METRON - *International Journal of Statistics*
2010, vol. LXVIII, n. 3, pp. 275-298

ADELCHI AZZALINI – MARC G. GENTON – BRUNO SCARPA

Invariance-based estimating equations for skew-symmetric distributions

Summary - We develop estimating equations for the parameters of the base density of a skew-symmetric distribution. The method is based on an invariance property with respect to asymmetry. Various properties of this approach and the selection of a root are discussed. We also present several extensions of the methodology, namely to the regression setting, the multivariate case, and the skew-*t* distribution. The approach is illustrated on several simulations and a numerical example.

Key Words - Asymmetry; Distributional invariance; Generalized skew-normal distribution; Root selection; Semiparametric.

1. Introduction

Consider the general framework of distributions whose probability density function is, up to a location parameter, of the form

$$f_{\text{SS}}(z) = 2 f_0(z)\pi(z), \qquad z \in \mathbb{R}, \tag{1}$$

where $f_0(\cdot)$ is a symmetric density in $\mathbb{R}$, that is $f_0(-z) = f_0(z)$ for all $z \in \mathbb{R}$, and $\pi(\cdot)$ is a skewing function satisfying $\pi(-z) = 1 - \pi(z) \geq 0$ for all $z \in \mathbb{R}$. Location and scale parameters, ξ and ω respectively, can be introduced by means of $Y = \xi + \omega Z$, where Z is a random variable with density (1). In the construction (1), the base density $f_0(z)$ is recovered when $\pi(z) \equiv 1/2$. Wang *et al.* (2004a) adopted the term *skew-symmetric* (SS) for distributions whose density is of the form given by (1). Azzalini & Capitanio (2003) considered a class of distributions equivalent to (1) but with the parametrization $\pi(z) = G(w(z))$, where $G(\cdot)$ is a one-dimensional cumulative distribution function whose derivative G' exists and satisfies $G'(-z) = G'(z)$, and $w(\cdot)$ is an odd function, that is $w(-z) = -w(z)$ for all $z \in \mathbb{R}$. The choices

Received June 2009 and revised January 2010.

$G = F_0$, the cumulative distribution function associated with the base density f_0, and $w(z) = \alpha z$, where $\alpha \in \mathbb{R}$ is a shape parameter controling the skewness in (1), have received particular attention in the literature, see Genton (2004) and Azzalini (2005) for overviews. The most popular distribution is obtained on setting $f_0(z) = \phi(z)$, the $N(0, 1)$ density, and $\pi(z) = \Phi(\alpha z)$, where Φ is the $N(0, 1)$ cumulative distribution function. It yields the skew-normal (SN) distribution studied by Azzalini (1985) and denoted by $\mathrm{SN}(\xi, \omega^2, \alpha)$, with density

$$f_{\mathrm{SN}}(y) = 2\omega^{-1}\,\phi(z)\,\Phi(\alpha z), \qquad y \in \mathbb{R}, \tag{2}$$

where $z = \omega^{-1}(y - \xi)$.

The class of distributions (1) can be obtained via a suitable censoring mechanism, regulated by $\pi(\cdot)$, applied to samples generated by the base density f_0, see Arellano-Valle *et al.* (2006) and references therein. Under this perspective, it is of interest to estimate the parameters of f_0 via a method which does not depend, or depends only to a limited extent, on the component $\pi(\cdot)$, which in many cases is not known, or is not of interest to be estimated. This is the key motivation of the present work.

The class of skew-symmetric distributions possesses a property of distributional invariance. It has been noted by several authors in various particular settings and is stated in the context of (1) in the following proposition.

Proposition 1 (Distributional Invariance). *Denote by X and Z two random variables having density function f_0 and* (1)*, respectively. If $T(\cdot)$ is an even function, that is $T(-x) = T(x)$ for all $x \in \mathbb{R}$, then*

$$T(X) \stackrel{d}{=} T(Z). \tag{3}$$

The proof of Proposition 1 can be found in Wang *et al.* (2004a, Prop. 6, p. 1264) and in Azzalini & Capitanio (2003, Prop. 2, p. 374) for the equivalent parametrization $\pi(z) = G(w(z))$. Other authors have derived similar proofs for particular cases; see for instance Azzalini (1986), Azzalini & Capitanio (1999), Genton *et al.* (2001), Wang *et al.* (2004b), Genton & Loperfido (2005).

The extension of the previous developments to the multivariate case is straightforward. In particular, Proposition 1 still holds for multivariate skew-symmetric distributions. The choices $f_0(z) = \phi_d(z)$, the multivariate normal density, and $\pi(z) = \Phi(\alpha^\top z)$, $\alpha z \in \mathbb{R}^d$, yield the multivariate skew-normal distribution studied by Azzalini & Dalla Valle (1996). Further generalizations of Proposition 1 to skewed distributions arising from arbitrary selection mechanisms can be found in Arellano-Valle & Genton (2010).

Ma *et al.* (2005) have derived locally efficient semiparametric estimators of the parameters of the base density $f_0(\cdot)$ in (1) based on the theory of regular asymptotically linear (RAL) estimators and their influence functions. The resulting estimators are consistent and asymptotically normal regardless of the possible misspecification of the function $\pi(z)$. The construction of these optimal estimators relies on the identification of the space tangent to the nuisance parameter $\pi(\cdot)$ and its orthogonal complement. The latter, not surprisingly, consists of even functions as suggested by Proposition 1. Further results along those lines, based on constrained local likelihood estimators, were reported by Ma & Hart (2007).

In this paper, we consider invariance-based estimating equations (IBEE) of the parameters of the base density $f_0(\cdot)$ in (1). Although IBEE can be considered a particular case of the RAL formulation, we view their introduction via the conceptually simpler motivation provided by the property of distributional invariance, without involving more elaborate concepts of the RAL formulation; this also explains the different term adopted. As a consequence of the selected scheme, we do not consider optimality aspects which are extensively discussed by Ma *et al.* (2005). Instead, we focus on the construction of simple estimating equations and on the identification of their roots, a problem which has been discussed only very briefly by Ma *et al.* (2005).

The paper is organized as follows. In Section 2, we derive estimating equations for the parameters of the base density f_0 based on the distributional invariance given in Proposition 1. We discuss various properties of this approach and the selection of a root. Several extensions of our methodology are described in Section 3, namely to the regression setting, the multivariate case, and the skew-*t* distribution. We illustrate our methodology on several simulations and a numerical example in Section 4 and conclude with a discussion in Section 5. Proofs are given in an Appendix.

2. Invariance-based estimating equations

The property of distributional invariance (3) leads immediately to the construction of quantities which are pivotal with respect to the location and scale parameters of f_0 as well as the skewing factor $\pi(z)$. From this key fact, one can build an estimation method for the parameters of f_0. To explore this path, we shall proceed by examining initially the simplest possible setting, and extend later the methodology to more elaborate cases.

A point to bear in mind, and the key challenge of the problem, is that ξ and ω represent different entities for f_0 and for f. For instance, if $f_0 = \phi$, then ξ and ω correspond to the mean and the standard deviation of the distribution, respectively, but this is not the case for f, where they represent another location and another scale parameter; this explains why we do not use the traditional symbols μ and σ.

