

Article

A Selective Overview of Skew-Elliptical and Related Distributions and of Their Applications

Chris Adcock ^{1,†} and Adelchi Azzalini ^{2,*,†} ¹ SOAS, University of London, London WC1H 0XG, UK; c.j.adcock@soas.ac.uk² Dipartimento di Scienze Statistiche, Università di Padova, 35121 Padova, Italia; adelchi.azzalini@unipd.it

* Correspondence: adelchi.azzalini@unipd.it

† These authors contributed equally to this work.

Received: 7 November 2019; Accepted: 2 January 2020; Published: 7 January 2020



Abstract: Within the context of flexible parametric families of distributions, much work has been dedicated in recent years to the theme of skew-symmetric distributions, or symmetry-modulated distributions, as we prefer to call them. The present contribution constitutes a review of this area, with special emphasis on multivariate skew-elliptical families, which represent the subset with more immediate impact on applications. After providing background information of the distribution theory aspects, we focus on the aspects more relevant for applied work. The exposition is targeted to non-specialists in this domain, although some general knowledge of probability and multivariate statistics is assumed. Given this aim, the mathematical profile is kept to the minimum required.

Keywords: probability distributions; elliptical distributions; skew-elliptical distributions; flexible parametric distributions

1. Context and Motivation

1.1. The Wider Perspective

The development of parametric families of distributions and the study of their properties is a fundamental component of statistical methodology. Given the very long tradition of this area, a vast arsenal of distributions have been accumulated over time, suitable for tackling a great variety of applied problems. Consequently, one may legitimately wonder whether additional effort in this direction is really motivated by scientific requirements or it is just the outcome of nice mathematical exercises providing secondary refinements of an already well-established area. Put in another way, one may question the motivation for the particularly intense activity developed on these themes in the last two decades or so, sometimes associated with the phrase ‘flexible parametric [families of] distributions’.

In our view, there are at least two, connected but conceptually separate, reasons which motivate the perpetual efforts in this domain. One such reason is the increase in size of the commonly available data records, which is continuously growing as time goes by, and typically a larger sample calls for a more flexible parametric family to fit the data. The basic factor is that a sample of limited size can easily appear compatible with different assumptions on the underlying distribution, while with larger samples departures from the expected pattern cannot so easily be dismissed as the effect of random variability. This implies that either a more specific distributional assumption is required, possibly motivated by subject matter consideration in real applications, or a more flexible parametric family must be employed.

The second fact, intersecting with the previous one, is that in more recent times the dimensionality of available data has also increased, not only the number of records. This is not to say that consideration

of multivariate distributions is a recent event—quite the opposite; it has a long history. As a case in point, recall that the extensive review of such literature by Pretorius [1] starts off, after the preliminary pages, recalling “a brief account of Schols’ treatment [2] of errors in space, and of Perozzo’s analysis [3] of Italian marriage statistics”, where multivariate data are examined. Notwithstanding this, it remains true that multivariate analysis has long remained largely linked to the assumption of the normal distribution. Surely the limited computational resources available, at least to most researchers, until a few decades ago had a major role in this situation. Consider that departure from normality implies non-linear behavior of various ingredients and thus involves a higher computational effort.

Currently, the availability of larger data sets requires consideration of more flexible families of distributions and, at the same time, more powerful computing resources allow researchers to explore new avenues. An overview of recent developments in this domain can be obtained by the accounts of Jones [4] and contributions to the ensuing discussion, as well as that of Babić et al. [5]; review in [4] focuses mainly on the univariate case, while [5] deals with the multivariate case.

1.2. *The Specific Target*

Within the broad above-delineated context, a number of directions have been considered and continue to generate lively activity. Here, we focus on one such theme which has received much attention since about the turn of the millennium, generating tremendous activity both on the theoretical and the applied side. Our chief target is the set of continuous multivariate distributions which includes the ‘skew-elliptically contoured’ (SEC), briefly skew-elliptical distributions, and a number of closely related constructions, since much work has dealt with this domain. To outline the area of interest, we take the book edited by Genton [6] and the monograph of Azzalini and Capitanio [7] as the primary landmarks of the territory.

Since the SEC class constitutes a prominent subset of the wider family of skew-symmetric distributions or symmetry-modulated distributions, as we prefer to call them, this wider family will also feature here but only to allow a more complete exposition of the SEC class, without attempting to discuss fully the wider class. In addition, we discuss further extensions of the SEC class, which go under the headings SUN (unified skew-normal), SUEC (unified skew-elliptical), and others to be described later.

Predictably, we shall provide arguments explaining why the SEC distributions constitute an interesting class to consider, and we illustrate its great potential in various aspects. However, the original motivation for writing the present review comes from another side. In the last 15–20 years, the delineated domain has witnessed an energetic and fruitful development, and many constructions have been generated, often having similar appearance but actually different. In addition, terminology has been employed in a non-systematic manner. The overall product is a conglomerate which almost inevitably appears confusing for the newcomer to this area. Just to indicate a key example, the popular term ‘skew- t distribution’ does not refer to a unique entity; in the literature, there are several such things sharing the same name. Very understandably, this ambiguity is often not perceived and the implications of adopting this or that variant construction are not grasped, although these differences are often non-negligible, sometimes substantial.

One of our aims is, therefore, to examine how various SEC formulations differ in their underlying mechanism and their suitability as building blocks for subsequent work. The emphasis is primarily on qualitative criteria, such as existence of reasonably simple generating mechanisms of the underlying random variables, potential links with some subject-matter theory, overall mathematical tractability, and others alike. Undoubtedly, this type of scrutiny involves some degree of subjective preference for the criteria being prioritized. Since our focus is primarily on modeling, we do not dwell on aspects of statistical inference. Throughout the discussion, mathematical aspects are kept to the minimum required for exposition, referring to original sources for more technical points.

Another aim of the review is to present a range of applications of the formal constructions. The term ‘application’ takes on two conceptually separate meanings, although in the actual

contributions, this separation is inevitably blurred. In Section 4, the term ‘application’ refers to extensions of existing statistical methods to increase their suitability for handling real data thanks to the increased flexibility and the more realistic assumption on the data distributions. In Section 5, the term ‘application’ refers to publications where the primary motivation was some real application or a set of real applications; this does not exclude that the possibility that the motivating problems lead to methodological advances. We have, instead, not included those publications where real data were employed as a purely numerical illustration of some eminently theoretical development.

The reader familiar with the monograph of Azzalini and Capitanio [7] can recognize that, to some extent, the present review can be regarded as an update of Chapters 7 and 8 of that work. For space reasons, we have opted to avoid extensive duplication of that material, apart for those parts which are necessary for exposition of contributions of interest. We have focused mainly on newer contributions, especially in a multivariate setting. Even with this restrictions, a full account of all publications would have led to a too long exposition. This has implied to make a selection, which again involves some subjective judgment. In this respect, a special note is due for the theme of stochastic processes, time series and, especially, spatio-temporal processes, where there seem to persist a debate among specialists on the correctness of certain formulations; our choice has been not to include here many such publications, given the intended non-specialist audience of this review.

Our emphasis is primarily on the multivariate setting. This choice is largely due to the growing emphasis on this context, as underlined in Section 1.1. Obviously, most facts hold in the univariate case, as well, but some of them become void there. Occasionally, we shall deal with univariate distributions, especially so when the point to be highlighted carries on directly from the univariate to the multivariate setting and discussion in the univariate case facilitates exposition.

2. Background Concepts

2.1. Basics of Elliptical Distributions

As the term skew-elliptical distributions suggests, these are obtained as extensions of elliptically contoured (EC) distributions, briefly called ‘elliptical distributions’. To start with, establish then notation and recall some basic facts.

We shall only deal with absolutely continuous EC distributions, which represent the only case relevant for the developments to follow. Consider a function $\tilde{p}$ from $\mathbb{R}^+$ to $\mathbb{R}^+$ such that, for a positive integer d ,

$$k_d = \int_0^\infty r^{d-1} \tilde{p}(r^2) dr < \infty. \quad (1)$$

Then, a d -dimensional continuous random variable X is said to have an elliptical distribution, with density generator $\tilde{p}$, if its density is of the form

$$p(x; \mu, \Sigma) = \frac{c_d}{\det(\Sigma)^{1/2}} \tilde{p}\{(x - \mu)^\top \Sigma^{-1} (x - \mu)\}, \quad x \in \mathbb{R}^d, \quad (2)$$

where $\mu \in \mathbb{R}^d$ is a location parameter, Σ is a symmetric positive-definite $d \times d$ scale matrix and $c_d = \Gamma(d/2) / (2 \pi^{d/2} k_d)$. In this case, we write $X \sim \text{EC}_d(\mu, \Sigma, \tilde{p})$.

Since $(x - \mu)^\top \Sigma^{-1} (x - \mu) = \text{constant}$ is the equation of an ellipsoid centered at point μ , the level curves of density (2) constitutes ellipsoids, which explains the name ‘elliptically contoured distributions’.

The EC class enjoys a remarkable number of attractive formal properties, including closure under marginalization, affine transformations, and conditioning on the values taken on by a subset of the d components. For a precise statement of these facts, as well as many others, we refer to standard accounts, such as the book by Fang et al. [8].

The most prominent representative of the EC class is the multivariate normal distribution, which occurs when $\tilde{p}(u) = \exp(-u/2)$. The normal family plays an important role within the EC

class, as it generates the subclass of *scale mixtures of normal variates*, defined as follows. If $X \sim N_d(0, \Sigma)$ and S is an independent positive variable which plays the role of ‘random scale factor’, it is easy to check that the scale mixture

$$Y = \mu + S X \quad (3)$$

has distribution of EC type. In the important instance, where

$$S = (V/\nu)^{-1/2}, \quad V \sim \chi_\nu^2, \quad (4)$$

we obtain the multivariate Student’s t distribution on ν degrees of freedom; in this case, we write $Y \sim t_d(\mu, \Sigma, \nu)$ in an obvious notation.

For future reference, it is appropriate to recall a relatively lesser known fact about the EC class and the scale mixtures of normals. While the p -dimensional marginal of a d -dimensional EC variable is still of EC type, as already recalled, this does not ensure that its marginal density functions belong to the same parametric family of the original density.

Kano [9] has studied necessary and sufficient conditions to ensure that a marginalization operation keeps the distribution within the same parametric class, which he calls ‘consistency property’. In operational terms, the required condition is that the EC distribution allows a stochastic representation of type (3) where the distribution of the mixing variable S does not depend on d . As a key example, this condition holds for S in (4); hence, the marginals of a Student’s t distribution are still of the same type. There are instead several EC families where, although a representation of type (3) exists, the distribution of S depends on d ; hence, marginalization consistency does not hold. The motivating instance of this sort in Kano [9] is the class of multivariate exponential power distributions, which fails to be consistent under marginalization.

2.2. The Multivariate Skew-Normal Distribution

Like the multivariate normal distribution is at the core of the EC class, so is the multivariate skew-normal (SN) with respect to the SEC class and many other related formulations. We present the SN distribution following essentially the constructive route of Azzalini and Dalla Valle [10], also in view of its relevance for the subsequent discussion.

Start from independent variables $U_0 = (U_{01}, \dots, U_{0d})^\top \sim N_d(0, \bar{\Psi})$ and $U_1 \sim N(0, 1)$, where $\bar{\Psi}$ is a positive-defined $d \times d$ correlation matrix. Next, for any vector $\delta = (\delta_1, \dots, \delta_d)^\top \in (-1, 1)^d$, define $Z = (Z_1, \dots, Z_d)^\top$ where

$$Z_j = (1 - \delta_j^2)^{1/2} U_{0j} + \delta_j |U_1| \quad (5)$$

for $j = 1, \dots, d$. Equivalently, write in a more compact form

$$U = \begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{d+1} \left(0, \begin{pmatrix} \bar{\Psi} & 0 \\ 0 & 1 \end{pmatrix} \right) \quad Z = D_\delta U_0 + \delta |U_1|, \quad (6)$$

where $D_\delta = (I_d - \text{diag}(\delta^2))^{1/2}$. For future use, define also the term-by-term transform

$$\lambda = (\lambda_1, \dots, \lambda_d)^\top = D_\delta^{-1} \delta, \quad (7)$$

which is invertible via $\delta_j = (1 + \lambda_j^2)^{-1/2} \lambda_j$. Some algebraic work yields the density function of Z , which is

$$\varphi_d(x; \bar{\Omega}, \alpha) = 2 \varphi_d(x; \bar{\Omega}) \Phi(\alpha^\top x), \quad x \in \mathbb{R}^d, \quad (8)$$

where $\varphi_d(\cdot; \bar{\Omega})$ denotes the $N_d(0, \bar{\Omega})$ density, $\Phi(\cdot)$ the $N(0, 1)$ distribution function, and

$$\bar{\Omega} = D_\delta (\bar{\Psi} + \lambda \lambda^\top) D_\delta, \quad (9)$$

$$\alpha = \left(1 + \lambda^\top \bar{\Psi}^{-1} \lambda\right)^{-1/2} D_\delta^{-1} \bar{\Psi}^{-1} \lambda = \left(1 - \delta^\top \bar{\Omega}^{-1} \delta\right)^{-1/2} \bar{\Omega}^{-1} \delta. \quad (10)$$

Clearly, the distribution of Z is regulated by the pair $(\bar{\Psi}, \delta)$, or equivalently by $(\bar{\Psi}, \lambda)$. If $\delta = 0 = \lambda$, then Z coincides with the normal variable U_0 ; otherwise, the non-linear transformation of U_1 in (6) produces a non-normal distribution of Z . Correspondingly, in the latter case, the vector α in (10) is non-zero, so that the density (8) is the product of the EC factor φ_d times a perturbation factor $\Phi(\alpha^\top x)$. The result is a density function with skew-elliptical contour level curves. The contour level plots in Figure 1 display two examples of SN densities (8) with $d = 2$.

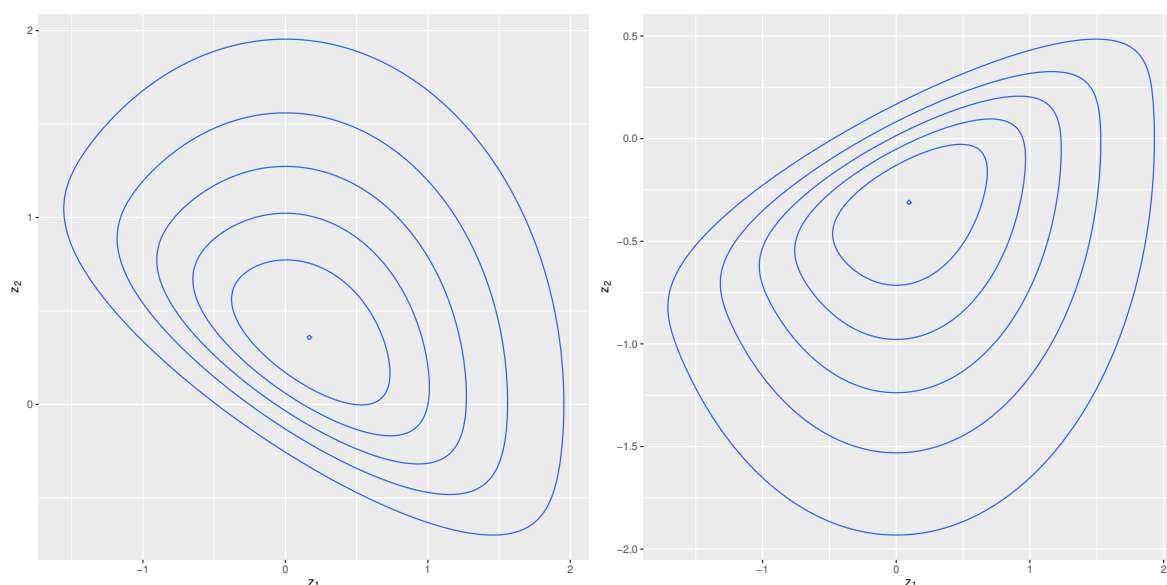


Figure 1. Contour level plots of two bivariate skew-normal densities given by (8); in both cases, $\bar{\Omega}$ is equal to the identity matrix, α is $(2, 4)$ in the first plot, it is $(2, -6)$ in the second plot.

If it was later shown by Azzalini and Capitanio [11] that the pair $(\bar{\Omega}, \alpha)$ uniquely identifies $(\bar{\Psi}, \delta)$, for any choice of the correlation matrix $\bar{\Omega}$ and any d -vector α . So $(\bar{\Omega}, \alpha)$ can equally be adopted as a legitimate parameterization of the family. Since in applied work we need to regulate location and scale, introduce the additional transformation

$$Y = \xi + \omega Z, \quad (11)$$

where $\xi \in \mathbb{R}^d$ and $\omega = \text{diag}(\omega_1, \dots, \omega_d)$ has positive diagonal elements. We correspondingly write $Y \sim \text{SN}_d(\xi, \Omega, \alpha)$, where $\Omega = \omega \bar{\Omega} \omega$.

In a broad sense, the role of the various regulating parameters is as follows: ξ controls location, ω scale, $\bar{\Omega}$ the dependence structure, α departure from normality. The phrase ‘in a broad sense’ is there to caution on the fact that, when we come to commonly used summary measures, these are actually influenced by all terms. For instance, the mean value of Y is not just ξ , but a function of all components parameters. This is the reason why the symbol ξ has been adopted, instead of the more popular μ . However, it is true that, for a given choice of (Ω, α) , an addition $+u$, say, to the ξ vector produces precisely the same addition $+u$ on $\mathbb{E}\{Y\}$, which justifies use of the name ‘location parameter’ for ξ .

Besides the stochastic representation (6) of additive form, the SN distribution allows other representations. An especially important one is the following, based on a conditioning mechanism. Consider the multivariate normal variable

$$X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix} \sim N_{d+1}(0, \Omega^*), \quad \Omega^* = \begin{pmatrix} \bar{\Omega} & \delta \\ \delta^\top & 1 \end{pmatrix}, \quad (12)$$

where Ω^* is a positive-definite correlation matrix. Then, both variables

$$Z' = (X_0 | X_1 > 0), \quad Z = \begin{cases} X_0 & \text{if } X_1 > 0, \\ -X_0 & \text{otherwise} \end{cases} \quad (13)$$

have density function (8). Strictly speaking, the equality sign following the random variable Z' should be the one of equality in distribution, but the simpler notation matches the one for Z . In addition, this notation expresses how the result is going to be employed in practice, for simulation purposes, to sample pseudo-random numbers from the target distribution.

