Scandinavian Journal of Statistics Theory and Applications
doi: 10.1111/j.1467-9469.2012.00822.x

Characteristic Function-based Semiparametric Inference for Skew-symmetric Models

CORNELIS J. POTGIETER

Department of Statistical Science, Southern Methodist University

MARC G. GENTON

CEMSE Division, King Abdullah University of Science and Technology

ABSTRACT. Skew-symmetric models offer a very flexible class of distributions for modelling data. These distributions can also be viewed as selection models for the symmetric component of the specified skew-symmetric distribution. The estimation of the location and scale parameters corresponding to the symmetric component is considered here, with the symmetric component known. Emphasis is placed on using the empirical characteristic function to estimate these parameters. This is made possible by an invariance property of the skew-symmetric family of distributions, namely that even transformations of random variables that are skew-symmetric have a distribution only depending on the symmetric density. A distance metric between the real components of the empirical and true characteristic functions is minimized to obtain the estimators. The method is semiparametric, in that the symmetric component is specified, but the skewing function is assumed unknown. Furthermore, the methodology is extended to hypothesis testing. Two tests for a null hypothesis of specific parameter values are considered, as well as a test for the hypothesis that the symmetric component has a specific parametric form. A resampling algorithm is described for practical implementation of these tests. The outcomes of various numerical experiments are presented.

Key words: asymmetry, characteristic function, distributional invariance, hypothesis testing, semiparametric inference, skew-symmetric distribution

1. Introduction

In recent years, there has been a growing interest in flexible parametric classes of distributions. This has led to several generalizations of known distributions, for example the skew-normal distribution as a generalization of the normal, see Genton (2004) and Azzalini (2005) for an overview. Any symmetric density can be generalized to a class of distributions that capture a wide variety of shapes. This class of skew-symmetric distributions has probability density function of the form

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

where $f_0(z) = f_0(-z)$ for all $z \in \mathbb{R}$ is the symmetric base density function and the skewing function $\pi(z)$ satisfies $0 \le \pi(z) = 1 - \pi(-z) \le 1, z \in \mathbb{R}$. Location parameter $\xi \in \mathbb{R}$ and scale parameter $\omega > 0$ can be introduced through a simple linear transform which leads to the density

$$f_{SS}(x \mid \xi, \omega) = 2\omega^{-1} f_0\{\omega^{-1}(x-\xi)\}\pi\{\omega^{-1}(x-\xi)\}, \quad x \in \mathbb{R}. \tag{2}$$

The well-known case of the skew-normal distribution is obtained when $f_0(z) = \phi(z)$ and $\pi(z) = \Phi(\alpha z)$, where $\phi(z)$ is the standard normal density function, $\alpha \in \mathbb{R}$ is a parameter that controls the skewness and $\Phi(z)$ denotes the standard normal distribution function. This family, first proposed by Azzalini (1985), has been extensively studied. If we let $t(z \mid v)$ and $T(z \mid v)$ denote the Student's t density and distribution functions with v degrees of freedom, then the skew-t

distribution is obtained by taking $f_0(z)=t(z\mid \nu)$ and $\pi(z)=T[\alpha z\{(\nu+1)/(\nu+z^2)\}^{1/2}\mid \nu+1]$, see Azzalini & Genton (2008) for a recent account and references therein. When the skewing function $\pi(z)$ is equal to 1/2 for all $z\in\mathbb{R}$, the symmetric base density is recovered. A class of densities of the form (1) was considered by Azzalini & Capitanio (2003), with parameterization $\pi(z)=G\{w(z)\}$ where G is a univariate symmetric cumulative distribution function and $w(z)$ is an odd function. The term *skew-symmetric* is used for families of the form (1) (see e.g. Wang *et al.*, 2004). In the present study, we will refer to density functions of the form (1) with $f_0(z)=\phi(z)$ as generalized skew-normal, and those with $f_0(z)=t(z\mid \nu)$ as generalized skew-t (see Genton & Loperfido, 2005).

A skew-symmetric distribution can be viewed as a perturbation of a given symmetric distribution regulated by the skewing function π with the resulting skewed distribution still retaining several of the properties of its symmetric counterpart. The class of distributions in (1) can also be obtained by applying a suitable censoring mechanism (or selection scheme) to the base density f_0. Here, the censoring is regulated by the function π, which can be viewed as the probability of inclusion in the sample. This censoring approach has been studied in detail in Arellano-Valle *et al.* (2006). When adopting this perspective, one of the main questions of interest is estimation of the parameters (ξ,ω) in the base density f_0 in such a way that the skewing function π does not have to be known in order to do so. Semiparametric efficient estimators of the location and scale parameters were first considered by Ma *et al.* (2005) and Ma & Hart (2007). In addition, Frederic (2011) considered estimation of skew-symmetric distributions through the use of B-splines and penalty functions.

One very useful property of the skew-symmetric family is that of distributional invariance. Let Z_1 and Z_2 be two random variables, the first having density f_0 and the second having density (1). Let $T(\cdot)$ be an even function, that is, $T(-z)=T(z)$ for all $z\in\mathbb{R}$. Then,

$$T(Z_1)\overset{d}{=}T(Z_2). \tag{3}$$

The proof can be found in Wang *et al.* (2004). A method of parameter estimation based on this distributional invariance property of the skew-symmetric family was considered by Azzalini *et al.* (2010). What makes this invariance property so useful is that the expectations of all even functions of Z_2 can be calculated without requiring any knowledge of π. The method of Azzalini *et al.* (2010) is therefore essentially a method of moments approach, using the first and second absolute moments of Z_2 to construct estimating equations. The difficulty encountered there is not so much with the method of estimation, although it does happen that the pair of estimating equations does not have a zero. Rather, there are multiple roots and the problem becomes one of selecting the correct root. The root selection problem was also encountered by Ma *et al.* (2005), who considered the optimality aspects of invariance-based estimating equations.

Parameter estimation using characteristic functions has been considered in contexts other than skew-symmetric distributions. Feuerverger & McDunnough (1981) considered efficient parameter estimation where the type of distribution was completely specified and the characteristic function known. Koutrouvelis & Kellermeier (1981) performed both parameter estimation and goodness-of-fit tests using the empirical characteristic function. Yu (2004) also considered empirical likelihood estimation as an alternative to maximum likelihood with specific application to diffusion models. Xu & Knight (2011) used empirical characteristic functions to estimate parameters in distributions that are finite mixtures of normal distributions. Kim & Genton (2011) recently provided a comprehensive description of the characteristic functions of scale mixtures of skew-normal distributions, in particular for the skew-normal and skew-t distributions.

In the current study, we propose a method based on minimizing a criterion function based on the real part of the characteristic function. Formulation as a minimization problem has the advantage that at least one minimum point always exists. However, the selection problem is still encountered, in that there are typically two local minima. The global minimum is generally not well separated from the other local minimum. In addition to this, we have found through extensive simulations that choosing the root corresponding to the global minimum does not, in general, lead to selecting the root closest to the true parameter values. It should be noted that there is a similarity between our proposed method and an application of the generalized method of moments (GMM) when there is a continuum of moment conditions as in Carrasco & Florens (2000). Braun *et al.* (2008) also considered GMM estimation by way of empirical transforms, of which the empirical characteristic function is a special case.

The article is organized as follows. In section 2, we discuss the characteristic function-based estimation method and derive asymptotic properties of the estimators. The newly proposed estimators are compared with the invariant-based estimating equation (IBEE) estimators of Azzalini *et al.* (2010) in section 3, both asymptotically and in a finite sample setting. The root selection problem is addressed in section 4. In section 5, the new methodology is used to construct two tests for specific parameter values for (ξ, ω) and the power properties of these are explored through simulations. A test is also proposed for testing whether the symmetric component of the generalized skew-symmetric distribution is correctly specified. Section 6 is a numerical investigation of the proposed estimators when root selection is applied. Two data applications are also presented. The methodology extends to both the multivariate case and the regression setting. These extensions are described in the online supplemental material. The derivation of the covariance matrix of the proposed estimators can also be found in the supplemental material.

2. Parameter estimation with characteristic functions

2.1. The empirical characteristic function

The characteristic function plays a central role in statistics. Specifically, the characteristic function of a random variable X is the function $\Psi : \mathbb{R} \longrightarrow \mathbb{C}$ defined as

$$\Psi(t) = \mathrm{E}\{\exp(itX)\}, \quad t \in \mathbb{R},$$

where $i = \sqrt{-1}$. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables. The empirical characteristic function (ecf) is

$$\Psi_n(t) = n^{-1} \sum_{1 \le j \le n} \exp(itX_j) = n^{-1} \sum_{1 \le j \le n} \cos(tX_j) + in^{-1} \sum_{1 \le j \le n} \sin(tX_j).$$

This empirical estimator of the characteristic function has been extensively studied in the literature (see Csörgő, 1981; Marcus, 1981). The following two results follow from these. For any fixed value $t \in \mathbb{R}$, the random variable $n^{1/2}\{\Psi_n(t) - \Psi(t)\}$ converges to a zero-mean complex-valued normal random variable with variance $\Psi(t)\overline{\Psi(t)}$, where the bar is used to denote complex conjugate. It is also known that $n^{1/2}\{\Psi_n(t) - \Psi(t)\}$ converges as a stochastic process in t on any finite interval $[-A, A]$. The empirical characteristic function as defined does not include location and scale parameters, which can be introduced by defining functions

$$c_n(t \mid \xi, \omega) = n^{-1} \sum_{1 \le j \le n} \cos\{t(X_j - \xi)/\omega\}, \quad s_n(t \mid \xi, \omega) = n^{-1} \sum_{1 \le j \le n} \sin\{t(X_j - \xi)/\omega\}. \tag{4}$$

These are unbiased estimators of $c(t)$ and $s(t)$, the real and imaginary components of the characteristic function of $Z = (X - \xi)/\omega$.