2.1. Estimation equations in the basic case

Consider initially the case of a simple random sample $y_1, \ldots, y_n$ drawn from a random variable Y with density function (1). Property (3) ensures that, for any choice of even functions $T_1, T_2, \ldots,$ the expected values

$$\mathbb{E}\left\{T_k\left(\frac{Y-\xi}{\omega}\right)\right\} = c_k, \qquad (k = 1, 2, \ldots), \tag{4}$$

depend on f_0 only, provided they exist.

To build a set of estimating equations, we replace the above expected values by sample averages for two such functions, T_1 and T_2, evaluated with Y replaced by the observed values. Hence we consider the solutions of

$$\frac{1}{n}\sum_{i=1}^{n} T_k\left(\frac{y_i-\xi}{\omega}\right) - c_k = 0, \qquad (k = 1, 2), \tag{5}$$

as estimates of ξ and ω.

To completely define the estimation method, we must select two specific functions $T_k(z)$'s. Among the many options, a quite natural one is to consider powers of $|z|$, in particular the first two of such functions, hence setting $T_1(z) = |z|$, $T_2(z) = z^2$. Other options could be considered, and some of them will be discussed later.

At first sight, we are now within the classical framework of unbiased estimating equations, with an apparently standard development, but a closer look reveals immediately some peculiar aspects. To ease discussion of this point and because of its central role, we shall now focus on the specific case when f_0 in (1) is the normal density.

Assume then that the variable $Z = (Y - \xi)/\omega$ has density function of "generalized skew-normal" type, namely of the form

$$\bar{f}_{\mathrm{GSN}}(z) = 2\,\phi(z)\,\pi(z), \qquad z \in \mathbb{R}, \tag{6}$$

for some arbitrary skewing function $\pi(\cdot)$, provided that $\pi(-z) = 1 - \pi(z) \geq 0$. Hence $|Z| \sim \chi_1$ and it is well known that

$$c_k = \mathbb{E}\left\{|Z|^k\right\} = \frac{2^{k/2}\,\Gamma(k/2+1/2)}{\Gamma(1/2)}, \qquad (k > 0),$$

which provides the coefficients (4) when $T_k(z) = |z|^k$. From the second equation in (5), the estimate $\hat{\omega}$ of ω is such that

$$\hat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\xi})^2, \tag{7}$$

where $\hat{\xi}$ is the estimate of ξ, which must be a solution of the equation

$$q(t) = \frac{1}{n}\sum_{i=1}^{n}|y_i - t| - \sqrt{\frac{2}{\pi}}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - t)^2\right)^{1/2} = 0. \tag{8}$$

To study the large-sample behaviour of this equation, we consider the corresponding equation obtained by replacing the sample averages by their expected values. Hence we are led to consider

$$Q(t) = \mathbb{E}\{|Z - t|\} - \sqrt{\frac{2}{\pi}}\,\mathbb{E}\left\{(Z - t)^2\right\}^{1/2} = 0. \tag{9}$$

In principle, one should consider the broader class of densities generated by a transformation of type $Y = \xi + \omega Z$, but it suffices to examine the case with $\xi = 0$ and $\omega = 1$, and notice that the solutions of (8) and of (9) are location and scale equivariant. By construction, (9) has always as root $t = 0$ which corresponds to the value of ξ, since $Z = (Y - \xi)/\omega$ in (9) has location parameter 0.

Computation of the moments involved by $Q(t)$ requires to work with a specific function $\pi(\cdot)$. In the context recalled in Section 1, the skew-normal distribution plays a central role, and in the following discussion we adopt $\pi(z) = \Phi(\alpha z)$; in this case, we denote the density function of Z by $\phi_\alpha(\cdot)$. It is known that

$$\mathbb{E}\{Z\} = \sqrt{\frac{2}{\pi}}\frac{\alpha}{\sqrt{1+\alpha^2}} = \mu_\alpha, \qquad \mathrm{var}\{Z\} = 1 - {\mu_\alpha}^2 = \sigma_\alpha^2, \tag{10}$$

and the distribution function $\Phi_\alpha(\cdot)$ has a known expression; see Azzalini (1985). After some algebraic work reported in detail in the Appendix, $Q(t)$ takes the specific form

$$\begin{aligned} Q_\alpha(t) = \mu_\alpha - t + 2\{\phi_\alpha(t) + t\,\Phi_\alpha(t) - \mu_\alpha\,\Phi(t\sqrt{1+\alpha^2})\} \\ - \sqrt{\frac{2}{\pi}}\left\{\sigma_\alpha^2 + (\mu_\alpha - t)^2\right\}^{1/2} \end{aligned} \tag{11}$$

for which it is immediate to confirm that $t = 0$ is always a zero. Further results proved in the Appendix ensure that there is at least a second root of (11), if $\alpha \neq 0$. This second root has the same sign of α, and it is numerically close to μ_α. When $\alpha = 0$, $t = 0$ is a stationary point. As a by-product of a more general result, the Appendix includes an expression for the mean absolute deviation of a skew-normal variate.

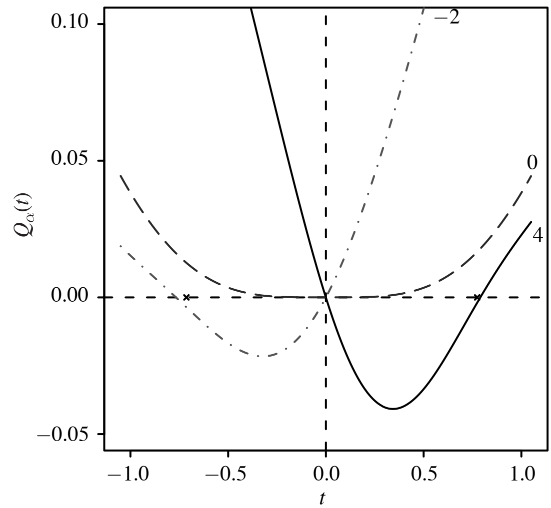


Figure 1. Plot of function $Q_\alpha(t)$ for $\alpha = -2, 0, 4$; the $\times$ signs on the horizontal axis mark the mean μ_α for $\alpha = -2, 4$.

Figure 1 illustrates graphically $Q_\alpha(t)$ for $\alpha = -2$, $\alpha = 0$ and $\alpha = 4$. From here, it is visible that $Q_\alpha(t)$ has exactly two zeroes except in the case $\alpha = 0$ when there is a single root. The fact that there are no more than two roots has been confirmed numerically in all cases which we have inspected, although it has not been proved analytically.

Since (9) represents the limiting behaviour of the sample version (8), there is clearly a radical difference from more familiar cases of estimating equations, where one has often reasons to believe that the equations have a single solution, at least for large samples.

As already noticed, the root of (9) at $t = 0$ corresponds to the true value of ξ and any other root represents a spurious solution to be discarded. In actual use of the method, one deals with the sample version (8), working with the unnormalized variable Y instead of Z, and the problem of recognising the "good root" arises. After $\hat{\xi}$ has been obtained as a root of (8), the corresponding estimate of ω^2 is given by (7).

The problem of selecting a root of (8) is inevitably exacerbated in finite samples, where $q(t)$ exhibits a more irregular behaviour than $Q(t)$, including possibly the presence of more than two roots or alternatively of no roots. Examples of this sort are illustrated in Figure 2 which refers to two samples of size $n = 50$ from a SN(0,1,2) variate; in the present context, $n = 50$ must be regarded to be a rather small sample size, and the behaviour of $q(t)$ is still affected by much random fluctuations. The meaning of the circles in the plots will be explained later on.

