$. □

Proof of Propositions 3 and 4. Consider the partition $Y^{\mathrm{T}} = (Y_1^{\mathrm{T}}, Y_2^{\mathrm{T}})$ and the corresponding partitions of ξ, Ω, ω, $\bar{\Omega}$, λ, where $\omega_i = \mathrm{diag}(\Omega_{ii})^{1/2}$ and $\bar{\Omega}_{ij} = \omega_i^{-1}\Omega_{ij}\omega_j^{-1}$ for $i, j = 1, 2$. The proof will be based on the factorization of $f_Y(y) = f_{Y_1,Y_2}(y_1, y_2)$ as $f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1)f_{Y_2|Y_1=y_1}(y_2)$. In fact, by applying first this factorization to the symmetric t density, we have

$$
t_p(y; \xi, \Omega, \nu) = t_{p_1}(y_1; \xi_1, \Omega_{11}, \nu)t_{p_2}(y_2; \xi_{2\cdot1}, \alpha_{Q_1}\Omega_{22\cdot1}, \nu_{2\cdot1}). \tag{24}
$$

Let now $z_{2\cdot1} = \omega_{2\cdot1}^{-1}(y_2 - \xi_{2\cdot1})$, and $Q_{2\cdot1}(z_{2\cdot1}) = z_{2\cdot1}^{\mathrm{T}}\bar{\Omega}_{22\cdot1}^{-1}z_{2\cdot1}$. By noting after some straightforward algebra that $\lambda^{\mathrm{T}}z = \sqrt{1 + \lambda_{2\cdot1}^{\mathrm{T}}\bar{\Omega}_{22\cdot1}\lambda_{2\cdot1}}\,\lambda_{1(2)}^{\mathrm{T}}z_1 + \lambda_{2\cdot1}^{\mathrm{T}}z_{2\cdot1}$, $Q(z) = Q_1(z_1) + Q_{2\cdot1}(z_{2\cdot1})$ and $\frac{\nu+p}{\nu+Q(z)} = \frac{\nu+p_1}{\nu+Q_1(z_1)}\,\frac{\nu_{2\cdot1}+p_2}{\nu_{2\cdot1}+Q_{2\cdot1}(\alpha_{Q_1}^{-1/2}z_{2\cdot1})}$, we obtain

$$
\begin{aligned}
&T_1\left(\sqrt{\frac{\nu + p}{\nu + Q(z)}}(\lambda^{\mathrm{T}}z + \tau); \nu + p\right) \\
&\quad = T_1\left(\sqrt{\frac{\nu_{2\cdot1} + p_2}{\nu_{2\cdot1} + Q_{2\cdot1}(z_{2\cdot1}^*)}}(\lambda_{2\cdot1}^{\mathrm{T}}z_{2\cdot1}^* + \tau_{2\cdot1}^*); \nu_{2\cdot1} + p_2\right),
\end{aligned}
\tag{25}
$$

where $z_{2\cdot1}^* = \alpha_{Q_1}^{-1/2}z_{2\cdot1}$. Hence, by replacing (24) and (25) in (1), and using also that

$$
T_1\left(\tau_{2\cdot1}^*/\sqrt{1 + \lambda_{2\cdot1}^{\mathrm{T}}\bar{\Omega}_{22\cdot1}\lambda_{2\cdot1}}; \nu_{2\cdot1}\right) = T_1\left(\sqrt{\frac{\nu + p_1}{\nu + Q_1(z_1)}}(\lambda_{1(2)}^{\mathrm{T}}z_1 + \tau_{1(2)}); \nu + p_1\right),
$$

we can rewrite the density of $Y = (Y_1^{\mathrm{T}}, Y_2^{\mathrm{T}})^{\mathrm{T}}$ as

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= \frac{t_{p_1}(y_1; \xi_1, \Omega_{11}, \nu)}{T_1(\tau/\sqrt{1+\lambda^{\mathrm{T}}\bar{\Omega}\lambda}; \nu)} T_1\left(\sqrt{\frac{\nu+p_1}{\nu+Q_1(z_1)}}(\lambda_{1(2)}^{\mathrm{T}} z_1 + \tau_{1(2)}); \nu+p_1\right) \\ &\quad\times \frac{t_{p_2}(y_2; \xi_{2\cdot 1}, \alpha_{Q_1}\Omega_{22\cdot 1}, \nu_{2\cdot 1})}{T_1(\tau_{2\cdot 1}^*/\sqrt{1+\lambda_{2\cdot 1}^{\mathrm{T}}\bar{\Omega}_{22\cdot 1}\lambda_{2\cdot 1}}; \nu_{2\cdot 1})} \\ &\quad\times T_1\left(\sqrt{\frac{\nu_{2\cdot 1}+p_2}{\nu_{2\cdot 1}+Q_{2\cdot 1}(z_{2\cdot 1}^*)}}(\lambda_{2\cdot 1}^{\mathrm{T}} z_{2\cdot 1}^* + \tau_{2\cdot 1}^*); \nu_{2\cdot 1}+p_2\right) \\ &= f_{Y_1}(y_1) f_{Y_2|Y_1=y_1}(y_2). \end{aligned}$$