2.2. Characteristic function-based semiparametric estimators

The real and imaginary components of the characteristic function of Z will be denoted by $c(t\,|\,\boldsymbol{\theta})$ and $s(t\,|\,\boldsymbol{\theta})$, where $\boldsymbol{\theta}\in\mathbb{R}^q$ represents unknown model parameters, in addition to the location and scale parameters, which may need to be estimated from the data. For example, in the class of generalized skew-t distributions, the degrees of freedom v may be unknown. As the empirical estimator $c_n(t\,|\,\xi,\omega)$ converges to $c(t\,|\,\boldsymbol{\theta})$ in both probability and mean square, the minimum distance criterion function

$$D_n(\xi,\omega,\boldsymbol{\theta})=\int_{\mathbb{R}}\{c_n(t\,|\,\xi,\omega)-c(t\,|\,\boldsymbol{\theta})\}^2 w(t)\,\mathrm{d}t \tag{5}$$

for $w(t)$ a symmetric weight function is proposed for the estimation of the location and scale parameters ξ and ω. We make use of an L_2 distance metric, but other metrics could also be employed. Beyond computational convenience associated with this metric, it also leads to the inequality $0\le D_n(\xi,\omega,\boldsymbol{\theta})\le 4\int_{\mathbb{R}}w(t)\,\mathrm{d}t$. The upper bound of the statistic depends on the weight function. In practice, the observed sample values are typically all rational, leading to an empirical characteristic function that is periodic in nature. When this is the case, the integral in (5) may not converge without an appropriate choice of weight function. When the parameter $\boldsymbol{\theta}$ is unknown, it may also be estimated by minimizing this criterion. At this point, we note that (5) is a function of only the real part of the characteristic function, as the skewing function π is assumed unknown. However, when a parametric model is assumed, this methodology can be extended and the minimization problem can be formulated in terms of both the real and imaginary parts of the characteristic function. We refer to this as the minimum distance characteristic function (MDCF) estimator.

The IBEE method of obtaining estimators proposed by Azzalini *et al.* (2010) typically uses the first two absolute moments of the underlying distribution. The proposed minimization method also has a connection to a method of moments. Assume that the underlying distribution has finite even moments of all order. The characteristic function therefore admits the expansion

$$c(t\,|\,\boldsymbol{\theta})=1+\sum_{k\ge 1}\frac{(-1)^k t^{2k}}{(2k)!}\mu_{2k}(\boldsymbol{\theta}) \tag{6}$$

with $\mu_k(\boldsymbol{\theta})=E\{(Z-\xi)^k/\omega^k\}$. The ecf also admits an expansion

$$c_n(t\,|\,\xi,\omega)=1+\sum_{k\ge 1}\frac{(-1)^k t^{2k}}{(2k)!}\hat{\mu}_{2k}(\xi,\omega) \tag{7}$$

with $\hat{\mu}_k(\xi,\omega)=n^{-1}\sum_{1\le j\le n}(X_j-\xi)^k/\omega^k$. Substitution of (6) and (7) into (5) and some straightforward calculations gives

$$D_n(\xi,\omega,\boldsymbol{\theta})=\sum_{j\ge 1}\sum_{k\ge 1}w^*_{jk}\{\mu_{2j}(\boldsymbol{\theta})-\hat{\mu}_{2j}(\xi,\omega)\}\{\mu_{2k}(\boldsymbol{\theta})-\hat{\mu}_{2k}(\xi,\omega)\}$$

with

$$w^*_{jk}=\frac{(-1)^{j+k}}{(2j)!(2k)!}\int_{\mathbb{R}}t^{2(j+k)}w(t)\,\mathrm{d}t.$$

We therefore see that there is an equivalent quadratic form involving the even moments of the underlying symmetric distribution. In the case where all even moments exist, the proposed MDCF method therefore does not use just two absolute population moments, but finds the estimator that minimize the weighted distance between all even moments and their empirical counterparts.

The minimization problem (5) is equivalent to solving the following equations upon setting them equal to zero,

$$\begin{aligned}\frac{\partial D_n(\xi,\omega,\boldsymbol{\theta})}{\partial\xi} &= 2\int_{\mathbb{R}}\{c_n(t\mid\xi,\omega)-c(t\mid\boldsymbol{\theta})\}\frac{\partial c_n(t\mid\xi,\omega)}{\partial\xi}w(t)\,\mathrm{d}t\\ &= \frac{2}{\omega}\int_{\mathbb{R}}\{c_n(t\mid\xi,\omega)-c(t\mid\boldsymbol{\theta})\}ts_n(t\mid\xi,\omega)w(t)\,\mathrm{d}t,\\ \frac{\partial D_n(\xi,\omega,\boldsymbol{\theta})}{\partial\omega} &= -\frac{2}{\omega}\int_{\mathbb{R}}\{c_n(t\mid\xi,\omega)-c(t\mid\boldsymbol{\theta})\}tc_n'(t\mid\xi,\omega)w(t)\,\mathrm{d}t,\\ \frac{\partial D_n(\xi,\omega,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= -2\int_{\mathbb{R}}\{c_n(t\mid\xi,\omega)-c(t\mid\boldsymbol{\theta})\}\frac{\partial c(t\mid\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}w(t)\,\mathrm{d}t.\end{aligned}$$

In the online supplemental material, the asymptotic covariance matrix of these minimum distance estimators is derived. The covariance matrix is of the form $\boldsymbol{\Gamma}=\omega^2\boldsymbol{\Sigma}(c(t\mid\boldsymbol{\theta}), s(t\mid\boldsymbol{\theta}), w(t))$ as it is proportional to the square of the scale parameter and is a functional of the real and imaginary components of the characteristic function of Z, as well as the weight function $w(t)$. No closed-form expression for this covariance matrix is available, but it can easily be computed numerically for all of the generalized skew-symmetric distributions considered here. In practice, the functional form of $c(t\mid\boldsymbol{\theta})$ is known, as it depends only on the symmetric density function f_0. On the other hand, the function $s(t\mid\boldsymbol{\theta})$ is not known, as it also depends on the unknown skewing function π. Once parameter estimates $\hat{\xi}$ and $\hat{\omega}$ have been obtained, the latter function can be estimated by $s_n(t\mid\hat{\xi},\hat{\omega})$ from (4). The estimated covariance matrix is then

$$\hat{\boldsymbol{\Gamma}}=\hat{\omega}^2\boldsymbol{\Sigma}(c(t\mid\hat{\boldsymbol{\theta}}), s_n(t\mid\hat{\xi},\hat{\omega}), w(t)). \tag{8}$$

The last issue to be addressed is the choice of the weight function. The optimal minimum variance choice of the weight function depends on the underlying distribution, which is not known in practice. If one specifies a fully parametric form for the underlying distribution, for instance that it belongs to the class of skew-normal distributions, it is possible to find the best possible weight function within a family of weight functions, for instance $w_1(t\mid\beta)=\exp(-\beta^2t^2)$, $t\in\mathbb{R}$ or $w_2(t\mid\beta)=(1-t^2)^\beta$, $|t|<1$ and $\beta>0$. However, all our numerical work suggests the choice of the parameter β within the family of weight functions is more important than the weight function itself. We illustrate in Fig. 1 the choice of the parameter β in $w_2(t\mid\beta)$ for minimizing $\det(\boldsymbol{\Gamma})$ when the underlying distribution is skew-normal with shape parameter α. Denote this covariance matrix by $\boldsymbol{\Gamma}_{\alpha,\beta}$ to indicate its dependence on the skewness parameter α and the weight function parameter β. The top panel of Fig. 1 shows $\arg\min_\beta\{\det(\boldsymbol{\Gamma}_{\alpha,\beta})\}$ as a function of $0.8\le\alpha\le 5$. The bottom panel reports $\min_\beta\{\det(\boldsymbol{\Gamma}_{\alpha,\beta})\}^{1/2}$ as well as the asymptotic standard deviations of the location and scale parameters for the same choice of β. For $0<\alpha\le 0.8$, the values are easily computed numerically, but because of the scale the standard deviations do not display well.

In practice, we do not assume a fully parametric form for the underlying family. In this instance, one can proceed as follows: Find initial estimators $(\hat{\xi}_0, \hat{\omega}_0, \hat{\boldsymbol{\theta}}_0)$ and plug these into (8). Call this estimator $\hat{\boldsymbol{\Gamma}}_0$ and note that it is still a function of the weight function. For a given family of weight functions, one can then find the value of β that minimizes $\det(\hat{\boldsymbol{\Gamma}}_0)$, say β_0. Use the minimizer β_0 to update the estimators. The performance of this estimator will be evaluated in future work.