The additive stochastic representation (6) and the two variants of the conditioning (13) establish direct connections between the SN family and various subject-matter motivated constructions, as elucidated later on in more detail. A further type of stochastic representation is presented at pp. 129–130 of Azzalini and Capitanio [7].

A direct implication of the second variant of representation (13) is the property of *perturbation invariance*: any even transformation of Z has the same distribution of the same transformation applied to X . Some noteworthy implications are

$$\mathbb{E}\{Z^\top Z\} = \bar{\Omega}, \quad (Y - \mu)^\top \Omega^{-1} (Y - \mu) \sim \chi_d^2. \quad (14)$$

The perturbation invariance property of the SN family is a special instance of a more general statement presented in the next subsection.

The SN distribution enjoys a high level of mathematical tractability, which approaches the one of the multivariate normal distribution. Among its many formal properties, a special mention is due for closure of the family with respect to affine transformations. Specifically, if $Y \sim \text{SN}_d(\xi, \Omega, \alpha)$, then

$$a + A^\top Y \sim \text{SN}_q(a + A^\top \xi, A^\top \Omega A, \tilde{\alpha})$$

for a q -vector a and a full-rank $d \times q$ matrix A ; here $\tilde{\alpha}$ is a q -vector, in which the explicit expression is given, for instance, on p. 133 of Azzalini and Capitanio [7]. A direct implication is that any q -dimensional marginal distribution of Y still has SN distribution. The moment generating function of Y takes the simple form

$$M(t) = 2 \exp\left(t^\top \xi + \frac{1}{2} t^\top \Omega t\right) \Phi(\delta^\top \omega t), \quad t \in \mathbb{R}^d. \quad (15)$$

Then, from the cumulant generating function $K(t) = \log M(t)$, we obtain

$$\mathbb{E}\{Y\} = \xi + \omega \mu_z, \quad \text{var}\{Y\} = \Omega - \omega \mu_z \mu_z^\top \omega,$$

where $\mu_z = \mathbb{E}\{Z\} = (2/\pi)^{1/2} \delta$.

One of the few limitations of the multivariate SN family is the lack of closure under conditioning, that is, if $Y = (Y_1, Y_2)$ has a SN distribution, then the distribution of Y_1 conditionally on the value taken on by Y_2 is not of SN type. This property can be achieved by considering a simple extension of

the family, denoted ‘extended skew-normal’, which includes an extra parameter, $\tau \in \mathbb{R}$. The density function corresponding to (8) is now

$$\varphi_d(x; \tilde{\Omega}, \alpha, \tau) = \varphi_d(x; \tilde{\Omega}) \frac{\Phi(\alpha_0 + \alpha^T x)}{\Phi(\tau)}, \quad x \in \mathbb{R}^d,$$

where

$$\alpha_0 = \tau \left(1 + \alpha^T \tilde{\Omega} \alpha \right)^{1/2}. \quad (16)$$

When Y is defined as at (11), the corresponding density function at $x \in \mathbb{R}^d$ is

$$\varphi_d(x - \xi; \Omega) \frac{\Phi\{\alpha_0 + \alpha^T \omega^{-1}(x - \xi)\}}{\Phi(\tau)}, \quad (17)$$

and we write $Y \sim \text{SN}_d(\xi, \Omega, \alpha, \tau)$. Clearly, if $\tau = 0$, we return to the original SN distribution. The corresponding moment generating function is

$$M(t) = \exp(t^T \xi + \frac{1}{2} t^T \Omega t) \Phi(\tau)^{-1} \Phi(\tau + \delta^T \omega t),$$

which shows that the moments of the distribution of Y are non-linear functions of τ .

The extended skew-normal distribution arises under the constructive route describe above, if the term $|U_1|$ in (5) is replaced by one having distribution $N(-\tau, 1)$ truncated below 0. Similarly, the distribution arises from a conditioning mechanism similar to the first expression of (13) if the condition $X_1 > 0$ is replaced by $X_1 + \tau > 0$.

The introduction of the additional parameter τ widens the ranges of the measures of asymmetry and kurtosis with respects to the original SN family. However, the increase of flexibility in this sense is not substantial, as visible from Figure 2.5 of Azzalini and Capitanio [7] for the univariate case. The main advantages of the extended SN family are arguably two others: (a) the more general generating mechanisms indicated in the previous paragraph, which make the distribution a more plausible ‘physically motivated’ model in a number of applications; (b) as anticipated, the extended family is closed under a conditioning operation, a property which turns out to be useful in some applications; the explicit expression of the corresponding parameters after conditioning are available, for instance, on p. 151 of Azzalini and Capitanio [7].

The price to pay for these attractive formal properties is the lack of an analogue of the second stochastic representation in 13. In turn, this removes the property of perturbation invariance, so that (14) and similar facts do not hold any longer.

Many other formal results can be obtained on the multivariate SN distribution, many of them with relatively little effort. A self-contained account of the main results, built collecting a number of contributions from the literature, is presented in Chapter 5 of Azzalini and Capitanio [7]. See, specifically, the results of Genton et al. [12], Capitanio [13], Balakrishnan and Scarpa [14], and Balakrishnan et al. [15], among others.

2.3. A General Result

The overall target of the rest of our presentation is to introduce a number of variants and extensions, of various degrees of generality, of the SN distribution. Many of these other constructions replace the normal density appearing in (8) by some alternative, but other directions are also considered. A substantial fraction of these extensions can be embraced in the following result presented by Azzalini and Capitanio [11] (p. 599). Here, and in the following, the phrase ‘ X is symmetric about 0’ referred to a random variable X is a shorthand for ‘ X is symmetrically distributed about 0’.

Proposition 1. Denote by T a continuous real-valued random variable with distribution function G_0 , symmetric about 0, and by Z_0 a d -dimensional variable with density function f_0 , independent of T , such that the real-valued variable $W = w(Z_0)$ is symmetric about 0. Then,

$$f(x) = 2 f_0(x) G_0\{w(x)\}, \quad x \in \mathbb{R}^d, \quad (18)$$

is a density function.

Proof. Note that the random variable $T - W$ is symmetric about 0, and expand $\mathbb{P}\{T - W \leq 0\}$ as the expected value of conditional probabilities, leading to

$$\frac{1}{2} = \mathbb{P}\{T \leq W\} = \mathbb{E}_{Z_0}\{\mathbb{P}\{T \leq w(Z_0)|Z_0 = x\}\} = \int_{\mathbb{R}^d} G_0\{w(x)\} f_0(x) dx.$$

□

An interesting fact is that the re-normalizing constant in (18) is universally 2. In addition, in the light of the prominent role played in the subsequent discussion by the condition of symmetry for f_0 , it is worth emphasizing that f_0 here can be *any* density function.

Implicit in the proof of Proposition 1, there is an acceptance-rejection argument which, combined with symmetry of $w(Z_0)$, leads to the following stochastic representations. If Z_0 and T are random variables as in Proposition 1, then both variables

$$Z' = (Z_0 | T \leq w(Z_0)), \quad Z = \begin{cases} Z_0 & \text{if } T < w(Z_0) \\ -Z_0 & \text{otherwise} \end{cases} \quad (19)$$

have density function (18). A corollary of the second representation in (19) is the property of *perturbation invariance* stated next.

Proposition 2. Under the assumptions of Proposition 1, denote by Z_0 a random variable with density function f_0 and by Z a random variable with density f . Then, for any function $t(\cdot)$ taking values in $\mathbb{R}^q$ ($q \geq 1$), such that $t(x) = t(-x)$ for all $x \in \mathbb{R}^d$, the transformed variables $t(Z)$ and $t(Z_0)$ have the same distribution.

In the univariate case, densities (18) can be viewed as weighted forms of f_0 , in the sense examined by Rao [16]. This particular sub-class of weighted distributions enjoys special features, namely the stochastic representations (19) and the implied property of perturbation invariance.

2.4. Symmetry-Modulated Distributions

After the publication of the Azzalini and Dalla Valle [10] paper, subsequent contributions in the literature have generated a number of extensions at progressive level of generality. While we shall focus mostly on the stream labeled ‘SEC distributions’, it is appropriate to delineate at least the main traits of the broader picture, for a more comprehension view. We set our exposition at a level which encompasses the majority of proposed formulations, especially those more directly connected to applied work, while retaining a relatively simple mathematical level.

Even more general formulations, moving along two different paths, are those of Arellano-Valle et al. [17] and Jupp et al. [18]. However, these treatments encompass also many situations which for technical reasons cannot be pursued operationally. The discussion below covers the vast majority of the practically workable constructions, as known at the time of writing.

Proposition 3 below, obtained by Azzalini and Capitanio [19], represents an easy-to-use restricted version of Proposition 1, since under conditions (20) it is immediate that the condition of symmetry of $w(Z_0)$ required in Proposition 1 is fulfilled. The result provides a simple method to build variant forms of a *baseline* density, f_0 say, which is required to be centrally symmetric density; that is, such that

$f_0(-x) = f_0(x)$. Multiplication of f_0 by a modulation factor produces *perturbed* or *modulated* versions of f_0 .

Proposition 3. Denote by f_0 a probability density function on $\mathbb{R}^d$, by $G_0(\cdot)$ a continuous distribution function on the real line, and by $w(\cdot)$ a real-valued function on $\mathbb{R}^d$, such that

$$f_0(-x) = f_0(x), \quad w(-x) = -w(x), \quad G_0(-y) = 1 - G_0(y) \quad (20)$$

for all $x \in \mathbb{R}^d, y \in \mathbb{R}$. Then, (18) is a density function on $\mathbb{R}^d$.

In independent work, Wang et al. [20] have obtained an essentially equivalent formulation, which can be expressed as follows. On setting $G(x) = G_0\{w(x)\}$, we can rewrite conditions (20) as

$$f_0(-x) = f_0(x), \quad G(x) \geq 0, \quad G(x) + G(-x) = 1 \quad (21)$$

for all $x \in \mathbb{R}^d$, and write the density as

$$f(x) = 2 f_0(x) G(x).$$

Not only if w and G_0 satisfy (20), then $G(x) = G_0\{w(x)\}$ satisfies (21), but also the essentially converse statement can be shown to hold: for any function $G(x)$ satisfying (21), there exist functions w, G_0 satisfying (20) leading to the same G ; hence, the same f , is in fact a multitude of them.

The conclusion is that, for any centrally symmetric density $f_0(x)$, Proposition 3 and the variant form based on (21) generate the same set of distributions. In this sense, they are equivalent formulations. Which of the two forms to adopt is a matter of convenience and taste. Version (21) is more suitable for mathematical work, but the actual construction of functions G with the required properties is more naturally approached using form (20).

Clearly, the SN density (8) is of type (18) with $f_0 = \varphi_d$, $w(x) = \alpha^\top x$, $G_0 = \Phi$. The vast majority of the distributions to be discussed later are of type (18) or some ‘extended’ version of them. The term *skew-symmetric distributions* is often used to identify these constructions, but we prefer the term *symmetry-modulated* for the following reason: skewness is typically the form of departure from symmetry associated with linear or mildly non-linear choices of $w(x)$, as it was the case in earlier constructions of this logic, but more elaborate functions $w(x)$ lead to distributions where skewness is not the more prominent feature. Figure 2 illustrates this point by using a strongly non-linear functions $w(x)$ to modulate, via $\Phi(w(x))$, a standard bivariate normal density function with independent components.

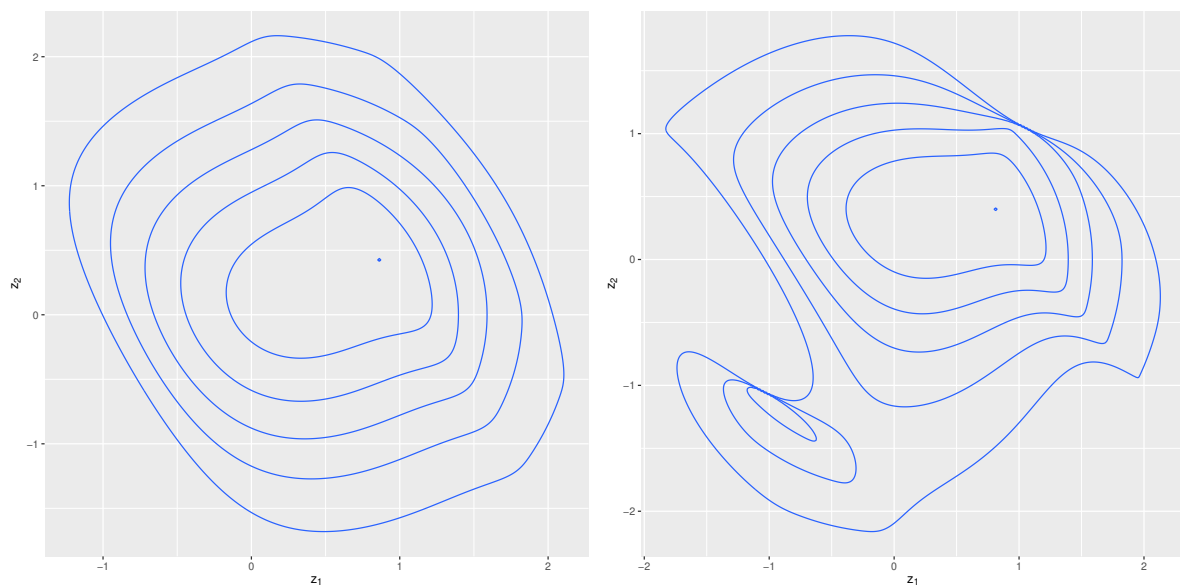


Figure 2. Examples of symmetry-modulated density generated from a baseline bivariate standard normal density with independent components.

The extension of the stochastic representations in (13), after orthogonalization to independent variables, is as follows. Under the conditions of Proposition 3, denote by X_0 a random variable with density f_0 and by T an independent variable with distribution function G_0 . Then, both variables

$$Z' = (X_0 | T \leq G_0(X_0)), \quad Z = \begin{cases} X_0 & \text{if } T \leq G_0(X_0), \\ -X_0 & \text{otherwise} \end{cases} \quad (22)$$

have density (18). The second form in (22) is more convenient for random number generation, since no rejection of a generated outcome X_0 can occur; the first form allows us to draw connections with a number formulations in applied domains where a selection mechanism of sample values occurs, regulated by a condition of type $T \leq G_0(X)$.

Similarly to the SN distribution and Proposition 2, the second variant of representation (22) leads to the perturbation invariance property: under the above conditions, if X_0 and Z have density function f_0 and f , respectively, and t is a even function from $\mathbb{R}^d$ to $\mathbb{R}^q$, then $t(X_0)$ and $t(Z)$ have the same distribution. For instance, if $Z = (Z_1, Z_2)$ is a random variable where density is depicted in each of the panels of Figure 2, then $Z_1^2 + Z_2^2 \sim \chi_2^2$, since this fact holds for the associated variable X_0 .

3. SEC and Other Modulations of EC Distributions

Any application of Proposition 3 with a baseline density f_0 of EC type (2) produces a density commonly labeled ‘skew-elliptical’, or SEC in our terminology; the terms apply even after a possible a location shift of the distribution. In this section, we review the main constructions of this sort and some closely related ones. The prefix ‘skew’ attached to many construction of this sort arose from the earlier constructions, where the most noticeable form of departure from a symmetric or elliptical shape was as asymmetry. However, there are cases, such as those depicted in Figure 2, where the asymmetry is not the more prominent feature. In these case, a term like ‘modulated elliptical contoured’ (MEC) distributions may be more appropriate.

3.1. SEC Distributions with a Linear-Combination Argument

An early form of SEC distribution has been put forward by Azzalini and Capitanio [11] as a corollary of Proposition 1. If we take f_0 to be of EC type, the most immediate SEC construction sets $w(x)$ to be a linear combination of the x components, so that

$$f(x) = 2 f_0(x) G_0(\alpha^\top x), \quad x, \alpha \in \mathbb{R}^d, \quad (23)$$

where α is a vector of constants. The linear form $w(x) = \alpha^\top x$ has the appeal of conceptual simplicity; it also mimics the structure of the multivariate SN density (8). In fact, when $f_0 \equiv \varphi_d$, we go back to the SN density.

For any given f_0 , (23) allows a wide range of choices for G_0 , as long as they meet the condition $G_0(t) + G_0(-t) = 1$. This flexibility is appealing but also problematic, in a sense, since it is there that there is no naturally preferable choice of G_0 .

By analogy with the SN case, we may think of taking G_0 equal to the distribution function of the univariate marginal of f_0 . Unfortunately, this analogy does not seem to convey any special virtue or simplification. This point is confirmed by consideration of univariate instances of (23), as examined by Gupta et al. [21] and Nadarajah and Kotz [22]. Even computation of basic quantities, like lower-order moments, becomes appreciably more complicated than in the SN case.

A prominent instance in this respect is provided in Section 4 of Nadarajah and Kotz [22], where f_0 is taken to be the univariate Student's t on ν degrees of freedom and G_0 the corresponding distribution function, leading to what could we denote 'linear skew- t distribution'. The expression of odd moments in their Theorem 1 is undoubtedly intricate.

Besides the SN case, densities (23) do not lead, as far as we know, to results matching stochastic representations (6),(13) and their useful implications. What remains valid are representations (22) and the connected property of perturbation invariance. From the latter, a key implication is that even-order moments are equal to those of f_0 , when they exist.

3.2. SEC Distributions via Conditioning

Branco and Dey [23] have put forward an alternative SEC formulation which has proved to be very fruitful and has, since then, received much attention. Explicitly or implicitly, it appears to be regarded by many authors as the standard form of SEC distributions. The key idea in Branco and Dey [23] is to extend the first generating mechanism in (13) by applying it when the normality assumption for X in (12) is replaced by the one of an elliptical distribution.

More specifically, we start from $X = (X_0, X_1^\top)^\top$ with density of EC type (2) and Σ equal to Ω^* in (12) and, for simplicity, set $\mu = 0$. The distribution of $(X_0 | X_1 > 0)$ can then be written as

$$f(z) = 2 f_0(z) F_{q(z)}(\alpha^\top z), \quad z \in \mathbb{R}^d, \quad (24)$$

where f_0 is the marginal density of X_0 induced by $p(x; 0, \Omega^*)$, and $F_{q(z)}$ is the associated conditional univariate distribution function of the remaining component. Except in the normal case, this conditional distribution depends on $q(z) = z^\top \Omega^{-1} z$. An alternative integral form of density $f(z)$ is given in expression (2.5) of Branco and Dey [23].