□

Proof of Proposition 5. By Proposition 1, we have $Y = \xi + \omega Z$, where by (7) $Z \stackrel{d}{=} (X|X_0 < \lambda^{\mathrm{T}} X + \tau)$, with X and X_0 being jointly distributed as in (8). Thus, we obtain $AY + b = \xi_A + \omega_A Z_A$, where $Z_A = \omega_A^{-1} A\omega Z$. Let now $X_A = \omega_A^{-1} A\omega X$ and $X_{0\cdot A} = \gamma_A^{-1}(X_0 - \lambda^{\mathrm{T}} X + \bar{\lambda}_A^{\mathrm{T}} X_A)$. From the properties of the symmetric t distribution, we have

$$\begin{pmatrix} X_A \\ X_{0\cdot A} \end{pmatrix} \sim t_{r+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \bar{\Omega}_A & 0 \\ 0 & 1 \end{pmatrix}, \nu\right).$$

The equivalence of the events $\{X_0 < \lambda^{\mathrm{T}} X + \tau\}$ and $\{X_{0\cdot A} < \lambda_A^{\mathrm{T}} X_A + \tau_A\}$ implies that $Z_A \stackrel{d}{=} (X_A|X_{0\cdot A} < \lambda_A^{\mathrm{T}} X_A + \tau)$, which by (23) has a density given by

$$f_{Z_A}(z) = \frac{1}{T_1(\tau_A/\sqrt{1+\lambda_A^{\mathrm{T}}\bar{\Omega}_A\lambda_A})} t_r(z; 0, \bar{\Omega}_A, \nu) T_1(\lambda_A^{\mathrm{T}} z + \tau_A; \nu + r).$$

Thus, the proof follows by using that $f_{Y_A}(y) = |\omega_A|^{-1} f_{Z_A}(\omega_A^{-1}(y - \xi_A))$. □

Proof of Proposition 6. The moments of $N_{(-a,\infty)}$ can be computed by means of its moment generating function:

$$M_{N_{(-a,\infty)}}(t) = e^{t^2/2}\frac{\Phi_1(a+t)}{\Phi_1(a)}.$$

The rest of the proof is a particular case of Proposition 7. □

Proof of Lemma 1. The proof of (14) is straightforward by considering the well-known fact $E(V^{-k/2}) = (\nu/2)^{k/2}\Gamma((\nu-k)/2)/\Gamma(\nu/2)$, for $\nu > k \geq 0$, where $V \sim Gamma(\nu/2, \nu/2)$. The proof of (15) follows by noting that $E(V^{-k/2}\Phi_1(\sqrt{V}\, a)) = c_k E(\Phi_1(\sqrt{V_k}\, a_k) = c_k T_1(a_k; \nu - k)$, where $c_k = (\nu/2)^{k/2}\Gamma((\nu-k)/2)/\Gamma(\nu/2)$, $a_k = \sqrt{(\nu-k)/\nu}\, a$ and $V_k \sim Gamma((\nu -$

$k)/2, (\nu-k)/2)$, for $\nu > k \geq 0$. Now, since $E(h(T_{(a,b)})) = E(h(T)|a < T < b)$ and $t_1(x;\nu) = \int_0^\infty \sqrt{v}\phi_1(\sqrt{v}x)g(v)dv$, where $g(v)$ is the density of V, we have

$$
\begin{aligned}
E(h(T_{(a,b)})) &= \frac{\int_a^b h(x)t_1(x;\nu)dx}{T_1(b;\nu) - T_1(a;\nu)} = \frac{\int_a^b h(x)\int_0^\infty \sqrt{v}\phi_1(\sqrt{v}x)g(v)dvdx}{T_1(b;\nu) - T_1(a;\nu)} \\
&= \frac{\int_0^\infty \left\{\int_a^b h(x)\sqrt{v}\phi_1(\sqrt{v}x)dx\right\} g(v)dv}{T_1(b;\nu) - T_1(a;\nu)} = \frac{\int_0^\infty \left\{\int_{\sqrt{v}a}^{\sqrt{v}b} h(z/\sqrt{v})\phi_1(z)dz\right\} g(v)dv}{T_1(b;\nu) - T_1(a;\nu)} \\
&= \frac{\int_0^\infty \{\Phi_1(\sqrt{v}\,b) - \Phi_1(\sqrt{v}\,a)\}E(h\left(N/\sqrt{v}\right)|\sqrt{v}a < N < \sqrt{v}b, V = v)g(v)dv}{T_1(b;\nu) - T_1(a;\nu)},
\end{aligned}
$$

which concludes the proof. □

Proof of Proposition 7. From the stochastic representation (12) with $p = 1$, we have for $\nu > k$:

$$
E(Z^k) = \sum_{j=0}^{k}\binom{k}{j}\delta^j(1-\delta^2)^{(k-j)/2}(\nu+1)^{-(k-j)/2}E(\bar{X}_1^{k-j})E\left(T^j_{(-\bar{\tau},\infty)}(\nu+T^2_{(-\bar{\tau},\infty)})^{(k-j)/2}\right),
$$