Using the general theory of M-estimators, one can easily show that the minimization method estimators are both consistent and asymptotically normal – see for example sections 5.2 and 5.3 of van der Vaart (2000). One should note that there is a pathological scenario in which the choice of weight function may affect the consistency of the estimators. Specifically, assume

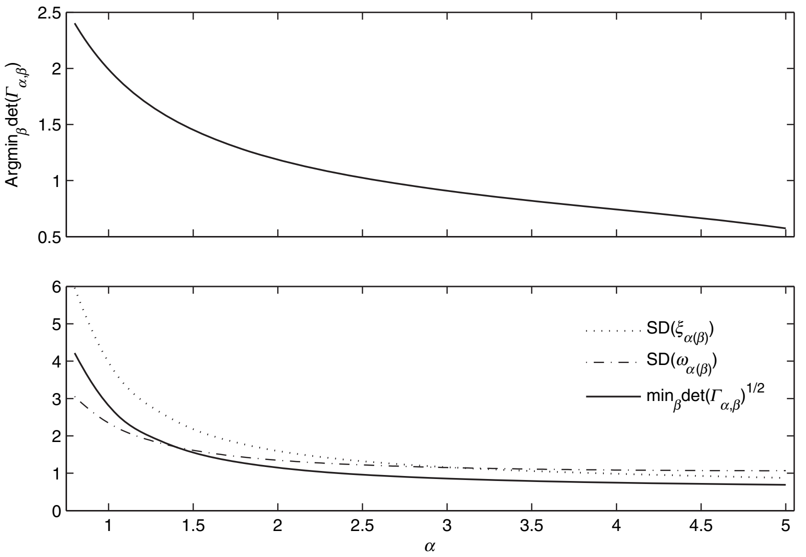

Fig. 1. Top panel: Optimal choice of β for an underlying SN(α) distribution with $0.8 \leq \alpha \leq 5$. Bottom panel: Minimum standard deviations obtained by optimal choice of β.

that the characteristic function of the true symmetric component is $c_0(t)$, but that the characteristic function is incorrectly specified to be $c_1(t)$. Also assume that the chosen weight function is bounded on some interval $[-A, A]$. When we have $\sup_{t\in[-A,A]} |c_0(t) - c_1(t)| = 0$, in other words the true characteristic function and incorrectly specified characteristic function agree on the interval $[-A, A]$, consistency of the estimators obtained is suspect.

2.3. *A hybrid estimator*

Consider for a moment a scenario where the skewing function π is known and maximum likelihood is used to obtain the location and scale estimators. The log-likelihood function for a sample of size n from a skew-symmetric distribution is

$$\ell(\boldsymbol{\theta} \mid X_1, \ldots, X_n) = n \log 2 - n \log \omega + \sum_{1\leq j\leq n} \log f_0\left(\frac{X_j - \xi}{\omega} \,\middle|\, \boldsymbol{\theta}\right) + \sum_{1\leq j\leq n} \log \pi\left(\frac{X_j - \xi}{\omega}\right).$$

The partial derivative with respect to ω is

$$\frac{\partial \ell}{\partial \omega} = -\frac{n}{\omega} + \frac{1}{\omega}\sum_{1\leq j\leq n}\left(\frac{X_j - \xi}{\omega}\right)\frac{f_0'\{(X_j-\xi)/\omega \mid \boldsymbol{\theta}\}}{f_0\{(X_j-\xi)/\omega \mid \boldsymbol{\theta}\}} + \frac{1}{\omega}\sum_{1\leq j\leq n}\left(\frac{X_j - \xi}{\omega}\right)\frac{\pi'\{(X_j-\xi)/\omega\}}{\pi\{(X_j-\xi)/\omega\}}.$$

This can be written as

$$\frac{\partial \ell}{\partial \omega} = -\frac{n}{\omega} + \frac{1}{\omega}\sum_{1\leq j\leq n} k_1\left(\frac{X_j - \xi}{\omega}\right) + \frac{1}{\omega}\sum_{1\leq j\leq n} k_2\left(\frac{X_j - \xi}{\omega}\right)$$

where $k_1(z \mid \boldsymbol{\theta}) = z f_0'(z \mid \boldsymbol{\theta})/f_0(z|\boldsymbol{\theta})$ is a symmetric function and $k_2(z) = z\pi'(z)/\pi(z)$. Using the invariance property (3), we have $\mathrm{E}\{k_1(Z \mid \boldsymbol{\theta})\} = 1$ for any random variable Z from a skew-symmetric distribution with symmetric component f_0. This also corresponds to the likelihood equation in the case of an underlying symmetric distribution. Moreover, since $\mathrm{E}\{k_2(Z)\} = 0$, one can argue that k_2 contributes little to the likelihood estimation in the asymmetric case.

The solution obtained for ω by solving the sample version of $\mathrm{E}\{k_1(Z\,|\,\boldsymbol{\theta})-1\}=0$ is

$$n^{-1}\sum_{1\le j\le n} k_1\left(\frac{X_j-\xi}{\omega}\,\middle|\,\boldsymbol{\theta}\right)-1=0 \tag{9}$$

and will be close to the likelihood estimator $\hat{\omega}$ (as a function of ξ and $\boldsymbol{\theta}$). It seems intuitive that this even function k_1 found using likelihood principles may give a more efficient estimate of the scale parameter than that found using other arbitrary even functions. However, we can only find such a function for the scale parameter, and not for the location parameter. Depending on the underlying symmetric model, it may also be possible to use likelihood principles to find even functions for estimating some subset of $\boldsymbol{\theta}$.

Using this, we propose a hybrid method for estimating the model parameters. Let $\hat{\omega}(\xi,\boldsymbol{\theta})$ denote the estimator of ω obtained by solving (9) as a function of ξ and $\boldsymbol{\theta}$. Now, find the remaining parameters by minimizing

$$D_n^*(\xi,\boldsymbol{\theta})=\int_{\mathbb{R}}[c_n\{t\,|\,\xi,\hat{\omega}(\xi,\boldsymbol{\theta})\}-c(t\,|\,\boldsymbol{\theta})]^2 w(t)\,\mathrm{d}t.$$

One of the advantages of following this approach is that the dimension of the problem has been reduced, as we only have to find the minimum in terms of $q+1$, rather than $q+2$ parameters. The asymptotic behaviour of these hybrid estimators is equivalent to the asymptotic behaviour of the following equations,

$$\begin{aligned}
0&=\frac{2}{\omega}\int_{\mathbb{R}}\left\{\cos\left(t\cdot\frac{X-\xi}{\omega}\right)-c(t\,|\,\boldsymbol{\theta})\right\}ts(t\,|\,\boldsymbol{\theta})w(t)\,\mathrm{d}t,\\
0&=-\frac{1}{\omega}+\frac{1}{\omega}\left(\frac{X-\xi}{\omega}\,\middle|\,\boldsymbol{\theta}\right)\frac{f_0'\{(X-\xi)/\omega\,|\,\boldsymbol{\theta}\}}{f_0\{(X-\xi)/\omega\,|\,\boldsymbol{\theta}\}},\\
0&=-2\int_{\mathbb{R}}\left\{\cos\left(t\cdot\frac{X-\xi}{\omega}\right)-c(t\,|\,\boldsymbol{\theta})\right\}\frac{\partial c(t\,|\,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}w(t)\,\mathrm{d}t.
\end{aligned}$$

As this development is very similar to that of the minimum distance method, outlined in the online supplemental material, details are omitted. Extensive simulation results, some of which are reported in section 6 of this paper, suggest that the hybrid method of estimation results in estimators that perform better in a mean squared error sense than those obtained by minimizing (5).

3. Comparing estimators

We present here a comparison of the different estimators in an asymptotic sense. Numerical results comparing the finite-sample performance of estimators can be found in section 6. The IBEE method of finding estimators proposed by Azzalini *et al.* (2010) is based on solving estimating equations of the type

$$n^{-1}\sum_{1\le j\le n} T_k\left(\frac{X_j-\xi}{\omega}\right)-c_k=0$$

with T_k typically being of the form $T_k(z)=|z|^k$ and $c_k=\mathrm{E}\{T_k(Z)\}$, $k=1,2,\ldots$. In order to effectively compare this method to the proposed characteristic function-based method, we consider the special case where the underlying distribution is skew-normal with skewness parameter $\delta=\alpha/(1+\alpha^2)^{1/2}\in[-1,1]$. In this instance, define

$$\mathbf{B}_1 = \begin{pmatrix} (2/\pi)\ \arctan(\delta/\sqrt{1-\delta^2}) & (2/\pi)^{1/2}\omega^{-1} \\ (8/\pi)^{1/2}\delta\omega & 2\omega \end{pmatrix} \quad \text{and} \quad \mathbf{B}_2 = \begin{pmatrix} 1-2/\pi & \omega^2\sqrt{2/\pi} \\ \omega^2\sqrt{2/\pi} & 2\omega^4 \end{pmatrix}$$

from which the asymptotic covariance matrix of the IBEE estimators can be calculated, $\mathbf{\Psi} = \mathbf{B}_1^{-1}\mathbf{B}_2\mathbf{B}_1^{-1}$ with determinant

$$\det(\mathbf{\Psi}) = \frac{1}{8}\frac{\omega^4\pi(\pi-3)}{\{\delta - \omega\ \arctan(\delta/\sqrt{1-\delta^2})\}^2}. \tag{10}$$