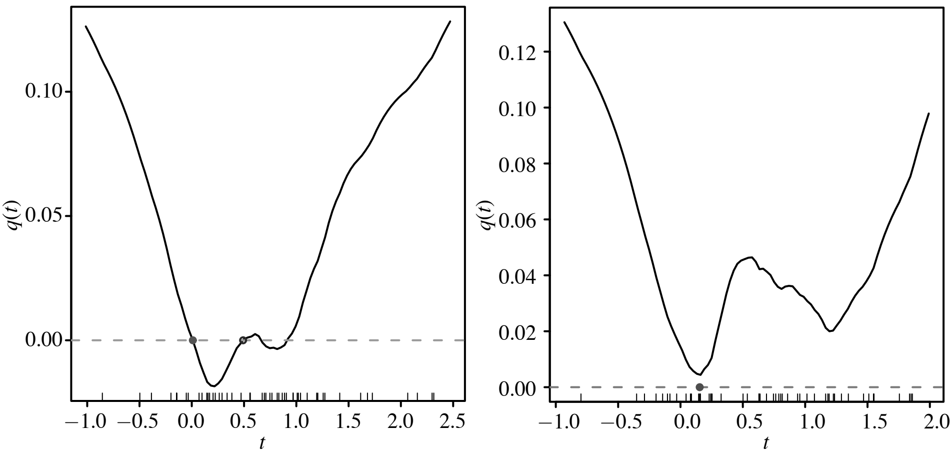


Figure 2. Plot of the function $q(t)$ associated with two samples of size $n = 50$ from SN(0,1,2) whose component values are indicated by the ticks at the bottom. The function of the left panel has four zeroes, of which the circles mark those which include the minimum of the function, and the solid circle denotes the one elected to be $\hat{\hat{\xi}}$. The function of the right panel has no zeroes and the circle denotes the point of minimum.

Before moving to the next stage, we discuss briefly alternative choices for the function T_1 appearing in (5). The choice $T_2(z) = z^2$ has been kept fixed, since it allows a simple solution (7) and hence it ensures the reduction to consideration of a single non-linear function which has to be solved numerically.

The alternative choices to $T_1(z) = |z|$ which have been considered are the following even functions:

$T_1(z)$:	$\mathbb{E}\{T_1(Z)\}$:
$\lvert z\rvert^3$	$\sqrt{8/\pi}$
$I_{(-a,a)}(z)$	$\Phi(a) - \Phi(-a)$
$\cos(a\,z)$	$\exp(-a^2/2)$
$\cosh(a\,z)$	$\exp(a^2/2)$

where $I_A(z)$ is the indicator function of the set A and a is an arbitrary constant, with the condition $a > 0$ for the second option. Using the property of distributional invariance, $\mathbb{E}\{T_1(Z)\}$ can be computed under the assumption $Z \sim N(0, 1)$, even if the parent distribution of Z is (6).

Of these alternative choices to $T_1 = |z|$, $|z|^3$ appears the less appealing because it is known that sample moments of higher moments are affected by higher variability of the lower moments. A similar remark holds for cosh. For three of the above functions $T_1(z)$, a value of a must be specified. Some exploratory work has indicated that a choice not too far from $a = 1$ is reasonable for all of them. The other indication which emerged was that these alternatives behave broadly in a similar way of $T_1 = |z|$, and they did not ease identification of the appropriate root $\hat{\xi}$, which soon emerged as the more critical aspect of the current approach. We therefore retained our original choice of T_1, and focused our attention on the far more critical issue which is discussed in the next subsection.

2.2. The problem of selecting one root

As explained in Section 2.1, a criterion is required to distinguish the "good root" of (8) from the spurious ones. On the other hand, when there are no solutions, a replacement for the "missing root" must be provided. Notice that the issue of selecting the appropriate root is logically separated from the one of establishing the estimating equations, in the sense that, once the estimating equations (7)-(8) have been adopted, there could be different options for the problem to be discussed in the present section. To some extent, also the opposite statement holds, in the sense that the criteria to follow could be applied to different estimating equations, as it will be clarified later.

The following appear to be reasonable criteria to dismiss the more peculiar cases, and reduce the problem to a standard form. There is some subjective component in these choices, but we have found that alternative variant forms do not change much the behaviour of the overall procedure.

- R1. If there are more than two roots, like in the left panel of Figure 2, argue that the lower portion of the function is the one approaching more closely the large-sample behaviour exhibited in Figure 1, and take into account only those two roots which include the minimum value of $q(t)$, discarding the others.
- R2. If there are no roots, like in the right panel of Figure 2, argue that the point of minimum of $q(t)$ is where the function is closer to having a zero, and take this point as the estimate.

In a few cases, where R2 has been applied, an estimate $\hat{\xi}$ has already been selected at this stage, and so also $\hat{\omega}$. In the majority of cases, however, we are then left with the problem of choosing among two roots, $\hat{\xi}_1$ and $\hat{\xi}_2$, say; this situation could possibly hold after application of rule R1. This is the problem to be discussed for the rest of this subsection.

Our first solution to this problem was as follows. Recall that, from the results of Section 2.1, the spurious root of (9) is close to the mean value of the distribution. This fact indicates that, in the sample version (8) of the equation, the root closer to the sample mean must be discarded. Therefore, after computing from (7) the two values, $\hat{\omega}_1$ and $\hat{\omega}_2$, say, associated with $\hat{\xi}_1$ and $\hat{\xi}_2$, we select the root $\hat{\xi}_j$ whose matching $\hat{\omega}_j$ is larger.

In the example exhibited in the left panel of Figure 2, two roots have been discarded by application of rule R1, retaining only those marked by circles.

Of these two roots, the $\hat{\xi}_j$ associated to higher $\hat{\omega}_j^2$ has been selected, denoted by the solid circle. In the case of the right panel, where no roots exist, rule R2 has produced the estimate indicated by the solid circle.

While the above-described criterion, based on the value of $\hat{\xi}_j$ with larger $\hat{\omega}_j$, has a simple motivation and implementation, it suffers of the major drawback of being strongly linked to the assumption that $\pi(z) = \Phi(\alpha\, z)$, for some parameter α. When this assumption and hence (11) do not hold, the support for this criterion disappears. One would hope that the spurious root of $Q(t)$ remains closer to the smaller value of $\hat{\omega}_j^2$, a fact which would keep the criterion operating successfully. Unfortunately, this is not the case. An explicit computation of (9) is difficult to achieve for most plausible alternative choices of $\pi(z)$, but it is easy to compute $Q(t)$ by numerical quadrature, and this has shown that, for many other choices of $\pi(z)$, the required condition does not hold. The alternative forms of $\pi(z)$ considered were of the form

$$\pi(z) = \Phi\left(\sum_{k=1}^{K} \alpha_{2k-1}\, z^{2k-1}\right), \tag{12}$$

for some small integer K and various choices of the coefficients α_{2k-1}. This class of functions leads to a family of densities of type (6) with a large variety of shapes, even for $K = 2$ only; see Ma & Genton (2004) for formal properties of this family, and for a set of graphical illustrations.

Clearly, if $K = 1$ in (12), we return to the skew-normal density, and in this case the method worked well, since it has been tailored for this case. If $K > 1$ and the coefficients α_{2k-1} are chosen to depart from the skew-normal shape, then the method failed.

