Ann. Inst. Statist. Math.
Vol. 48, No. 4, 687–709 (1996)

ESTIMATION OF A MULTIVARIATE BOX-COX TRANSFORMATION TO ELLIPTICAL SYMMETRY VIA THE EMPIRICAL CHARACTERISTIC FUNCTION*

ADOLFO J. QUIROZ[1], MIGUEL NAKAMURA[2] AND FRANCISCO J. PÉREZ[2]

[1] *Universidad Simón Bolívar, Departamento de Matemáticas, Apartado Postal 89000, Valle de Sartenejas, Baruta, Caracas, Venezuela*
[2] *Centro de Investigación en Matemáticas, Apartado Postal 402, Guanajuato, Gto. 36000, Mexico*

(Received February 6, 1995; revised October 16, 1995)

Abstract. Let $\boldsymbol{X} = (X_1, X_2, \dots, X_d)^t$ be a random vector of positive entries, such that for some $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_d)^t$, the vector $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ defined by $X_i^{(\lambda_i)} = (X_i^{\lambda_i} - 1)/\lambda_i$, $i = 1, \dots, d$ is elliptically symmetric. We describe a procedure based on the multivariate empirical characteristic function for estimating the λ_i's. Asymptotic results regarding consistency of the estimators are given and we evaluate their performance in simulated data. In a one-dimensional setting, comparisons are made with other available transformations to symmetry.

Key words and phrases: Elliptically contoured distributions, empirical characteristic function, Box-Cox transformations.

1. Introduction

The family of elliptically contoured distributions is a natural semi-parametric generalization of the multivariate Gaussian distribution. There exists a large literature on the subject of elliptically contoured distributions. For an overview, we refer the reader to the book edited by Fang and Anderson (1990). See also Cambanis *et al.* (1981), Devlin *et al.* (1976), and Fang *et al.* (1990). Among other things, a theory for the distribution of correlation coefficients, sample covariance matrix, Hotelling T^2, and other important statistics, that parallel the classical theory for the multivariate normal, is available for the family of elliptically contoured distributions. This, and the well known fact that many statistical procedures yield superior performance when data supports elliptical symmetry (see Nelson *et al.* (1989)), motivates the consideration of transformations to elliptical symmetry (instead of more general forms of multivariate symmetry).

* Adolfo Quiroz and Miguel Nakamura's research was partially supported by CONACYT (Mexico) grants numbers 1858E9219 and 4224E9405, while Dr. Quiroz was visiting Centro de Investigación en Matemáticas at Guanajuato, Mexico.

Here we consider the problem of transforming a vector of positive entries $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^t$ to elliptical symmetry via a multivariate version of the Box-Cox transformation family. For positive X and real λ let

$$X^{(\lambda)} = \begin{cases} (X^\lambda - 1)/\lambda, & \text{for } \lambda \neq 0, \\ \log(X), & \text{for } \lambda = 0. \end{cases} \tag{1.1}$$

For $\boldsymbol{X}$ as above and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_d)^t \in R^d$, let

$$\boldsymbol{X}^{(\boldsymbol{\lambda})} = (X_1^{(\lambda_1)}, X_2^{(\lambda_2)}, \ldots, X_d^{(\lambda_d)})^t. \tag{1.2}$$

We will assume that for some $\boldsymbol{\lambda}_0 \in R^d$, $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$ has a density $f = f_{\boldsymbol{\lambda}_0}$ (with respect to Lebesgue measure, as usual) which is elliptically symmetric, that is, for some real positive definite symmetric matrix A and vector $\boldsymbol{\mu} \in R^d$, and for some function $g : [0, \infty) \to [0, \infty)$, f is of the form

$$f(x) = g(\|A(x - \boldsymbol{\mu})\|). \tag{1.3}$$

If $E\|\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}\|^2$ exists, then the matrix A can be taken to be (by redefining g) the square root inverse of the covariance matrix Σ of $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$, and $\boldsymbol{\mu}$ will correspond to the mean of $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$.

The case $d = 1$ has been addressed extensively in the literature. Estimation of the transformation parameter may be achieved by techniques based on the notion of residual skewness or other, more general measures of one-dimensional symmetry (Taylor (1985)), minimum distance ideas (Nakamura and Ruppert (1990)), or by comparing residual means and medians (Hinkley (1977)). In the present paper, we discuss a method which is based on a characterization of elliptical symmetry through characteristic functions, which is defined from the outset for general d.

Maximum likelihood for the transformation parameter when f is the multivariate normal density, is considered in the work of Andrews *et al.* (1971), and Velilla (1993). The latter work compares maximum likelihood estimation and marginal estimation of each λ_i obtained by treating each coordinate separately as a one-dimensional problem, showing that the efficiency loss in the second method may be substantial when the components of $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$ are correlated. The method which we propose is non-parametric in that f is not assumed to be multivariate normal or even known, and is truly multidimensional in the sense that no marginalization takes place in estimating the vector $\boldsymbol{\lambda}$. The method is described as follows.

Given an i.i.d. sample $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from the distribution of $\boldsymbol{X}$ and $\boldsymbol{\lambda} \in R^d$, let

$$\bar{\boldsymbol{X}}_{\boldsymbol{\lambda}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^{(\boldsymbol{\lambda})}, \quad \text{and} \quad S_{\boldsymbol{\lambda}} = \frac{1}{n-1}\sum_{i=1}^n (\boldsymbol{X}_i^{(\boldsymbol{\lambda})} - \bar{\boldsymbol{X}}_{\boldsymbol{\lambda}})(\boldsymbol{X}_i^{(\boldsymbol{\lambda})} - \bar{\boldsymbol{X}}_{\boldsymbol{\lambda}})^t. \tag{1.4}$$

Also let

$$e_j(\boldsymbol{\lambda}) = S_{\boldsymbol{\lambda}}^{-1/2}(\boldsymbol{X}_j^{(\boldsymbol{\lambda})} - \bar{\boldsymbol{X}}_{\boldsymbol{\lambda}}), \tag{1.5}$$

for $j = 1, \ldots, n$ be the *standardized residual vectors* obtained by the transformation being set at $\boldsymbol{\lambda}$. We refer to $\boldsymbol{e}_j(\boldsymbol{\lambda})$ as the j-th $\boldsymbol{\lambda}$-residual. The *empirical characteristic function* for the $\boldsymbol{\lambda}$-residuals is given by

$$\frac{1}{n}\sum_{j=1}^{n}\exp(iu^t\boldsymbol{e}_j(\boldsymbol{\lambda})) = \exp(-iu^tS_{\boldsymbol{\lambda}}^{-1/2}\bar{\boldsymbol{X}}_{\boldsymbol{\lambda}})\hat{\varphi}_{n,\boldsymbol{\lambda}}(S_{\boldsymbol{\lambda}}^{-1/2}u),$$

where

$$\hat{\varphi}_{n,\boldsymbol{\lambda}}(v) = \frac{1}{n}\sum_{j=1}^{n}\exp(iv^t\boldsymbol{X}_j^{(\boldsymbol{\lambda})}).$$