Here, $\det(\mathbf{\Psi}) \to \infty$ as $\delta \to 0$, illustrating how the method fails when the underlying distribution is symmetric. This is not just a problem that occurs when using the IBEE approach. The proposed method based on empirical characteristic functions suffers from the same drawback as $\delta \to 0$. In fact, this is a special case of (2) in Ma & Hart (2007) that shows there is no semiparametric-efficient estimator of the location parameter in a generalized skew-symmetric distribution when $\pi(z) = 1/2$ for all $z \in \mathbb{R}$. When multiple parameters are being estimated, the determinant of the covariance matrix is a good measure of overall performance. It is of interest to note that the determinant here is inherently a function of the scale parameter ω and not just proportional to ω^4 as is the determinant of the covariance matrix of the characteristic function-based semiparametric estimators. When ω is close to 0, we have $\det(\mathbf{\Psi}) \approx \omega^4\pi(\pi-3)/(8\delta^2)$, while for ω large, we have $\det(\mathbf{\Psi}) \approx \omega^4\pi(\pi-3)/[8\{\arctan(\delta/\sqrt{1-\delta^2})\}^2]$. In Fig. 2, we plot the determinant of $\mathbf{\Psi}$ from the IBEE method as a function of δ, as well as the determinant of $\mathbf{\Gamma}$ from the MDCF method (5) for two choices of weight function, $w_1(t) = \exp(-t^2/2)$, $t \in \mathbb{R}$ and $w_2(t) = 1 - t^2$, $|t| \leq 1$. These specific weight functions were chosen for illustration purposes only. The methods are compared by considering the log-differences of the determinants of the

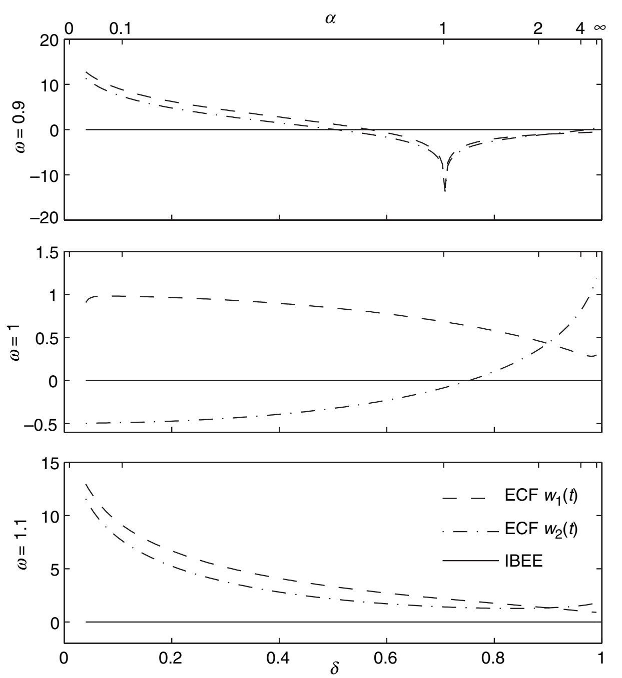

Fig. 2. Difference of the log-determinants of the covariance matrices $\mathbf{\Gamma}$ (for the IBEE method) and $\Psi(w_j)$(for the MDCF method with weight functions $w_1(t) = \exp\left(-t^2/2\right)$ and $w_2(t) = (1-t^2)I(|t| \leq 1)$). Top panel: $\omega = 0.9$; middle panel: $\omega = 1$; bottom panel: $\omega = 1.1$.

covariance matrices. The IBEE method is used as a benchmark. When a curve falls below the benchmark line, it indicates better performance, whereas a curve above the benchmark indicates poorer performance. While the asymptotic covariance matrix for the MDCF method only depends on the scale parameter ω in that it is a constant of proportionality, with the IBEE method ω features intrinsically in the covariance matrix of the estimators.

We note that there is no estimator which consistently shows the best performance. For $\omega = 0.9$, the MDCF method outperforms the IBEE method for moderate to large values of the skewness parameter δ. There is an anomalous peaked minimum which is a result of (10) being an intrinsic function of ω. Similar behaviour can be seen for other choices of $\omega < 1$. When $\omega = 1$, there is very little difference between the IBEE and MDCF estimators. For $\omega = 1.1$, the IBEE method outperforms the MDCF method. In general, when the underlying distribution is skew-normal, the IBEE estimator performs better when $\omega > 1$, while the characteristic function method yields better results over a large range of δ-values when $\omega < 1$.

4. Root selection

A difficulty when estimating (ξ, ω) is the occurrence of multiple solutions. This was observed by both Ma *et al.* (2005) and Azzalini *et al.* (2010). In the latter study, particular attention was paid to the generalized skew-normal distribution and a way of selecting the correct root when the underlying distribution is skew-normal was proposed. The authors also discussed an approach which was based on the complexity of the models associated with each of the roots. This method gave mixed results. Their approach, as well as the new implied skewness method proposed below, is based on estimation of the skewing function π. Let $(\hat{\xi}, \hat{\omega}, \hat{\boldsymbol{\theta}})$ be one of the possible estimators obtained from a sample $X_1, \ldots, X_n$. Define the standardized values $\tilde{Z}_j = (X_j - \hat{\xi})/\hat{\omega}$ and kernel estimator

$$\tilde{f}(z) = (nh)^{-1} \sum_{1 \le j \le n} K\left(\frac{z - \tilde{Z}_j}{h}\right)$$

for K a symmetric density function, from which an estimator for π is $\tilde{\pi}(z) = \tilde{f}(z)/\{2f_0(z \mid \hat{\boldsymbol{\theta}})\}$. It should be noted that while this is a consistent estimator of $\pi(z)$ for any given $z \in \mathbb{R}$, it does not satisfy the constraints imposed on the skewing function, namely that $\pi(z) = 1 - \pi(-z) \le 1$. One can therefore improve upon the above estimator by defining $\hat{\pi}(z) = \tilde{\pi}(z)/\{\tilde{\pi}(z) + \tilde{\pi}(-z)\}$, which does satisfy said constraints. The method of choosing between estimators suggested by Azzalini *et al.* (2010) was based on a measure of the complexity of the estimated π-functions. They favoured ‘simpler’ models over more ‘complex’ ones. A typical measure of complexity is

$$C(\hat{\pi}) = \int_{\mathbb{R}} \{\hat{\pi}''(z)\}^2 \, \mathrm{d}z, \tag{11}$$

and when confronted with competing estimators $(\hat{\xi}_1, \hat{\omega}_1, \hat{\pi}_1)$ and $(\hat{\xi}_2, \hat{\omega}_2, \hat{\pi}_2)$, one would choose the first triple if $C(\hat{\pi}_1) \le C(\hat{\pi}_2)$ and the second otherwise.

We propose selection based on comparing the sample skewness to the implied skewness associated with a particular π-function. For an estimated π-function $\hat{\pi}_j$, the implied mean is

$$\hat{\mu}_{j,1} = 2\int_{\mathbb{R}} z f_0(z \mid \hat{\boldsymbol{\theta}}_j)\hat{\pi}_j(z)\, \mathrm{d}z$$

and the implied kth central moment is

$$\hat{\mu}_{j,k} = 2\int_{\mathbb{R}} (z - \hat{\mu}_{j,1})^k f_0(z \mid \hat{\boldsymbol{\theta}}_j)\hat{\pi}_j(z)\, \mathrm{d}z, \quad k \ge 2.$$

The implied skewness associated with $\hat{\pi}_j$ is therefore $\hat{\kappa}_j = \hat{\mu}_{j,3}/\hat{\mu}_{j,2}^{3/2}$. Let $\hat{\kappa}$ denote the observed sample skewness. The solution for which the implied skewness is closest to the observed sample skewness is taken as the estimator.

The suggested root selection algorithm is therefore as follows:

(i) If there is only one root $(\hat{\xi}, \hat{\omega}, \hat{\boldsymbol{\theta}})$, it is taken as the estimate.
(ii) When there are competing estimates, calculate $d_j = |\hat{\kappa}_j - \hat{\kappa}|$ and select the estimate with smallest d_j.

We investigate the performance of this root selection algorithm in our numerical studies in section 6.

5. Hypothesis testing

5.1. Two tests

We consider tests of the hypothesis $H_0 : (\xi, \omega) = (\xi_0, \omega_0)$ with the parameter $\boldsymbol{\theta}$ in the model known. We will also discuss how the assumption of $\boldsymbol{\theta}$ known can be relaxed in testing. In the present development, as the latter parameter is assumed known, it is suppressed in the notation of this section. The statistic D_n in (5) being minimized to obtain parameter estimates $(\hat{\xi}, \hat{\omega})$ is based on the process $P_n(t) = n^{1/2}\{c_n(t \mid \xi_0, \omega_0) - c(t)\}$ where ξ_0 and ω_0 are the true parameter values. By the multivariate central limit theorem, the process $P_n(t)$ is asymptotically Gaussian with covariance function $K(t_1, t_2) = c(t_2 - t_1)/2 + c(t_2 + t_1)/2 - c(t_1)c(t_2)$. Our interest here is in the limiting process $P(t) = \lim_{n\to\infty} P_n(t)$, multiplied by the weight function $w^{1/2}(t)$, which has orthogonal decomposition

$$w^{1/2}(t)P(t) = \sum_{j\geq 1} Z_j e_j(t)$$

by the Karhunen–Loève representation theorem, see Grenander (1981, chapter 1.4, theorem 2).