Another approach to select one of $\hat{\xi}_1$ and $\hat{\xi}_2$ is based on the quadratic inference function (QIF) proposed by Qu *et al.* (2000). In its original formulation, this is an estimation method which in the present context leads to estimate ξ by minimizing

$$q^*(t) = \begin{pmatrix} g_1 \\ g_3 \end{pmatrix}^T \left\{ \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} g_{i1}^2 & g_{i1} g_{i3} \\ g_{i3} g_{i1} & g_{i3}^2 \end{pmatrix} \right\}^{-1} \begin{pmatrix} g_1 \\ g_3 \end{pmatrix}$$

where g_1 is the left side of the first estimating equation in (5), evalutated at $\xi = t$, and g_3 can be chosen either in the form

$$g_3(t) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - t}{\hat{\omega}} \right|^3 - c_3,$$

or equal to some of the alternative functions proposed in Section 2.1. We have however considered the QIF criterion as a supplement to our IBEE method, to select the preferred root of (8). Some numerical exploration, not reported here, has however shown that $q^*(t)$ has several minima, and they coincide almost perfectly with the roots of (8). While in a sense it is reassuring that the two approaches provide reciprocally consistent indications, we were left with our selection problem unsolved.

A further criterion for choosing between $\hat{\xi}_1$ and $\hat{\xi}_2$ is based on the use of a nonparametric estimate $\hat{f}(y)$ of the density function $f(y)$ of the observed variable Y and on the adequacy of $(\hat{\xi}_j, \hat{\omega}_j)$ to match $\hat{f}$ once these estimates are inserted in the semiparametric formulation

$$f_{\mathrm{GSN}}(y) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\,\pi\left(\frac{y-\xi}{\omega}\right), \tag{13}$$

which corresponds to (6) after the transformation $Y = \xi + \omega Z$. More specifically, the method is as follows:

1. on the basis of the sample $y_1, \ldots, y_n$, construct a nonparametric estimate $\hat{f}$ of f, for instance via the kernel method;
2. from $\hat{\xi}_j, \hat{\omega}_j$ obtained earlier, produce the "standardized" samples $z_{ij} = (y_i - \hat{\xi}_j)/\hat{\omega}_j$ (for $i = 1, \ldots, n$, $j = 1, 2$), and the corresponding nonparametric estimates $\tilde{f}_1(z)$, $\tilde{f}_2(z)$;
3. on setting
$$r_j(z) = \frac{\tilde{f}_j(z)}{2\,\phi(z)}, \qquad \hat{\pi}_j(z) = \frac{r_j(z)}{r_j(z) + r_j(-z)},$$
obtain estimates $\hat{\pi}_j(z)$ which fulfill the condition $\hat{\pi}_j(z) + \hat{\pi}_j(-z) = 1$, and then the semiparametric estimates
$$\hat{f}_j(y) = \frac{2}{\hat{\omega}_j}\phi\left(\frac{y-\hat{\xi}_j}{\hat{\omega}_j}\right)\,\hat{\pi}_j\left(\frac{y-\hat{\xi}_j}{\hat{\omega}_j}\right); \tag{14}$$
4. compare the semiparametric and the nonparametric estimates by a distance criterion such as
$$D_j = \int_{\mathbb{R}} |\hat{f}_j(y) - \hat{f}(y)|^2 \mathrm{d}y, \qquad j = 1, 2,$$
and adopt the pair $(\hat{\xi}_j, \hat{\omega}_j)$ with smaller D_j.

This procedure was tested via a set of simulations, sampling data from various alternatives of type (12)-(13) and examining when the new criterion led to selection of the appropriate root $\hat{\xi}_j$. Unfortunately, even this new procedure did not provide a satisfactory methodology in the sense of producing reliable results over a wide set of alternatives. It does not make much sense to allocate space here for details on the numerical outcomes, but we can summarize the essence by saying that the new method worked well where the first one did not, that is with choices of (12) where $K > 1$, and vice versa.

Clearly there are various ways in which the above procedure could be better tuned, such as changing the distance criterion of the densities entering in the final step, or modifying the rule for choosing the smoothing parameter of the nonparametric estimate. However these variations do not alter substantially the final outcome.

One feature which we observed systematically in the estimates (14) was that, even when the "wrong" pair $(\hat{\xi}_j, \hat{\omega}_j)$ was used, still the density estimate was adequate, due to a suitable compensation of the skewing factor $\hat{\pi}_j(\cdot)$. The latter one was grossly different from the real $\pi(\cdot)$, and typically much more erratic, but the final $\hat{f}_j$ was close to the actual f.

There are other criteria for selecting $\hat{\xi}_j$ which we have considered, but with performance which was similar to the one just described, in the sense of working well when $K > 1$ and the shape is different from the normal or skew-normal density, but poorly in these more regular cases. Therefore, it is not appealing to give a more detailed description.

The overall message from these attempts seems to be that the formulation is trapped in some form of non-identifiability. This reading is plausible if one considers that the set of admissible functions $\pi(\cdot)$ is extremely wide. On the basis of this remark, and on the observation that the estimated $\hat{f}_j$ associated to the wrong estimate is typically more erratic than the correct $\hat{f}_j$, we have adopted a final strategy which aims at selecting the "less complex" model, where model complexity is related to the degree of variability of the estimated function. There are many possible choices for a numerical quantifier of the idea of model complexity, but a very simple and commonly employed one is

$$C_j = \int_{\mathbb{R}} \left(\hat{\pi}_j''(y)\right)^2 \,\mathrm{d}y. \tag{15}$$

The finally adopted criterion was therefore to select the root $\hat{\xi}_j$ leading to the smaller C_j $(j = 1, 2)$. Detailed outcome of some numerical work using the above criterion are reported in Section 4.

2.3. Some properties

To obtain the asymptotic distribution of the estimate $\hat{\theta}$ of $\theta = (\xi, \omega)$, we regard the estimate as a solution of the pair of estimating equations which are the sum of n independent and identically distributed components whose generic term is formed by the pair

$$g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = 0,$$

where

$$g_1 = \frac{|Y - \xi|}{\omega} - \sqrt{\frac{2}{\pi}},$$
$$g_2 = (Y - \xi)^2 - \omega^2,$$

which can be viewed within the general framework of Godambe (1991). The corresponding theory works by expanding locally $g = 0$ near the true parameter point; this effectively corresponds to the requirement that the consistent root of the equations has been selected. Under this assumption, $\hat{\theta}$ is asymptotically normal with mean θ and asymptotic variance

$$\Sigma(\hat{\theta}) = \mathbb{E}\{-\dot{g}\}^{-1}\, \mathbb{E}\Big\{g\, g^\top\Big\} \Big[\mathbb{E}\{-\dot{g}\}^{-1}\Big]^\top, \tag{16}$$

where

$$\dot{g} = \left(\frac{\partial g_i}{\partial \theta_j}\right).$$

See also the related problem considered by Cox (1993), specifically his formula (4).