where $\bar{X}_1 = (1-\delta^2)^{-1/2}X_1 \sim t(0,\nu+1)$, implying that $E(\bar{X}_1^k) = E(V_1^{-k/2})E(N_1^k)$, with $N_1 \sim N(0,1)$ and $V_1 \sim Gamma[(\nu+1)/2, (\nu+1)/2)$ being independent. For the kth moment of $T_{(-\bar{\tau},\infty)}$, we have by Lemma 1 that:

$$
E(T^k_{(-\bar{\tau},\infty)}) = \frac{E(V^{-k/2}\Phi(V^{1/2}\bar{\tau})E[N^k_{(-V^{1/2}\bar{\tau},\infty)}|V])}{T_1(\bar{\tau};\nu)}, \quad \nu > k,
$$

where $N_{(-a,\infty)} \stackrel{d}{=} (N|N + a > 0)$. The moments of $N_{(-a,\infty)}$ are given in Proposition 6. Considering also (14) the proof follows. □

Proof of Proposition 8. Since $Y = \xi + \omega Z$, where $Z \sim EST_p(0, \bar{\Omega}, \nu, \lambda, \tau)$, we have by letting $Z_0 = \bar{\Omega}^{-1/2}Z$ that

$$
(Y - \mu_Y)^{\mathrm{T}}\Sigma^{-1}(Y - \mu_Y) = (Z_0 - \mu_0)^{\mathrm{T}}\Sigma_0^{-1}(Z_0 - \mu_0),
$$

where by Proposition 6, $Z_0 \sim EST_p(0, I_p, \nu, \bar{\eta}, \tau)$ with $\bar{\eta} = \bar{\Omega}^{1/2}\lambda$, $\mu_0 = E(Z_0)$ and $\Sigma_0 = Var(Z_0)$. Note from the results in Section 2.3 for the mean an covariance matrix of Z that $\mu_0 = \bar{\delta}\mu_*$ and $\Sigma_0 = a_*^2 I_p - (a_*^2 - \sigma_*^2)\bar{\delta}\bar{\delta}^{\mathrm{T}}$, where $\bar{\delta} = \bar{\Omega}^{-1/2}\delta$, it follows that

$$
\Sigma_0^{-1} = \frac{1}{a_*^2}\left\{I_p + \alpha_*\bar{\delta}\bar{\delta}^{\mathrm{T}}\right\}, \quad \text{with} \quad \alpha_* = \frac{a_*^2 - \sigma_*^2}{a_*^2 - (a_*^2 - \sigma_*^2)\|\bar{\delta}\|^2}.
$$

Consider now the functions $\bar{X}_* = X_* - E(X_*)$ and $S_* = \sqrt{(\nu + X_*^2)/(\nu + 1)}$, where $X_* = T_{(-\bar{\tau},\infty)}$, and also the matrix $\bar{\Delta} = (I_p - \bar{\delta}\bar{\delta}^{\mathrm{T}})^{1/2}$. From (12) we have $Z_0 - \mu_0 \overset{d}{=} S_*\bar{\Delta}\bar{X} + \bar{\delta}\bar{X}_*$, where $\bar{X} \sim t_p(0, I_p, \nu + 1)$ and is independent of X_* and so of $\bar{X}_*$ and S_*. Thus, by noting that $\bar{\delta}^{\mathrm{T}}\bar{X} \overset{d}{=} \|\bar{\delta}\|\bar{X}_1$, and $\bar{\delta}^{\mathrm{T}}\bar{\Delta}\bar{X} \overset{d}{=} \|\bar{\delta}\|\sqrt{1 - \|\bar{\delta}\|^2}\,\bar{X}_1$, where $\bar{X}_1 \sim t(0, 1, \nu + 1)$, we obtain after some algebra

$$\begin{aligned}(Z_0 - \mu_0)^{\mathrm{T}}\Sigma_0^{-1}(Z_0 - \mu_0) &= \frac{1}{a_*^2}\{\|Z_0 - \mu_0\|^2 + \alpha_*[\bar{\delta}^{\mathrm{T}}(Z_0 - \mu_0)]^2\}\\ &\overset{d}{=} \frac{1}{a_*^2}\left\{\|S_*\bar{\Delta}\bar{X} + \bar{\delta}\bar{X}_*\|^2 + \alpha_*(S_*\bar{\delta}^{\mathrm{T}}\bar{\Delta}\bar{X} + \|\bar{\delta}\|^2\bar{X}_*)^2\right\}\\ &\overset{d}{=} \frac{1}{a_*^2}\left\{S_*^2\|\bar{X}_{(1)}\|^2 + (1 + \alpha_*\|\bar{\delta}\|^2)(\bar{Z}_1 - \bar{\mu}_1)^2\right\},\end{aligned}$$