Our estimator $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}_0$ is the vector minimizing

$$W_n(\boldsymbol{\lambda}) = \int_{R^d}\Im^2(\exp(-iu^tS_{\boldsymbol{\lambda}}^{-1/2}\bar{\boldsymbol{X}}_{\boldsymbol{\lambda}})\hat{\varphi}_{n,\boldsymbol{\lambda}}(u^tS_{\boldsymbol{\lambda}}^{-1/2}))w(u)du, \tag{1.6}$$

where $\Im(\cdot)$ denotes imaginary part and $w(u)$ is a weight function to be defined below. The intuition for considering a statistic of the form (1.6), comes from the well known fact that, in the univariate case, symmetry of the distribution is equivalent to having a null imaginary part in the characteristic function. In the one-dimensional case, the statistic in (1.6) for any fixed λ corresponds to the Feuerverger and Mureika (1977) statistic for testing a hypothesis of symmetry. For general dimension, our statistic is a member of (if w is chosen to be a *radial* function) the large family considered by Ghosh and Ruymgaart (1992) for testing the hypothesis of spherical symmetry (see also Csörgő (1986) and Csörgő and Heathcote (1987)). In the following section we discuss further analytical justification for this choice of estimator. Section 3 is non-technical, and contains a few examples and real data applications which illustrate the behavior of $\hat{\boldsymbol{\lambda}}$. Implementation issues are discussed and empirical comparisons with other estimators are made. In Section 4, the consistency and $\sqrt{n}$-consistency of this estimator is established, using empirical processes results for U-statistics. The theory given in Sections 2 and 4 does not cover all the examples included in Section 3: the examples with "heavier tails" are not covered. Nevertheless, these were included in Section 3 for the sake of practical evaluation of the proposed methodology.

2. The empirical characteristic function and transformations to elliptical symmetry

A random vector $\boldsymbol{X} \in R^d$ with probability density f, is said to possess elliptical symmetry when f can be written as in (1.3). Let us say that $\boldsymbol{X}$ has *symmetry of marginals* or *marginal symmetry* when for every $u \in R^d$, the random variable $u^t\boldsymbol{X}$ is symmetric (about some center), in the usual univariate sense. It is not difficult to prove that

PROPOSITION 2.1. *Elliptical symmetry implies symmetry of marginals.*

Denote by φ and $\varphi_{u^t\boldsymbol{X}}$ the characteristic functions of $\boldsymbol{X}$ and $u^t\boldsymbol{X}$, respectively. Let $\boldsymbol{\mu} = E(\boldsymbol{X})$. For each coordinate X_j of $\boldsymbol{X}$ and $k \geq 1$, let $\mu_{j,k} = E(X_j^k)$ when this moment exists. We have

PROPOSITION 2.2. (a) *The random vector* $\boldsymbol{X}$ *with density* f, *has symmetry of marginals if, and only if* $\Im(\varphi(u)\exp(-iu^t\boldsymbol{\mu})) = 0$ *for every* $u \in R^d$.

(b) *If, for each* $j \leq d$ *there exists an* $\alpha_j > 0$ *such that*

$$\sum_{1\leq k<\infty} \frac{|\mu_{j,k}|\alpha^k}{k!} < \infty, \qquad \forall |\alpha| < \alpha_j, \tag{2.1}$$

then $\boldsymbol{X}$ *has symmetry of marginals if, and only if* $\Im(\varphi(u)\exp(-iu^t\boldsymbol{\mu})) = 0$ *in a neighborhood of the origin.*

PROOF. Proposition 2.2(a) is well known, see, for instance, Section 12.4 of Loève (1955). The only fact which remains to be shown is the "if" part of 2.2(b). Let $u \in R^d$. (2.1) implies that all moments of $u^t\boldsymbol{X}$ exist. Then, using the formula for moments in terms of derivatives of the characteristic function at zero, we conclude that all odd moments of $u^t(\boldsymbol{X} - \boldsymbol{\mu})$ are zero. Now, for a distribution satisfying (2.1), Carleman's condition holds (Shohat and Tamarkin (1943) or Serfling (1980), Section 1.13) and therefore the problem of moments has a unique solution. That is, the distribution of $u^t(\boldsymbol{X} - \boldsymbol{\mu})$ is completely determined by its moments. Let Z be a variable with the same distribution of $u^t(\boldsymbol{X} - \boldsymbol{\mu})$, and let ϵ be a Rademacher variable independent of Z. Then, $\epsilon|Z|$ has the same moments of $u^t(\boldsymbol{X} - \boldsymbol{\mu})$ and is clearly symmetric about zero. Thus, by the uniqueness of the distribution given by Carleman's result, $u^t(\boldsymbol{X} - \boldsymbol{\mu})$ is symmetric about zero, completing the proof. □

Another elementary fact, more directly related to the transformations we are currently considering, is the following.

PROPOSITION 2.3. *Let* $\boldsymbol{X}$ *be a random vector of positive coordinates, such that the support of the distribution of each coordinate* X_i $(i \leq d)$ *contains a non-degenerate (possibly infinite) interval, and let* $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ *be as defined in the previous section. If there exists a* $\boldsymbol{\lambda}$ *in* R^d *such that* $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ *is marginally symmetric, then it is unique.*

PROOF. Suppose that for $\boldsymbol{\lambda}$, $\boldsymbol{\beta} \in R^d$, $\boldsymbol{\lambda} \neq \boldsymbol{\beta}$, $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ and $\boldsymbol{X}^{(\boldsymbol{\beta})}$ are marginally symmetric. Since $\boldsymbol{\lambda} \neq \boldsymbol{\beta}$, then, at least for one coordinate, we must have $\lambda_i \neq \beta_i$. Without loss of generality assume $\lambda_1 \neq \beta_1$. Take $u = (1, 0, \ldots, 0) \in R^d$. It follows, by definition of marginal symmetry, that $X_1^{(\lambda_1)}$ and $X_1^{(\beta_1)}$ are symmetric univariate variables. Thus, it suffices to show that if X is univariate with support in a non-degenerate interval, $X^{(\lambda)}$ and $X^{(\beta)}$ symmetric implies $\lambda = \beta$. Suppose that for $\lambda \neq \beta$ both $X^{(\lambda)}$ and $X^{(\beta)}$ are symmetric. To simplify notations, without affecting the argument, assume that $\lambda = 1$. We will also assume $\beta > 0$. The other cases can be treated similarly. We have that X and X^β are symmetric, with non-zero centers of symmetry. Let t and s be the centers of symmetry of X and X^β, respectively, and let $\delta = s - t$. For $r \in R$ we must have, using the symmetry of X^β: $\Pr(X^\beta > s + r) = \Pr(X^\beta < s - r)$. For $r > -s$, this can be rewritten as

$$\Pr(X > t + ((t + \delta + r)^{1/\beta} - t)) = \Pr(X < t - (t - (t + \delta - r)^{1/\beta})). \tag{2.2}$$

By the assumption on the support of X, there exist infinitely many values of r in an interval, such that $(t+\delta+r)^{1/\beta}$ is an increasing point of the distribution of X. For each such r, (2.2) implies that the equality $(t+\delta+r)^{1/\beta}-t=t-(t+\delta-r)^{1/\beta}$ must hold, and this is an impossibility. □

Let $\boldsymbol{X}$ be, as above, a random vector of positive coordinates, such that the support of the distribution of each coordinate X_i $(i \le d)$ contains a non-degenerate interval. Denote by $\varphi_{\boldsymbol{\lambda}}$ the characteristic function of $\boldsymbol{X}^{(\boldsymbol{\lambda})}$. Let $w : R^d \to [0,\infty)$ be a weight function such that $w(0)>0$, $\int_{R^d} w(u)du$ is finite, and is continuous at the origin. Let Λ be a compact set in R^d. Assume that for $\boldsymbol{\lambda}_0 \in \Lambda$, $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$ has an elliptically symmetric density. For each $\boldsymbol{\lambda} \in R^d$ let $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$ and $\Sigma_{\boldsymbol{\lambda}}$ denote, the mean and covariance matrix of $\boldsymbol{X}^{(\boldsymbol{\lambda})}$, respectively. These are assumed to exist for all $\boldsymbol{\lambda} \in \Lambda$. The three previous propositions have the following key result as a corollary.