For the specific choice of g adopted here, computations are much simplified if one takes into account that $|Y - \xi|$ is distributed as $|U|$ where $U \sim N(0, \omega^2)$. Then some simple algebra leads to

$$\mathbb{E}\Big\{g_1^2\Big\} = 1 - 2/\pi, \quad \mathbb{E}\{g_1\, g_2\} = \omega^2 \sqrt{2/\pi}, \quad \mathbb{E}\Big\{g_2^2\Big\} = 2\,\omega^4,$$

while computation of

$$\mathbb{E}\{-\dot{g}\} = \mathbb{E}\begin{pmatrix} \dfrac{\operatorname{sgn}(Y - \xi)}{\omega} & \dfrac{|Y - \xi|}{\omega^2} \\ 2\,(Y - \xi) & 2\,\omega \end{pmatrix}$$

involves specification of the distribution of Y for computing the elements of the first column. Under the assumption that Y is of skew-normal type, one obtains

$$\mathbb{E}\{-\dot{g}\} = \begin{pmatrix} \dfrac{2}{\pi\,\omega}\arctan\alpha & \sqrt{\dfrac{2}{\pi}}\dfrac{1}{\omega} \\ 2\omega\mu_\alpha & 2\,\omega \end{pmatrix},$$

where we have used the fact that

$$\mathbb{P}\{Y-\xi>0\} = \mathbb{P}\{Z>0\} = \frac{1}{2}+\frac{1}{\pi}\arctan\alpha.$$

To appreciate what is the practical effect of replacing the maximum likelihood estimate (MLE) by the semiparametric IBEE as well as the validity of the above asymptotic results, some numerical evidence is reported in Table 1, as a part of a somewhat more extended numerical work not reported here in full. The table reports first the asymptotic variance matrix for IBEE (16) computed under the assumption of skew-normality for Y with shape parameter $\alpha = 5$, and compares it with the analogous matrix for the MLE when the parametric formulation is exploited and the shape parameter is regarded as known. In addition, Table 1 shows the outcome of a simulation experiment with samples of size $n = 250$ by reporting n times the sample variance matrix of the estimates over 5000 replicates, both for IBEE and MLE.

TABLE 1: *Asymptotic variance matrices of* $(\hat{\xi}, \hat{\omega}^2)$ *sampling from a SN(0,1,5) variate.*

Theoretical			
IBEE		MLE	
0.721	–0.564	0.527	–0.412
–0.564	0.941	–0.412	0.826
Empirical for $n = 250$			
IBEE		MLE	
0.782	–0.610	0.592	–0.467
–0.610	0.974	–0.467	0.867

3. EXTENSIONS

3.1. Regression

It is possible to extend the above approach to the regression case setting where

$$y_i = \xi_i + \omega\, z_i, \quad (i = 1, \ldots, n), \tag{17}$$

$$\xi_i = x_i^\top \beta = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j, \tag{18}$$

where x_i denotes a $(p+1)$-dimensional vector of covariates whose first component x_{i0} is assumed to be identically 1, $\beta = (\beta_0, \dots, \beta_p)^\top$ is a vector of parameters, and the error terms $z_1, \dots, z_n$ are independently sampled values from (6).

To estimate the set of parameters $(\beta_0, \beta_1, \dots, \beta_p, \omega)^\top$, reconsider the relationships used in Section 2.1, namely

$$\mathbb{E}\{|Y - \xi|\} - \sqrt{2/\pi}\,\omega = 0,$$
$$\mathbb{E}\Big\{(Y - \xi)^2\Big\} - \omega^2 = 0,$$

which imply

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{|Y_i - \xi_i|\,x_{ij}\} - \sqrt{2/\pi}\,\omega\,\bar{x}_j = 0,$$
$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big\{(Y_i - \xi_i)^2\Big\} - \omega^2 = 0,$$

for $j = 0, \dots, p$; here $\bar{x}_j$ denotes the sample mean of the j-th covariate. Obviously, we obtain estimates of the parameters from solution of the sample analogue of (6)-(7), namely

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \xi_i|\,x_{ij} - \sqrt{2/\pi}\,\omega\,\bar{x}_j = 0, \tag{19}$$
$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \xi_i)^2 - \omega^2 = 0. \tag{20}$$

For solving these equations, it is convenient to assume that $\bar{x}_j = 0$ for $j = 1, \dots, p$. If this is not the case, we effectively work with shifted variables such that the required conditions hold; at the end of the estimation process, an appropriate adjustment of the intercept term provides the estimate referred to the original variables. With the above condition, (19) becomes

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \xi_i|\,x_{ij} = \begin{cases}\sqrt{2/\pi}\,\hat{\omega} & \text{if } j = 0,\\ 0 & \text{if } j = 1, \dots, p,\end{cases} \tag{21}$$

where ξ_i is expressed by (18) and of course

$$\hat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \xi_i)^2.$$

For solving (21), we can make use of a Gauss-Seidel type of procedure, which works as follows: at each stage of the iterative solution process and for the j-th component of β, define the j-th working variable as

$$u_i^{(j)} = y_i - \sum_{r \neq j} \beta_r\, x_{ir}, \qquad (i = 1, \ldots, n),$$

where the β_r's on the right-hand side denote the current values of the remaining parameters, and consider the $p + 1$ equations

$$\frac{1}{n}\sum_{i=1}^{n} |u_i^{(j)} - \beta_j x_{ij}|\, x_{ij} = \begin{cases} \sqrt{2/\pi}\,\hat{\omega} & \text{if } j = 0, \\ 0 & \text{if } j = 1, \ldots, p, \end{cases} \tag{22}$$

which are all equations in one variable each. These equations must be solved cyclically in turn, for $j = 0, 1, \ldots, p, 0, 1, \ldots, p, 0, 1, \ldots$ until convergence.

Of the equations in (22), the one for $j = 0$ is identical to the one considered for the case of a simple sample, hence we make use of the same strategy developed earlier. If $j > 0$, the terms x_{ij}'s take on positive and negative values. This remark suggests that there is always a solution to these equations, a fact which appears supported by some numerical experimentation, not reported here. In a large proportion of cases, this solution was also unique.

3.2. Multivariate case

Consider a random sample $y_1, \ldots, y_n$ drawn from a random vector $Y = (Y_1, \ldots, Y_d)^\top$ with multivariate generalized skew-normal density function given by

$$f_{\mathrm{GSN}}(y) = 2\phi_d(y - \xi; \Omega)\pi\{\omega^{-1}(y - \xi)\}, \quad y \in \mathbb{R}^d, \tag{23}$$

where $\xi \in \mathbb{R}^d$ is a location vector, $\Omega = (\omega_{ij})$ is a $p \times p$ covariance matrix, $\pi(\cdot)$ is a skewing function, and $\omega = \mathrm{diag}(\omega_1, \ldots, \omega_d)$ with $\omega_i = \sqrt{\omega_{ii}}$. Hence $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$ is a correlation matrix. Let $Z = \omega^{-1}(Y - \xi)$. As mentioned in the introduction, the distributional invariance in Proposition 1 still holds in the multivariate setting. Therefore, we start by solving estimating equations of the type (5) component-wise, that is, we treat the solutions of

$$\frac{1}{n}\sum_{i=1}^{n} T_k\left(\frac{y_{ij}-\xi_j}{\omega_j}\right) - c_k = 0, \qquad (k = 1, 2;\ j = 1, \dots, d), \tag{24}$$

as estimates of ξ and ω, where the choice of the even functions T_1 and T_2 has been discussed in Section 2.1. Then, we solve for the off-diagonal elements of Ω from the equation $\mathbb{E}(\omega Z Z^\top \omega) = \Omega$. This leads to the following explicit estimator for the $d(d-1)/2$ off-diagonal elements of Ω:

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\xi})(y_i - \hat{\xi})^\top. \tag{25}$$

Note that, since only the off-diagonal elements of Ω are estimated by (25), the estimate of the normalized matrix, $\hat{\bar{\Omega}} = \hat{\omega}^{-1}\hat{\Omega}\hat{\omega}^{-1}$, is guaranteed to be a correlation matrix by construction, that is, it has 1's on the diagonal.