Here, the Z_j are independent standard normal random variables and the functions $e_j(t)$ are continuous real-valued functions that are pairwise orthogonal in L_2. The functions $e_j(t)$ are the eigenfunctions of the kernel function $w^{1/2}(t_1)w^{1/2}(t_2)K(t_1, t_2)$. We then have

$$\begin{aligned} n \cdot D_n(\xi_0, \omega_0) &= n \cdot \int_{\mathbb{R}} w(t)\{c_n(t \mid \xi_0, \omega_0) - c(t)\}^2 \,\mathrm{d}t \\ &\longrightarrow \int_{\mathbb{R}} w(t)P^2(t)\,\mathrm{d}t = \int_{\mathbb{R}} \left\{\sum_{j\geq 1} Z_j e_j(t)\right\}^2 \mathrm{d}t = \sum_{j\geq 1} \lambda_j Z_j^2 \end{aligned}$$

where

$$\lambda_j = \int_{\mathbb{R}} e_j^2(t)\,\mathrm{d}t$$

and it is assumed without loss of generality that $\lambda_1 > \lambda_2 > \cdots > 0$. Asymptotically, $n \cdot D_n(\xi_0, \omega_0)$ is therefore distributed as an infinite sum of weighted χ_1^2 random variables. In practice, this is usually approximated by the sum of the first M components, where M is chosen in such a way that $\sum_{j\geq M+1} \lambda_j$ is small. While there are no closed-form expressions for the eigenvalues, these can be computed numerically if one wishes to use the asymptotic distribution of the test. Numerical calculation of the eigenvalues may be costly, but methodology as outlined in Matsui & Takemura (2008) is applicable. In practice, one may wish to rely on a bootstrap procedure for implementation of the test.

An alternative test is found by using the estimators obtained by minimizing $D_n(\xi, \omega)$. The asymptotic covariance matrix of these estimators is of the form $\omega^2 \boldsymbol{\Sigma}(c(t), s(t), w(t))$, where the present notation is used to indicate that the covariance is a functional of $c(t)$, $s(t)$ and $w(t)$. This can be approximated by replacing the unknown function $s(t)$ with its empirical counterpart $s_n(t \mid \hat{\hat{\xi}}, \hat{\omega})$, and substituting $\hat{\omega}$ for ω. The test statistic

$$S_n = \left(\frac{\hat{\hat{\xi}} - \xi_0}{\hat{\omega}}, 1 - \frac{\omega_0}{\hat{\omega}} \right) \boldsymbol{\Sigma}^{-1}(c(t), s_n(t \mid \hat{\hat{\xi}}, \hat{\omega}), w(t)) \left(\frac{\hat{\hat{\xi}} - \xi_0}{\hat{\omega}}, 1 - \frac{\omega_0}{\hat{\omega}} \right)^{\top}$$

has an asymptotic χ^2_2 distribution. We therefore have two different tests for inference about specific parameter values. The first of these is not asymptotically distribution-free, but does not require any knowledge of the skewing function in the underlying distribution. The second is asymptotically distribution-free, but requires an estimate of the imaginary component of the characteristic function.

5.2. Power comparison

A small simulation study was performed to compare the power of the two proposed test statistics. Data were generated from a distribution of the form (2) with the symmetric component taken to be standard normal and the skewing function $\pi(z) = \Phi(\alpha_1 z + \alpha_3 z^3)$, $\alpha_1, \alpha_3 \in \mathbb{R}$. The three different parameter specifications were $(\alpha_1, \alpha_3) \in \{(2, 0); (0, 2); (2, -1)\}$. The critical values of the null distribution were obtained using Monte Carlo resampling with sample size taken to be $n = 100$. The weight function used for the characteristic function-based semiparametric estimators was $w(t) = 1 - t^2$, $|t| \leq 1$. Thereafter, samples were drawn from distributions under the alternatives $H_1 : \xi = \xi_1$ and $H_1 : \omega = \omega_1$. The estimated power is taken to be the proportion of test statistics under the alternative that exceed the critical value under the null distribution, with $M = 1000$ samples taken under the alternative. The results are shown in Fig. 3. As can be seen, the statistic D_n consistently performs better than S_n, with the

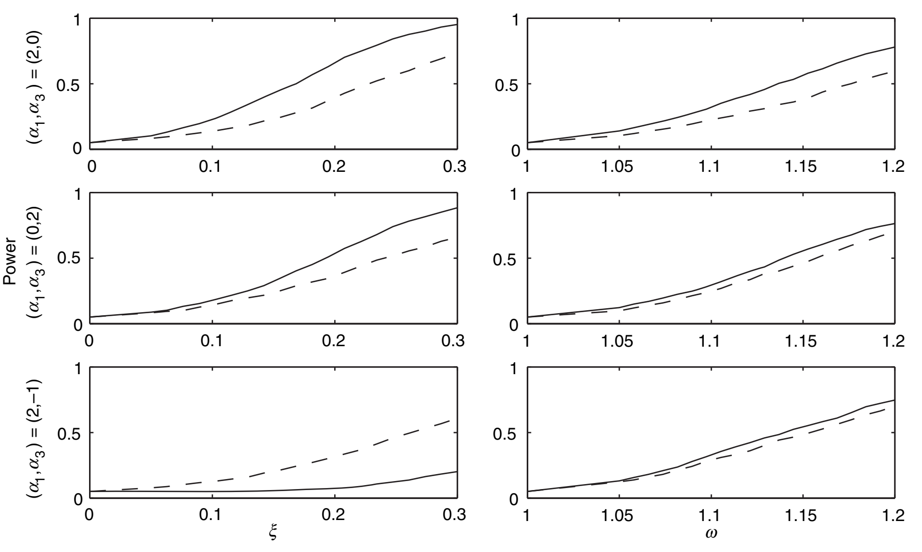

Fig. 3. Estimated power curves for location shift (left column) and scale change (right column). The solid curve represents the statistic D_n and the dashed curve represents the statistics S_n. The power was estimated at the 5% level of significance using Monte Carlo sampling.

exception of a location-shift alternative for the model with $(\alpha_1, \alpha_3)=(2, -1)$. This parameter specification corresponds to a bimodal distribution, which may explain the different performance under this scenario. In this instance, the statistic S_n performs much better than D_n.

5.3. Other tests

In the previous section, the parameter $\boldsymbol{\theta}$ was assumed known. This may not be realistic. It is still possible to test the hypothesis about specific values of ξ and ω. Let ξ_0 and ω_0 denote the parameter values under the null hypothesis, and let $\hat{\boldsymbol{\theta}}_0$ be the minimizer of

$$D_n^+(\boldsymbol{\theta} \mid \xi_0, \omega_0) = \int_{\mathbb{R}} \{c_n(t \mid \xi_0, \omega_0) - c(t \mid \boldsymbol{\theta})\}^2 w(t)\,\mathrm{d}t.$$

One can use $D_n^+(\hat{\boldsymbol{\theta}}_0 \mid \xi_0, \omega_0)$ to test $H_0: (\xi, \omega) = (\xi_0, \omega_0)$ when the parameter $\boldsymbol{\theta}$ is unknown.

Another hypothesis of interest is whether the symmetric component of the distribution has been correctly specified, for example $H_0: f_0(z) = \phi(z)$, where $\phi(z)$ is the normal density function, or $H_0: f_0(z) = t(z \mid v)$ with $t(z \mid v)$ the Student's t density with degrees of freedom v. Here, v can be assumed known, or can be estimated from the data. To test this hypothesis, the parameter estimates $(\hat{\xi}, \hat{\omega})$ are plugged into the statistic D_n in (5). The argument for using this as test statistic is as follows: when the null hypothesis is true, the real component of the characteristic function $c(t \mid \boldsymbol{\theta})$ will be close to the empirical $c_n(t \mid \hat{\xi}, \hat{\omega})$. The statistic $D_n(\hat{\xi}, \hat{\omega})$ measures this distance.

While asymptotic theory can be developed for the test statistics here, it is of greater practical value to find a bootstrap approach. First, standardize the sample $Z_j = (X_j - \xi)/\omega$, $j = 1, \ldots, n$. The standardization is performed using either the null values ξ_0 and ω_0, or the estimates $\hat{\xi}$ and $\hat{\omega}$, with the appropriate choice depending on the hypothesis being considered. Next, using the Z-values, define $\hat{\pi}(z) = \tilde{f}(z)/\{\tilde{f}(z) + \tilde{f}(-z)\}$ where $\tilde{f}$ is a kernel estimator of the standardized sample. To sample from a skew-symmetric distribution that is 'close' to the sample, proceed as follows. Generate a value Y from the symmetric density $f_0(z)$, calculate the value $p = \hat{\pi}(Y)$ and include the value Y in the sample with probability p. The bootstrap sample obtained in this way comes from a skew-symmetric distribution with specified symmetric component and skewing component closely matching that of the sample. Theoretical properties of a similar test are considered in Jiménez–Gamero (2012).