Within the class of distributions which can be built from this paradigm, special interest lies on the wide sub-class where the parent $(d+1)$ -dimensional density $p(x; 0, \Omega^*)$ corresponds to a scale mixture of normal variates as specified in (3). It can be shown that (24) corresponds then to a scale mixture of SN variates, retaining the same mixing distribution for the scale factor. With the inclusion of location and scale parameters, these scale mixtures can then be represented more conveniently as

$$Y = \xi + \omega S Z_0, \quad (25)$$

where $Z_0 \sim \text{SN}_d(0, \bar{\Omega}, \alpha)$, and ξ, ω are as in (11). Besides its conceptual significance, (25) facilitates computation of moments of Y , hence of (24), when moments of S up to the same order are known, given that the SN moments are available.

The paper of Azzalini and Capitanio [19] has explored further the distributions generated by Proposition 3 and, within this frame, the SEC class (24) in more detail. Among other results, their Proposition 9 states that the SEC distributions (24) can also be obtained via a different route, that is, an additive construction like in (5) and (6) but with (U_0, U_1) having now a joint $(d + 1)$ -dimensional EC distribution with a scale matrix like the one of U in (6).

Another question examined in Azzalini and Capitanio [19] was: do densities of type (24) belong to the SEC or MEC family as described as the beginning of the present section? That is, is (24) an EC density modulated as in Proposition 3? Among other facts, this condition is important to ensure the property of perturbation invariance. It is not immediate that the final factor of (24) can be written in the form of the matching factor in (18). The answer turned out to be ‘yes’ for a range of parametric families examined in the paper. Confirmation that the answer is ‘yes’ in all cases was provided later on by Azzalini and Regoli [24]. In this paper, it is also proved that densities (24) are quasi-concave, which means that all their level curves delineate convex regions. This property rules out twisted shapes like the one on the right side of Figure 2, so that the term SEC appears appropriate in this case.

An important instance of (25) occurs when S is chosen as in (4). The random variable so obtained is said to have a skew- t (ST) distribution, and we write $Y \sim \text{ST}_d(\xi, \Omega, \alpha, \nu)$, where $\Omega = \omega \bar{\Omega} \omega$. Some basic implications are: (i) if $\alpha = 0$, the Z_0 variable in (25) reduces to a normal one and the ST distribution reduces to a regular Student’s t ; (ii) if $\nu \rightarrow \infty$, then $S \rightarrow 1$ in probability, and the ST distribution reduces to the SN. The ST distribution has been the focus of much subsequent work, both on the theoretical and the applied side. In this section, we summarize the theoretical aspects; subsequent sections deal with the more applied side.

For the ST family, compliance with conditions (20) is visible from an alternative and simpler expression of the density function: in the normalized case $\text{ST}_d(0, \bar{\Omega}, \alpha, \nu)$ with null location and unit scale parameters, this is

$$f_{\text{ST}}(z) = 2 t_d(z; \bar{\Omega}, \nu) T \left(\alpha^\top z \sqrt{\frac{\nu + d}{\nu + q(z)}}; \nu + d \right), \quad (26)$$

where $t_d(z; \bar{\Omega}, \nu)$ is the regular multivariate Student’s t density on ν degrees of freedom, $T(\cdot; \rho)$ is the univariate t distribution function of ρ degrees of freedom, and $q(z) = z^\top \bar{\Omega}^{-1} z$. Since the argument of $T(\cdot; \nu + d)$ is an odd function of z , $w(z)$ say, it is immediate that (26) falls within the conditions of Proposition 3. Expression (26) was also obtained independently by Gupta [25].

Besides a simple expression of the ST density function, the quoted papers include results for its distribution function, moments, the distribution of affine transformations and of quadratic forms. Additional results on the multivariate ST distribution, or more generally on scale mixtures of multivariate SN variables, are provided by Kim and Mallick [26], Kim [27], Capitanio [13], Giorgi [28], Arevalillo and Navarro [29], and Kim and Zhao [30], among others.

Although the $\text{ST}_d(\xi, \Omega, \alpha, \nu)$ distribution features only a scalar ν parameter in addition to the $\text{SN}_d(\xi, \Omega, \alpha)$ counterpart, this ingredient provides the possibility to regulate the decay of the density in the tails of the distribution, in the univariate and the multivariate domain, where in the latter case the term ‘tail’ must be suitably intended. The substantial increase in flexibility so achieved in practical data fitting has been demonstrated by Azzalini and Genton [31], considering data from a variety of application areas and data types, and it has subsequently been confirmed by other authors in a range of applied problems.

Wang and Genton [32] introduce another multivariate SEC family: the skew-slash distribution. Additional work on the SEC class has been presented by Kim and Genton [33], Shushi [34], and Arellano-Valle et al. [35], among others.

Similarly to the SN distribution and with similar motivations, an ‘extended’ ST distribution has been studied, in independent work of Adcock [36] and of Arellano-Valle and Genton [37]. Both formulations start from a $(d + 1)$ -dimensional Student’s t variable as the building ingredient, but interestingly, they follow different paths. The development of Adcock [36] is based on the additive construction, qualitatively similar to (6), with the U_1 variable is centered at τ ; Arellano-Valle and Genton [37] use instead a conditioning mechanism, like for (24), but with condition $X_1 > 0$ replaced by $X_1 + \tau > 0$. Up to minor differences in the adopted parameterization, the density of the variable, once equipped with location and scale parameters like in (25), turns out by either route to be

$$t_d(y - \xi; \Omega, \nu) T \left[\sqrt{\frac{\nu + d}{\nu + Q(y - \xi)}} \left\{ \alpha_0 + \alpha^T \omega^{-1}(y - \xi) \right\}; \nu + d \right] T(\tau; \nu)^{-1}, \quad y \in \mathbb{R}^d, \quad (27)$$

where $Q(x) = x^T \Omega^{-1} x$ and α_0 is as in (16). Additional related work, directly motivated by applications, has been presented by Lee et al. [38] and Marchenko and Genton [39], to be discussed later on. Additional work of Adcock [40] concerns both the theoretical and the applied side, with special emphasis on quantitative finance.

It might be noticed that, in the developments presented above, there is a prevalence of a limited number of SEC families, as compared to the large number of conceptual possibilities. There is especially much emphasis in the ST distribution and its extended version (27), and a few more instances, such as the skew-slash distribution, while there is very little consideration of, say, the multivariate exponential power distribution and other families. One aspect which comes into play is the result of [9], recalled in Section 2.1, implying that many families, including the multivariate exponential power distribution, are not closed under the marginalization operation; that result was stated for the EC class, but it carries on for the SEC class. Another advantageous property is closure under a linear transformation of a multivariate variable, or at least the existence of a tractable distribution for such transformation.

More complex formulations, such as some of those discussed in the review paper [5], allow for a separate tail parameter for each component variable. This offers an increased level of flexibility, but at some cost in terms of mathematical tractability. When one is setting-up the mathematical formulation for a given applied problem, a balance must be struck between various components, especially flexibility, mathematical tractability and subject-matter interpretability. The appropriate point of balance depends on the problem at hand.

3.3. Constructions Based on Several Latent Variables

In some form or another, all d -dimensional distributions discussed so far, except for the general result of Proposition 1, are obtained starting from a $(d+1)$ -dimensional symmetrically distributed random variable of which one component is not observed but in which its role is to modify the symmetric distribution of the others. This fact is most easily seen in the SEC construction by conditioning leading to (24) where the component X_1 of the vector X constitutes a latent variable, where the effect is to perturb the symmetric distribution of X_0 . In other instances, this component may be slightly less obvious to identify but it is there. For example, in the additive construction (5)–(6), the univariate variable which generates asymmetry is $|U_1|$.

The constructions to be examined below follow similar logics to the earlier ones, but they involve use of multiple latent variables. This extension should allow to model in a more realistic way a wider range of practical problems. Moreover, in most cases, although not all of them, the distributions so obtained are regulated by a larger number of parameters, for a given dimensionality d and equivalent structure of the other components. Therefore, a more adequate fit to observed data is expected. At the same time, when such an elaborate construction is not really necessary in a given applied problem, there is a risk of over-parameterization, not necessarily in the formal sense of a non-identifiable model, but as one with a nearly-flat log-likelihood function, and, consequently, poorly estimated parameters.

An early instance of this scheme is the one of Sahu et al. [41], which is closely linked to the earlier SEC formulation of Branco and Dey [23]. Instead of a 1-dimensional conditioning operated on a $(d+1)$ -dimensional EC variable, now a d -dimensional conditioning is applied to a $(2d)$ -dimensional EC variable. However, given the required constraints on the EC variance structure, the number of parameters regulating the distributions is the same of the analogous construction using the earlier scheme. Hence, for instance, the new form of skew- t distribution so constructed is regulated by $d(d+1)/2 + 2d + 1$, the same number of (26). Note that here, when $d > 1$, the term skew-normal and skew- t distribution refer to different distributions of those presented in Section 2.2,3.2, respectively.

Shortly afterwards, a number of generalizations of the SN distribution (8) have been put forward by González-Farías et al. [42] and González-Farías et al. [43] with the name ‘closed skew-normal’ (CSN) to underline the property of closure under convolution, as well as under conditioning. In this formulation, one starts from a $(d+m)$ -dimensional normal variable (X_0, X_1) and considers the distribution of X_0 conditionally on $X_1 + \tau > 0$, where X_1 and τ are m -dimensional and the inequality must be intended to hold for each component. Similar constructions have been put forward independently by Liseo and Loperfido [44] under the name ‘hierarchical skew-normal’ and by Arellano-Valle and Genton [45] under the name ‘fundamental skew-normal’. The not-so-obvious interconnection among these three proposals have been examined by Arellano-Valle and Azzalini [46], who have shown their essential equivalence up to a change of parameterizations, once some redundancies are removed. This clarification has led to a ‘unified skew-normal’ (SUN) formulation.

An additional development of Arellano-Valle and Azzalini [46] has been the extension of this type of construction when the distribution of (X_0, X_1) is of EC type instead of normal. This line of development has been carried on by Arellano-Valle and Genton [47] and the associated term is ‘unified skew-elliptical’ (SUE or SUEC) distributions. Further work in this direction has been presented by Arellano-Valle et al. [48], Abtahi and Towhidi [49], Aziz and Gupta [50], Gupta et al. [51], and Richter and Venz [52].

To facilitate the comprehension of the relative position of the main classes of distributions discussed so far, plus some others to be presented later, it is useful to examine Figure 3. At the core of the diagram seats the multivariate SN distribution with density (8). All other constructions define supersets of the SN family at various level of generality, often moving in different directions. The set ‘Lemma 1999’ refers to the distributions generated by Proposition 1; ‘Sym-Mod 2003/04’ denotes the set of distributions generated by Proposition 3; ‘SEC-1999’ denotes the distributions in equation (23); ‘SEC-2001’ refers to the skew-elliptical class introduced in Branco and Dey [23]; ‘SEC-2003’ denotes the class of distributions introduced by Sahu et al. [41]. The label ‘MEC-2018’ will be described later on.

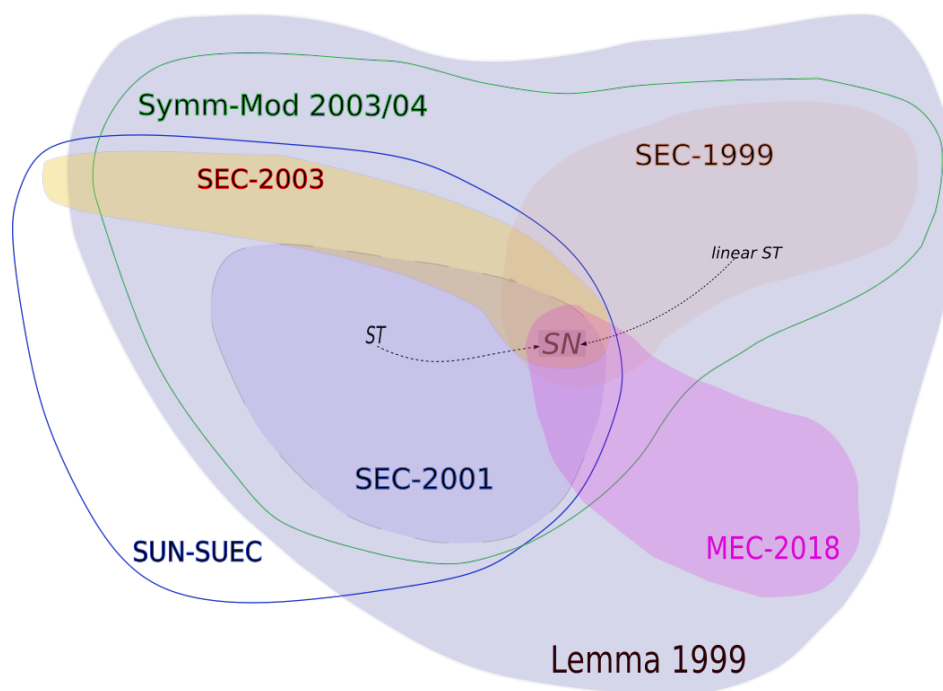


Figure 3. Diagram representing the relative position of some selected classes of distributions. SUN = unified skew-normal; SEC = skew-elliptically contoured; SUEC = unified skew-elliptical; SN = skew-normal; MEC = modulated elliptical contoured.

3.4. Other Mixtures of SN Variables

Since scale mixtures of SN variables with representation (25) constitute the more interesting subclass of SEC distributions (24), it is not surprising that various authors have examined other forms of SN mixtures.

Arellano-Valle et al. [53] and Arellano-Valle et al. [54] have studied shape mixtures of multivariate SN variables, in a number of variant constructions. Proceeding further in this direction, Arellano-Valle et al. [55] have examined the more general construction with random mixtures of both the scale and the slant parameter of SN variates. The SSMSN class so generated encompasses a number of existing formulations, forming a wide subset of those generated by Proposition 3; see their Proposition 1 which must be interpreted in the sense stated in its preceding paragraph. Further work in the same direction, but only for the univariate case, has been presented by Tamandi and Jamalizadeh [56].

Arsilan [57] has introduced the idea of variance-mean mixture of the multivariate SN variables, that is, with representation

$$Y = \xi + \gamma V + \sqrt{V} \omega Z_0, \quad (28)$$

where ξ, ω and Z_0 are as in (25) and γ is a d -dimensional parameter. The positive random variable V , independent of Z_0 , operates a mixing operation on the SN variable Z_0 in a way which replicates the one in the well-known construction of the variance-mean mixture of a normal variable; see Barndorff-Nielsen et al. [58]. Similarly to the mixing of normal variables, the more prominent special case of (28) occurs when V has a generalized inverse gamma distribution.

3.5. Miscellanea

In the above exposition, a recurring scheme to achieve perturbation of symmetry has been via conditioning events, such as $(X_1 > 0)$ in (13) or $(X_1 + \tau > 0)$, for the extended SN and CSN/SUN distribution or some similar sort of one-sided constraints in the more general case of EC and other centrally symmetric distributions. Some qualitatively similar event, such as $T \leq w(Z_0)$,

is involved in (19). The general treatment of Arellano-Valle et al. [17] deals instead with an arbitrary conditioning event. Starting from a $(d + m)$ -dimensional variable (U, V) , the marginal density f_V of the V component, as modulated by the conditioning event $\{U \in C\}$, is

$$f_V(x) \frac{\mathbb{P}\{U \in C | V = x\}}{\mathbb{P}\{U \in C\}}, \quad x \in \mathbb{R}^d. \quad (29)$$

This expression is extremely general, since C is an arbitrary set. However, to translate (29) into operational constructions, one must obtain expressions for the two probabilities appearing there. These computations are amenable in some favorable cases only, many of which have already been examined in the formulations discussed earlier. Still, some additional tractable cases exist. One of them has been considered briefly by Genton [59] and more systematically by Kim [60] where the conditioning event involves a set of bilateral constraints of type $(a < U < b)$, say, in a normal distribution context.

In another direction of work, Jamalizadeh and Kundu [61] apply a multivariate version of the Birnbaum-Sanders transformation of variables to the multivariate SN distribution.

To demonstrate the potential of Proposition 1 beyond the domain of validity of Proposition 3, Azzalini and Regoli [62] have constructed densities with baseline f_0 of EC type but without enforcing the condition $w(-x) = -w(x)$ in (20). The resulting distributions enjoy some interesting formal properties and they can have similar flexibility of those generated from Proposition 3. In the diagram of Figure 3, the set of these distributions is labeled ‘MEC-2018’.

An account on matrix-variate SN distributions and extensions, with related formulations for applied work, is provided by Gupta et al. [63] (Chapter 11); this includes references to earlier work in the same direction.

3.6. Cave Nomina

An unpleasant aspect of the stream of literature under consideration is represented by certain aspects of its terminology, especially confusing to the newcomer who is hoping to grasp the key concepts. The problem comes in two variants. In the first form, we encounter different names attached to the same object, or very nearly the same object. A good case in point is represented by the names closed SN, hierarchical SN, fundamental SN, and unified SN. They all refer to the same class of distribution up to inessential details, but, as explained in Section 3.3, the first three of them have been developed independently from each other, and the attempt to unify them under the fourth name has not succeeded. So now there exist four names, and at least three of them in current use, all referring to the same entity.

The second form of the problem is the reverse of the first one: the same name is attached to different objects. Arguably, the most prominent instance of this sort is represented by the term ‘skew- t ’ and its acronym ST. As far as we know, the first occurrence of the term is in Branco and Dey [23], later adopted by many authors with the same meaning; we regard this version, with density (26), as the more convincing ST construction. However, shortly afterwards, a different distribution, again called skew- t , has been presented in Sahu et al. [41]; this ST distribution coincides with the earlier one only when $d = 1$. But these are only two examples of skew- t distributions among the many, not all developed inside the present stream of literature. We do not list all such instances, but we must at least mention the paper by Giorgi and McNeil [64], where two types of ST distributions coexist in the same publication: one with density (26), and the second one coming from an entirely different construction, linked to the GH distribution.

Those mentioned here are only some examples of confusing terminology; unfortunately, others exist. Given the persistence of the current nomenclature, it seems difficult to imagine that the terminology will be rationalized, except perhaps after a long time. Therefore, readers of the literature must not rely simply on the names of the distributions: a direct inspection of the formal expressions of these densities, or some equivalent ingredient, is required.