PROPOSITION 2.4. (a) *With the hypothesis of the previous paragraph, there exists only one value of $\boldsymbol{\lambda}$ in Λ, namely $\boldsymbol{\lambda}_0$, such that*

$$\int_{R^d} \Im^2(\exp(-iu^t\Sigma_{\boldsymbol{\lambda}}^{-1/2}\boldsymbol{\mu}_{\boldsymbol{\lambda}})\varphi_{\boldsymbol{\lambda}}(\Sigma_{\boldsymbol{\lambda}}^{-1/2}u))du = 0. \tag{2.3}$$

(b) *If, in addition to the hypothesis above we assume that the coordinates of $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ for $\boldsymbol{\lambda} \in \Lambda$, satisfy the moment condition* (2.1) (*with constants α_j possibly depending on $\boldsymbol{\lambda}$*), *then there exists only one value of $\boldsymbol{\lambda}$ in Λ, namely $\boldsymbol{\lambda}_0$ such that*

$$\int_{V} \Im^2(\exp(-iu^t\Sigma_{\boldsymbol{\lambda}}^{-1/2}\boldsymbol{\mu}_{\boldsymbol{\lambda}})\varphi_{\boldsymbol{\lambda}}(\Sigma_{\boldsymbol{\lambda}}^{-1/2}u))w(u)du = 0, \tag{2.4}$$

where V denotes any neighborhood of the origin.

PROOF. We only give proof of Proposition 2.4(a). 2.4(b) is obtained in the same fashion, using Proposition 2.2(b) instead of 2.2(a). By Propositions 2.1 and 2.3, $\boldsymbol{\lambda}_0$ is the only $\boldsymbol{\lambda}$ such that $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ is marginally symmetric. For an invertible matrix such as $\Sigma_{\boldsymbol{\lambda}}^{-1/2}$, it is easy to see that $\Sigma_{\boldsymbol{\lambda}}^{-1/2}(\boldsymbol{X}^{(\boldsymbol{\lambda})}-\boldsymbol{\mu}_{\boldsymbol{\lambda}})$ is marginally symmetric if and only if $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ is marginally symmetric. Now, an application of Proposition 2.2(a) finishes the proof. □

Since the integral in (2.4) is being approximated by the integral of the $\boldsymbol{\lambda}$-residual empirical characteristic function in (1.6), Proposition 2.4 justifies the proposed methodology. Proposition 2.4 will also be instrumental in Section 3 in establishing consistency of the estimator.

For computational reasons, in the applications we make a restriction to the context of Proposition 2.4(b) by setting the function w in (1.6) equal to zero outside a neighborhood of the origin. In general, certain care must be taken in the choice of the function w in (1.6). Because of the Riemann-Lebesgue theorem, $\varphi_{\boldsymbol{\lambda}}(u)$ will *always* approach zero when $\|u\| \to \infty$, even in the absence of elliptical symmetry. Thus it would not make sense to give much weight to large values of u. On the other hand, if integration were restricted to a very small vicinity of the origin, we could be loosing information present in data. We would also like to choose w in such way that the integral $W_n(\boldsymbol{\lambda})$ may be computed in closed form (see Section 3). These considerations (and some trial and error) have led us to the choice:

$$w(u) = \chi_B(u), \tag{2.5}$$

where χ denotes indicator function and B is a box $[-A, A]^d$, with the value of A to be specified in Section 3.

It is worth noting that in the case $\boldsymbol{\lambda}_0 = \mathbf{0}$ with $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$ having a Gaussian distribution, the moment conditions of Proposition 2.4(b) do not hold: If $\boldsymbol{X}$ is log-normal and we consider the moments of a coordinate of $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ for any $\boldsymbol{\lambda} \neq \mathbf{0}$, then the series in (2.1) diverges for every $\alpha > 0$. Still, in this case, direct calculation shows that the integrand in (2.4) is not zero for $\boldsymbol{\lambda} \neq \mathbf{0}$, so that the thesis of 2.4(b) does hold, which is what is needed to prove consistency of $\hat{\boldsymbol{\lambda}}$. In this case the proof of Theorem 4.1 goes through with some minor modifications. The moment conditions (2.1) do hold when the distribution of $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$ is truncated Gaussian (with elliptical symmetry) and for any $\boldsymbol{X}$ with support in a box of the form $[\epsilon, M]^d$, for some $0 < \epsilon < M < \infty$. This is not a stringent assumption in the present setting, since we are interested in modeling $\boldsymbol{X}$ with positive entries; transformation (1.3) would not be viable otherwise.

3. Behavior of $\hat{\lambda}$ on simulated and real data

We consider the weight function as defined by (2.5). Algebraic manipulation shows that the following explicit formula for $W_n(\boldsymbol{\lambda})$ holds:

$$W_n(\boldsymbol{\lambda}) = \frac{2^{d-1}}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \left[\prod_{m=1}^{d} \frac{\sin\{(e_{jm} - e_{km})A\}}{e_{jm} - e_{km}} - \prod_{m=1}^{d} \frac{\sin\{(e_{jm} + e_{km})A\}}{e_{jm} + e_{km}} \right], \tag{3.1}$$

where $\boldsymbol{e}_j = (e_{j1}, \ldots, e_{jd})^t$ is defined through (1.5). If $e_{jm} - e_{km} = 0$ or if $e_{jm} + e_{km} = 0$, the terms in the products above are interpreted to be equal to A. This formula enables us to easily minimize $W_n(\boldsymbol{\lambda})$ using numerical methods, without actually having to compute an integral for each trial value of $\boldsymbol{\lambda}$.

3.1 *Symmetry in one dimension*

In this subsection we assume that $X^{(\lambda_0)}$ is a symmetric random variable having mean μ_0 and variance σ_0^2. This simple setting provides interesting comparisons with other available methods for estimating transformations to symmetry.