6. Numerical studies

6.1. Small-sample comparison

To compare the IBEE method of Azzalini *et al.* (2010) with the new minimum distance method, a simulation study was performed. For each sample generated, the estimators were calculated using three different methods: the IBEE estimators, the characteristic function-based semiparametric estimators and the hybrid estimators. In the case of multiple roots, which will be more thoroughly addressed in the next section, the pair $(\hat{\xi}, \hat{\omega})$ closest (in L_2 norm) to the true value $(\xi, \omega) = (0, 1)$ was selected. This was done to compare the methods in the case where the correct root is always selected. Figure 4 shows the square-root of the determinant of the covariance matrices which was estimated from the pairs $(\hat{\xi}_j, \hat{\omega}_j)$, $j = 1, \ldots, M$ for $M = 1000$. This was done for four different underlying distributions, all of the generalized skew-t form, with degrees of freedom taken to be $v = 3, 4, 5$ and ∞, respectively, and $\delta = \alpha/(1 + \alpha^2)^{1/2} \in [-1, 1]$. The characteristic function and hybrid estimators were obtained using the weight function $w(t) = \exp(-t^2/10)$. In all instances, the sample size was $n = 250$.

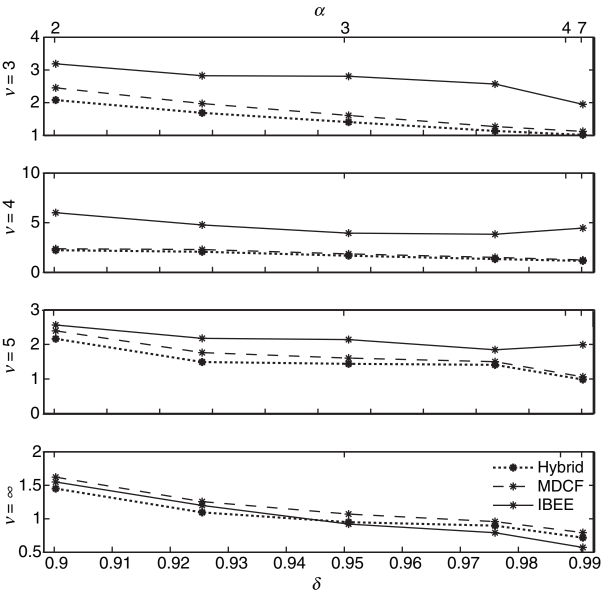

Fig. 4. Logarithm of the determinant of the covariance matrices of three estimators (IBEE, MDCF and hybrid estimators) obtained using Monte Carlo simulation for various skew-t distributions with degrees of freedom ν.

As can be seen in Fig. 4, both the characteristic function and the hybrid methods outperform the IBEE method for the skew-t distributions with heavy tails ($\nu = 3, 4, 5$). The hybrid method also performs consistently better than the characteristic function method. In the skew-normal case ($\nu = \infty$), mixed results are obtained. For $0.9 < \delta < 0.95$ ($2.06 < \alpha < 3.04$), the hybrid method has the best performance, while for $0.95 < \delta < 0.99$ ($3.04 < \alpha < 7.02$) the IBEE method performs best. In the case of the skew-normal distribution, however, the difference in performance is relatively very small.

A second simulation study was done to compare Monte Carlo standard errors of the estimators for the different methods for various choices of skewing functions in the generalized skew-normal case. Samples were generated from distributions with skewing function $\pi(z) = \Phi(\alpha_1 z + \alpha_3 z^3)$, see Ma & Genton (2004) for the effect of various choices of (α_1, α_3). In the simulations, the sample size was $n = 250$.

In Table 1, we present the Monte Carlo root mean square errors (RMSE) of the location and scale estimators,

$$\mathrm{RMSE} = M^{-1/2} \sqrt{\sum_{1 \le m \le M} (\hat{\gamma}_m - \gamma_{\mathrm{true}})^2} \tag{12}$$

where $\hat{\gamma}$ and γ_{true} are, respectively, the estimator and true value of the parameter of interest. We also present the determinant of the matrix

$$\hat{\boldsymbol{\Omega}} = M^{-1} \sum_{1 \le m \le M} (\hat{\xi}_m - \xi, \hat{\omega}_m - \omega)^\top (\hat{\xi}_m - \xi, \hat{\omega}_m - \omega)$$

as a global measure of performance. The minimum values in each column are highlighted in **bold**. Upon inspection of Table 1, it is clear that when estimators are compared in terms of the marginal performance only, that is, one compares the standard errors of individual estimators, the hybrid and minimum distance estimators of location $\hat{\xi}$ have smaller RMSE than

Table 1. *Monte Carlo standard errors of parameter estimates for samples of size $n=250$ from various generalized skew-normal distributions with skewing function $\pi(z)=\Phi(\alpha_1 z+\alpha_3 z^3)$. The three methods are invariance-based estimating equations (IBEE), minimum distance characteristic function (MDCF) and the hybrid estimators.*

		$n^{1/2}\widehat{\text{RMSE}}(\hat{\xi})$			$n^{1/2}\widehat{\text{RMSE}}(\hat{\omega})$			$\widehat{\det}^{1/2}(n\cdot\hat{\mathbf{\Omega}})$		
(α_1,α_3)	Density	IBEE	MDCF	Hybrid	IBEE	MDCF	Hybrid	IBEE	MDCF	Hybrid
(2, 0)		2.21	2.33	**2.14**	**1.58**	1.72	1.59	1.63	1.79	**1.52**
(0, 0.3)		3.13	2.95	**2.85**	1.43	1.51	**1.33**	2.77	2.90	**2.37**
(0, 1)		3.34	**2.31**	2.34	1.66	1.76	**1.47**	3.58	2.08	**1.82**
(0, 2)		3.53	2.43	**2.40**	1.93	1.78	**1.61**	3.26	1.93	**1.67**
(1, 1)		3.05	2.92	**2.82**	1.86	1.86	**1.77**	2.43	2.48	**2.25**
(1, 2)		2.64	2.29	**2.10**	1.73	1.61	**1.54**	1.90	1.82	**1.49**
(2, − 1)		0.84	**0.76**	1.01	**0.74**	0.79	0.81	0.62	**0.60**	0.78

the IBEE estimator, with the hybrid estimator having smallest RMSE in most of the scenarios considered. When comparing the estimated RMSE for $\hat{\omega}$, the hybrid estimator still performs best in most scenarios, with the exception of the model configurations $(\alpha_1,\alpha_3)=(2,0)$ and $(2,-1)$ in which case the IBEE estimator has smallest RMSE. When comparing the determinants of the matrix $\hat{\mathbf{\Omega}}$, the hybrid method outperforms the IBEE and minimum distance methods in six of the seven scenarios reported, the exception being $(\alpha_1,\alpha_3)=(2,0)$ where the minimum distance method performs best. While these results only represent a small fraction of the possible distributions that could be encountered in practice, it is encouraging to note the good performance of the newly proposed estimators.

6.2. Simulations

Samples of size $n=250$ were generated from various skew-symmetric distributions of the form (2) with $f_0(z)$ taken to be the normal density function $\phi(z)$ and the t-density $t(z\,|\,v)$. In each instance, estimators of $(\hat{\xi},\hat{\omega})$ were obtained using the hybrid characteristic function method. Below are reported the results from a simulation study wherein the proposed root selection algorithms of section 4 are investigated. The two root selection methods, model complexity and implied skewness, were both implemented and a comparison is provided. The hybrid method of parameter estimation was used with weight function $w(t)=1-t^2, |t|<1$. In Tables 2 and 3, P_{M} represents the proportion of samples in which multiple roots were obtained and the selection criterion had to be applied. P_{C} represents the proportion of times that the root closest to the true parameter value was selected when using model complexity as criterion, given that there were multiple solutions to choose from. Similarly, P_{S} represent the proportion of times the closest root was selected when using implied skewness as selection criterion. The estimates resulting from the two different selection mechanisms are not only compared in terms of the selection proportions above, but also in terms of the RMSE (12) based on $M=1000$ Monte Carlo samples. For each specification, there are three different RMSE values, the column 'True' contains the Monte Carlo estimate of RMSE when the solution closest to the true value is always selected, the column 'Comp' corresponds to using model complexity as criterion, and the column 'Skew' corresponds to using skewness as selection criterion.