where $\bar{Z}_1 = S_*\sqrt{1 - \|\bar{\delta}\|^2}\,\bar{X}_1 + \|\bar{\delta}\|\bar{X}_*$ and $\bar{\mu}_1 = E(\bar{Z}_1) = \|\bar{\delta}\|\mu_*$. Here $\bar{X}_1$ represents the first component of $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_p)^{\mathrm{T}}$ while $\bar{X}_{(1)} = (\bar{X}_2, \ldots, \bar{X}_p)^{\mathrm{T}}$, and they are uncorrelated. Moreover, they are independent of X_*, so that $\bar{Z}_1 \sim EST_1(0, 1, \nu, \|\bar{\eta}\|, \tau)$. Hence, by using that

$$E(\|\bar{X}_{(1)}\|^{2k}) = \frac{\Gamma(k + (p-1)/2)}{\Gamma((p-1)/2)}\frac{(\nu + 1)^k\Gamma(-k + (\nu + 1)/2)}{\Gamma[(\nu + 1)/2]},\ \nu > 2k,$$

we obtain for $\beta_{2,p}$

$$a_*^4\beta_{2,p} = E(\|\bar{X}_{(1)}\|^4)E(S_*^4) + 2(1 + \alpha_*\|\bar{\delta}\|^2)E(\|\bar{X}_{(1)}\|^2)E(S_*^2[\bar{Z}_1 - \bar{\mu}_1]^2).$$

Similarly, to compute $\beta_{1,p}$ we use

$$(Z_0 - \mu_0)^{\mathrm{T}}\Sigma_0^{-1}(Z_0' - \mu_0) \overset{d}{=} \frac{1}{a_*^2}\left\{S_*S_*'\bar{X}_{(1)}^{\mathrm{T}}\bar{X}_{(1)}' + (1 + \alpha_*\|\bar{\delta}\|^2)(\bar{Z}_1 - \bar{\mu}_1)(\bar{Z}_1' - \bar{\mu}_1')\right\},$$

to obtain

$$a_*^6\beta_{1,p} = 3(1 + \alpha_*\|\bar{\delta}\|^2)\left(\frac{\nu + 1}{\nu - 1}\right)^2[E(S_*^2(\bar{Z}_1 - \bar{\mu}_1))]^2 + (1 + \alpha_*\|\bar{\delta}\|^2)^3[E((\bar{Z}_1 - \bar{\mu}_1)^3)]^3.$$

Thus, the proof follows by using the relations

$$E[(\bar{Z}_1 - \bar{\mu}_1)^k] = E[(\sqrt{1 - \|\bar{\delta}\|^2}\,S_*\bar{X}_1 + \|\bar{\delta}\|\bar{X}_*)^k], \qquad k = 1, 2, 3, 4,$$

and

$$E(S_*^2[\bar{Z}_1 - \bar{\mu}_1]) = \|\bar{\delta}\| E(S_*^2 \bar{X}_*)$$
$$E(S_*^2[\bar{Z}_1 - \bar{\mu}_1]^2) = (1 - \|\bar{\delta}\|^2) \left(\frac{\nu+1}{\nu-1}\right) E(S_*^4) + \|\bar{\delta}\|^2 E(S_*^2 \bar{X}_*^2). \qquad \square$$

Proof of Proposition 9. (a) Consider the log-likelihood function (21) associated to the sample $y_1, \dots, y_n$. The partial derivatives of order one of the log-likelihood are

$$\frac{\partial \ell}{\partial \xi} = \frac{1}{\omega}\left[\sum_{i=1}^n z_i - \lambda \sum_{i=1}^n \zeta_1(\lambda z_i + \tau)\right],$$
$$\frac{\partial \ell}{\partial \omega} = -\frac{1}{\omega}\left[n - \sum_{i=1}^n z_i^2 + \lambda \sum_{i=1}^n z_i \zeta_1(\lambda z_i + \tau)\right],$$
$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^n z_i \zeta_1(\lambda z_i + \tau) + n\tau\lambda(1+\lambda^2)^{-3/2}\zeta_1(\tau/\sqrt{1+\lambda^2}),$$
$$\frac{\partial \ell}{\partial \tau} = \sum_{i=1}^n \zeta_1(\lambda z_i + \tau) - n(1+\lambda^2)^{-1/2}\zeta_1(\tau/\sqrt{1+\lambda^2}),$$

where $z_i = \omega^{-1}(y_i - \xi)$ and $\zeta_1(a) = \phi_1(a)/\Phi_1(a)$. For $\lambda = 0$ to be a solution to the score equations from the above partial derivatives requires $\bar{z} = 0$ from the first and third equations, hence $\xi = \bar{y}$. From the second equation we must have $\omega = s^2$. The fourth equation is satisfied for all $\tau \in \mathbb{R}$.