3.3. Generalized skew-t

So far we have considered only cases where the base function f_0 in (1) is the normal density, possibly multivariate. The above formulation can however easily be adapted to consider other distributions for f_0, and a special interest lies in families of distributions which allow regulation of their tail behaviour, which is what we consider next. In the recent related literature, interest has focused on the so-called skew-t distribution; see specifically Azzalini & Capitanio (2003) and Azzalini & Genton (2008). The common ingredient with these papers is that f_0 is taken to be the Student's t distribution. The key difference is that here the skewing factor $\pi(z)$ is treated nonparametrically rather than parametrically, which explains the term "generalized skew-t", similarly to the term "generalized skew-normal" for (6).

Consider then the case that f_0 in (1) is the t density with ν degrees of freedom, where ν is a parameter to be estimated alongside the location ξ and the scale factor ω. Hence we need to consider three equations similar to those in (5). If we choose again $T_k(z) = |z|^k$, then the equations become

$$\mathbb{E}\{|Y-\xi|\} = c'_\nu\, \mathbb{E}\Big\{(Y-\xi)^2\Big\}^{1/2},$$
$$\mathbb{E}\Big\{(Y-\xi)^2\Big\} = \frac{\nu}{\nu-2}\,\omega^2,$$
$$\mathbb{E}\Big\{|Y-\xi|^3\Big\} = c''_\nu\, \mathbb{E}\Big\{(Y-\xi)^2\Big\}^{3/2},$$

where the coefficients c'_ν, c''_ν can be computed by exploiting the property of distributional invariance plus the fact that T^m is distributed as $F^{m/2}$ if $T \sim t_\nu$ and $F \sim F(1,\nu)$, leading to

$$c'_\nu = \frac{(\nu-2)^{1/2}}{\pi^{1/2}} \frac{\Gamma\left(\dfrac{\nu-1}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)}, \qquad c''_\nu = \frac{(\nu-2)^{3/2}}{\pi^{1/2}} \frac{\Gamma\left(\dfrac{\nu-3}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)},$$

where $\nu > 3$ is assumed. In practical work, the above equations must clearly be replaced by the sample counterparts.

4. Simulations and a numerical example

To illustrate the validity of the procedure, we simulated data from various distributions. We mainly selected specifications of the generalized skew-normal type (12)-(13), with a number of different parameters set, (α_1, α_3), in order to consider a wide range of possible shapes for the distribution function and therefore for the censoring mechanism. As already mentioned, a polynomial with $K = 2$ in (12) is sufficient to produce a rich variety of density shapes. The numerical work is based on the R package `sn` (Azzalini, 2008).

For each choice of (α_1, α_3), we generated 3000 samples of $n = 250$ values from the sampling distribution and, for each sample, we applied rules R1-R2 followed by selection of the root $\hat{\xi}_j$ leading to the smaller C_j, when at least two roots were observed. Therefore we examined when the criterion led to selection of the appropriate root.

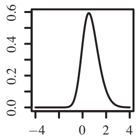
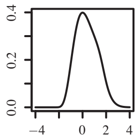
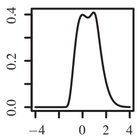
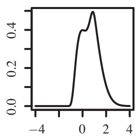
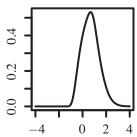
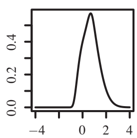
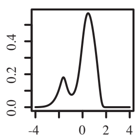
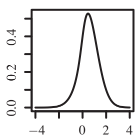
Table 2 shows the results obtained for some choices of the coefficients (α_1, α_3). A small plot of each density function is shown, as well as the proportion ($P1$) of samples for which rules R1 has been applied to obtain an estimate, that is, samples for which at least two roots of the estimating equation (8) exist.

By assuming the correct root to be the one closest to the true value, $\xi = 0$ in our simulation, we report the proportion ($P2$) of correct choices for each simulation as measure of appropriate selection. Mean and standard deviation of the sampling distributions of the parameters are also presented.

We evaluated, also, the performance of the procedure in situations not coherent with model (12). Therefore, a second set of simulations has been performed by considering as cumulative distribution function in (12) the Cauchy distribution rather than the normal one; these are reported in Table 2.

Another set of simulations has been produced in order to illustrate the performance of the procedure for the extension related to the generalized skew-t

TABLE 2: *Simulation study. For each sampling distribution we report the polynomials coefficients α_1 and α_3, a plot of the density function and the values of mean and standard error of the sampled random variable, the proportion of samples which yield at least two roots of the estimating equation ($P1$) and the percentage of estimates well selected among the roots ($P2$) , as well as the mean and standard deviation of the simulated distribution of the estimates for the location and scale parameters.*

$G(\cdot)$	(α_1, α_3)	Density	μ	σ	$P1$	$P2$	ξ		ω	
							Mean	SD	Mean	SD
Normal	(2, 0)		0.71	0.70	0.97	0.75	0.2	0.32	0.91	0.13
Normal	(0, 0.3)		0.36	0.93	0.76	0.49	−0.23	0.35	1.14	0.15
Normal	(0, 1)		0.55	0.71	0.59	0.63	0.07	0.42	1.04	0.12
Normal	(0, 2)		0.63	0.78	0.58	0.54	0.26	0.43	0.95	0.12
Normal	(1, 1)		0.67	0.74	0.83	0.71	0.22	0.35	0.92	0.13
Normal	(1, 2)		0.70	0.71	0.89	0.75	0.2	0.31	0.91	0.14
Normal	(2, −1)		0.05	1.00	1.00	0.87	0.19	0.46	1.08	0.22
Cauchy	(2, 0)		–	–	0.98	0.48	0.57	0.58	1.01	0.09

distribution presented in Section 3.3. We simulated 3000 samples of $n = 250$ values from a skew-t distribution with a variety of values for the skewness parameter ($\alpha = 0, 0.2, 0.5, 1, 3, 5, 7.5, 10$) and of the degree of freedom ($\nu = 4, 5, 10, 20$). Mean and standard deviation of the estimates of the parameters ξ and ω are plotted in Figure 3.