Table 2 provides a summary of the simulation results in the generalized skew-normal case with $\pi(z)=\Phi(\alpha_1 z+\alpha_3 z^3)$. When comparing the two proposed methods for selecting between competing solutions, there is not a clearly preferred approach. Both model complexity and implied skewness have instances where they show the best performance in terms of selecting

Table 2. *Investigating root selection for generalized skew-normal distributions. In the table, P_M denotes the propotion of samples in which multiple solutions were obtained and selection had to be applied, P_C denotes the proportion of samples in which model complexity gave the correct solution and P_S denotes the proportion of samples in which implied skewness gave the correct solution.*

				RMSE($\hat{\xi}$)			RMSE($\hat{\omega}$)		
(α_1, α_3)	P_M	P_C	P_S	True	Comp	Skew	True	Comp	Skew
(2, 0)	0.90	0.90	0.68	0.18	0.29	0.41	0.12	0.14	0.19
(0, 0.3)	0.70	0.31	0.72	0.21	0.46	0.28	0.07	0.16	0.13
(0, 1)	0.81	0.48	0.73	0.20	0.51	0.26	0.10	0.11	0.11
(0, 2)	0.72	0.72	0.88	0.21	0.38	0.26	0.11	0.13	0.12
(1, 1)	0.81	0.91	0.87	0.20	0.28	0.28	0.12	0.13	0.14
(1, 2)	0.84	0.92	0.83	0.19	0.27	0.30	0.12	0.14	0.16
(2, −1)	0.98	0.89	0.76	0.11	0.39	0.62	0.06	0.19	0.31

Table 3. *Investigating root selection for generalized skew-t distributions. For the estimated degrees of freedom $\hat{v}$, we provide the median as measure of location, difference between the* 65th *and* 35th *quantiles as measure of spread. We also report the estimated probability of $\hat{v}=\infty$.*

				RMSE($\hat{\xi}$)			RMSE($\hat{\omega}$)		
(α_1, α_3)	P_M	P_C	P_S	True	Comp	Skew	True	Comp	Skew
(2, 0)	1.00	0.74	0.70	0.47	0.67	0.60	0.44	0.52	0.40
(0, 0.3)	1.00	0.71	0.61	0.53	0.62	0.65	0.22	0.52	0.52
(0, 1)	1.00	0.86	0.72	0.53	0.62	0.58	0.36	0.52	0.45
(0, 2)	1.00	0.78	0.77	0.41	0.58	0.54	0.37	0.47	0.41
(1, 1)	1.00	0.57	0.74	0.35	0.61	0.55	0.40	0.42	0.32
(1, 2)	1.00	0.47	0.74	0.30	0.71	0.58	0.36	0.57	0.43
(2, −1)	1.00	0.92	0.76	0.16	0.32	0.71	0.13	0.19	0.44
	$\hat{v}$								
	Median			$Q_{0.65}-Q_{0.35}$			$\hat{P}(\hat{v}=\infty)$		
(α_1, α_3)	True	Comp	Skew	True	Comp	Skew	True	Comp	Skew
(2, 0)	6.96	7.04	8.14	8.31	14.77	6.10	0.32	0.30	0.24
(0, 0.3)	6.97	7.68	8.78	4.61	5.41	4.40	0.21	0.27	0.25
(0, 1)	7.08	7.66	8.34	6.33	9.03	7.05	0.27	0.31	0.27
(0, 2)	7.33	7.13	7.72	4.71	5.16	4.78	0.27	0.29	0.28
(1, 1)	8.24	6.66	7.59	10.90	5.21	4.12	0.30	0.27	0.22
(1, 2)	7.42	5.88	6.64	5.89	5.56	3.24	0.27	0.26	0.17
(2, −1)	5.22	5.46	6.36	2.19	2.26	2.56	0.08	0.12	0.06

the solution closest to the true parameter values. The proportion of times that implied skewness selects the correct solution is consistently between 0.7 and 0.8, while the same proportion varies greatly for model complexity. In one instance, when $(\alpha_1, \alpha_3)=(0, 0.3)$, model complexity only selects the correct solution in 31% of the samples. On the other hand, when $(\alpha_1, \alpha_3)=(1, 2)$, model complexity selects the correct solution in 92% of the samples. As an alternative to comparing the selection mechanisms in terms of how often they select the correct solution, the methods can be compared in terms of RMSE as defined in (12). Of the seven parameter configurations considered, model complexity results in smaller RMSE values three times, implied skewness results in smaller RMSE three times, and in one instance their performance is virtually identical.

The same type of simulation was done for a generalized skew-*t* distribution with skewing function $\pi(z)=T\left[(\alpha_1 z+\alpha_3 z^3)\{(v+1)/(v+z^2)\}^{1/2}\middle| v+1\right]$. The degrees of freedom was also treated as an unknown parameter which needed to be estimated. This was done for $v=5$, and results are presented in Table 3.

When the degrees of freedom are also estimated as a parameter in the model, the multiple solution problem grows in scale. When only location and scale parameters are estimated, most samples have two potential solutions to which some selection criteria has to be applied. When degrees of freedom are also estimated, most samples result in either three or four possible solutions. This highlights the importance of having an effective way of selecting the 'best' solution. In Table 3, it appears that model complexity (P_C) tends to select the true solution more often than using implied skewness (P_S) as a selection criterion. However, when comparing the RMSE for the solutions selected, using model skewness consistently leads to a lower RMSE for both $\hat{\xi}$ and $\hat{\omega}$ than using model complexity. Based on this assessment, implied skewness may be the preferred selection method. Assessing the performance when it comes to estimating the degrees of freedom is more difficult. In the simulations, data were generated from a model with $v=5$ degrees of freedom. The distribution of the estimator tends to be very skewed to the right and often the estimated value of v is infinity. The proportion of times this happens is reported in the table. Because of this, RMSE cannot be used to compare the different methods. Also in Table 3, the median of the simulations, as well as a measure of spread, the difference between the 65th and 35th quantiles, is reported. This measure of spread is used because even the interquartile range is equal to infinity for some of the parameter specifications. In most of the scenarios considered, using implied skewness leads to a larger median bias when compared with model complexity. However, the estimator of v resulting from implied skewness has a much smaller spread than that selected using model complexity. The proportion of samples in which $\hat{v}$ is equal to infinity ranges between 0.2 and 0.3 in most of the scenarios considered.

6.3. The frontier data

The frontier data, a simulated data set of size $n=50$ from a skew-normal distribution with $(\xi, \omega, \alpha)=(0, 1, 5)$ and first reported in Azzalini & Capitanio (1999), has become quite infamous in the literature as being an example of a data set that presents difficulties when estimating the parameters. The profile likelihood function of the shape parameter α is unbounded, leading to maximum likelihood estimate $\hat{\alpha}=\infty$. We applied both the characteristic function minimum distance method and the hybrid method to the frontier data to estimating the location and scale parameters, each with both $w_1(t)=\exp(-t^2)$ and $w_2(t)=1-t^2, |t|<1$. The skewness method for choosing between competing solutions was implemented. Also, while plug-in estimates of the covariance matrices can be obtained from (8), we recognize that this does not take into account the additional variability introduced by performing root selection. We therefore implement the semiparametric bootstrap with $B=1000$ and using implied skewness to choose between competing solutions to estimate the covariance matrices. We also use these same bootstrap replicates to construct marginal 95% bootstrap confidence intervals for the location and scale parameters. In Table 4, we report for each method and choice of weight function the estimated location and scale parameters $\hat{\xi}$ and $\hat{\omega}$, the bootstrap covariance matrix $\hat{\boldsymbol{\Gamma}}^*$, the determinant of the covariance matrix as an overall measure of variability, the 95% confidence intervals and the lengths of these intervals, and also the estimated shape parameter if one assumes a skew-normal distribution, but use the MDCF or hybrid estimates of location and scale substituted in the likelihood function. The estimated skew-normal density corresponding to $(\hat{\xi}, \hat{\omega}, \hat{\alpha})=\left(0.0003, 1.155, 7.548\right)$ is shown in Fig. 5, along with the true density and a histogram of the frontier data. When testing the hypothesis that f_0, the symmetric component, is normal, the bootstrap p-value (resulting from $B=1000$ boot-

Table 4. *Summary of parameter estimates for MDCF and hybrid parameter estimates for the frontier data*

Weight	$w_1(t)=\exp(-t^2)$		$w_2(t)=(1-t^2)I(t^2\leq 1)$	
Method	MDCF	Hybrid	MDCF	Hybrid
$\hat{\xi}$	−0.113	−0.098	0.0003	−0.060
$\hat{\omega}$	1.254	1.236	1.155	1.206
$50\hat{\boldsymbol{\Gamma}}^*$	$\begin{bmatrix} 10.29 & -4.10 \\ -4.10 & 2.82 \end{bmatrix}$	$\begin{bmatrix} 9.33 & -3.90 \\ -3.90 & 2.79 \end{bmatrix}$	$\begin{bmatrix} 9.98 & -4.37 \\ -4.37 & 2.79 \end{bmatrix}$	$\begin{bmatrix} 9.98 & -4.37 \\ -4.37 & 2.79 \end{bmatrix}$
$\det(50\hat{\boldsymbol{\Gamma}}^*)$	12.150	10.825	10.860	12.909
95% CI ξ	[−0.429, 1.282]	[−0.414, 1.213]	[−0.309, 1.292]	[−0.449, 1.257]
95% CI ω	[0.716, 1.572]	[0.699, 1.551]	[0.633, 1.477]	[0.656, 564]
length CI ξ	1.712	1.628	1.601	1.706
length CI ω	0.856	0.852	0.844	0.908
$\hat{\alpha}$	1,028.750	87.129	7.548	15.032

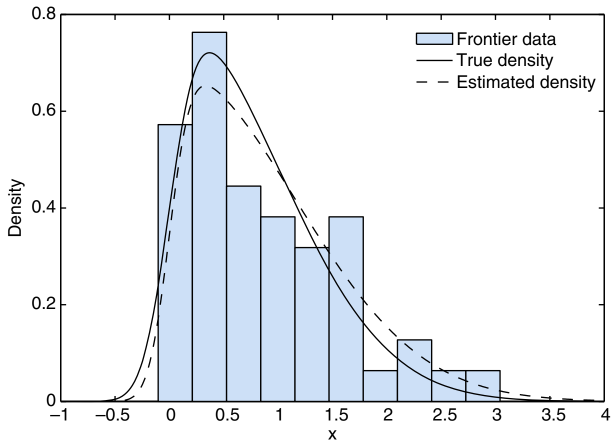

Fig. 5. A histogram of the frontier data, along with the true skew-normal density with $(\xi,\omega,\alpha)=(0,1,5)$ and the estimated density using the hybrid method with $w(t)=1-t^2$, which has $(\hat{\xi},\hat{\omega},\hat{\alpha})=(0.0003, 1.155, 7.548)$.

strap samples) is 0.320. On the other hand, when testing the hypothesis that f_0 is Cauchy, the bootstrap p-value is only 0.018.