Let $\theta_0^t = (\lambda_0, \mu_0, \sigma_0^2)$. Let us firstly analyze the issue of choice of the constant A via simulation. We consider samples of size 100 under the four configurations

Table 1.

case no.		$A=0.6$	$A=0.8$	$A=1.0$	$A=1.5$	$A=2.0$
1	mean	0.4989	0.4917	0.5061	0.5140	0.5087
	variance	0.0845	0.0849	0.0846	0.0981	0.1114
	skewness	0.1963	0.0060	0.0099	−0.0928	−0.0961
	kurtosis	0.1862	0.2679	0.3796	0.0910	0.3076
	m.s.e.	0.0847	0.0847	0.0846	0.0981	0.1114
2	mean	−0.2375	−0.2374	−0.2337	−0.2424	−0.2459
	variance	0.0387	0.0336	0.0391	0.0412	0.0493
	skewness	−0.0323	−0.0150	0.0816	0.1830	0.0137
	kurtosis	0.2960	0.1447	0.1673	0.1295	0.1558
	m.s.e.	0.0387	0.0336	0.0391	0.0413	0.0493
3	mean	0.9339	0.9322	0.9373	0.9432	0.9377
	variance	0.1633	0.1872	0.1599	0.1937	0.2182
	skewness	0.0030	−0.0378	−0.1064	−0.2100	−0.1443
	kurtosis	0.6626	0.3368	0.2985	0.2898	0.2224
	m.s.e.	0.1635	0.1875	0.1600	0.1937	0.2184
4	mean	0.0010	0.0154	−0.0095	−0.0059	−0.0044
	variance	0.0332	0.0301	0.0324	0.0330	0.0379
	skewness	−0.0215	−0.0152	0.0589	0.0129	−0.0620
	kurtosis	0.1680	−0.1054	−0.0674	0.2625	0.3404
	m.s.e.	0.0332	0.0303	0.0325	0.0330	0.0380

$\theta_0^t = (0.51, 4, 0.81)$, $(-0.24, -5, 1)$, $(0.95, 4, 1)$, and $(0, 2, 0.25)$, when the distribution of $X^{(\lambda_0)}$ is standard normal. Table 1 shows results which describe the distribution of $\hat{\lambda}$, based on 1000 trials, letting A take on the values 0.6, 0.8, 1.0, 1.5, and 2.0.

An interesting feature in Table 1 is that minimum variance within the proposed class of estimators is achieved selecting A near 1.0, regardless of the value of θ_0. It is also apparent that values of A which are either too small or too large do not favor estimating the transformation; this fact was pointed out earlier when discussing the weight function. On the other hand, the values obtained for sample skewness and kurtosis suggest asymptotic normality of the estimator.

We conducted a second small Monte Carlo exploratory study aimed at comparing the estimator based on the empirical characteristic function with other estimators of λ defined for one dimensional data. Some of these alternative estimators have appeared previously in the literature, while others are adapted here by minimizing existing tests for symmetry. In describing these alternative estimators, we use the notation $e_k(\lambda)$ for the k-th λ-residual. One family of tests for symmetry which we consider is

$$R_{n,p}(\lambda) = \sum_{j=1}^{[n/2]} |2 \cdot \operatorname{med}\{e_k(\lambda), 1 \le k \le n\} - e_{(j)}(\lambda) - e_{(n-j+1)}(\lambda)|^p,$$

with $p > 0$ (suggested by David Ruppert in personal communication). The list of estimators included in our study is the following:

1. The estimator due to Berry (1987), which selects λ to minimize

$$K_n(\lambda) = \left| \frac{1}{n} \sum_{j=1}^{n} e_j^3(\lambda) \right| + \left| \frac{1}{n} \sum_{j=1}^{n} e_j^4(\lambda) - 3 \right| .$$

2. The minimizer of the test due to Boos (1982), which is

$$B_n(\lambda) = n \left\{ \frac{\frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} |e_j(\lambda) + e_k(\lambda) - 2\hat{m}|}{\sum_{j<k} |e_j(\lambda) - e_k(\lambda)|} - 1 \right\},$$

where $\hat{m} = \text{med}\{[e_j(\lambda) + e_k(\lambda)]/2, 1 \le j \le k \le n\}$.

3. The estimator based on the empirical characteristic function, setting $A = 0.8$ and minimizing $W_n(\lambda)$ defined by (3.1).
4. The minimizer of $R_{n,1}(\lambda)$.
5. The minimizer of $R_{n,2}(\lambda)$.
6. The minimizer of $R_{n,3}(\lambda)$.
7. The *skewness estimator*, obtained by minimizing

$$T_n(\lambda) = \left| \frac{1}{n} \sum_{j=1}^{n} e_j^3(\lambda) \right| .$$

(Actually, the value of λ is found which makes $T_n(\lambda) = 0$.) This estimator is investigated by Taylor (1985). He shows that this skewness estimator has an optimality property for normal errors, and compares it with other estimators of λ as well.

8. A minimum distance type of estimator (Nakamura and Ruppert (1990)) obtained by minimizing

$$V_n(\lambda) = \int_{-\infty}^{\infty} \{F_n(x, \lambda) + F_n(-x, \lambda) - 1\}^2 x^2 dx,$$

where $F_n(x, \lambda)$ is the empirical distribution function of $\{e_j(\lambda), 1 \le j \le n\}$.

9. The maximum likelihood estimator assuming f is standard normal. Although this estimator actually attempts to transform to normality, not general symmetry, it was included in this study for comparative reasons. The estimate of λ minimizes

$$L_n(\lambda) = n \log\{\hat{\sigma}^2(\lambda)\} + \sum_{j=1}^{n} e_j^2(\lambda) - 2(\lambda - 1) \sum_{j=1}^{n} \log\{X_j\}.$$

We simulated one-dimensional samples with transformation (1.1) using three symmetric densities for f: the standard normal, the t-distribution with 3 degrees of freedom, and a double exponential distribution. Sample sizes considered $n = 30$, 50 and 80, again generating 1000 Monte Carlo replications in each setting. The values used for λ_0 were 0.0, 0.51, and 1.0, with $(\mu_0, \sigma_0^2) = (2, 0.5^2)$ for the normal and double exponential settings, and $\theta_0^t = (0, 2, 0.4^2)$, $(.51, 2, 0.45^2)$, and $(1, 2, 0.35^2)$ for the t-distribution. Table 2 summarizes Monte Carlo results for a few representative cases, where estimators within a fixed case have been ranked according to empirical mean squared error. A number of conclusions are noteworthy.

For the normal distribution and large sample sizes, the method based on the empirical characteristic function is amongst the most efficient after (normal theory) maximum likelihood.

In the non-normal symmetric situations, none of the estimators considered is uniformly best across all distributions and sample sizes. Which method occupies the top spot seems to depend strongly on the particular setting as well as the sample size. That is, a winner in one situation may rank last in another. However, the empirical characteristic function estimator never ranks worse than fourth (according to mean squared error). No other estimator maintains a high ranking across all situations.

Some of the transformation estimators that minimize a test for symmetry are comparatively inefficient. Examples are estimators numbered 2 and 4. There could be a connection between these results and the power these statistics would exhibit in testing symmetry.

There is a substantial improvement in the estimators based on $R_{n,p}(\lambda)$ when $p = 1$ is switched to $p = 2$, and then to $p = 3$. This point motivates further investigation of this estimator for other values of p, but we do not address this issue or any of the asymptotic properties of this estimator here.

Our results also suggest that the sample skewness, $T_n(\lambda)$, yields a very competitive estimator in this one-dimensional setting. However, our interest here lies in multivariate statistics and, in that setting, the skewness estimator is not directly applicable.