4. Applications as Methodological Developments

There exists two logically distinct meanings of the term ‘applications’, at least in the present context. One such meaning refers to the use of the above-described distributions as a formal reference set for improving of existing statistical methods by building them on more flexible and realistic distributional assumptions, as compared to traditional methods, which are typically built on the normality assumption. The other meaning of ‘applications’ refers to more specifically application-oriented work. Clearly, these two sorts of activity are not completely separate: methodological developments are usually illustrated by some numerical work, sometimes on non-trivial practical problems, and vice versa application-motivated work may involve substantial theoretical development. However, in most cases there is one component which drives the author’s work and we organize our exposition separating the material accordingly. The rest of the present section deals with the more methodology-oriented developments, while the next section focuses on the more applied side.

4.1. Regression Models and Variants

In univariate regression models, a set of p , say, explanatory variables x_i is observed alongside the response value on the i th individual. The x_i vector, regarded as non-random, is used to model some quantity of interest in the distribution of the response variable, Y_i . In the present context, the distribution of Y_i is assumed to be one of those examined in Section 2.3, most often the SN or the ST distribution but sometimes some other one. The quantity of interest is the location parameter for the i th response, denoted ξ_i , which is expressed as a function of x_i and some vector parameter, β say, which then becomes the primary parameter of interest. In the basic setting, the connecting function is a linear one so that we write $\xi_i = x_i^\top \beta$, but non-linear functions have also been considered. In the case of multivariate response variables, this scheme is generalized in the same way as within the classical theory.

Much work in this area has dealt with univariate response variables, and it can broadly be separated in two parts. A set of contributions have developed diagnostic methods and similar tools which extend popular methods from the standard methodology; these include Bolfarine et al. [65], Ferreira and Arellano-Valle [66], and Labra et al. [67]. Other contributions work with formulations similar to but different from classical regression, such as measurement in error models, change-point detection, and others; these include Arellano-Valle et al. [68], Figueiredo et al. [69], and Arellano-Valle et al. [70].

A more limited set of contributions have considered similar problems in the multivariate context. Expressions of the score functions of the (profile) log-likelihood and other results useful for maximum likelihood estimation (MLE) are provided by Azzalini and Capitanio [11] for the SN case, and by Azzalini and Capitanio [19] for the ST case. Multivariate measurement error models have been considered by Lachos et al. [71] and by Arellano-Valle et al. [72].

For the analysis of binary data in a context of generalized linear models, Chen et al. [73] have proposed to use the distribution function of the univariate SN distribution to extend the probit model. This direction has been developed further in a Bayesian framework by Bazán et al. [74]. In a similar logic, item response theory has been extended to a more general formulation by Bazán et al. [75]; see also [76,77]. For Bayesian inference of a probit model, Durante [78] shows that the exact posterior distribution of the generalized regression parameters under Gaussian prior is of SUN type.

4.2. Finite Mixtures and Model-Based Clustering

Finite mixtures of g , say, component distributions provide a way to introduce a further layer of flexibility to data fitting. In their essential formulation, a density of interest, $f(x)$ say, is written as

$$f(x; \theta) = \sum_{j=1}^g \pi_j f_j(x; \theta_j) \quad (30)$$

where f_j is a component density regulated by a set of parameters θ_j , and π_j is a positive probability such that $\pi_1 + \dots + \pi_g = 1$. While not conceptually necessary, it is usually the case that the g component densities are assumed to belong to the same parametric family, each with a different set of parameters, $\theta_1, \dots, \theta_g$. Suitable conditions are necessary to avoid non-identifiability of the model. More detailed information is provided by classical general accounts, such as the monograph of McLachlan and Peel [79].

Construction (30) is particularly attractive in the context of model-based clustering, where it lends itself to a natural interpretation by regarding f_j as the density of the j th subpopulation, having relative weight π_j in the overall population which comprises g groups or subpopulations. Consequently, the literature on finite mixtures has a major overlap with the one on model-based clustering, especially in the multivariate context.

As with many other research areas, this one has also long focused on the assumption of normality, which means that each component density of (30) was assumed to be Gaussian. Later on, more flexible parametric families have been adopted as component densities, to introduce a more realistic formulation and to reduce the number g of components, since in many instances the normal assumption requires more than one component to accommodate non-Gaussian data patterns even of evidently clustered points. In this development, the proposals discussed in Section 2.3 have been among the earlier alternatives to the Gaussian assumption which have been considered, and their role is still relevant. An exhaustive list of all pertaining contributions would be very long, cumbersome to compile, and almost instantaneously obsolete. Therefore, we highlight a selection of titles, which seem to us of special relevance for one reason or another.

Earlier work of this type has naturally started in the univariate context: Lin et al. [80] have developed a methodology for fitting finite mixtures of SN distributions; Lin et al. [81] carried on this to ST distributions. Otiniano et al. [82] have proved the identifiability of such mixtures.

The next step was to extend these constructions to the multivariate case; see Lin [83], Lin [84], Pyne et al. [85], Frühwirth-Schnatter and Pyne [86], Vrbik and McNicholas [87], and Vrbik and McNicholas [88], among others. The monograph Lachos Dávila et al. [89] collects a set of papers, with each paper targeting one classical problem in this domain.

More recent contributions have often adopted additionally flexible parametric families, such as the CSN/SUN or SUEC families recalled in Section 3.3. A notable instance is Lee and McLachlan [90] and the connected software tools described in Lee and McLachlan [91]. Alternative component distributions have been considered by Tamandi and Jamalizadeh [56].

The idea of finite mixture can be combined with the one of a regression model, when the relationship connecting variables varies in different subgroups of the population. A construction of this type under SN distributional assumption has been examined by Liu and Lin [92], while Zeller et al. [93] replace the SN assumption with a scale mixtures of SN variates, and Doğru and Arslan [94] use ST distributions.

4.3. Spatial and Spatio-Temporal Models

Several contributions have proposed formulations in the context of spatial statistics. This involves consideration of the distribution of a variable of interest, Y , observed at a set of points $s = (s_1, \dots, s_d)$ belonging to a given space $\mathcal{S}$; denote by $Y(s) = (Y(s_1), \dots, Y(s_d))$ the set of observed variables. In the simpler setting, the elements of $\mathcal{S}$ represent geographical locations, but they can represent coordinates

in a 3-dimensional space or a combination of spatial and time coordinates. In the classical formulation, it is assumed that a stationary spatial Gaussian process is defined over S , with a certain dependence structure which depends of the relative position of the points. This implies that the joint distribution of $Y(s)$ is multivariate normal, for any choice of the set of points s . Note that, for the relevant applications, it is important that, once the parameters of the spatial process have been estimated, we can compute the conditional distribution at $Y(s')$, where s' represents some other set of points, conditionally on the values observed on $Y(s)$. Think, for instance, of Y as the level of some pollutant over a certain region S and of s as the coordinates where d measuring stations are located.

Since in certain cases the Gaussian assumption appears unrealistic, one alternative is to replace the normal distribution by a more flexible parametric family. The SN assumption, possibly in its extended form (17), is a natural candidate here, given its simplicity and mathematical tractability. This represents the formulation proposed by Kim and Mallick [95] and adopted by some subsequent authors. Unfortunately, their formulation has been shown by Minozzo and Ferracuti [96] to be invalid, since the required spatial process cannot exist in a mathematically coherent sense with respect to marginalization operations. This problem prevents the conditioning operation indicated above. Furthermore, in Minozzo and Ferracuti [96] it is also shown that various other existing proposals have similar coherence problems.

A construction free from the above inconsistency problem has been proposed by Zhang and El-Shaarawi [97]. Their core ingredient is the process

$$Z(s) = \sqrt{1 - \delta^2} U_0(s) + \delta |U_1(s)|, \quad \delta \in (-1, 1), \quad (31)$$

where $U_0(s)$ and $U_1(s)$ are independent stationary Gaussian spatial processes with standard marginals. The process so constructed exists and, from (5), we see that each marginal of $Z(s_0)$ for a single point s_0 is skew-normal. The drawback is that, when s_0 denotes a set of points, not just one, the distribution of $Z(s_0)$ is not jointly skew-normal. The basic component $Z(s)$ is then supplemented by additional terms to regulate location and scale. An extension of this model to accommodate observed temporal replicates at each geographical location, illustrated by applied work, has been presented by Schmidt et al. [98]; see also the subsequent comments by Genton and Hering [99]. Additional related work includes Hering and Genton [100], Hosseini et al. [101], Karimi and Mohammadzadeh [102], Zareifard and Jafari Khaledi [103], Ahmad and Pinoli [104], Rimstad and Omre [105], Rezaie et al. [106], and Boojari et al. [107], among others.

4.4. Methods for Biostatistics and Medical Statistics

There is a range of statistical methods targeted largely, if not entirely, to biological and medical applications which have been improved by consideration of the families of distributions described above. At the basic level, there is, as in other areas, an improvement in data fitting when flexible parametric families are adopted in place of the normal and other traditional distributions, especially so in the multivariate context. At another level, there is a role in enhancing existing methodologies and even in helping to develop new ones.

Longitudinal data exist in a range of application areas, but they are especially important in medical statistics, with a particularly distinctive imprint when in combination with time-to-event data. A substantial number of contributions in this domain has been summarized by Azzalini and Capitanio [7] (pp. 218–220), and (also for space reasons) we do not reproduce them here. More recent contributions, focusing especially on those connected with multivariate response, include the work of Baghfalaki et al. [108], Huang and Dagne [109], Chang and Zimmerman [110], and Jana et al. [111].

For the joint modeling of longitudinal data and time-to-event outcomes, the skew-normal and related distributions provide ingredients for the log-likelihood expressions of Barrett et al. [112] and Baghfalaki and Ganjali [113]; in the latter case, the term ‘skew-normal distribution’ refers to its variant form in [41], recalled in Section 3.3, not to distribution (8).

The current trend in design of clinical trials favors interim analysis and adaptive designs. The work of Shun et al. [114] shows the relevance of the skew-normal distribution in this context. The formulation of Azzalini and Bacchieri [115] for adaptive designs hinges firmly on the CSN/SUN.

4.5. Methods for Observational Studies and the Social Sciences

In observational studies, especially frequent in the social sciences, a recurrent issue is the presence of a selection mechanism in the sampling process, leading to the problem of selection bias. Under assumption of normality, the proposal of Heckman [116] provides a method to correct for this bias, which has been widely applied in the social sciences and related fields. This selection has a direct link with the stochastic representation via a conditioning mechanism of the extended skew-normal distribution and the CSN/SUN construction. See Grilli and Rampichini [117] for a detailed discussion of the connection. The link between these two domains has indicated a way to reduce, if not overcome completely, a key limitation of Heckman's method, namely its strong dependence on the validity of the normality assumption. By replacing the Gaussian assumption with a multivariate t distribution, so that the post-selection distribution is of skew- t type, Marchenko and Genton [39] have extended Heckman's method to a more flexible form, more robust to failure of the distributional assumption. Proceeding further along these lines, Azzalini et al. [118] have considered response variables with discrete distributions, hence moving towards the edge of the scope of the present review.

Small area estimation is another theme often, although not exclusively, connected to the social sciences in its broad sense. Recently, a number of authors have employed flexible parametric distributions to replace the normality assumption of existing model-based formulations. An early exploration by Fabrizi and Trivisano [119] has been followed by formulations using either univariate SN (Ferraz and Moura [120]) or multivariate SN distributions (Ferrante and Pacei [121]), for so-called area-level models. The work of Diallo and Rao [122] is on unit-level modeling based on the CSN/SUN distribution.

Another methodology often linked to social sciences application is factor analysis. Montanari and Viroli [123] make use of the convenient mathematical properties of the multivariate SN distribution to develop a skew-normal factor model. Additional work in factor analysis has been presented by Liu and Lin [124], Lin et al. [125], Lin et al. [126], and Wang et al. [127].

4.6. A Popular Benchmark

There is a data set, known as the Australian Institute of Sport (AIS) data, which is employed in a vast set of publications to illustrate numerically some methodological proposal. The data represent observations on 13 variables, mostly of biomedical nature, on 202 Australian athletes. A full description of the AIS data and the data themselves are available within the R package *sn* which provides computational tools for SN and ST model fitting; see [128]. Given the purely numerical nature of these exercises, they cannot be described as 'applied work', but they provide illustrative examples for a wide range of methods. Notable references which make use of the AIS data include [10,11,19,32,53,55,59,68], but this list could be far longer. There are seldom cross-paper comparisons of the various methods that are applied to the AIS data.

Having mentioned computation tools for data fitting in the R computing environment, it is appropriate to mention also similar tools for the Stata environment, presented in [129].

5. Applications to Real Problems

This section of the paper is concerned with applications, the majority of which employ the SN or the ST distributions, possibly in the multivariate form; however, other families have been used, especially but not only the CSN/SUN family. In most cases, authors do not set out to propose or develop new theories or methods pertaining to the application. Two exceptions are portfolio theory

and risk, for which the parsimony of both the SN and ST distributions has led to new insights and methods. Papers are grouped under application theme headings.

5.1. Economics and Applied Financial Economics

There is a growing number of studies in applied financial economics that are mainly empirical in content. Harvey et al. [130] is concerned with portfolio selection using the version [41] of the SN distribution. Barbi and Romagnoli [131] employ the extended skew normal distribution for similar applications. Carmichael and Coën [132] use the SN distribution to derive explicit expressions for the skewness premium of a financial asset. Alodaat and Al-Rawwash [133] apply the extended SN distribution to the EURO/USD exchange rate. This is in contrast to results in Adcock [134] and Adcock [135] which indicate that estimated values of the extension parameter, denoted τ in Section 2.2, are often numerically different from zero.

Papers which use SN or ST distributions in conjunction with econometric methods *per se* are rare. An exception is due to Abanto-Valle et al. [136], who embed the ST distribution within the stochastic volatility framework of Heston [137]. They apply the resulting model to daily returns on the NASDAQ index. A second exception is in De Luca and Loperfido [138], who describe a multivariate SGARCH model, which is applied to MSCI equity indices.

There are other applications that are concerned with finance and economics in a broader sense. Chen et al. [139] present a regularization method for the multivariate ST distribution. This is applied to the hourly average electricity spot price data set for New South Wales (from Panagiotelis and Smith [140]). There are applications to credit card data, such as Brito and Duarte Silva [141] and Azzalini et al. [118]. Ozaki and Silva [142] use the SN distribution to rate crop insurance contracts in Brazil.

There exists a close connection between the theme of stochastic frontier analysis (SFA) for production efficiency and additive representations of type (5) for the SN distribution and related ones. A general account on SFA is [143]; the link between these two literatures is elucidated in its simpler SN setting at [7] (pp. 91–93), but it carries on in more advanced constructions indicated next. A detailed discussion with a sketch of an extension to SEC distributions has been presented by Domínguez-Molina et al. [144]; see also further work at [145]. Tchumtchoua and Dey [146] apply the ST distribution of Sahu et al. [41] and models for SFA to the United States primary metal industry data set of Aigner and Schmidt [147]. For data presenting repeated observations on each production unit, Colombi [148] and Brorsen and Kim [149] develop methods for SFA based on the CSN distribution; see also [150]. Filippini and Greene [151] describe several developments of SFA methods based on the SN distribution; these are applied to a study of the cost efficiency of Swiss railways.

5.2. Quantitative Finance

The applications referred to in other sub-sections are concerned with the use of statistical methods based on skew-elliptical distributions. There may be some technical developments specific to an application, but in general, there is no attempt to contribute to theory as such. Quantitative finance is different in a number of respects. First, Adcock and Shutes [152] suggest that a motivation for use of skew-elliptical distributions is that the non-negative variable in the stochastic representation represents a departure from financial market efficiency in the sense of Fama [153]. Secondly, Adcock [154] and Adcock [40] show that the multivariate extended SN and ST distributions play an important role in the theory of portfolio selection. Thirdly, as described in the next subsection, several authors have used the SN and ST distributions to develop measures of risk, typically to be applied to portfolios or to insurance data. The important early work of Simaan [155] employs a formulation where an elliptical random variable is summed to a positive variable multiplied by a vector of parameter, qualitatively quite close to the idea of skew-elliptical discussed in the earlier sections, although technically not quite the same. However, when the positive component has half-normal distribution, what we get is precisely a multivariate SN variable. The first formal treatment of the portfolio selection problem using

skew-elliptical distributions is due to Adcock and Shutes [152], who employ the extended version of the multivariate SN. They argue that the extended version is required theoretically, because the distribution of the non-negative variable is not necessarily a half-normal.

The multivariate extended SN and ST distributions both lead to extensions to Stein's lemma, Stein [156]. For normally distributed returns, the consequence of Stein's lemma is that, subject to some regularity conditions, the portfolios of all expected utility maximizers will lie on Markowitz' mean-variance efficient frontier. This important result was extended to cover elliptically symmetric distributions by Landsman and Nešlehová [157]. It was further extended in Adcock [154] and Adcock [40] to the multivariate extended SN and ST distributions. For expected utility maximizers, there is a single mean-variance-skewness efficient surface. In the same papers the results are extended to the closed version of the distributions, for which there are mean-variance-skewness hyper-surfaces.

The papers above are implicitly concerned with portfolio selection for a single investment period. When single period returns follow a multivariate SN or ST distribution, the corresponding distributions for multiple periods are sums of log SN or ST variables. Roch and Valdez [158] provide lower convex order bound approximations for such sums. Their work extends that due to Dhaene et al. [159] and Dhaene et al. [160]. To the best of our knowledge, there are few other papers based on skew-elliptical distributions that are concerned with finance theory. One exception is due to Blasi and Scarlatti [161], who study stochastic dominance rules for the SN distribution. In a contribution to efficient set mathematics, Bodnar and Gupta [162] study the inference procedures for testing the weights in the global minimum variance portfolio when asset returns follow the CSN distribution.

The market model, which is motivated at least in part by portfolio selection, is an important component of finance theory, and provides the foundations for later developments most notably by Fama and French [163] and in subsequent papers by them and many others. Conventionally, the market model is a simple regression model in which the dependent and independent variables are, respectively, the returns on a risky asset and the so-called market portfolio, usually represented by an appropriate stock market index. Assuming multivariate normality, or elliptical symmetry more generally, the market model is the conditional distribution of the dependent variable given a value of the independent. These results are extended in Adcock [134] and Adcock [36] in which the market model is derived when asset returns follow the multivariate extended SN and ST distributions, respectively. The resulting models are non-linear in the return on the market portfolio. Beta, the traditional measure of systematic risk, is no longer defined in the conventional manner.