6.4. Ambulatory expenditure data

A practical example where a distribution of type (2) arises is illustrated by an ambulatory expenditure data from the 2001 Medical Expenditure Panel Survey analysed by Cameron and Trivedi (2010). The decision to spend is assumed to be related to the spending amount, hence the observations form a selected sample. Cameron and Trivedi (2010) considered a sample-selection model based on the assumption of normality, hence leading to a parametric skew-normal distribution with $f_0(z)=\phi(z)$ and $\pi(z)=\Phi(\alpha z)$.

The data consist of 2802 observations, considered on a logarithmic scale for this analysis since the variable, ambulatory expenditure, is strictly positive. The observed data $Y_i=\log(\text{Ambex}_i)$, $i=1,\ldots,2802$, has mean 6.555, standard deviation 1.411 and skewness coefficient −0.341. When fitting a generalized skew-normal model using the characteristic function method, there are two possible solutions for the location and scale parameters ξ and ω, namely $(\hat{\xi}_1, \hat{\omega}_1) = (6.547, 1.409)$ and $(\hat{\xi}_2, \hat{\omega}_2) = (7.816, 1.891)$. The first of these solutions is nearly identical to the mean and standard deviation of the data. However, both model complexity and implied skewness result in the selection of the second solution. If one uses the hybrid characteristic function method which partially incorporates the likelihood function, the selected solution is $(7.823, 1.897)$, which is almost the same as the standard solution. For the standard solution, the plug-in estimate of covariance is

$$\hat{\mathbf{\Gamma}} = \frac{1}{2802} \begin{bmatrix} 22.112 & 14.717 \\ 14.717 & 11.584 \end{bmatrix}.$$

If one fits a skew-normal model to the data, the maximum likelihood estimates are $(\hat{\xi}, \hat{\omega}, \hat{\alpha}) = (7.907, 1.954, -1.730)$. Inverting the estimated information matrix, the estimated covariance matrix of these parameter estimates is

$$\hat{\mathbf{\Gamma}}_{ML} = \frac{1}{2802} \begin{bmatrix} 11.507 & 7.949 & -23.045 \\ 7.949 & 7.400 & -17.609 \\ -23.045 & -17.609 & 57.338 \end{bmatrix}.$$

As a measure of efficiency, we consider the square root of the ratio of the determinants of the covariance matrices for the location and scale parameters only. This ratio, when taking that of the characteristic function-based semiparametric estimators relative to the maximum likelihood estimators, is 1.342. This indicates that there is an approximate 34% loss in efficiency when the true model is, in fact, skew-normal, but the estimation is done for a generalized skew-normal model. If one alternatively decides to fit a generalized skew-t distribution, there are three possible solutions. Using model complexity as criterion, the selected solution is $(\hat{\xi}, \hat{\omega}, \hat{\nu}) = (7.541, 1.655, 26.491)$. In this instance, implied skewness does not give the same solution, which is $(\hat{\xi}, \hat{\omega}, \hat{\nu}) = (7.342, 1.528, 19.083)$. It should be noted that these two solutions have estimates of location and scale that are very similar.

7. Discussion

This study proposed a method of estimating location and scale parameters in skew-symmetric distributions assuming a specific parametric form for the symmetric base density f_0, but without knowledge of the skewing function π. The method of estimation relies on both the invariance property (3) of even functions of skew-symmetric random variables, as well as the characteristic function of said random variables.

The characteristic function methodology was compared with the IBEE approach suggested by Azzalini *et al.* (2010). In many instances, specifically when the underlying symmetric component is heavy-tailed, the new estimators performed better than the IBEE estimators, while they were comparable in the generalized skew-normal case.

A question that arises both in the context of the IBEE estimators, as well as the characteristic function-based semiparametric estimators, is that of selecting between competing solutions. Azzalini *et al.* (2010) suggested using model complexity, while an approach using implied skewness was proposed in this study. These approaches were compared through Monte Carlo simulations in both the generalized skew-normal and generalized skew-t scenarios. Both approaches performed well under certain conditions. It would appear that implied skewness consistently leads to selecting the appropriate root between 70% and 80% of the time. While model complexity performed much better for some of the model configurations, it also performed far worse in other instances.

The methodology developed here extends very naturally to both the multivariate skew-symmetric setting and also to a regression setting where the errors are assumed to come from a generalized skew-symmetric distribution. These extensions are explored in greater detail in the supplemental material available online.

Acknowledgements

This research was supported by NSF grants DMS-1007504 and DMS-0914951, and also by Award No. KUS-C1-016-04 made by King Abdullah University of Science and Technology (KAUST).

Supporting Information

Additional information for this article is available online, including discussions of how the methodology can be extended to the multivariate setting and to a regression framework where the noise follows a generalized skew-symmetric distribution. The derivation of the asymptotic covariance matrix of the semiparametric estimators is also presented.

References

Arellano-Valle, R. B., Branco, M. D. & Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *Can. J. Statist.* **34**, 581–601.

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

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Statist.* **32**, 159–188.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *J. Roy. Statist. Soc. Ser. B* **61**, 579–602.

Azzalini, A. & 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.

Azzalini, A. & Genton, M. G. (2008). Robust likelihood methods based on the skew-*t* and related distributions. *Inter. Statist. Rev.* **76**, 106–129.

Azzalini, A., Genton, M. G. & Scarpa, B. (2010). Invariance-based estimating equations for skew-symmetric distributions. *Metron*, **68**, 353–377, Special issue on 'Skew-symmetric and flexible distributions'.

Braun, M. P., Meintanis, S. G. & Melas, V. B. (2008). Optimal design approach to GMM estimation of parameters based on empirical transforms. *Inter. Statist. Rev.* **76**, 387–400.

Cameron, A. C. & Trivedi, P. K. (2010). *Microeconometrics using stata*, Revised edition. Stata Press, College Station TX.

Carrasco, M. & Florens, J. (2000). Generalization of GMM to a continuum of moment conditions. *Econometr. Theor.* **16**, 797–834.

Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Prob.* **9**, 130–144.

Feuerverger, A. & McDunnough, P. (1981). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Ser. B* **43**, 20–27.

Frederic, P. (2011). Modeling skew-symmetric distributions using B-spline and penalties. *J. Statist. Plann. Inference* **141**, 2878–2890.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*, Edited Volume, Chapman & Hall CRC, Boca Raton, FL.

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

Grenander, U. (1981). *Abstract inference*. John Wiley and Sons, New York.

Jiménez-Gamero, M. D. (2012). *Testing for the symmetric component of skew-symmetric distributions*. (forthcoming).

Kim, H.-M. & Genton, M. G. (2011). Characteristic functions of scale mixtures of multivariate skew-normal distributions. *J. Multi. Anal.* **102**, 1105–1117.

Koutrouvelis, I. A. & Kellermeier, L. (1981). A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated. *J. Roy. Statist. Soc. Ser. B* **43**, 173–176.

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

Ma, Y., Genton, M. G. & Tsiatis, A. A. (2005). Locally efficient semiparametric estimators for generalized skew-elliptical distributions. *J. Amer. Statist. Assoc.* **100**, 980–989.

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

Marcus, M. B. (1981). Weak convergence of the empirical characteristic function. *Ann. Prob.* **9**, 194–201.

Matsui, M. & Takemura, A. (2008). Goodness-of-fit tests for symmetric stable distributions – Empirical characteristic function approach. *TEST* **17**, 546–566.

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press, Cambridge.

Wang, J., Boyer, J. & Genton, M. G. (2004). A skew-symmetric representation of multivariate distributions. *Statist. Sinica* **14**, 1259–1270.

Xu, D. & Knight, J. (2011). Continuous empirical characteristic function estimation of mixtures of normal parameters. *Econometr. Rev.* **30**, 25–50.

Yu, J. (2004). Empirical characteristic function estimation and its applications. *Econometr. Rev.* **23**, 93–123.

Received July 2011, in final form August 2012

Cornelis J. Potgieter, Department of Statistical Science, Southern Methodist University, PO Box 750332, Dallas, TX 75275-0332, USA.
E-mail: cpotgieter@smu.edu