(b) Lengthy computations yield expressions for the second-order derivatives of the log-likelihood which, at $\lambda = 0$, simplify to

$$\frac{\partial^2 \ell}{\partial \xi^2} = -\frac{n}{\omega^2}, \quad \frac{\partial^2 \ell}{\partial \omega^2} = -\frac{2n}{\omega^2}, \quad \frac{\partial^2 \ell}{\partial \lambda^2} = -n\zeta_1^2(\tau), \quad \frac{\partial^2 \ell}{\partial \xi \partial \lambda} = -\frac{n}{\omega}\zeta_1(\tau),$$
$$\frac{\partial^2 \ell}{\partial \tau^2} = \frac{\partial^2 \ell}{\partial \xi \partial \omega} = \frac{\partial^2 \ell}{\partial \xi \partial \tau} = \frac{\partial^2 \ell}{\partial \omega \partial \lambda} = \frac{\partial^2 \ell}{\partial \omega \partial \tau} = \frac{\partial^2 \ell}{\partial \lambda \partial \tau} = 0.$$

Therefore, the observed information matrix is singular at $\lambda = 0$. $\square$

Proof of Proposition 10. (a) From (20) with $p = 1$, $n = 1$, and using that $\partial z/\partial \xi = -1/\omega$ and $\partial z/\partial \omega = -z/\omega$, where $z = (y - \xi)/\omega$, we have for the score functions that

$$
\begin{aligned}
S_\xi &= \frac{\partial l}{\partial \xi} = (1/\omega)(\nu+1)\left(1+z^2/\nu\right)^{-1}(z/\nu) + r_1(q;\nu+1)(\partial q/\partial \xi),\\
S_\omega &= \frac{\partial l}{\partial \omega} = -(1/\omega)+(1/\omega)(\nu+1)\left(1+z^2/\nu\right)^{-1}(z^2/\nu)+r_1(q;\nu+1)(\partial q/\partial \omega),\\
S_\lambda &= \frac{\partial l}{\partial \lambda} = r_1(q;\nu+1)(\partial q/\partial \lambda) - r_1(\bar\tau;\nu)(\partial \bar\tau/\partial \lambda),\\
S_\nu &= \frac{\partial l}{\partial \nu} = c'_\nu - (1/2)\log\left(1+z^2/\nu\right) + (1/2)(1+1/\nu)\left(1+z^2/\nu\right)^{-1}(z^2/\nu)\\
&\quad + (1/T_1(q;\nu+1))(\partial T_1(q;\nu+1)/\partial \nu) - (1/T_1(\bar\tau;\nu))(\partial T_1(\bar\tau;\nu)/\partial \nu),\\
S_\tau &= \frac{\partial l}{\partial \tau} = r_1(q;\nu+1)(\partial q/\partial \tau) - r_1(\bar\tau;\nu)(\partial \bar\tau/\partial \tau),
\end{aligned}
$$

where

$$
\begin{aligned}
&\partial\bar\tau/\partial\lambda = -(1+\lambda^2)^{-1}\lambda\bar\tau, \quad (\partial\bar\tau/\partial\tau) = (1+\lambda^2)^{-1/2},\\
&\partial c_\nu/\partial\nu = c'_\nu = (1/2)\left\{\psi\left((\nu+1)/2\right) - \psi\left(\nu/2\right) - 1/\nu\right\},\\
&\partial q/\partial\xi = (1+1/\nu)^{1/2}\left\{\left(1+z^2/\nu\right)^{-3/2}(z/\nu)(\lambda z+\tau) - \left(1+z^2/\nu\right)^{-1/2}\lambda\right\}(1/\omega),\\
&\partial q/\partial\omega = (1+1/\nu)^{1/2}\left\{\left(1+z^2/\nu\right)^{-3/2}(z^2/\nu)(\lambda z+\tau) - \left(1+z^2/\nu\right)^{-1/2}\lambda z\right\}(1/\omega),\\
&\partial q/\partial\lambda = (1+1/\nu)^{1/2}\left(1+z^2/\nu\right)^{-1/2} z,\\
&\partial q/\partial\nu = -(1/2)\,(1+1/\nu)^{-1/2}(1/\nu^2)\left(1+z^2/\nu\right)^{-1/2}(\lambda z+\tau)\\
&\qquad\qquad + (1/2)\,(1+1/\nu)^{1/2}\left(1+z^2/\nu\right)^{-3/2}(z^2/\nu^2)(\lambda z+\tau),\\
&\partial q/\partial\tau = (1+1/\nu)^{1/2}\left(1+z^2/\nu\right)^{-1/2},
\end{aligned}
$$

and $\psi(x)$ is the digamma function. By using now that $T_1(a;\nu) = \int_{-\infty}^{0} t_1(x+a;\nu)dx$ and by applying appropriately some of the properties given in the proof of Proposition 9 to obtain truncated moments under the Student's t model, we can show that