5.3. Risk

The word risk has many interpretations. The number of papers and other publications with the word in the title is large and their scope is broad. For the purpose of this review, there are three topics of interest: (i) the probability distribution of the losses arising from a portfolio of insurance policies, or their equivalent; (ii) the computation of measures of risk attributable to a portfolio of any type of financial asset or liability arising from the tail of the probability distribution; (iii) the distribution theory of extreme returns. In this sub-section, there are two measures of risk that appear commonly in relevant literature. The first is value at risk, abbreviated to VaR. The second is conditional value at risk, abbreviated to CVaR. This measure is also referred to as expected shortfall (ES) or tail conditional expectation (TCE). The definition of both measures is well known.

Bernardi et al. [164] employ the well-used Danish fire data (first used in McNeil [165]) to exemplify the use of a finite mixture of SNs to model loss distributions. This paper builds on earlier work by in Bernardi [166] in which various measures of risk are derived corresponding to the finite mixture of SNs. Pigeon et al. [167] use the simplified form of the CSN distribution due to Gupta and Chen [168] to study insurance losses. Eling [169] reports an investigation fitting insurance claims data using a number of benchmark probability distributions and the Danish fire and US indemnity data sets. He reports that “..... the skew-normal and skew student are reasonably good models compared to 17 benchmark

distributions.”. Using the data set of Antonio and Plat [170], Pigeon et al. [167] model the occurrence of insurance claims.

Moving to VaR and CVaR and relevant distribution theory, Vernic [171] derives an expression for CVaR for the SN distribution. The analogous expression for the ST is given in Adcock et al. [172]. The paper by Tian et al. [173] extend results in Wang [174] and Wang [175] to develop tail-preserving and coherent distortion risk measures, which are intended to overcome shortcoming in both VaR and CVaR. The SN and several other skew-elliptical distributions are used to illustrate the general methods. Kollo et al. [176] examine the tail dependence of the bivariate skew t-copula. They show that dependent on the values of the skewing parameter the tail dependence coefficient may differ considerably from that derived using the Student t-copula. In an earlier paper, Pettere and Kollo [177] use the ST copula to model future cash flows from insurance claims. Padoan [178] presents multivariate extreme models based on underlying multivariate SN and ST distributions. This is a work that is similar to the paper by Lysenko et al. [179] which is concerned with generalized SN distributions.

5.4. Biology and Life Sciences

There are many studies in these categories, addressing a wide variety of applications. Not surprisingly, there are numerous studies concerned with AIDS or HIV. Lin and Wang [180] use multiple regression based on the multivariate SN distribution to study AIDS related data. Huang and Dagne [181] use the [41] version of the multivariate SN distribution in a Bayesian setting to model AIDS data in the presence of errors in covariates. They report that *“The proposed method may have a significant impact on AIDS research because, in the presence of skewness in the data and measurement errors in covariates, appropriate statistical inference for HIV dynamics is important for making reliable clinical decisions”*. The same data is studied in Huang and Dagne [182] which is concerned with semi-parametric mixed-effects models with SN errors. Both Baghfalaki and Ganjali [113] and Doğru and Arslan [94] employ the CSN distribution in a longitudinal study of the HIV data set from Goldman et al. [183]. Huang et al. [184] report a study using the AIDS clinical trial study data of Lederman et al. [185]. For the analysis of a longitudinal study of AIDS, Yu et al. [186] use a linear mixed model and introduce a special version of multivariate ST distribution where each component variable has a different degrees of freedom to regulate the tail behavior; see Ghosh and Hanson [187] for further details. In the same paper they also report a study of behavioral factors related to sexually transmitted infections.

The papers listed in this paragraph provide a sense of numerous issuers in health where the data exhibits skewness and for which empirical studies employ one or more of the distributions covered in this review. Zadkarami and Rahimi [188] use SN regression to carry out a study of factors affecting birth weight based on a British data set from 1958. Mahmud et al. [189] apply a probit log-SN mixture model to the Diary of Asthma and Viral Infections Study (known as DAVIS). Ho et al. [190] use factor mixture models based on the SN and ST distributions to study fatigue in breast cancer patients. Azzalini and Bacchieri [115] report the analysis of the results of a study into clinical trials on patients affected by intermittent claudication which is a symptom that describes muscle pain on mild exertion. Jamalizadeh et al. [191] use the visual acuity data from Fishman et al. [192] to exemplify the distributions of order statistics and linear combinations thereof from a bivariate selection normal distribution. Mansourian et al. [193] apply the SN distribution to a study of patients with back pain. Liu and Lin [92] employ a finite mixture regression model with SN errors to examine data from the National Survey of Child and Adolescent Well-Being (U.S. Department of Health and Human Services, Administration for Children, Youth and Families, 2003). The motivation for this paper is that assuming that errors are normally distributed in the presence of a substantial degree of skewness can lead to inaccuracy in estimating the number of classes in the mixture. Ismail et al. [194] use regression models with ST residuals in a spatial analysis of resting state brain activity in healthy controls and individuals with Down’s syndrome. See also Kline and Tobias [195] (relationship between body mass and earnings using the 1970 British Cohort Study), Fernandes et al. [196] (study in genetics using mice), Chai and Bailey [197] (coronary artery calcification), Bandyopadhyay et al. [198] (linear mixed

model with Bayesian Estimation applied to periodontal disease), Bandyopadhyay et al. [199] (also linear mixed models for censored responses with applications to HIV viral loads), Dagne [200] (a tobit model with ST errors applied to measles vaccine and HIV/AIDS data). Pyne et al. [85] apply a finite mixture of multivariate ST distributions to high-dimensional flow cytometric data analysis. In addition, concerned with measurements by flow cytometry, Ho et al. [201] use a ST-normal distribution to modeling cellular state transitions.

Similar to health, there is a wide variety of papers devoted to applications in agriculture. InBaghfalaki et al. [108] the multivariate SN distribution is employed in regression models with SN errors and non-random dropouts. The methods are applied to a data set concerned with mastitis in cattle and an experiment studying the effect of the inhibition of testosterone production in rats (see Diggle and Kenward [202] for details of the data set). Contreras-Reyes [203] analyzes a fish condition factor index using the SN distribution. This paper also presents information theory measures for the SN distribution. Contreras-Reyes et al. [204] use the ST distribution to study log-transformed length-at-age data of the southern blue whiting.

5.5. Environmental Issues

Lee and Skinner [205] use the SN distribution to testing fossil calibrations for vertebrate molecular trees. Stoy et al. [206] also use the SN distribution in an application to arctic ecosystem carbon exchange, as do Simolo et al. [207] in an application concerning the evolution of extreme temperatures in a warming climate. There is a related application to temperatures in China in Brito and Duarte Silva [141]. Rimstad and Omre [105] apply the CSN distribution to seismic data from the North Sea. Anagreh et al. [208] apply the SN distribution to model the potential of wind and solar energy resources in Aqaba, Jordan.

Bartoletti and Loperfido [209] use the SN distribution to model air pollution data recorded in 2006 from monitoring stations in Italy. Karimi and Mohammadzadeh [102] use the CSN distribution with a spatial regression model to analyze air pollution data were collected in Tehran on 1 March 2007. The model described in the paper has the facility to deal with missing data. Morris et al. [210] consider daily observations of maximum 8-hour ozone measurements for the 31 days of July 2005 at 1089 Air Quality System (AQS) monitoring sites in the United States and introduce a spatio-temporal model for threshold exceedances based on the ST distribution. The short theoretical note by Guttorp and Loperfido [211] demonstrates that the extended SN distribution offers a feasible model to study network bias in air quality monitoring design. Hashemi et al. [212] extend the Birnbaum-Saunders distribution by replacing a standard normal variate with its SN equivalent, and they illustrate the use of the new distribution with daily ozone concentrations data set. Zakaria et al. [213] simulate monthly rainfall data and fit a bivariate ST copula. Nathoo [214] reports a study of juvenile white spruce growth using the ST distribution.

5.6. Industrial and Technological Applications

Li et al. [215] develop Shewhart-type control charts under the SN distribution. They provide tables of charting and an illustrative example from polarizer manufacturing is described. There is similar work in Figueiredo and Gomes [216], who also provide tables and describe an application to cork stopper manufacture. Rendao et al. [217] use multivariate SN regression to examine the effect of technological progress, industrial structure and energy consumption structure on energy intensity (defined as energy consumption per unit of GDP) in China from 2000 to 2011. Youssef [218] develops methods based on the SN distribution to improve the estimation hand-set locations, a hand-set being, for example, a mobile telephone or a GPS receiver. Zadkarami and Rowhani [219] uses the SN in a discriminant analysis to classify the pixels of a remotely sensed satellite image. Balef et al. [220] use the log SN distribution to model delay variation for supply voltages in an electricity network. Tsai and Lin [221] use a non-linear SN model and develop procedures to assess the reliability information of products in which quality degrades over time. Data on traffic speed have been modeled by Zou

and Zhang [222] using the SN and ST distributions. Ramprasath et al. [223] use the SN distribution for the analysis of integrated circuits working. Müller et al. [224] use the ST distribution for statistical trilateration in communication networks.

Although not within the technological area, but rather the scientific one, we include here a mention to applications in astronomy by Eyer and Genton [225] and Simola et al. [226].

6. Two Illustrations in Applied Domains

To facilitate the perception of the potential usefulness of the distributions presented above, and how their formal properties can be linked to subject-matter considerations and theory, we present two illustrations in applied domains.

6.1. A Finance Application

The multivariate extended skew-normal and skew- t distributions both lead to extensions to Stein's lemma, [156]. For normally distributed returns, the consequence of Stein's lemma is that, subject to some regularity conditions, the portfolios of all expected utility maximizers will lie on Markowitz' mean-variance efficient frontier. This important result was extended to cover elliptically symmetric distributions by [157]. It was further extended in [40,154] to the multivariate extended skew-normal and skew- t distributions. For expected utility maximizers, there is a single mean-variance-skewness efficient surface. In the same papers the results are extended to the closed version of the distributions, for which there are mean-variance-skewness hyper-surfaces.

Consider an extended skew-normal distribution, $Y \sim \text{SN}_d(\xi, \Omega, \alpha, \tau)$, say, with density function (16) and define the associated quantities δ, ω as in Section 2.2. The extension to Stein's lemma is

$$\text{cov}\{Y, h(Y)\} = \Omega \mathbb{E}\{\nabla h(Y)\} + \omega \delta \zeta_1(\tau) [\mathbb{E}\{h(W)\} - \mathbb{E}\{h(Y)\}],$$

where $W \sim N_d(\xi - \tau\omega\delta, \Omega - \omega\delta\delta^\top\omega)$ and $\zeta_1(x) = \varphi(x)/\Phi(x)$ is the inverse Mills ratio, having set $\varphi(x) = \Phi'(x)$. For portfolio selection, the elements of the d -vector Y now denote the return on risky financial assets. The aim of an expected utility maximizer is to choose a vector of portfolio weights, commonly denoted w , that maximize the expected value of the investor's utility function $U(w^\top Y)$, where $w^\top Y$ is portfolio return. For conventional portfolio selection, the expected value of $U(\cdot)$ is maximized subject to the budget constraint that the elements of w sum to unity and that each element satisfies $w_i \geq 0$; that is, no short positions are allowed. In keeping with standard economic theory, it is assumed that $U'(\cdot) \geq 0$ and that $U''(\cdot) < 0$. Using the notation above, and omitting terms involving Lagrange multipliers, the central component of the first order conditions (FoC's) for the constrained maximization obtained from $\mathbb{E}\{U(\cdot)\}$ is

$$\mathbb{E}\{U'(w^\top Y)\} \xi + \omega \delta \zeta_1(\tau) (\mathbb{E}\{U'(w^\top W)\} - \mathbb{E}\{U'(w^\top Y)\}) + \mathbb{E}\{U''(w^\top Y)\} \Omega w \quad (32)$$

It is common practice to replace the terms involving $U'(\cdot)$ and $U''(\cdot)$ with constants. On dividing through by $\mathbb{E}\{U''(w^\top Y)\}$, the FoC's are $\theta\xi + \eta\omega\delta - \Omega w$, with $\theta, \eta \geq 0$. For given values of θ and η portfolio selection may be performed using quadratic programming. Noting that the FoCs do not depend explicitly on $\mathbb{E}\{Y\}$ or $\text{cov}\{Y\}$, it is also common practice to rearrange (32) and write it as

$$\theta \mathbb{E}\{Y\} + \eta\omega\delta - \text{cov}\{Y\} w. \quad (33)$$

By varying the values of θ and η different portfolios are generated corresponding to different degrees of risk appetite and preference for skewness, respectively. As shown in [36], this results in a mean-variance-skewness efficient surface, which is a generalization of Markowitz's mean-variance efficient frontier. When $\eta = 0$, corresponding to no interest in skewness on the part of the investor, the solution to (33) is the mean-variance efficient frontier. The setup at equation (33) is as described in [227], who summarize a number of related formulations by other authors. There is an important

difference between the mean-variance efficient frontier and the mean-variance-skewness efficient surface. For the former, the expected return on a portfolio is an increasing function of portfolio variance, or of the risk appetite θ . When skewness is included, expected return depends *inter alia* on the scalar quantity $Y = \eta \mu^\top D \omega \delta$, where $\mu = \mathbb{E}\{Y\}$ and D is a symmetric positive semi-definite matrix, which is determined by the optimization process. Depending on the values of μ and δ , Y may assume both positive and negative values. Thus, in some cases excessive preference for skewness may result in lower portfolio expected return. An example of such a surface is shown in [40].

To illustrate the effect of varying risk aversion and preference for skewness, we describe a simulation loosely based on the returns on the shares of financial services companies traded on the Shanghai stock exchange. Stocks from one of the main Chinese stocks markets are employed for this illustration as it is a view reported in the finance literature that skewness is a feature of emerging markets; see, for example, the paper by [228]. The illustration is based on 30 stocks all which had daily price data from January 2009 to December 2018, giving 2606 observations. Portfolio selection was performed based on Equation (33). It follows standard practice and includes the budget and non-negativity constraints.

The left panel of Figure 4 contains a sketch of the computed surface. The diagram shows that increasing values of θ and η lead a flat surface; that is, the entire budget is invested in the single security with the maximum expected return. For a fixed value of η , the resulting curve resembles the efficient frontier and for $\eta = 0$ it is the efficient frontier itself. If the investor has zero preference for expected return, $\theta = 0$, expected return increases with η , but at the maximum value shown in Figure 4, the expected return is less than the maximum achievable. The right panel of Figure 4 shows the number of non-zero holdings. As the plot illustrates, as both θ and η increase the number of non-zero holdings decreases to one. It should be noted that, while this example illustrates the effect of skewness on portfolio selection, a real world application would include other constraints on the weights in order to maintain a degree of diversification. In the interests of brevity, resulting from the use of such design considerations, the parameters used and the simulations are omitted, but they are available on request.

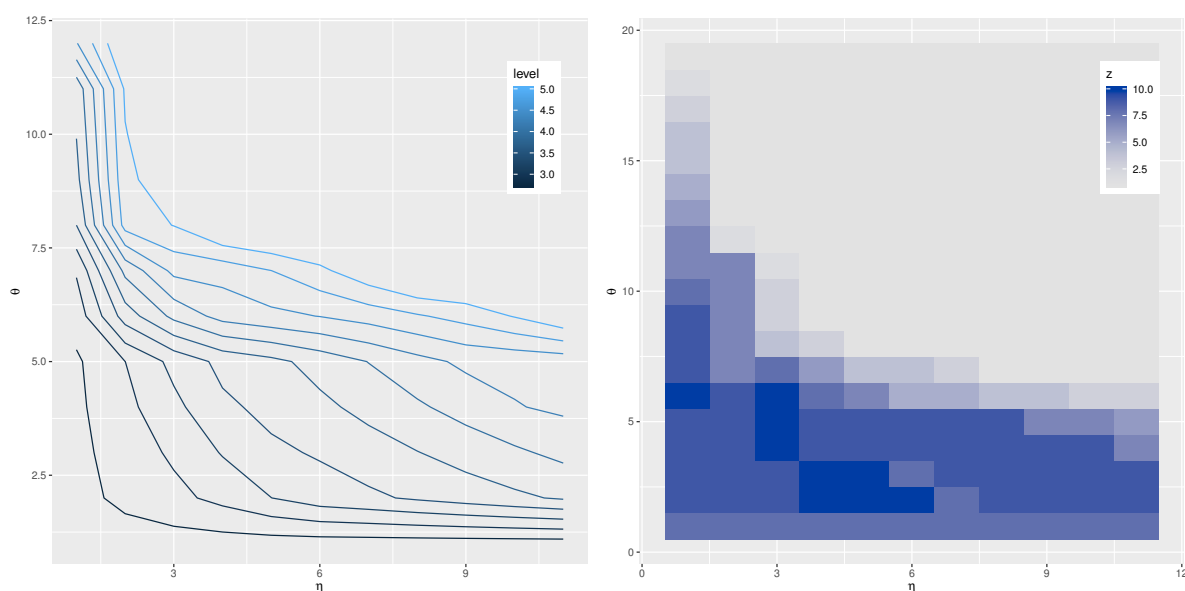


Figure 4. Finance application. Left panel: mean-variance-skewness efficient surface; right panel: number of non-zero weights.

6.2. Stochastic Frontier Analysis: An Simple Case

Consider more closely the connection with the SFA theme mentioned in Section 5.1. As its general target, SFA involves the study of the relationship between the output from a group of production

units to a number of input factors for these units, with the inclusion of an ingredient which represents the inefficiency of the units with respect to the optimal level allowed by the current technological development. In a simple and classical setting, this situation can be expressed as

$$Y_i = \beta_0 + x_i^\top \beta + V_i - U_i, \quad i = 1, \dots, n, \quad (34)$$

where Y_i is a random variable representing the output of the i th production unit, typically expressed on the logarithmic scale, x_i is a p -dimensional vector of input levels for that unit (again, usually expressed on the log scale), and V_i, U_i are independent random variables, where V_i represents a purely random error and U_i is a positive random variable, representing the inefficiency level of the i th unit. Classical specifications for U, V , where we now drop the subscript i for simplicity, are as follows:

$$V \sim N(0, \sigma_v^2), \quad U \sim \sigma_u \chi_1 \quad (35)$$

where χ_1 denotes the square root of a χ_1^2 random variable, also called half-normal distribution.