3.2 *Symmetry in two and three dimensions*

Let us turn to a bivariate scenario, and exemplify estimation of $\boldsymbol{\lambda}$ in simulated data sets of size 200, letting $\boldsymbol{\lambda}_0 = (0.5, 0.1)$ and setting the density f to be one of four elliptically symmetric bivariate distributions, labeled D1–D4. D1 is $\begin{pmatrix} 8 & 2 \\ 2 & 1 \end{pmatrix}(U + \binom{5}{8})$, where U has a uniform distribution on the unit circle on the plane. D2 is a bivariate normal having covariance $\begin{pmatrix} 1 & .8 \\ .8 & 2 \end{pmatrix}$ and mean $\binom{5}{5}$. D3 is a bivariate normal having covariance $\begin{pmatrix} 1 & .1 \\ .1 & 2 \end{pmatrix}$ and mean $\binom{5}{5}$. D4 is $\begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}(V + \binom{13}{13})$, where V has the (spherically symmetric) bivariate distribution corresponding to the density characterized by $\|V\| \sim \chi_4^2$ and $\arctan(V_1/V_2) \sim \text{Uniform}(0, 2\pi)$. D4 is intended to represent a distribution with "heavier tails" than the bivariate normal. D2 and D3 differ only in that one has a greater covariance term; for likelihood-based methods, this may induce a difference when compared to marginal symmetrization (see Velilla (1993)).

Table 3 reports results obtained by minimizing (3.1) as a function of the bivariate transformation parameter, where we have set $A = 1$.

For each setting, Figs. 1(a) and 1(b) show scatter plots of original simulated

Table 2.

	$n = 30$ $\lambda = .5$				$n = 80$ $\lambda = 0$		
	bias	variance	m.s.e.		bias	variance	m.s.e.
		NORMAL				NORMAL	
m.l.e.	−0.0250	0.4589	0.4595	m.l.e.	0.0060	0.0347	0.0347
min.dist.	−0.0414	0.5941	0.5958	skewness	−0.0062	0.0385	0.0385
emp.ch.fn	−0.0202	0.5989	0.5993	emp.ch.fn	−0.0016	0.0391	0.0391
skewness	−0.0084	0.6170	0.6170	min.dist.	0.0037	0.0454	0.0454
R2	−0.0085	0.6678	0.6678	R3	−0.0011	0.0459	0.0460
R3	0.0577	0.7171	0.7204	R2	0.0019	0.0520	0.0520
Boos	−0.0169	0.8104	0.8106	Boos	0.0063	0.0633	0.0633
R1	0.0102	0.9802	0.9803	R1	0.0049	0.0742	0.0742
Berry	−0.2610	1.6209	1.6890	Berry	−0.0098	0.1341	0.1342
		t 3 d.f.				t 3 d.f.	
Berry	−0.0204	1.6206	1.6210	R2	−0.0110	0.1744	0.1745
R3	0.0263	1.7439	1.7446	skewness	0.0007	0.1852	0.1852
R2	−0.1368	1.7697	1.7884	R3	0.0071	0.1945	0.1946
emp.ch.fn	−0.0573	1.8248	1.8280	emp.ch.fn	0.0137	0.1959	0.1961
m.l.e.	−0.0381	1.8361	1.8375	Berry	0.0028	0.2106	0.2106
skewness	−0.1168	1.9127	1.9268	min.dist.	0.0224	0.2185	0.2190
R1	−0.0156	2.0084	2.0086	Boos	0.0044	0.2314	0.2314
min.dist.	−0.0640	2.0119	2.0160	m.l.e.	−0.0027	0.2348	0.2348
Boos	−0.0394	2.1979	2.1994	R1	0.0180	0.3571	0.3574
		D. EXPON.				D. EXPON.	
m.l.e.	−0.0381	0.7604	0.7618	Berry	−0.0014	0.0599	0.0599
skewness	−0.0185	0.8501	0.8504	skewness	0.0099	0.0599	0.0600
Berry	−0.0727	0.9196	0.9248	emp.ch.fn	−0.0148	0.0612	0.0614
emp.ch.fn	−0.0127	0.9477	0.9478	R2	−0.0146	0.0663	0.0665
min.dist.	−0.0342	0.9942	0.9953	R3	0.0033	0.0682	0.0682
R3	0.0274	1.0169	1.0176	min.dist.	0.0068	0.0693	0.0693
R2	−0.0500	1.0241	1.0266	m.l.e.	0.0141	0.0747	0.0749
Boos	−0.0219	1.2210	1.2214	Boos	0.0051	0.1092	0.1092
R1	−0.0546	1.2618	1.2647	R1	−0.0089	0.1523	0.1524

data $\boldsymbol{X}$, and data after having applied the estimated transformation, $\boldsymbol{X}^{(\hat{\boldsymbol{\lambda}})}$, respectively. These figures suggest good performance of the empirical characteristic function estimator when elliptical symmetry is achievable by a Box-Cox transformation.

We found that an application on a real-life bivariate data set is also interesting to look at. The set regards industrial pollution in 558 water samples in Leon, Mexico. Two selected variables, VAR1 and VAR2, are considered here for illustrative purposes; both are concentrations of chemical substances, *i.e.* are pos-

Table 2. (continued).

	$n = 50$ $\lambda = 1$				$n = 80$ $\lambda = .5$		
	bias	variance	m.s.e.		bias	variance	m.s.e.
		NORMAL				NORMAL	
m.l.e.	−0.0906	0.5278	0.5360	m.l.e.	−0.0256	0.1374	0.1381
emp.ch.fn	−0.0117	0.5793	0.5794	skewness	0.0074	0.1613	0.1614
skewness	0.0019	0.6084	0.6084	emp.ch.fn	−0.0162	0.1691	0.1694
R3	0.0176	0.7745	0.7748	min.dist.	−0.0129	0.1764	0.1766
R2	−0.0207	0.7797	0.7801	R3	0.0036	0.1824	0.1824
min.dist.	−0.0601	0.8321	0.8357	R2	0.0111	0.2243	0.2244
Boos	−0.0153	0.8376	0.8378	Boos	−0.0032	0.2432	0.2432
R1	0.0507	1.0342	1.0368	R1	0.0109	0.3000	0.3001
Berry	−0.3556	1.4653	1.5918	Berry	−0.0963	0.4402	0.4495
		t 3 d.f.				t 3 d.f.	
R2	−0.1860	2.6436	2.6781	skewness	−0.0392	0.5606	0.5621
R3	−0.1269	2.7379	2.7537	R3	−0.0024	0.5785	0.5785
emp.ch.fn	−0.1282	2.7434	2.7598	emp.ch.fn	0.0285	0.6017	0.6025
Boos	−0.1649	2.8166	2.8437	Berry	0.0188	0.6069	0.6073
Berry	−0.1801	2.8438	2.8762	R2	−0.0130	0.6534	0.6536
m.l.e.	−0.1015	2.9392	2.9495	Boos	−0.0321	0.6685	0.6695
min.dist.	−0.1607	2.9469	2.9727	min.dist.	−0.0267	0.6857	0.6864
skewness	−0.1748	3.1144	3.1449	m.l.e.	0.0274	0.7358	0.7366
R1	−0.1779	3.1451	3.1761	R1	−0.0354	0.8467	0.8480
		D. EXPON.				D. EXPON.	
Berry	−0.0409	0.9310	0.9327	R3	0.0106	0.2350	0.2351
skewness	−0.0216	0.9612	0.9617	Berry	−0.0067	0.2425	0.2425
emp.ch.fn	−0.0482	0.9609	0.9632	emp.ch.fn	−0.0189	0.2483	0.2487
m.l.e.	0.0552	0.9662	0.9692	skewness	−0.0119	0.2504	0.2505
R3	−0.0630	0.9745	0.9785	m.l.e.	0.0524	0.2626	0.2653
R2	−0.0827	0.9809	0.9877	R2	−0.0004	0.2719	0.2719
min.dist.	−0.0295	1.0171	1.0180	min.dist.	−0.0002	0.2813	0.2813
R1	−0.0627	1.2578	1.2617	Boos	−0.0025	0.3943	0.3943
Boos	−0.0242	1.3945	1.3951	R1	−0.0455	0.4390	0.4411