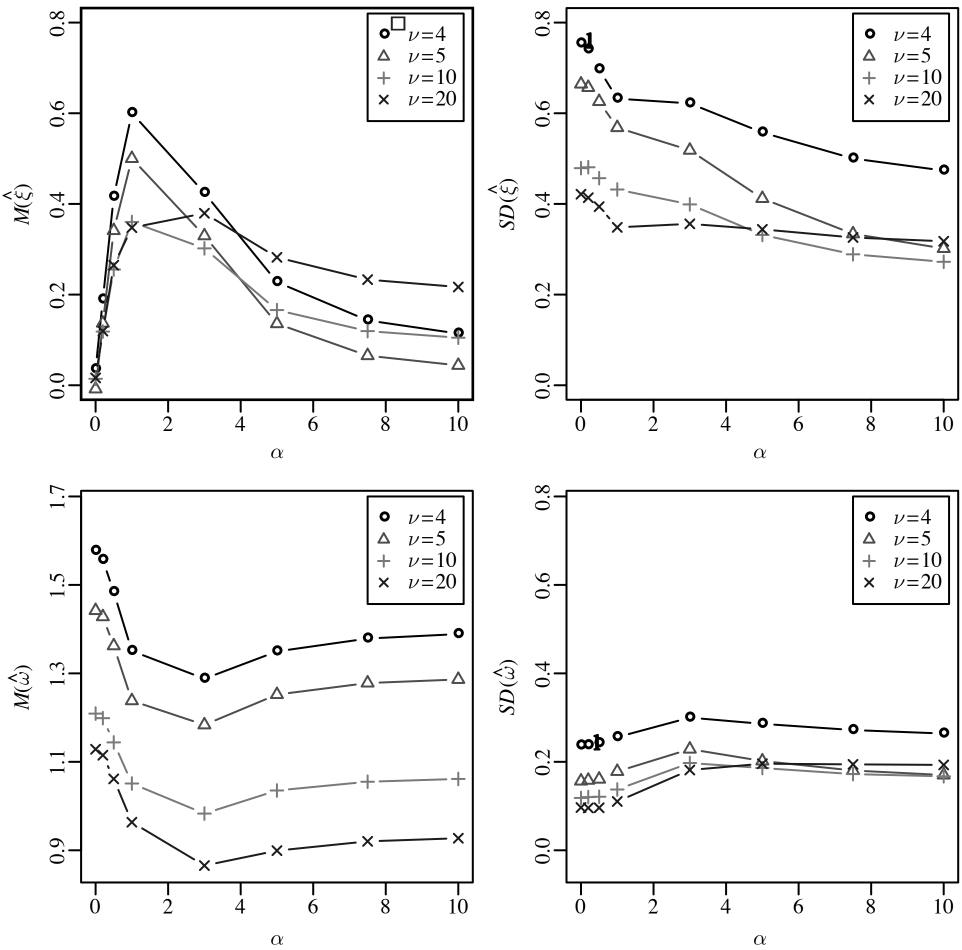


Figure 3. Simulation study. Plot of means and standard deviations of the estimates of ξ and ω obtained by applying the invariance-based estimating equations to simulated data from skew-t distributions with $\alpha = 0, 0.2, 0.5, 1, 3, 5, 7.5, 10$ and $\nu = 4, 5, 10, 20$.

An example of application of the method on real data has been performed by studying the relationship between lean body mass (LBM) and body mass index (BMI) for a subset of the athletes collected at the Australian Institute of Sport (AIS). The data have been examined by Cook & Weisberg (1994) and Azzalini

& Capitanio (1999), among others. Figure 4 displays the scatter plot of the data and the fit of the line obtained by estimating parameters using IBEE. This line is compared with the corresponding lines fitted by least squares regression and MLE. The IBEE and MLE lines are fairly similar.

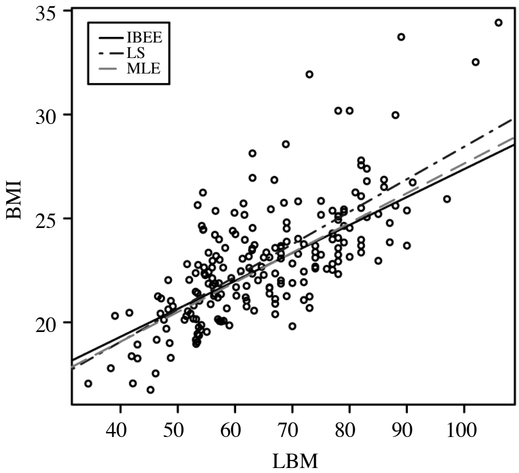


Figure 4. AIS data. Fitted lines obtained by using IBEE, least squares and MLE which assumes a SN error distribution.

It is also possible to include the generalized skew-t case in the regression setting (17)-(18). However, the actual implementation of the method in this case proved to be quite awkward, due to high multiplicity of the solutions of the estimating equations. Therefore we omit a detailed discussion of this case.

5. Discussion

The IBEE methodology aims at estimating the location and scale parameters (plus possibly others) of the base function f_0 in (1) avoiding a parametric specification of the perturbation factor $\pi(\cdot)$. The essential idea of the formulation is to construct pivotal quantities which exploit the distributional invariance of Proposition 1 to derive a set of estimating equations for the parameters of f_0.

From a technical viewpoint, the formulation is a special case of the RAL methodology of Ma *et al.* (2005). The IBEE proposal is however based on a simpler, more intuitive idea, and in this sense it can be regarded as having its own motivation behind it; on the other hand, IBEE does not attempt to obtain optimal estimating equations.

The analysis carried out in this paper has however shown that the more critical point of this sort of approach lies not so much in the search for the most efficient variant form of the estimating equations, but rather in the recurrent presence of multiple roots of these equations. In a typical situation, two roots exist, with the consequent problem of choosing one of these roots as the actual estimate.

The construction of a scheme for selecting the appropriate root of the estimating equations turned out to be a very challenging problem. This difficulty can be seen to be a direct consequence of the flexibility of the semiparametric scheme (1) when $\pi(\cdot)$ is treated nonparametrically. In such a case, there appear to exist two parameters sets (ξ, ω, π) effectively equivalent in representing f, a fact which leads to a sort of nearly non-identifiability situation. These remarks apply equally to the RAL formulation of Ma *et al.* (2005), where the question of selection of the appropriate root is only mentioned very briefly.

Provided the problem of choosing the root can be solved satisfactorily, the methodology appears to provide a viable approach, which allows to estimate the parameters of interest without specification of the nuisance component $\pi(\cdot)$, with limited loss of efficiency compared with the MLE associated to a correct parametric specification of $\pi(\cdot)$. The earlier part of the discussion supports the adoption of the rule based on the minimal value of (15). This choice has proved to work fairly well in a range of cases, although not universally. It is not surprising that more difficulties have emerged with the generalized skew-t distribution, since the increased flexibility produced by choosing the base f_0 equal to the t density instead of ϕ, combined with the maximal flexibility of π, exacerbates the problem described earlier of near indeterminacy of the solution.

Our overall conclusion is that the IBEE approach is worth to be considered when the parameters of interest are those of f_0 and one does not wish to specify the form of the nuisance component π. The approach must however be used with appropriate care, since a mindless application of the methodology can be seriously misleading. Even if we have put forward a criterion, (15), for choosing one of the existing roots, it is wise to inspect directly each of the candidate densities (14), and compare them with background information on the specific problem under consideration, as well as with the outcome from alternative approaches associated to parametric forms of $\pi(z)$. Elaborated forms of f_0 involving several parameters are not recommended.

Appendix: Proof of some properties of the curve $Q_\alpha(t)$

We study some properties of the function $Q(t)$ defined in (9) under the assumption that $Z \sim SN(0, 1, \alpha)$. For the first term of $Q(t)$, write

$$\mathbb{E}\{|Z - t|\} = \int_{-\infty}^{t} (t - x)\phi_\alpha(x)\mathrm{d}x + \int_{t}^{\infty} (x - t)\phi_\alpha(x)\mathrm{d}x,$$

and then make use of the expression given by Chiogna (1998) for the incomplete moments of a skew-normal variable. After some simple algebra, one arrives at

$$\mathbb{E}\{|Z-t|\} = \mu_\alpha - t + 2\{\phi_\alpha(t) + t\,\Phi_\alpha(t) - \mu_\alpha\,\Phi(t\sqrt{1+\alpha^2})\},$$

where μ_α and $\Phi_\alpha(\cdot)$ have been introduced in (10) and the subsequent passage. For the second term of $Q(t)$, the general fact

$$\mathbb{E}\left\{(Z-t)^2\right\} = \text{var}\{Z\} + (t - \mathbb{E}\{Z\})^2$$

applies, where in the present case the mean and variance are given in (10). On combining the above ingredients, one obtains $Q_\alpha(t)$ as given in (11).