$$\begin{aligned}
\partial T_1(q;\,\nu+1)/\partial\nu = {} & c'_{\nu+1}T_1(q;\,\nu+1) - (1/2)\int_{-\infty}^{q} t_1(x;\,\nu+1)\log\left(1+x^2/(\nu+1)\right)dx \\
& + \frac{\nu+2}{2(\nu+1)^2}\int_{-\infty}^{q} x^2\left(1+x^2/(\nu+1)\right)^{-1} t_1(x;\,\nu+1)dx \\
& - \frac{\nu+2}{\nu+1}\int_{-\infty}^{q} x\left(1+x^2/(\nu+1)\right)^{-1} t_1(x;\,\nu+1)dx(\partial q/\partial\nu) \\
= {} & c'_{\nu+1}T_1(q;\,\nu+1) - (1/2)J\,(q;\,\nu+1) + \frac{1}{2(\nu+3)}\int_{-\infty}^{\sqrt{\frac{\nu+3}{\nu+1}}q} x^2 t_1(x;\,\nu+3)dx \\
& - \sqrt{\frac{\nu+1}{\nu+3}}\int_{-\infty}^{\sqrt{\frac{\nu+3}{\nu+1}}q} x t_1(x;\,\nu+3)dx(\partial q/\partial\nu) \\
= {} & c'_{\nu+1}T_1(q;\,\nu+1) - (1/2)J\,(q;\,\nu+1) + \frac{1}{2(\nu+1)}\left[T_1(q;\,\nu+1) - q t_1(q;\,\nu+1)\right] \\
& + t_1(q;\,\nu+1)(\partial q/\partial\nu) \\
= {} & c'_{\nu+1}T_1(q;\,\nu+1) - (1/2)J\,(q;\,\nu+1) \\
& + (1/2)(1/\nu)(1+1/\nu)^{-1}\left[T_1(q;\,\nu+1) - q t_1(q;\,\nu+1)\right] \\
& - (1/2)(1/\nu^2)t_1(q;\,\nu+1)\,(1+1/\nu)^{-1/2}\left(1+z^2/\nu\right)^{-1/2}(\lambda z+\tau), \\
\partial T_1(\bar\tau;\,\nu)/\partial\nu = {} & c'_\nu T_1(\bar\tau;\,\nu) - (1/2)\int_{-\infty}^{\bar\tau} t_1(x;\,\nu)\log\left(1+x^2/\nu\right)dx \\
& + \frac{\nu+1}{2\nu^2}\int_{-\infty}^{\bar\tau} x^2\left(1+x^2/\nu\right)^{-1} t_1(x;\,\nu)dx \\
= {} & c'_\nu T_1(\bar\tau;\,\nu) - (1/2)J(\bar\tau;\,\nu) + \frac{1}{2(\nu+2)}\int_{-\infty}^{\sqrt{\frac{\nu+2}{\nu}}\bar\tau} x^2 t_1(x;\,\nu+2)dx \\
= {} & c'_\nu T_1(\bar\tau;\,\nu) - (1/2)J(\bar\tau;\,\nu) + (1/2)(1/\nu)\left[T_1(\bar\tau;\,\nu) - \bar\tau t_1(\bar\tau;\,\nu)\right],
\end{aligned}$$

where $J(z;\,\nu) = \int_{-\infty}^{z} t_1(x;\,\nu)\log\left(1+x^2/\nu\right)dx$. This yields the score vector in (a).

(b) We note as $\nu\to\infty$ that

$$\begin{gathered}
q\to\lambda z+\tau,\quad r_1(x;\,\nu)\to\zeta_1(x)\equiv\phi_1(x)/\Phi_1(x),\\
c'_\nu\to 0,\quad \partial q/\partial\xi\to-\lambda/\omega,\\
\partial q/\partial\omega\to-\lambda z/\omega,\quad \partial q/\partial\lambda\to z,\quad \partial q/\partial\nu\to 0,\quad \partial q/\partial\tau\to 1\\
\partial T_1(\bar\tau;\,\nu)/\partial\nu\to 0,\quad \partial T_1(q;\,\nu+1)/\partial\nu\to 0.
\end{gathered}$$

Then the Fisher information matrix can be computed as $I(\theta) = Var(S_\theta) = (E\{S_{\theta_i} S_{\theta_j}\})$. So $I(\theta) = Var(S) \to Var(\tilde{S}) = I(\tilde{\theta})$, as $\nu \to \infty$, where $\tilde{\theta} = (\xi, \omega, \lambda, \infty, \tau)^{\mathrm{T}}$. Because $\tilde{S}_\nu = 0$, $I(\tilde{\theta})$ is always singular. Also, if $\lambda = 0$, we have $\tilde{S}_\tau = 0$ and $S_\lambda = \omega\zeta(\tau)\tilde{S}_\xi$. Hence, the rank of $I(\tilde{\theta})$ is 3 when $\lambda = 0$ whatever the value of τ.