The distribution of $R = V - U$ under specification (35) is of the SN type, taking into account its additive representation (5). Simple algebra shows that the SN parameters are related to those in (35) by $\omega^2 = \sigma_u^2 + \sigma_v^2$, $\alpha = -\sigma_u / \sigma_v$. The regression parameters β maintain their meaning across the two settings. The connection of this simple SFA formulation with the SN family makes available extensions of the SN distributions developed in the context of symmetry-modulated distributions.

To illustrate this point in a simple case, we make use of some data, namely the data set `milkProd` available within the R package `Benchmarking`; see [229]. The data set refers to the dairy industry; the production units are Danish milk producers. For each of $n = 108$ units, four variables are available: `milk` (the amount produced, in kg), `energy` (energy expenses), `vet` (veterinary expenses), `cows` (number of cows).

A model of type (34) is fitted to the data, setting Y_i equal to the log-amount of milk produced and the $p = 3$ input values x_i equal to the log-transformed values of the other variables. As for the (U, V) variables, we start with the distributional assumptions (35). The data fitting outcome is the same using either the `FRONTIER` program, very popular in the SFA community, which is described in the standard reference text [143], or the R package `sn` [128], after suitable conversion of the parameters. The fitted model is then

$$Y = 7.521 + 0.1214 \log(\text{energy}) + 0.06276 \log(\text{vet}) + 0.8791 \log(\text{cows}) + R$$

where $R \sim \text{SN}(0, 0.21376^2, -3.5979)$.

The left panel of Figure 5 displays the scatter plot of the fitted values versus the observed milk log-production. It is visible that, for any given abscissa, the points are more elongated above the identity line than below it, hence indicating asymmetry of the distribution, which in turn indicates the presence of the inefficiency component U_i in (34). This feature is confirmed in the right panel of Figure 5, where the continuous blue line represents the fitted SN density of the R variable, clearly asymmetric.

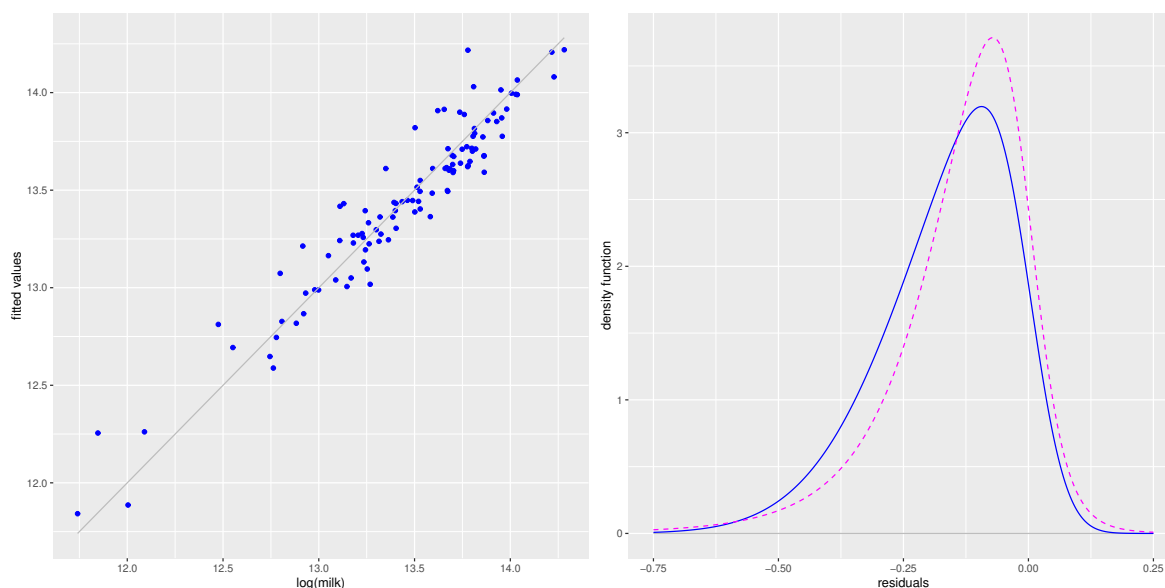


Figure 5. Fitting SFA models to milk production data. Left panel: scatter plot of fitted and observed data; right panel: fitted SN density (continuous blue line) and ST density (dashed magenta line).

Closer inspection of the residuals values $\hat{R}_i$ indicates somewhat longer tails than those expected under assumption (35). Various proposals for handling this sort of situation have been put forward in the SFA literature, but their discussion would require considerable space, and it would lead us away from our main theme. We confine ourselves to a brief sketch of a route stemming from the SEC class, specifically the ST distribution (26).

Besides the genesis via a conditioning argument presented in Section 3.2, the ST distribution can also be constructed via an additive mechanism analogous to (5)–(6) for the SN distribution; see Proposition 9 of [19] for a full specification. Using this result in the bivariate case, we can replace the assumption (35) by one where (U, V) is constructed from (U^*, V) with bivariate Student's t distribution and $U = |U^*|$. In this case, the variables (U^*, V) will be uncorrelated, but not independent, since the multivariate Student's t does not allow independence of the components. The corresponding variable $R = V - U$ will have a ST distribution. For more details about this formulation, see Tancredi [230]. For the data under consideration, we have fitted a ST distribution with $\nu = 5$ degrees of freedom. This value of ν was selected by simple inspection of the histogram of the earlier residuals, without using a formal estimation procedure, but subsequent numerical inspection of the log-likelihood function has confirmed the appropriateness of the choice. After maximization with respect to the other parameters we obtained a maximized log-likelihood value 68.99 on 6 degrees of freedom, having regarded ν as a fixed value. The corresponding ST density is represented by the dashed magenta line in the right plot of Figure 5. The log-likelihood value is 1.165 higher than the corresponding value for the SN fit. While a formal test procedure is not possible because the models are not-nested, there is a non-negligible improvement in the ST model, especially in the light of the limited number of observations. A more formal procedure would be to consider an information criterion, such as Akaike's AIC or similar; since the two models have the same number of parameters, we end up comparing the maximized log-likelihood functions, as already done.

7. Final Comments

Some sort of final remarks seem appropriate. We refrain from undertaking a hazardous forecasting exercise on the future evolution of the area and, instead, confine ourselves to the less impressive, but still useful, attempt of identifying the key aspects of the past evolution and of the present situation.

It is reasonable to say that the take-off of the domain occurred at the end of the past century, between 1996 and 1999. While before that time only a few sporadic publications existed, after that time,

papers started appearing in rapid succession, involving a gradually growing number of authors. For the subsequent 15 years or so, the focus of interest has been primarily on theoretical aspects, in two main directions: (i) formulation of more and more general families of distributions and (ii) development of statistical procedures for fitting these distributions, at least for the more realistic and tractable families.

In parallel, some applications of the newly proposed distributions have started appearing. In the initial stage, they naturally took the form of simple data fitting, often in a univariate context. However, in the last decade or so, we have witnessed an evolution of these applications which have evolved into more elaborate constructions, which often build on the formal properties of the distributions to link them with subject-matter considerations and theory.

This is not to say that the theoretical developments have stopped. New contributions still appear, although not at the tumultuous rate of earlier years, which after all may be not detrimental.

The overall perception emerging from this picture is the one of a domain which has entered its maturity stage. The theoretical developments of the early years have lead to applications, from the basic to the more elaborate ones, spanning a wide range of applied areas. This interaction of theoretical and applied work is culturally appealing and generally gratifying for all those that have contributed, and still contribute, to this project.

Author Contributions: The authors equally contributed to the preparation of the paper. Both authors have read and agreed to the published version of the manuscript

Funding: This research received no funding, neither external nor internal.

Acknowledgments: The authors are grateful to Judy Adcock for assembling the library of papers concerned with applications. Thanks are also due to the reviewers of the paper, for comments which have led to an improved presentation of the original version.

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

References

1. Pretorius, S.J. Skew bivariate frequency surfaces, examined in the light of numerical illustrations. *Biometrika* **1930**, *22*, 109–223.
2. Schols, C.M. *Over de Theorie der Fouten in de Ruimte en in Het Platte Vlak*; Verhandelingen der Koninklijke Akademie van Wetenschappen: Amsterdam, 1875; Volume XV, Part VI, pp. 1–67; Reprinted as ‘Théorie des erreurs dans le plan et dans l’espace’. In *Ann. de l’École Polytechnique de Delft*; 1886, pp. 123–175.
3. Perozzo, L. Nuove applicazioni del calcolo delle probabilità allo studio dei fenomeni statistici, e distribuzione dei matrimoni secondo l’età degli sposi. In *Reale Accademia dei Lincei, Serie 3a, Memorie Classe di Scienze Morali, Storiche e Filologiche*; 1882; Volume X, pp. 1–22.
4. Jones, M.C. On families of distributions with shape parameters (with discussion). *Int. Stat. Rev.* **2015**, *83*, 175–227, doi:10.1111/insr.12055.
5. Babić, S.; Ley, C.; Veredas, D. Comparison and classification of flexible distributions for multivariate skew and heavy-tailed data. *Symmetry* **2019**, *11*, 1216, doi:10.3390/sym11101216.
6. Genton, M.G. (Ed.) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004.
7. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; IMS monographs; Cambridge University Press: Cambridge, UK, 2014.
8. Fang, K.T.; Kotz, S.; Ng, K.W. *Symmetric Multivariate and Related Distributions*; Chapman & Hall: London, UK, 1990.
9. Kano, Y. Consistency property of elliptical probability density functions. *J. Multivar. Anal.* **1994**, *51*, 139–147.
10. Azzalini, A.; Dalla Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726, doi:10.1093/biomet/83.4.715.
11. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B* **1999**, *61*, 579–602.
12. Genton, M.G.; He, L.; Liu, X. Moments of skew-normal random vectors and their quadratic forms. *Stat. Probab. Lett.* **2001**, *51*, 319–325.

13. Capitanio, A. On the canonical form of scale mixtures of skew-normal distributions. *arXiv* **2012**, arXiv:1207.0797.
14. Balakrishnan, N.; Scarpa, B. Multivariate measures of skewness for the skew-normal distribution. *J. Multivar. Anal.* **2012**, *104*, 73–87, doi:10.1016/j.jmva.2011.06.017.
15. Balakrishnan, N.; Capitanio, A.; Scarpa, B. A test for multivariate skew-normality based on its canonical form. *J. Multivar. Anal.* **2014**, *128*, 19–32, doi:10.1016/j.jmva.2014.02.015.
16. Rao, C.R. Weighted distributions arising out of methods of ascertainment: what population does a sample represent? In *A celebration of statistics: the ISI centenary volume*; Atkinson, A.C., Fienberg, S.E., Eds.; Springer: Berlin, Germany, 1985; Chapter 24, pp. 543–569.
17. Arellano-Valle, R.B.; Branco, M.D.; Genton, M.G. A unified view on skewed distributions arising from selections. *Canad. J. Stat.* **2006**, *34*, 581–601.
18. Jupp, P.E.; Regoli, G.; Azzalini, A. A general setting for symmetric distributions and their relationship to general distributions. *J. Multivar. Anal.* **2016**, *148*, 107–199, doi:10.1016/j.jmva.2016.02.011.
19. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. R. Stat. Soc., ser. B* **2003**, *65*, 367–389, doi:10.1111/1467-9868.00391.
20. Wang, J.; Boyer, J.; Genton, M.G. A skew-symmetric representation of multivariate distributions. *Stat. Sin.* **2004**, *14*, 1259–1270.
21. Gupta, A.K.; Chang, F.C.; Huang, W.J. Some skew-symmetric models. *Random Oper. Stoch. Equ.* **2002**, *10*, 133–140, doi:10.1515/rose.2002.10.2.133.
22. Nadarajah, S.; Kotz, S. Skew models I. *Acta Appl. Math.* **2007**, *98*, 1–28, doi:10.1007/s10440-007-9087-0.
23. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113.
24. Azzalini, A.; Regoli, G. Some properties of skew-symmetric distributions. *Ann. Inst. Stat. Math.* **2012**, *64*, 857–879, doi:10.1007/s10463-011-0338-5.
25. Gupta, A.K. Multivariate skew t -distribution. *Statistics* **2003**, *37*, 359–363, doi:10.1080/715019247.
26. Kim, H.M.; Mallick, B.K. Moments of random vectors with skew t distribution and their quadratic forms. *Stat. Probab. Lett.* **2003**, *63*, 417–423; Corrigendum at **2009**, *79*, 2098–2099, doi:10.1016/S0167-7152(03)00121-4.
27. Kim, H.M. A note on scale mixtures of skew normal distribution. *Stat. Probab. Lett.* **2008**, *78*, 1694–1701; Corrigendum at **2013**, *83*, 1937, doi:10.1016/j.spl.2008.01.008.
28. Giorgi, E. Indici Non Parametrici per Famiglie Parametriche Con Particolare Riferimento Alla t Asimmetrica. Tesi di Laurea Magistrale, Università di Padova. 2012. Available online: <http://tesi.cab.unipd.it/40101/> (accessed on 6 January 2020).
29. Arevalillo, J.M.; Navarro, H. A note on the direction maximizing skewness in multivariate skew- t vectors. *Stat. Probab. Lett.* **2015**, *96*, 328–332, doi:10.1016/j.spl.2014.10.014.
30. Kim, H.M.; Zhao, J. Multivariate measures of skewness for the scale mixtures of skew-normal distributions. *Commun. Stat. Appl. Methods.* **2018**, *25*, 109–130, doi:10.29220/CSAM.2018.25.2.109.
31. Azzalini, A.; Genton, M.G. Robust likelihood methods based on the skew- t and related distributions. *Int. Stat. Rev.* **2008**, *76*, 106–129, doi:10.1111/j.1751-5823.2007.00016.x.
32. Wang, J.; Genton, M.G. The multivariate skew-slash distribution. *J. Stat. Plan. Inference* **2006**, *136*, 209–220, doi:10.1016/j.jspi.2004.06.023.
33. Kim, H.M.; Genton, M.G. Characteristic functions of scale mixtures of multivariate skew-normal distributions. *J. Multivar. Anal.* **2011**, *102*, 1105–1117, doi:10.1016/j.jmva.2011.03.004.
34. Shushi, T. A proof for the conjecture of characteristic function of the generalized skew-elliptical distributions. *Stat. Probab. Lett.* **2016**, *119*, 301–304, doi:10.1016/j.spl.2016.08.017.
35. Arellano-Valle, R.B.; Contreras-Reyes, J.E.; Genton, M.G. Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scand. J. Stat.* **2013**, *40*, 42–62, doi:10.1111/j.1467-9469.2011.00774.x.
36. Adcock, C.J. Asset pricing and portfolio selection based on the multivariate extended skew-Student- t distribution. *Ann. Oper. Res.* **2010**, *176*, 221–234, doi:10.1007/s10479-009-0586-4.
37. Arellano-Valle, R.B.; Genton, M.G. Multivariate extended skew- t distributions and related families. *Metron* **2010**, *LXVIII*, 201–234.
38. Lee, S.; Genton, M.G.; Arellano-Valle, R.B. Perturbation of numerical confidential data via skew- t distributions. *Manag. Sci.* **2010**, *56*, 318–333, doi:10.1287/mnsc.1090.1104.
39. Marchenko, Y.V.; Genton, M.G. A Heckman selection- t model. *J. Am. Stat. Assoc.* **2012**, *107*, 304–317.

40. Adcock, C.J. Mean-variance-skewness efficient surfaces, Stein's lemma and the multivariate extended skew-Student distribution. *Eur. J. Oper. Res.* **2014**, *234*, 392–401, doi:10.1016/j.ejor.2013.07.011.
41. Sahu, K.; Dey, D.K.; Branco, M.D. A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Stat.* **2003**, *31*, 129–150. Corrigendum: vol. 37 (2009) , 301–302.
42. González-Farías, G.; Domínguez-Molina, J.A.; Gupta, A.K. Additive properties of skew normal random vectors. *J. Stat. Plan. Inference* **2004**, *126*, 521–534, doi:10.1016/j.jspi.2003.09.008.
43. González-Farías, G.; Domínguez-Molina, J.A.; Gupta, A.K. The closed skew-normal distribution. In *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*; Genton, M.G., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004; Chapter 2, pp. 25–42.
44. Liseo, B.; Loperfido, N. A Bayesian interpretation of the multivariate skew-normal distribution. *Stat. Probab. Lett.* **2003**, *61*, 395–401.
45. Arellano-Valle, R.B.; Genton, M.G. On fundamental skew distributions. *J. Multivar. Anal.* **2005**, *96*, 93–116, doi:10.1016/j.jmva.2004.10.002.
46. Arellano-Valle, R.B.; Azzalini, A. On the unification of families of skew-normal distributions. *Scand. J. Stat.* **2006**, *33*, 561–574, doi:10.1111/j.1467-9469.2006.00503.x.
47. Arellano-Valle, R.B.; Genton, M.G. Multivariate unified skew-elliptical distributions. *Chil. J. Stat.* **2010**, *1*, 17–33.
48. Arellano-Valle, R.B.; Jamalizadeh, A.; Mahmoodian, H.; Balakrishnan, N. L-statistics from multivariate unified skew-elliptical distributions. *Metrika* **2014**, *77*, 559–583, doi:10.1007/s00184-013-0453-9.
49. Abtahi, A.; Towhidi, M. The new unified representation of multivariate skewed distributions. *Statistics* **2013**, *47*, 126–140, doi:10.1080/02331888.2011.577896.
50. Aziz, M.A.; Gupta, A.K. Quadratic forms in unified skew normal random vectors. *J. Probab. Stat. Sci.* **2013**, *11*, 1–15.
51. Gupta, A.K.; Aziz, M.A.; Ning, W. On some properties of the unified skew normal distribution. *J. Stat. Theory Pract.* **2013**, *7*, 480–495, doi:10.1080/15598608.2013.790233.
52. Richter, W.D.; Venz, J. Geometric representations of multivariate skewed elliptically contoured distributions. *Chil. J. Stat.* **2014**, *5*, 71–90.
53. Arellano-Valle, R.B.; Castro, L.M.; Genton, M.G.; Gómez, H.W. Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Anal.* **2008**, *3*, 513–539.
54. Arellano-Valle, R.B.; Genton, M.G.; Loschi, R.H. Shape mixtures of multivariate skew-normal distributions. *J. Multivar. Anal.* **2009**, *100*, 91–101, doi:10.1016/j.jmva.2008.03.009.
55. Arellano-Valle, R.B.; Ferreira, C.S.; Genton, M.G. Scale and shape mixtures of multivariate skew-normal distributions. *J. Multivar. Anal.* **2018**, *166*, 98–110, doi:10.1016/j.jmva.2018.02.007.
56. Tamandi, M.; Jamalizadeh, A. Finite mixture modeling using shape mixtures of the skew scale mixtures of normal distributions. *Commun. Stat. Simul. Comput.* **2019**, doi:10.1080/03610918.2018.1547397.
57. Arslan, O. Variance-mean mixture of the multivariate skew normal distribution. *Stat. Pap.* **2015**, *56*, 353–378, doi:10.1007/s00362-014-0585-7.
58. Barndorff-Nielsen, O.; Kent, J.; Sørensen, M. Normal variance-mean mixtures and z distributions. *Int. Stat. Rev.* **1982**, *50*, 145–159, doi:10.2307/1402598.
59. Genton, M.G. Discussion of “The skew-normal”. *Scand. J. Stat.* **2005**, *32*, 189–198.
60. Kim, H.J. A class of weighted multivariate normal distributions and its properties. *J. Multivar. Anal.* **2008**, *99*, 1758–1771, doi:10.1016/j.jmva.2008.01.008.
61. Jamalizadeh, A.; Kundu, D. A multivariate Birnbaum-Saunders distribution based on the multivariate skew normal distribution. *J. Jpn. Stat. Soc.* **2015**, *45*, 1–20.
62. Azzalini, A.; Regoli, G. On symmetry-modulated distributions: revisiting an old result and a step further. *Stat* **2018**, *7*, e171, doi:10.1002/sta4.171.
63. Gupta, A.K.; Varga, T.; Bodnar, T. *Elliptically Contoured Models in Statistics and Portfolio Theory*, II ed.; Springer: Berlin, Germany, 2013, doi:10.1007/978-1-4614-8154-6.
64. Giorgi, E.; McNeil, A.J. On the computation of multivariate scenario sets for the skew-*t* and generalized hyperbolic families. *Comput. Stat. Data Anal.* **2016**, *100*, doi:10.1016/j.csda.2014.06.024.
65. Bolfarine, H.; Montenegro, L.C.; Lachos, V.H. Influence diagnostics for skew-normal linear mixed models. *Sankhyā* **2007**, *69*, 648–670.