As a by-product of these computations, one obtains the expression of the mean absolute deviation of a skew-normal variate, by setting $t = \mu_\alpha$ in $\mathbb{E}\{|Z-t|\}$. This leads to an explicit expression for the mean absolute deviation

$$\mathbb{E}\{|Z-\mu_\alpha|\} = 2\{\phi_\alpha(\mu_\alpha) + \mu_\alpha[\Phi_\alpha(\mu_\alpha) - \Phi(\mu_\alpha\sqrt{1+\alpha^2})]\},$$

which is a result of independent interest. Some numerical work has shown that an approximation to the mean absolute deviation is provided by

$$\mathbb{E}\{|Z-\mu_\alpha|\} \approx \sqrt{\frac{2}{\pi}}\,\sigma_\alpha,$$

a fact which is well-known to hold exactly for the normal distribution, when $\alpha = 0$.

We now want to obtain some properties of the function $Q_\alpha(t)$. Specifically we are interested in locating the zeroes of $Q_\alpha(t)$. It is immediate to recognise that one zero always exists at $t = 0$; the rest of the discussion aims at locating at least another zero.

It can be checked that $Q_{-\alpha}(-t) = Q_\alpha(t)$; this follows easily by evaluating (11) with reversed sign of α and t, and taking into account the property

$$1 - \Phi_\alpha(-t) = \Phi_{-\alpha}(t),$$

(Azzalini, 1985). Therefore, in the following discussion, there is no loss of generality in assuming that $\alpha \geq 0$.

The derivative of $Q_\alpha(t)$, taking into account that $\phi'_\alpha(t) = -t\,\phi_\alpha(t) + 2\alpha\phi(t)\phi(\alpha t)$, can be written as

$$Q'_\alpha(t) = 2\,\Phi_\alpha(t) - 1 - \sqrt{\frac{2}{\pi}}\frac{t-\mu_\alpha}{\sqrt{1-2t\,\mu_\alpha+t^2}},$$

which converges to $\pm(1-\sqrt{2/\pi})$ when $t \to \pm\infty$. Evaluation of this derivative at $t = 0$, with use of the property

$$\Phi_\alpha(0) = \frac{1}{2} - \frac{1}{\pi}\arctan\alpha,$$

leads to

$$h(\alpha) = Q'_\alpha(0) = \frac{2}{\pi}\left(\frac{\alpha}{\sqrt{1+\alpha^2}} - \arctan\alpha\right),$$

such that $h(0) = 0$ and

$$h'(\alpha) = \frac{2}{\pi(1+\alpha^2)}\left(\frac{1}{\sqrt{1+\alpha^2}} - 1\right) \le 0,$$

where the equality sign holds only for $\alpha = 0$. For $\alpha > 0$, $h'(\alpha) < 0$ ensures that $Q'_\alpha(0) < 0$ and hence $Q_\alpha(t) < 0$ for some $t > 0$, since $Q_\alpha(0) = 0$. This fact, combined with $Q_\alpha(t) \to \infty$ when $t \to \pm\infty$, leads to the conclusion that a second zero of $Q_\alpha(t)$ exists for some $t > 0$. Similarly, when $\alpha < 0$, a second root exists for some $t < 0$.

ACKNOWLEDGMENTS

We would like to thank an anonymous reviewer for interesting and constructive comments. This research work has been supported by MIUR, Italy, under grant PRIN 2006132978. The work of Genton is supported in part by US NSF grants DMS-0504896 and CMG ATM-0620624, and by a Swiss NSF grant.

REFERENCES

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

ARELLANO-VALLE, R. B. and GENTON, M. G. (2010) An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators, *Ann. Inst. Statist. Math.*, 62, 363–381.

AZZALINI, A. (1985) A class of distributions which includes the normal ones, *Scand. J. Statist.*, 12, 171–178.

AZZALINI, A. (1986) Further results on a class of distributions which includes the normal ones, *Statistica*, 46, 199–208.

AZZALINI, A. (2005) The skew-normal distribution and related multivariate families. With discussion by Marc G. Genton and a rejoinder by the author, *Scand. J. Statist.*, 32, 159–200.

AZZALINI, A. (2008) R package `sn`: The skew-normal and skew-*t* distributions (version 0.4–6), URL `http://azzalini.stat.unipd.it/SN`.

AZZALINI, A. and CAPITANIO, A. (1999) Statistical applications of the multivariate skew normal distribution, *J. Roy. Statist. Soc. Ser. B*, 61, 579–602. Full version of the paper at `http://arXiv.org` (No. 0911.2093).

AZZALINI, A. and CAPITANIO, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *J. Roy. Statist. Soc. Ser. B*, 65, 367–389. Full version of the paper at `http://arXiv.org` (No. 0911.2342).

AZZALINI, A. and DALLA VALLE, A. (1996) The multivariate skew-normal distribution, *Biometrika*, 83, 715–726.

AZZALINI, A. and GENTON, M. G. (2008) Robust likelihood methods based on the skew-t and related distributions, *Internat. Statist. Rev.*, 76 , 106–129.

CHIOGNA, M. (1998) Some results on the scalar skew-normal distribution, *J. Italian Statist. Soc.*, 1, 1–14.

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

COX, D. R. (1993) Unbiased estimating equations derived from statistics which are functions of the data, *Biometrika*, 80 , 905–909.

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

GENTON, M. G., HE, L. and LIU, X. (2001) Moments of skew-normal random vectors and their quadratic forms, *Statist. Probab. Lett.*, 51, 319–325.

GENTON, M. G. and LOPERFIDO, N. (2005) Generalized skew-elliptical distributions and their quadratic forms, *Ann. Inst. Statist. Math.*, 57, 389–401.

GODAMBE, V. P., editor (1991) *Estimating Functions*, Oxford University Press.

MA, Y. and GENTON, M. G. (2004) A flexible class of skew-symmetric distributions, *Scand. J. Statist.*, 31, 459–468.

MA, Y., GENTON, M. G. and TSIATIS, A. A. (2005) Locally efficient semiparametric estimators for generalized skew-elliptical distributions, *J. Amer. Statist. Assoc.*, 100, 980–989.

MA, Y. and HART, J. D. (2007) Constrained local likelihood estimators for semiparametric skew-normal distributions, *Biometrika*, 94, 119–134.

QU, A., LINDSAY, B. and LI, B. (2000) Improving generalised estimating equations using quadratic inference functions, *Biometrika*, 87, 823–836.

WANG, J., BOYER, J. and GENTON, M. G. (2004a) A skew-symmetric representation of multivariate distributions, *Statist. Sinica*, 14, 1259–1270.

WANG, J., BOYER, J. and GENTON, M. G. (2004b) A note on an equivalence between chi-square and generalized skew-normal distributions, *Statist. Probab. Lett.*, 66, 395–398.

ADELCHI AZZALINI
Dipartimento di Scienze Statistiche
Università di Padova
Via C. Battisti, 241-243
35121 Padova (Italia)
azzalini@stat.unipd.it

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

BRUNO SCARPA
Dipartimento di Scienze Statistiche
Università di Padova
Via C. Battisti, 241-243
35121 Padova (Italia)
scarpa@stat.unipd.it