(c) When setting $\lambda = \tau = 0$, for the Fisher information matrix, we have by the symmetry (at zero) of the distribution of Z that $E\{S_\xi S_\omega\} = E\{S_\xi S_\nu\} = E\{S_\xi S_\tau\} = E\{S_\omega S_\lambda\} = E\{S_\lambda S_\nu\} = E\{S_\lambda S_\tau\} = 0$. For the non-null elements we have:

$$
\begin{aligned}
E\{S_\xi^2\} &= (1/\omega)^2 \nu\,(1 + 1/\nu)^2\, E\left\{\left(1 + Z^2/\nu\right)^{-2} (Z^2/\nu)\right\},\\
E\{S_\xi S_\lambda\} &= 2(1/\omega)\nu t_1(0; \nu+1)\,(1 + 1/\nu)^{3/2}\, E\left\{\left(1 + Z^2/\nu\right)^{-3/2} (Z^2/\nu)\right\},\\
E\{S_\omega^2\} &= (1/\omega)^2 \nu^2\,(1 + 1/\nu)^2\, E\left\{\left(1 + Z^2/\nu\right)^{-2} (Z^2/\nu)^2\right\} - (1/\omega^2),\\
E\{S_\omega S_\nu\} &= -(1/2)(1/\omega)\nu\,(1 + 1/\nu)\, E\left\{\left(1+Z^2/\nu\right)^{-1} (Z^2/\nu)\log\left(1+Z^2/\nu\right)\right\}\\
&\quad + (1/2)(1/\omega)\nu\,(1+1/\nu)^2\, E\left\{\left(1+Z^2/\nu\right)^{-1}\left(1+Z^2/\nu\right)^{-1} (Z^2/\nu)^2\right\},\\
E\{S_\omega S_\tau\} &= 2(1/\omega)\nu t_1(0; \nu+1)\,(1 + 1/\nu)^{3/2}\, E\left\{\left(1 + Z^2/\nu\right)^{-3/2} (Z^2/\nu)\right\},\\
E\{S_\lambda^2\} &= 4[t_1(0; \nu+1)]^2 \nu\,(1 + 1/\nu)\, E\left\{\left(1 + Z^2/\nu\right)^{-1} (Z^2/\nu)\right\},\\
E\{S_\nu^2\} &= (1/4) E\left\{\left[\log\left(1+Z^2/\nu\right)\right]^2\right\}\\
&\quad + (1/4)\,(1 + 1/\nu)^2\, E\left\{\left(1+Z^2/\nu\right)^{-2} (Z^2/\nu)^2\right\}\\
&\quad - (1/2)\,(1+1/\nu)\, E\left\{\left(1+Z^2/\nu\right)^{-1} (Z^2/\nu)\log\left(1 + Z^2/\nu\right)\right\} - K_\nu^2,\\
E\{S_\tau^2\} &= 4[t_1(0; \nu+1)]^2\,(1 + 1/\nu)\, E\left\{\left(1 + Z^2/\nu\right)^{-1}\right\} - 4[t_1(0; \nu)]^2,
\end{aligned}
$$

where

$$
\begin{aligned}
K_\nu &= c'_{\nu+1} + (1/2)(1/(\nu+1)) - J(0; \nu+1) - J(0; \nu) + (1/2)(1/\nu),\\
c'_{\nu+1} &= (1/2)\left\{\psi\left((\nu+2)/2\right) - \psi\left((\nu+1)/2\right) - 1/(\nu+1)\right\},\\
J(0; \nu) &= \int_{-\infty}^{0} t_1(x; \nu)\log(1 + x^2/\nu)\,dx = (1/2)\left\{\psi((\nu+1)/2) - \psi(\nu/2)\right\}.
\end{aligned}
$$

Thus, the proof follows by using the change of variable $u = (1 + z^2/\nu)^{-1}$ to obtain:

$$E\{(Z^2/\nu)^k(1 + Z^2/\nu)^{-m/2}[\log(1 + Z^2/\nu)]^l\}$$
$$= \frac{B[(\nu + m - 2k)/2, (1 + 2k)/2]}{B(\nu/2, 1/2)} E\{[-\log U]^l\},$$

for $\nu + m > 2k$, and where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma[(a + b)/2]$ is the Beta function and U is a $Beta((\nu + m - 2k)/2, (1 + 2k)/2)$ random variable. For $l = 0, 1, 2,$ we obtain

$$E\{(Z^2/\nu)^k(1+Z^2/\nu)^{-m/2}\} = \frac{B[(\nu + m - 2k)/2, (1 + 2k)/2]}{B(\nu/2, 1/2)},$$

$$E\{(Z^2/\nu)^k(1+Z^2/\nu)^{-m/2}[\log(1 + Z^2/\nu)]\} = -\frac{B[(\nu + m - 2k)/2, (1 + 2k)/2]}{B(\nu/2, 1/2)}$$
$$\times \{\psi[(\nu + m - 2k)/2] - \psi[(\nu + m + 1)/2]\},$$

$$E\{(Z^2/\nu)^k(1 + Z^2/\nu)^{-m/2}[\log(1 + Z^2/\nu)]^2\} = \frac{B[(\nu + m - 2k)/2, (1 + 2k)/2]}{B(\nu/2, 1/2)}$$
$$\times \Big\{(\psi[(\nu + m - 2k)/2] - \psi[(\nu + m + 1)/2])^2$$
$$+\psi'[(\nu + m - 2k)/2] - \psi'[(\nu + m + 1)/2]\Big\}.$$

From the structure of the Fisher information matrix, it can be checked that its columns are linearly independent, hence the matrix is non-singular. □

REFERENCES

ADCOCK, C. J. (2010) Asset pricing and portfolio selection based on the multivariate extended skew-Student-t distribution, *Annals of Operations Research*, 176, 221–234.