66. Ferreira, C.S.; Arellano-Valle, R.B. Estimation and diagnostic analysis in skew-generalized-normal regression models. *J. Stat. Comput. Simul.* **2018**, *88*, 1039–1059, doi:10.1080/00949655.2017.1419351.
67. Labra, F.V.; Garay, A.M.; Lachos, V.H.; Ortega, E.M.M. Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *J. Stat. Plan. Inference* **2012**, *142*, 2149–2165, doi:10.1016/j.jspi.2012.02.018.
68. Arellano-Valle, R.B.; Ozán, S.; Bolfarine, H.; Lachos, V.H. Skew-normal measurement error models. *J. Multivar. Anal.* **2005**, *96*, 265–281.
69. Figueiredo, C.C.; Bolfarine, H.; Sandoval, M.C.; Lima, C.R.O.P. On the skew-normal calibration model. *J. Appl. Stat.* **2010**, *37*, 435–451, doi:10.1080/02664760802715906.
70. Arellano-Valle, R.B.; Castro, L.M.; Loschi, R.H. Change point detection in the skew-normal model parameters. *Commun. Stat. Theory Methods* **2013**, *42*, 603–618, doi:10.1080/03610926.2011.611321.
71. Lachos, V.H.; Labra, F.V.; Bolfarine, H.; Ghosh, P. Multivariate measurement error models based on scale mixtures of the skew-normal distribution. *Statistics* **2010**, *44*, 541–556, doi:10.1080/02331880903236926.
72. Arellano-Valle, R.B.; Azzalini, A.; Ferreira, C.S.; Santoro, K. A two-piece normal measurement error model. *Comput. Stat. Data Anal.* **2020**, *144*, to appear, doi:10.1016/j.csda.2019.106863.
73. Chen, M.H.; Dey, D.K.; Shao, Q.M. A new skewed link model for dichotomous quantal response data. *J. Am. Stat. Assoc.* **1999**, *94*, 1172–1186, doi:10.1080/01621459.1999.10473872.
74. Bazán, J.L.; Bolfarine, H.; Branco, M.D. A framework for skew-probit links in binary regression. *Commun. Stat. Theory Methods* **2010**, *39*, 678–697, doi:10.1080/03610920902783849.
75. Bazán, J.L.; Branco, M.D.; Bolfarine, H. A skew item response model. *Bayesian Anal.* **2006**, *1*, 861–892.
76. Bazán, J.L.; Branco, M.D.; Bolfarine, H. Extensions of the skew-normal ogive item response model. *Braz. J. Probab. Stat.* **2014**, *28*, 1–23.
77. Sanros, J.R.S.; Azevedo, C.L.N.; Bolfarine, H. A multiple group item response theory model with centered skew-normal latent trait distributions under a Bayesian framework. *J. Appl. Stat.* **2013**, *40*, 2129–2149.
78. Durante, D. Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika* **2019**, *106*, 765–779, doi:10.1093/biomet/asz034.
79. McLachlan, G.J.; Peel, D. *Finite Mixture Models*; Description John Wiley & Sons: Hoboken, NJ, USA, 2000.
80. Lin, T.I.; Lee, J.C.; Yen, S.Y. Finite mixture modelling using the skew normal distribution. *Stat. Sin.* **2007**, *17*, 909–927.
81. Lin, T.I.; Lee, J.C.; Hsieh, W.J. Robust mixture modeling using the skew t distribution. *Stat. Comput.* **2007**, *17*, 81–92, doi:10.1007/s11222-006-9005-8.
82. Otiniano, C.E.G.; Rathie, P.N.; Ozelim, L.C.S.M. On the identifiability of finite mixture of skew-normal and skew- t distributions. *Stat. Probab. Lett.* **2015**, *106*, 103–108, doi:10.1016/j.spl.2015.07.015.
83. Lin, T.I. Maximum likelihood estimation for multivariate skew normal mixture models. *J. Multivar. Anal.* **2009**, *100*, 257–265, doi:10.1016/j.jmva.2008.04.010.
84. Lin, T.I. Robust mixture modeling using multivariate skew t distributions. *Stat. Comput.* **2010**, *20*, 343–356, doi:10.1007/s11222-009-9128-9.
85. Pyne, S.; Hu, X.; Wang, K.; Rossin, E.; Lin, T.I.; Maier, L.M.; Baecher-Alland, C.; McLachlan, G.J.; Tamayo, P.; Hafler, D.A.; De Jager, P.L.; Mesirov, J.P. Automated high-dimensional flow cytometric data analysis. *PNAS* **2009**, *106*, 8519–8524, doi:10.1073/pnas.0903028106.
86. Frühwirth-Schnatter, S.; Pyne, S. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew- t distributions. *Biostatistics* **2010**, *11*, 317–336, doi:10.1093/biostatistics/kxp062.
87. Vrbik, I.; McNicholas, P. Analytic calculations for the EM algorithm for multivariate skew t -mixture model. *Stat. Probab. Lett.* **2012**, *82*, 1169–1174, doi:10.1016/j.spl.2012.02.020.
88. Vrbik, I.; McNicholas, P.D. Parsimonious skew mixture models for model-based clustering and classification. *Comput. Stat. Data Anal.* **2014**, *71*, 196–210, doi:10.1016/j.csda.2013.07.008.
89. Lachos Dávila, V.H.; Cabral, C.R.B.; Zeller, C.B. *Finite Mixture of Skewed Distributions*; Briefs in Statistics; Springer: Berlin, Germany, 2018; doi:10.1007/978-3-319-98029-4.
90. Lee, S.X.; McLachlan, G.J. Finite mixtures of canonical fundamental skew t -distributions. *Stat. Comput.* **2016**, *26*, 573–596, doi:10.1007/s11222-015-9545-x.
91. Lee, S.X.; McLachlan, G.J. EMMIXskew: an R package for the fitting of a mixture of canonical fundamental skew t -distributions. *J. Stat. Software* **2018**, doi:10.18637/jss.v083.i03.

92. Liu, M.; Lin, T.I. A skew-normal mixture regression model. *Educ. Psychol. Meas.* **2014**, *74*, 139–162, doi:10.1177/0013164413498603.
93. Zeller, C.B.; Cabral, C.R.B.; Lachos, V.H. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *Test* **2016**, *25*, 375–396, doi:10.1007/s11749-015-0460-4.
94. Doğru, F.Z.; Arslan, O. Robust mixture regression based on the skew t distribution. *Rev. Colomb. Estad.* **2017**, *40*, 45–64. doi:10.15446/rce.v40n1.53580.
95. Kim, H.M.; Mallick, B.K. A Bayesian prediction using the skew Gaussian distribution. *J. Stat. Plan. Inference* **2004**, *120*, 85–101. doi:10.1016/S0378-3758(02)00501-3.
96. Minozzo, M.; Ferracuti, L. On the existence of some skew-normal stationary processes. *Chil. J. Stat.* **2012**, *3*, 159–172.
97. Zhang, H.; El-Shaarawi, A. On spatial skew-Gaussian processes and applications. *Environmetrics* **2010**, *21*, 33–47, doi:10.1002/env.982.
98. Schmidt, A.M.; Gonçalves, K.C.M.; Velozo, P.L. Spatiotemporal models for skewed processes. *Environmetrics* **2017**, *28*, e2411, doi:10.1002/env.2411.
99. Genton, M.G.; Hering, A.S. Comments on: Spatiotemporal models for skewed processes. *Environmetrics* **2017**, *28*, e2430. doi:10.1002/env.2430.
100. Hering, A.S.; Genton, M.G. Powering up with space-time wind forecasting. *J. Am. Stat. Assoc.* **2010**, *105*, 92–104, doi:10.1198/jasa.2009.ap08117.
101. Hosseini, F.; Eidsvik, J.; Mohammadzadeh, M. Approximate Bayesian inference in spatial GLMM with skew normal latent variables. *Comput. Stat. Data Anal.* **2011**, *55*, 1791–1806.
102. Karimi, O.; Mohammadzadeh, M. Bayesian spatial regression models with closed skew normal correlated errors and missing. *Stat. Pap.* **2012**, *53*, 205–218, doi:10.1007/s00362-010-0329-2.
103. Zareifard, H.; Jafari Khaledi, M. Non-Gaussian modeling of spatial data using scale mixing of a unified skew Gaussian process. *J. Multivar. Anal.* **2013**, *114*, 16–28, doi:10.1016/j.jmva.2012.07.003.
104. Ahmad, O.S.; Pinoli, J.C. Lipschitz-Killing curvatures of the excursion sets of skew Student's t random fields. *Stoch. Models* **2013**, *29*, 273–289, doi:10.1080/15326349.2013.783290.
105. Rimstad, K.; Omre, H. Skew-Gaussian random fields. *Spat. Stat.* **2014**, *10*, 43–62, doi:10.1016/j.spasta.2014.08.001.
106. Rezaie, J.; Eidsvik, J.; Mukerji, T. Value of information analysis and Bayesian inversion for closed skew-normal distributions: Applications to seismic amplitude variation with offset data. *Geophysics* **2014**, *79*, R151–R163, doi:10.1190/geo2013-0048.1.
107. Boojari, H.; Khaledi, M.J.; Rivaz, F. A non-homogeneous skew-Gaussian Bayesian spatial model. *Stat. Methods Appl.* **2016**, *25*, 55–73, doi:10.1007/s10260-015-0331-x.
108. Baghfalaki, T.; Ganjali, M.; Khounsivash, M. A non-random dropout model for analyzing longitudinal skew-normal response. *J. Iran. Stat. Soc.* **2012**, *11*, 101–129.
109. Huang, Y.; Dagne, G.A. Simultaneous Bayesian inference for skew-normal semiparametric nonlinear mixed-effects models with covariate measurement errors. *Bayesian Anal.* **2012**, *7*, 189–210, doi:10.1214/12-BA706.
110. Chang, S.C.; Zimmerman, D.L. Skew-normal antedependence models for skewed longitudinal data. *Biometrika* **2016**, *103*, 363–376, doi:10.1093/biomet/asw006.
111. Jana, S.; Balakrishnan, N.; S.Hamid, J. Estimation of the parameters of the extended growth curve model under multivariate skew normal distribution. *J. Multivar. Anal.* **2018**, *166*, 111–128, doi:10.1016/j.jmva.2018.02.008.
112. Barrett, J.; Diggle, P.; Henderson, R.; Taylor-Robinson, D. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *J. R. Stat. Soc. Ser. B* **2015**, *77*, 131–148, doi:10.1111/rssb.12060.
113. Baghfalaki, T.; Ganjali, M. A Bayesian approach for joint modeling of skew-normal longitudinal measurements and time to event data. *Rev. Stat. Stat. J.* **2015**, *13*, 169–191.
114. Shun, Z.; Gordon Lan, K.K.; Soo, Y. Interim treatment selection using the normal approximation approach in clinical trials. *Stat. Med.* **2008**, *27*, 597–618, doi:10.1002/sim.2990.
115. Azzalini, A.; Bacchieri, A. A prospective combination of phase II and phase III in drug development. *Metron* **2010**, *LXVIII*, 347–369.
116. Heckman, J.J. Sample selection bias as a specification error. *Econometrica* **1979**, *47*, 153–161.

117. Grilli, L.; Rampichini, C. Selection bias in linear mixed models. *Metron* **2010**, *LXVIII*, 309–329.
118. Azzalini, A.; Kim, H.M.; Kim, H.J. Sample selection models for discrete and other non-Gaussian response variables. *Stat. Methods Appl.* **2019**, *28*, 27–56, doi:10.1007/s10260-018-0427-1.
119. Fabrizi, E.; Trivisano, C. Robust models for mixed effects in linear mixed models applied to small area estimation. *J. Stat. Plan. Inference* **2010**, *140*, 433–443, doi:10.1016/j.jspi.2009.07.022.
120. Ferraz, V.R.S.; Moura, F.A.S. Small area estimation using skew normal models. *Comput. Stat. Data Anal.* **2012**, *56*, 2864–2874, doi:10.1016/j.csda.2011.07.005.
121. Ferrante, M.R.; Pacei, S. Small domain estimation of business statistics by using multivariate skew normal models. *J. R. Stat. Soc. Ser. A* **2017**, *180*, 1057–1088, doi:10.1111/rssa.12307.
122. Diallo, M.S.; Rao, J.N.K. Small area estimation of complex parameters under unit-level models with skew-normal errors. *Scand. J. Stat.* **2018**, *45*, 1092–1116, doi:10.1111/sjos.12336.
123. Montanari, A.; Viroli, C. A skew-normal factor model for the analysis of student satisfaction towards university courses. *J. Appl. Stat.* **2010**, *37*, 473–487, doi:10.1080/02664760902736737.
124. Liu, M.; Lin, T.I. Skew-normal factor analysis models with incomplete data. *J. Appl. Stat.* **2015**, *42*, 789–805, doi:10.1080/02664763.2014.986437.
125. Lin, T.I.; Wu, P.H.; McLachlan, G.J.; Lee, S.X. A robust factor analysis model using the restricted skew-*t* distribution. *TEST* **2015**, *24*, 510–531, doi:10.1007/s11749-014-0422-2.
126. Lin, T.I.; McLachlan, G.J.; Lee, S.X. Extending mixtures of factor models using the restricted multivariate skew-normal distribution. *J. Multivar. Anal.* **2016**, *143*, 398–413, doi:10.1016/j.jmva.2015.09.025.
127. Wang, W.L.; Castro, L.M.; Chang, Y.T.; Lin, T.I. Mixtures of restricted skew-*t* factor analyzers with common factor loadings. *Adv. Data Anal. Classifi.* **2018**, doi:10.1007/s11634-018-0317-2.
128. Azzalini, A. *The R package sn: The skew-normal and related distributions such as the skew-t (version 1.5-4)*; Università di Padova: Padova, Italia, 2019. Available online: <https://cran.r-project.org/package=sn> (accessed on 6 January 2020).
129. Marchenko, Y.V.; Genton, M.G. A suite of commands for fitting the skew-normal and skew-*t* models. *Stata J.* **2010**, *10*, 507–539.
130. Harvey, C.R.; Liechty, J.C.; Liechty, M.; Müller, P. Portfolio selection with higher moments. *Quant. Financ.* **2010**, *10*, 469–485.
131. Barbi, M.; Romagnoli, S. Skewness, basis risk, and optimal futures demand. *Int. Rev. Econ. Financ.* **2018**, *58*, 14–29.
132. Carmichael, B.; Coën, A. Asset Pricing with Skewed Returns. *Financ. Res. Lett.* **2013**, *10*, 50–57.
133. Alodaat, M.; Al-Rawwash, M.Y. The extended skew Gaussian process for regression. *Metron* **2014**, *72*, 317–330.
134. Adcock, C.J. Capital asset pricing for UK stocks under the multivariate skew-normal distribution. In *Skew Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Genton, M., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA 2004; Chapter 11, pp. 191–204.
135. Adcock, C.J. Stein's Lemma For Skew-Normal Distributions: a Comment and an Example. *J. Appl. Probab. Stat.* **2013**, *8*, 58–64.
136. Abanto-Valle, C.A.; Lachos, V.H.; Dey, D.K. Bayesian estimation of a skew-*t* stochastic volatility model. *Methodol. Comput. Appl. Probab.* **2015**, *17*, 721–738.
137. Heston, S.L. Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Rev. Financ. Stud.* **1993**, *6*, 327–343.
138. De Luca, G.; Loperfido, N. Modelling Multivariate Skewness in Financial Returns: a SGARCH Approach. *Eur. J. Financ.* **2015**, *21*, 1113–1132.
139. Chen, L.; Pourahmadi, M.; Maadooliat, M. Regularized Multivariate Regression Models with Skew-*t* Error Distributions. *J. Stat. Plan. Inference* **2014**, *149*, 125–129.
140. Panagiotelis, A.; Smith, M. Bayesian density forecasting of intraday electricity prices using multivariate skew *t* distributions. *Int. J. Forecast.* **2008**, *24*, 710–727.
141. Brito, P.; Duarte Silva, A.P. Modelling interval data with Normal and Skew-Normal distributions. *J. Appl. Stat.* **2012**, *39*, 3–20.
142. Ozaki, V.A.; Silva, R.S. Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *J. Appl. Stat.* **2009**, *36*, 443–452.
143. Coelli, T.J.; Prasada Rao, D.S.; O'Donnell, C.; Battese, G.E. *An Introduction to Efficiency and Productivity Analysis*, II ed.; Springer: Berlin, Germany, 2005.