Table 3.

distribution	$\hat{\lambda}_1$	$\hat{\lambda}_2$
D1	0.308	−0.015
D2	0.483	0.146
D3	0.293	0.120
D4	0.677	0.128

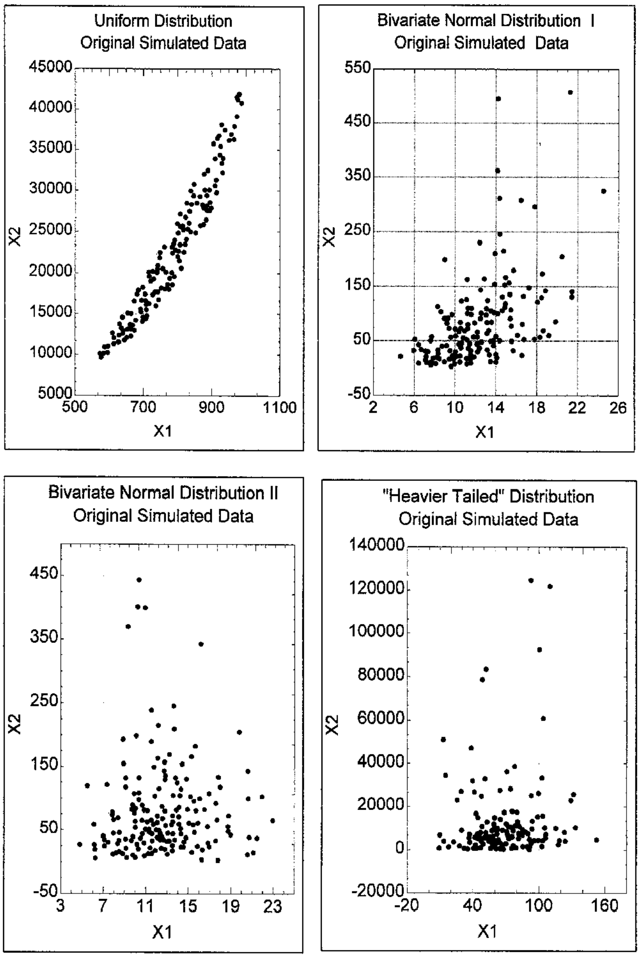

Fig. 1(a). Four simulated bivariate distributions before transformation to elliptical symmetry.

itive variables. Figure 2 shows scatterplots and marginal distributions of original data and after an estimated transformation obtained using the empirical characteristic function. It is striking how similar these are in appearance to some of the plots in Figs. 1(a) and 1(b). For comparisons, bivariate transformations are here not only estimated by using the empirical characteristic function, but also by maximum likelihood assuming bivariate normality (Andrews *et al.* (1971)). In addition, we also estimate both transformation parameters using *marginal* likelihood, that is, by individually transforming VAR1 and VAR2 to normality using normal likelihoods. The three pairs of estimates are very similar in this case (see Table 4).

A final example regards a three-dimensional subset of Urology data (Andrews and Herzberg (1985), p. 249), depicted in Fig. 3. Selected Variables are pH (PH), conductivity (MHO), and urea concentration (UREA), in 79 urine specimens. The

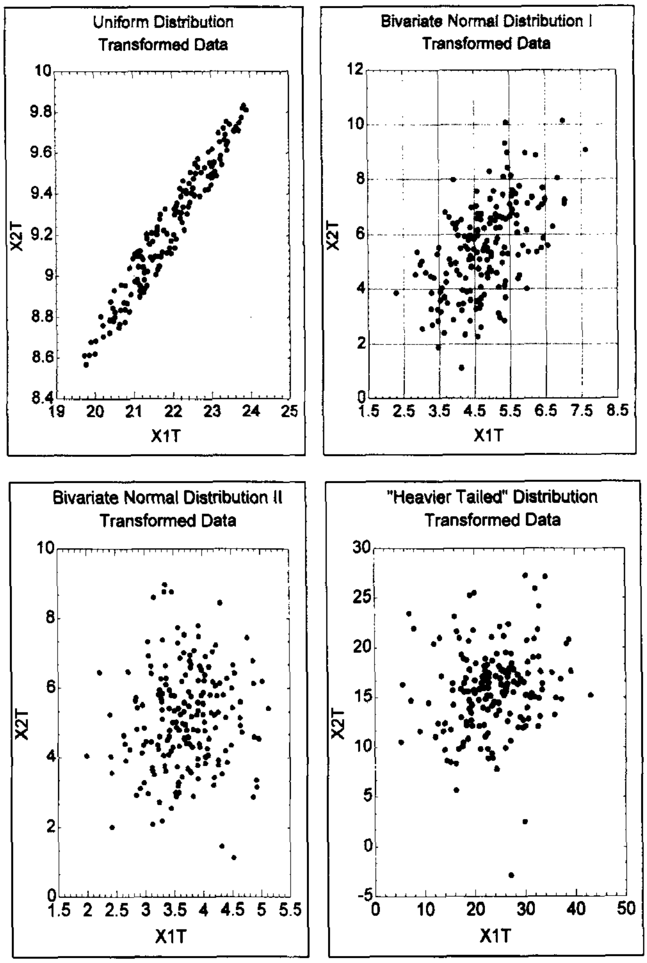

Fig. 1(b). Four simulated bivariate distributions after transformation to elliptical symmetry.

three-dimensional implementation of equation (3.1), again setting $A = 1$, yields the estimate $\hat{\boldsymbol{\lambda}} = (-1.46, 1.00, 0.60)$. Figure 3 also shows bivariate projections of the data after using this value in transforming all entries. We note that the marginal skewness displayed in original data no longer appears after transformation. In addition, judging by all two-dimensional scatterplots of transformed points, elliptical symmetry might have been achieved.

4. Consistency of $\hat{\lambda}$

Here we establish consistency and $\sqrt{n}$-consistency of $\hat{\boldsymbol{\lambda}}$ by means of empirical processes theory as described in Dudley (1984) and Pollard (1984), and empirical processes results for U-statistics as given in Arcones and Giné (1993). Let us begin by listing some assumptions needed to prove consistency.