ARELLANO-VALLE, R. B. (2010) On the information matrix of the multivariate skew-t model, *Metron*, 68, 371–386.

ARELLANO-VALLE, R. B. and AZZALINI, A. (2006) On the unification of families of skew-normal distributions, *Scandinavian Journal of Statistics*, 33, 561–574.

ARELLANO-VALLE, R. B., BRANCO, M. D. and GENTON, M. G. (2006) A unified view on skewed distributions arising from selections, *Canadian Journal of Statistics*, 34, 581–601.

ARELLANO-VALLE, R. B., DEL PINO, G. and SAN MARTÍN, E. (2002) Definition and probabilistic properties of skew distributions, *Statistics and Probability Letters*, 58, 111–121.

Arellano-Valle, R. B. and Genton, M. G. (2005) On fundamental skew distributions, *Journal of Multivariate Analysis*, 96, 93–116.

Arnold, B. C. and Beaver, R. J.(2000a) The skew-Cauchy distribution, *Statistics and Probability Letters*, 49, 285–290.

Arnold, B. C. and Beaver, R. J.(2000b) Hidden truncation models, *Sankhyā Ser. A*, 62, 22–35.

Azzalini, A. (1985) A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 12, 171–178.

Azzalini, A. (2005) The skew-normal distribution and related multivariate families (with discussion by Marc G. Genton and a rejoinder by the author), *Scandinavian Journal of Statistics*, 32, 159–200.

Azzalini, A. (2006) R package sn: The skew-normal and skew-t distributions (version 0.4-2), URL http://azzalini.stat.unipd.it/SN.

Azzalini, A. and Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distribution, *Journal of the Royal Statistical Society Series B*, 61, 579–602.

Azzalini, A. and Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society Series B*, 65, 367–389.

Azzalini, A. and Dalla Valle, A. (1996) The multivariate skew-normal distribution, *Biometrika*, 83, 715–726.

Azzalini, A. and Genton, M. G. (2008) Robust likelihood methods based on the skew-t and related distributions, *International Statistical Review*, 76, 106–129.

Branco, M. D. and Dey, D. K. (2001) A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis*, 79, 99–113.

Capitanio, A., Azzalini, A. and Stanghellini, E. (2003) Graphical models for skew-normal variates, *Scandinavian Journal of Statistics*, 30, 129–144.

Cook, R. D. and Weisberg, S. (1994) *An Introduction to Regression Graphics*, Wiley, New York.

Diciccio, T. and Monti, A. (2009) Inferential aspects of the skew-t distribution, Manuscript.

Fang, K.-T., Kotz, S. and Ng, K.-W. (1990) *Symmetric Multivariate and Related Distributions*, Monographs on Statistics and Applied Probability, 36, Chapman and Hall, Ltd., London.

Genton, M. G. (2004) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Edited Volume, Chapman & Hall/CRC, Boca Raton, FL, pp. 416.

Genz, A. (1992) Numerical computation of multivariate Normal probabilities, *Journal of Computational and Graphical Statistics*, 1, 141–149.

Genz, A. and Bretz, F. (2002) Methods for the computation of multivariate t-probabilities, *Journal of Computational and Graphical Statistics*, 11, 950–971.

González-Farías, G., Domínguez-Molina, J. A. and Gupta, A. K. (2004) The closed skew-normal distribution, In: *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, M. G. Genton (ed.), Chapman & Hall / CRC, Boca Raton, FL, 25–42.

Lee, S., Genton, M. G. and Arellano-Valle, R. B. (2010) Perturbation of numerical confidential data via skew-t distributions, *Management Science*, 56, 318–333.

Mardia, K. V. (1970) Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 36, 519–530.

Muralidhar, K. and Sarathy, R. (2003) A theoretical basis for perturbation methods, *Statistics and Computing*, 13, 329–335.

Park, J. W., Genton, M. G. and Ghosh, S. K. (2007) Censored time series analysis with autoregressive moving average models, *Canadian Journal of Statistics*, 35, 151–168.

PEWSEY, A. (2006) Some observations on a simple means of generating skew distributions, In: *Advances in Distribution Theory, Order Statistics, and Inference*, N. Balakrishnan, E. Castillo and J. M. Sarabia (eds.), Boston: Statistics for Industry and Technology, Birkhäuser 75–84.

STUART, A. and ORD, J. K. (1987) *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory*, Oxford University Press, Fifth Edition, New York.

REINALDO B. ARELLANO-VALLE
Departamento de Estadística
Facultad de Matemática
Pontificia Universidad Católica de Chile
Santiago 22 (Chile)
reivalle@mat.puc.cl

MARC G. GENTON
Department of Statistics
Texas A&M University
College Station, TX 77843-3143 (USA)
genton@stat.tamu.edu