144. Domínguez-Molina, J.A.; González-Farías, G.; Ramos-Quiroga, R. Skew-normality in stochastic frontier analysis. In *Skew-elliptical Distributions and Their Applications: a Journey Beyond Normality*; Genton, M.G., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA 2004; Chapter 13, pp. 223–242.
145. Domínguez-Molina, J.A.; González-Farías, G.; Ramos-Quiroga, R.; Gupta, A.K. A matrix variate closed skew-normal distribution with applications to stochastic frontier analysis. *Commun. Stat. Theory Methods* **2007**, *36*, 1671–1703, doi:10.1080/03610920601126126.
146. Tchumtchoua, S.; Dey, D.K. Bayesian estimation of stochastic frontier models with multivariate skew t error terms. *Commun. Stat. Theory Methods* **2007**, *36*, 907–916, doi:10.1080/03610920601041630.
147. Aigner, D. J., L.C.K.; Schmidt, P. Formulation and Estimation of Stochastic Production Function Model. *J. Econ.* **1977**, *12*, 21–37.
148. Colombi, R. Closed skew normal stochastic frontier models for panel data. In *Advances in Theoretical and Applied Statistics*; Torelli, N.; Pesarin, F.; Bar-Hen, A., Eds.; Studies in Theoretical and Applied Statistics; Springer: Berlin, Germany, 2013; Chapter 17, pp. 177–186, doi:10.1007/978-3-642-35588-2_17.
149. Brorsen, B.W.; Kim, T. Data aggregation in stochastic frontier models: the closed skew normal distribution. *J. Product. Anal.* **2013**, *39*, 27–34, doi:10.1007/s11123-012-0274-2.
150. Colombi, R.; Kumbhakar, S.C.; Martini, G.; Vittadini, G. Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *J. Product. Anal.* **2014**, *42*, 123–136, doi:10.1007/s11123-014-0386-y.
151. Filippini, M.; Greene, W. Persistent and transient productive inefficiency: A maximum simulated likelihood approach. *J. Product. Anal.* **2016**, *45*, 187–196.
152. Adcock, C.J.; Shutes, K. Portfolio Selection Based on The Multivariate Skew-Normal Distribution. In *Financial Modelling*; Skulimowski, A., Ed.; Progress and Business Publishers: Kraków, Poland, 1999.
153. Fama, E.F. Efficient Capital Markets II. *J. Financ.* **1991**, *46*, 1575–1617.
154. Adcock, C.J. Extensions of Stein's Lemma for the Skew-Normal Distribution. *Commun. Stat. Theory Methods* **2007**, *36*, 1661–1671.
155. Simaan, Y. Portfolio Selection and Asset Pricing-Three-Parameter Framework. *Manag. Sci.* **1993**, *39*, 568–577.
156. Stein, C.M. Estimation of The Mean of a Multivariate Normal Distribution. *Ann. Stat.* **1981**, *9*, 1135–1151.
157. Landsman, Z.; Nešlehová, J. Stein's Lemma for elliptical random vectors. *J. Multivar. Anal.* **2008**, *99*, 912–927.
158. Roch, O.; Valdez, E.A. Lower convex order bound approximations for sums of log-skew normal random variables. *Appl. Stoch. Models Bus. Ind.* **2011**, *27*, 487–502.
159. Dhaene, J.; Denuit, M.; Goovaerts, M.J.; Kaas, R.; Vyncke, D. The concept of comonotonicity in actuarial science and finance: applications. *Insur. Math. Econ.* **2002**, *31*, 133–161.
160. Dhaene, J.; Denuit, M.; Goovaerts, M.J.; Kaas, R.; Vyncke, D. The concept of comonotonicity in actuarial science and finance: theory. *Insur. Math. Econ.* **2002**, *31*, 1–33.
161. Blasi, F.; Scarlatti, S. From Normal vs Skew-Normal Portfolios: FSD and SSD Rules. *J. Math. Financ.* **2012**, *2*, 90–95.
162. Bodnar, T.; Gupta, A.K. Robustness of the inference procedures for the global minimum variance portfolio weights in a skew-normal model. *Eur. J. Financ.* **2015**, *21*, 1176–1194.
163. Fama, E.F.; French, K.R. The Cross-section of expected stock returns. *J. Financ.* **1992**, *47*, 427–465.
164. Bernardi, M.; Maruotti, A.; Lea, P. Skew mixture models for loss distributions: A Bayesian approach. *Insur. Math. Econ.* **2012**, *51*, 617–623.
165. McNeil, A. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bull.* **1997**, *27*, 117–137.
166. Bernardi, M. Risk measures for Skew Normal mixtures. *Stat. Probab. Lett.* **2013**, *83*, 1819–1824.
167. Pigeon, M.; Antonio, K.; Denuit, M. Individual Loss Reserving with the Multivariate Skew Normal Framework. *ASTIN Bull. J. Int. Actuar. Assoc.* **2013**, *43*, 399–428.
168. Gupta, A.K.; Chen, J. A class of multivariate skew normal models. *Ann. Inst. Stat. Math.* **2004**, *56*, 305–315.
169. Eling, M. Fitting Insurance Claims to Skewed Distributions: Are the Skew-Normal and Skew-Student Good Models? *Insur. Math. Econ.* **2012**, *51*, 239–248.
170. Antonio, K.; Plat, R. Micro-level stochastic loss reserving for general insurance. *Scand. Actuar. J.* **2014**, *2014*, 649–669.
171. Vernic, R. Multivariate skew-normal distributions with applications in insurance. *Insur. Math. Econ.* **2006**, *38*, 413–426.

172. Adcock, C.J.; Eling, M.; Loperfido, N. Skewed Distributions in Finance and Actuarial Science: A Review. *Eur. J. Financ.* **2015**, *21*, 1253–1281.
173. Tian, W.; Wang, T.; Baokun, L. Risk Measures with Wang Transforms under Flexible Skew-generalized Settings. *Int. J. Intell. Technol. Appl. Stat.* **2014**, *7*, 185–205.
174. Wang, S.S. A Class of Distortion Operators for Pricing Financial and Insurance Risks. *J. Risk and Insur.* **2000**, *67*, 15–36.
175. Wang, S.S. A universal framework for pricing financial and insurance risks. *Astin Bull.* **2002**, *32*, 213–234.
176. Kollo, T.; Pettere, G.; Valge, M. Tail Dependence of Skew t-Copulas. *Commun. Stat. Simul. Comput* **2017**, *46*, 1024–1034.
177. Pettere, G.; Kollo, T. Risk Modeling for Future Cash Flow using Skew T-Copula. *Commun. Stat. Theory Methods* **2011**, *40*, 2919–2925.
178. Padoan, S. Multivariate extreme models based on underlying skew-t and skew-normal distributions. *J. Multivar. Anal.* **2011**, *102*, 977–991.
179. Lysenko, N.; Roy, P.; Waeber, R. Multivariate Extremes of Generalized Skew-normal Distributions. *Stat. Probab. Lett.* **2009**, *79*, 525–533.
180. Lin, T.I.; Wang, W.L. Multivariate skew-normal linear mixed models for multi-outcome longitudinal data. *Stat. Model.* **2013**, *13*, 199–221.
181. Huang, Y.; Dagne, G. A Bayesian Approach to Joint Mixed-Effects Models with a Skew-Normal Distribution and Measurement Errors in Covariates. *Biometrics* **2011**, *69*, 260–269.
182. Huang, Y.; Dagne, G.A. Simultaneous Bayesian inference for skew-normal semiparametric nonlinear mixed-effects models with covariate measurement errors. *Bayesian Anal.* **2012**, *7*, 189–210.
183. Goldman, A.I.; Carlin, B.P.; Crane, L.R.; Launer, C.; Korvick, J.A.; Deyton, L.; Abrams, D.L. Response of CD4+ and Clinical Consequences to Treatment Using ddI or ddC in Patients with Advanced HIV Infection. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **1996**, *11*, 161–169.
184. Huang, Y.; Chen, J.; Yan, C. Mixed-Effects Joint Models with SkewNormal Distribution for HIV Dynamic Response with Missing and Mismeasured Time-Varying Covariate. *Int. J. Biostat.* **2012**, *8*, 1–28.
185. Lederman, M.; Connick, E.; Landay, A.; Kuritzkes, D.R.; Spritzler, J.; Clair, S.M.; Kotzin, B.L.; Fox, L.; Chiozzi, M.H.; Leonard, J.M.; et al. Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and zalcitabine: results of AIDS Clinical Trials Group Protocol 315. *J. Infect. Dis.* **1998**, *178*, 70–79.
186. Yu, B.; O'Malley, A.J.; Ghosh, P. Linear mixed models for multiple outcomes using extended multivariate skew-t distributions. *Stat Interface* **2014**, *7*, 101–111.
187. Ghosh, P.; Hanson, T. A semiparametric Bayesian approach to multivariate longitudinal data. *Aust. N. Z. J. Stat.* **2010**, *52*, 275–288.
188. Zadkarami, M.R.; Rahimi, A. Factors Associated with Birthweight: An Application of the Multiple Skew-Normal Regression. *Res. J. Obs. Gynecol.* **2008**, *1*, 9–17.
189. Mahmud, S.; Lou, W.; Johnston, N. A probit- log- skew-normal mixture model for repeated measures data with excess zeros, with application to a cohort study of paediatric respiratory symptoms. *BMC Med. Res. Methodol.* **2010**, *55*, 1–12.
190. Ho, R.; Fong, T.C.T.; Cheung, I.K.M. Cancer-related fatigue in breast cancer patients: factor mixture models with continuous non-normal distributions. *Qual. Life Res.* **2014**, *23*, 2909–2916.
191. Jamalizadeh, A.; Balakrishnan, N.; Salehi, M. Order statistics and linear combination of order statistics arising from a bivariate selection normal distribution. *Stat. Probab. Lett.* **2010**, *80*, 445–451.
192. Fishman, G.; Baca, W.; Alexander, K.R.; Derlacky, D.J.; Glenn, A.M.; Viana, M.A.G. Visual acuity in patients with best vitelliform macular dystrophy. *Ophthalmology* **1993**, *100*, 1665–1670.
193. Mansourian, M.; Mahdizadeh, Z.; Park, J.J.; Haghighi-Javanmard, S. Skew-symmetric Random Effect Models with Application to a Preventive Cohort Study: Improving Outcomes in Low Back Pain Patients. *Int. J. Prev. Med.* **2013**, *4*, 279–285.
194. Ismail, S.; Sun, W.; Nathoo, F.S.; Babul, A.; Moiseev, A.; Beg, M.; Virji-Babul, N. A Skew-t space-varying regression model for the spectral analysis of resting state brain activity. *Stat. Methods Med Res.* **2012**, *22*, 422–438.
195. Kline, B.; Tobias, J.L. The Wages Of BMI: Bayesian Analysis Of A Skewed Treatment—Response Model With Nonparametric Endogeneity. *J. Appl. Econ.* **2008**, *23*, 767–793.

196. Fernandes, E.; Pacheco, A.; Penha-Gonçalves, C. Mapping of quantitative trait loci using the skew-normal distribution. *J. Zhejiang Univ. Sci.* **2007**, *8*, 792–801.
197. Chai, H.; Bailey, K.R. Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. *Stat. Med.* **2008**, *27*, 3643–3655.
198. Bandyopadhyay, D.; Lachos, V.H.; Abanto-Valle, C.A.; Ghosh, P. Linear Mixed Models for Skew-Normal Independent bivariate responses with an application to Periodontal Disease. *Stat Med.* **2010**, *10*, 2643–2655.
199. Bandyopadhyay, D.; Lachos, V.H.; Castro, L.M.; Dey, D.K. Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biom. J.* **2012**, *54*, 405–425.
200. Dagne, G.A. NIH Bayesian Inference for Skew-Normal Mixture Models With Left Censoring. *J. Biopharm. Stat.* **2013**, *23*, 1023–1041.
201. Ho, H.; Lin, T.I.; Chang, H.H.; Haase, S.B.; Huang, S.; Pyne, S. Parametric modeling of cellular state transitions as measured with flow cytometry. *BMC Bioinform.* **2012**, *13*, .
202. Diggle, P.J.; Kenward, M.G. Informative dropout in longitudinal data analysis (with discussion). *Appl. Stat.* **1994**, *43*, 49–94.
203. Contreras-Reyes, J. Analyzing Fish Condition Factor Index Through Skew-Gaussian Information Theory Quantifiers. *Fluct. Noise Lett.* **2016**, *15*, 1–15 .
204. Contreras-Reyes, J.E.; Arellano-Valle, R.B.; Canales, T.M. Comparing growth curves with asymmetric heavy-tailed errors: application to the southern blue whiting (*Micromesistius australis*). *Fish. Res.* **2014**, *159*, 88–94.
205. Lee, M.S.Y.; Skinner, A. Testing fossil calibrations for vertebrate molecular trees. *Zool. Scr.* **2011**, *40*, 538–543.
206. Stoy, P.C.; Williams, M.; Disney, M.; Prieto-Blanco, A.; Huntley, B.; Baxter, R.; Lewis, P. Upscaling as ecological information transfer: A simple framework with application to Arctic ecosystem carbon exchange. *Landsc. Ecol.* **2009**, *24*, 971–986.
207. Simolo, C.; Brunetti, M.; Maugeri, M.; Nanni, T. Evolution of extreme temperatures in a warming climate. *Geophys. Res. Lett.* **2011**, *38*, 1–6.
208. Anagreh, Y.; Bataineh, A.; Al-Odat, M. Assessment of renewable energy potential, at Aqaba in Jordan. *Renew. Sustain. Energy Rev.* **2010**, *14*, 1347–1351.
209. Bartoletti, S.; Loperfido, N. Modelling air pollution data by the skew-normal distribution. *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 513–517.
210. Morris, S.A.; Reich, B.J.; Thibaud, E.; Cooley, D. A space-time skew- t model for threshold exceedances. *Biometrics* **2017**, *73*, 749–758, doi:10.1111/biom.12644.
211. Guttorp, P.; Loperfido, N. Network bias in air quality monitoring design. *Environmetrics* **2008**, *19*, 661–671.
212. Hashemi, F.; Amirzadeh, V.; Jamalizadeh, A. An Extension of the Birnbaum-Saunders Distribution Based on Skew-Normal- t Distribution. *J. Stat. Res. Iran* **2015**, *12*, 1–37.
213. Zakaria, R.; Metcalfe, A.V.; Howlett, P.; Piantadosi, J.; Boland, J. Using the skew- t copula to model bivariate rainfall distribution. *Anziam J.* **2010**, *51*, C231–C246.
214. Nathoo, F.S. Space-time regression modeling of tree growth using the skew- t distribution. *Environmetrics* **2010**, *21*, 817–833, doi:10.1002/env.1057.
215. Li, C.I.; Su, N.C.; Su, P.F.; Shyr, Y. The design of $\bar{X}$ and R control charts for skew normal distributed data. *Commun. Stat. Theory Methods* **2014**, *43*, 4908–4924, doi:10.1080/03610926.2012.717666.
216. Figueiredo, F.; Gomes, M.I. The Skew-Normal Distribution in SPC. *Rev. Stat. J.* **2013**, *11*, 83–104.
217. Renda, Y.; Longquan, Z.; Kun, L. An Empirical Study on the Energy Intensity in China Based on the Skew-normal Distribution. *Int. J. Smart Home* **2015**, *9*, 73–82.
218. Youssef, M. Outlier Detection Technique Based on Skew-Normal Distribution. *Int. J. Res. Wirel. Syst.* **2013**, *2*, 7–16.
219. Zadkarami, M.R.; Rowhani, M. Application of Skew-normal in Classification of Satellite Image. *J. Data Sci.* **2010**, *8*, 597–606.
220. Balef, H.A.; Kamal, M.; Afzali-Kusha, A.; Pedram, M. All-Region Statistical Model for Delay Variation Based on Log-Skew-Normal Distribution. *IEEE Trans. Comput. Aided Des. Integr. Circ. Syst.* **2016**, *35*, 1503–1508.
221. Tsai, C.C.; Lin, C.T. Lifetime inference for highly reliable products based on skew-normal accelerated destructive degradation test model. *IEEE Trans. Reliab.* **2015**, *64*, 1340–1355.
222. Zou, Y.; Zhang, Y. Use of Skew-Normal and Skew- t Distributions for Mixture Modeling of Freeway Speed Data. *J. Transp. Res. Board* **2011**, No. 2260, 67–75.

223. Ramprasath, S.; Vijaykumar, M.; Vasudevan, V. A skew-normal canonical model for statistical static timing analysis. *IEEE Trans. VLSI Syst.* **2016**, *24*, 2359–2368, doi:10.1109/TVLSI.2015.2501370.
224. Müller, P.; José A. del Peral-Rosado, R.P.; Seco-Granados, G. Statistical trilateration with skew- t distributed errors in LTE Networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 7114–7127, doi:10.1109/TWC.2016.2597836.
225. Eyer, L.; Genton, M.G. An astronomical distance determination method using regression with skew-normal errors. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Genton, M.G., Ed.; Chapman & Hall/CRC: Boca Raton, FL, USA 2004; Chapter 18, pp. 309–319.
226. Simola, U.; Dumusque, X.; Cisewski-Kehe, J. Measuring precise radial velocities and cross-correlation function line-profile variations using a skew normal density. *Astron. Astrophys.* **2019**, *622*, A131, doi:10.1051/0004-6361/201833895.
227. Lu, X.; Liu, Q.; Xue, F. Unique closed-form solutions of portfolio selection subject to mean-skewness-normalization constraints. *Oper. Res. Perspect* **2019**, *6*, 1–15.
228. Bekaert, G.; Harvey, C.R.; Erb, C.B.; Viskantam, T.E. Distributional Characteristics of Emerging Market Returns & Asset Allocation. *J. Portf. Manag.* **1998**, *24*, 201–116.
229. Bogetoft, P.; Otto, L. *Benchmarking with DEA and SFA*; R package version 0.27. 2018. Available online: <https://cran.r-project.org/package=Benchmarking> (accessed on 6 January 2020).
230. Tancredi, A. *Accounting for Heavy Tails in Stochastic Frontier Models*; Working paper 2002.16; Dipartimento di Scienze Statistiche, Università di Padova: Padova, Italia, 2002.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).