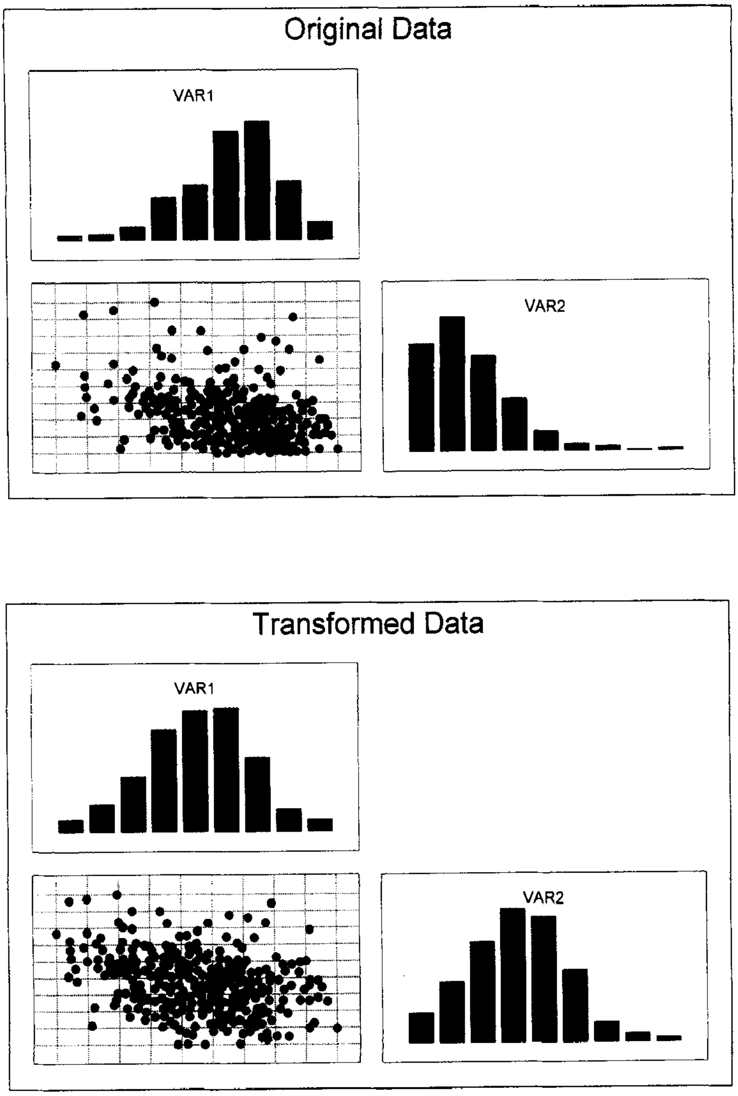

Fig. 2. Industrial pollution example: bivariate data plotted before and after transformation to elliptical symmetry.

Table 4.

estimator	$\hat{\lambda}_1$	$\hat{\lambda}_2$
char. function	1.890	0.539
max. likelihood	1.949	0.462
marginal likelihoods	2.076	0.443

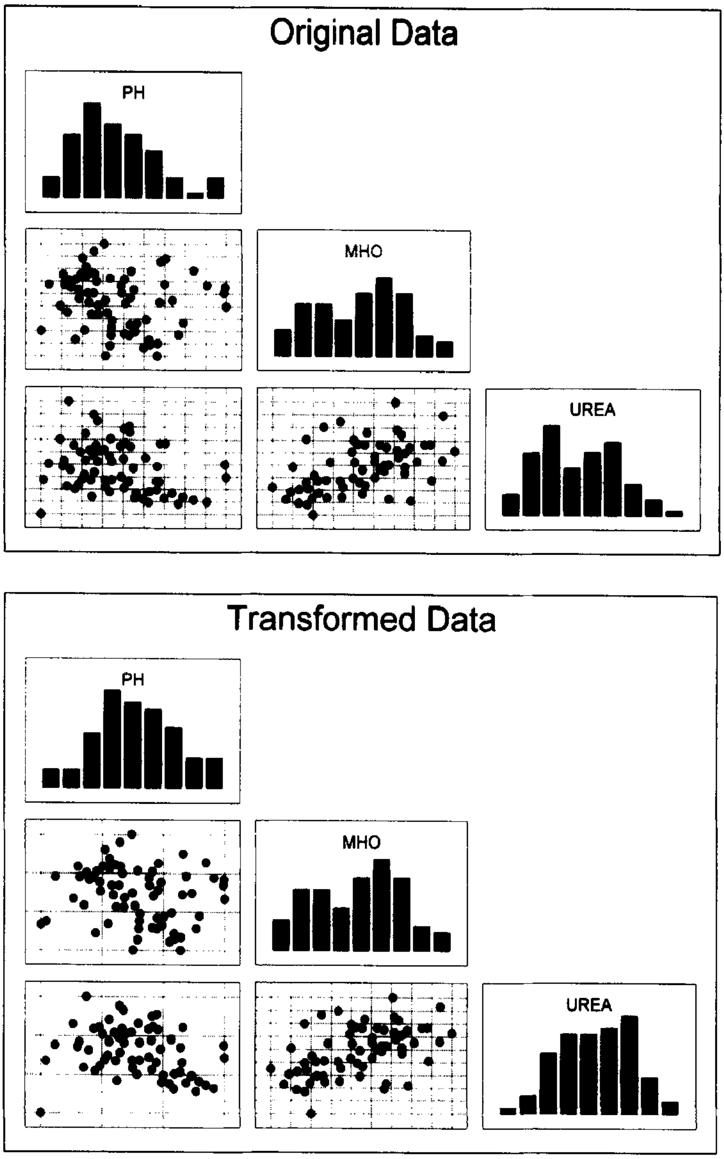

Fig. 3. Urology example: Pairwise scatter plots of trivariate data before and after transformation to elliptical symmetry.

ASSUMPTION 4.1. $\hat{\boldsymbol{\lambda}}$ is selected from a compact set $\Lambda \subset R^d$ such that $\boldsymbol{\lambda}_0 \in$ int(Λ) (the interior of Λ).

Since the parameters being estimated correspond to powers (in a power transformation), computing limitations prevent the consideration of very large values for the coordinates of $\hat{\boldsymbol{\lambda}}$. In practice, one will look for the value of the $\hat{\lambda}_i$'s in a relatively small interval (like $[-5, 5]$, say). It would not be realistic to consider powers of the order of 100 in the transformations. Thus, Assumption 4.1 does not represent a real limitation in practice.

ASSUMPTION 4.2. The setting of Proposition 2.4(b) holds, that is, we have an i.i.d. sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of vectors in R^d, with the same distribution of $\boldsymbol{X}$. $\boldsymbol{X}$ is a vector of positive elements, and such that the distribution of $\boldsymbol{X}^{(\boldsymbol{\lambda}_0)}$ is elliptically symmetric. The coordinates of $\boldsymbol{X}^{(\boldsymbol{\lambda})}$ for $\boldsymbol{\lambda} \in \Lambda$, satisfy the moment condition (2.1) and the supports of their distributions contain non-degenerate intervals.

ASSUMPTION 4.3. For $\boldsymbol{\lambda}$ in Λ, $\Sigma_{\boldsymbol{\lambda}}$, as defined before Proposition 2.4, is non-singular.

Write P for the probability law of $\boldsymbol{X}$ and P_n for its empirical version: If C is a measurable set in R^d and g a measurable function,

$$P_n(C) = \frac{1}{n}\sum_{i=1}^{n} \chi_{(\boldsymbol{X}_i \in C)} \quad \text{and} \quad P_n(g) = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i).$$

With Assumptions 4.1 and 4.2 we get what is called, in the empirical processes literature, an *envelope function*, namely, there exists a function F_1 on R^d such that F_1 is in $L^2(P)$ and satisfies $F_1(x) \geq \|x^{(\boldsymbol{\lambda})}\|$, for all x in the support of f and $\boldsymbol{\lambda} \in \Lambda$. The following two propositions are needed for the proof of consistency.

PROPOSITION 4.1. *$\bar{\boldsymbol{X}}_{\boldsymbol{\lambda}}$, as defined in* (1.4) *is uniformly consistent for $\boldsymbol{\lambda} \in \Lambda$:*

$$\sup_{\boldsymbol{\lambda} \in \Lambda} \|\bar{\boldsymbol{X}}_{\boldsymbol{\lambda}} - \boldsymbol{\mu}_{\boldsymbol{\lambda}}\| \to 0, \qquad a.s. \text{ as } \quad n \to \infty. \tag{4.1}$$

PROOF. For $\boldsymbol{\lambda} \in \Lambda$, write $T_{\boldsymbol{\lambda}}$ for the vector transformation $x \to x^{(\boldsymbol{\lambda})}$ and T_{λ_j} for the coordinate transformation $x_j \to x_j^{(\lambda_j)}$. Let π_j denote the projection $\pi_j(x) = x_j$. The j-th coordinate of the average $\bar{\boldsymbol{X}}_{\boldsymbol{\lambda}}$ is $P_n(T_{\lambda_j} \circ \pi_j)$. Let $\Lambda_j = \{\lambda_j, \boldsymbol{\lambda} \in \Lambda\}$. Without loss of generality we shall assume that Λ_j is an interval. Also let

$$\begin{aligned} C_{\lambda_j} &= \{(x,t) \in R^d \times R : 0 \leq t \leq (T_{\lambda_j} \circ \pi_j)(x) \\ &\qquad \text{or } (T_{\lambda_j} \circ \pi_j)(x) \leq t < 0\}, \qquad \text{and} \\ \Theta_{(x,t)} &= \{\lambda_j : (x,t) \in C_{\lambda_j}\}. \end{aligned}$$

Rewrite $\Theta_{(x,t)} = \{\lambda_j : 0 \leq t \leq (x_j^{\lambda_j} - 1)/\lambda_j \text{ or } (x_j^{\lambda_j} - 1)/\lambda_j \leq t < 0\}$, to see that $\Theta_{(x,t)}$ is the union of at most two intervals in Λ_j. It follows that the dual density (Assouad (1983)) of the class

$$\boldsymbol{C} = \{C_{\lambda_j} : \lambda_j \in \Lambda_j\}$$

is finite and, therefore, by Proposition 2.13 of Assouad (1983) $\boldsymbol{C}$ is a Vapnik-Chervonenkis class. We conclude that the class

$$\boldsymbol{F} = \{T_{\lambda_j} \circ \pi_j : \lambda_j \in \Lambda_j\}$$

is a VC-subgraph class (see Dudley (1984, 1987) for this and related notions). Then, applying Lemma II.25 and Theorem II.24 of Pollard (1984) we obtain (4.1). □

PROPOSITION 4.2. *With* $S_{\boldsymbol{\lambda}}$ *as defined in* (1.4),

$$\sup_{\boldsymbol{\lambda}\in\Lambda} \|S_{\boldsymbol{\lambda}} - \Sigma_{\boldsymbol{\lambda}}\| \to 0 \quad \text{a.s. as} \quad n \to \infty, \tag{4.2}$$

where $\|\cdot\|$ *denotes Euclidean norm with the matrices seen as vectors of length* d^2.

PROOF. Let $S_{\boldsymbol{\lambda},\infty} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i^{(\boldsymbol{\lambda})} - \boldsymbol{\mu}_{\boldsymbol{\lambda}})(\boldsymbol{X}_i^{(\boldsymbol{\lambda})} - \boldsymbol{\mu}_{\boldsymbol{\lambda}})^t$. By the previous proposition and the Cauchy-Schwarz inequality, we have

$$\sup_{\boldsymbol{\lambda}\in\Lambda} \|S_{\boldsymbol{\lambda}} - S_{\boldsymbol{\lambda},\infty}\| \to 0 \quad \text{a.s. as} \quad n \to \infty.$$

Therefore we can replace $S_{\boldsymbol{\lambda}}$ in (4.2) by $S_{\boldsymbol{\lambda},\infty}$, which will simplify the proof. For a probability law Q on R^d, a collection of measurable functions $\boldsymbol{G}$ and $p = 1, 2$ let $N_p(\epsilon, \boldsymbol{G}, Q)$ denote the metric entropy of $\boldsymbol{G}$ with respect to the $L^p(Q)$ norm, as defined in Pollard ((1984), Chapter 2). Let $\boldsymbol{F}$ be the class of functions defined in the previous proof. By the fact that $\boldsymbol{F}$ is a VC-subgraph class, established above, and Lemma II.25 of Pollard (1984)

$$N_p(\epsilon, \boldsymbol{F}, Q) \le a_p(1/\epsilon)^{b_p} \tag{4.3}$$

for positive constants a_p, b_p depending only on the Vapnik-Chervonenkis index of the class $\boldsymbol{C}$ and on $\int F_1^p dQ$. From the moment assumption on the coordinates of $\boldsymbol{X}^{(\boldsymbol{\lambda})}$, we obtain, via dominated convergence, that the components of $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$ and the entries of $\Sigma_{\boldsymbol{\lambda}}^{-1/2}$ are differentiable functions of $\boldsymbol{\lambda} \in \Lambda$. It follows that the class of functions

$$\boldsymbol{F}_{\boldsymbol{\mu}} = \{g : g(x) = (T_{\lambda_j} \circ \pi_j)(x) - (\boldsymbol{\mu})_j; j \le d, \boldsymbol{\lambda} \in \Lambda\}$$

also satisfies (4.3), for possibly different constants a_p and b_p. Then, we have (see, for example, Problem 24, Chapter 2 in Pollard (1984)) that the class

$$\boldsymbol{F}_{\boldsymbol{\mu}} * \boldsymbol{F}_{\boldsymbol{\mu}} = \{gh : g, h \in \boldsymbol{F}_{\boldsymbol{\mu}}\}$$

satisfies

$$N_1(\epsilon, \boldsymbol{F}_{\boldsymbol{\mu}} * \mathbf{F}_{\boldsymbol{\mu}}, Q) \le N_2(\epsilon, \boldsymbol{F}_{\boldsymbol{\mu}} * \boldsymbol{F}_{\boldsymbol{\mu}}, Q) \le a_2^2(1/\epsilon)^{2b_2},$$

for the constants a_2 and b_2 mentioned above. An application of Theorem II.24 of Pollard (1984) now gives that for each pair i, j,

$$\sup_{\boldsymbol{\lambda}\in\Lambda} |(S_{\boldsymbol{\lambda},\infty})_{ij} - (\Sigma_{\boldsymbol{\lambda}})_{ij}| \to 0, \quad \text{a.s.}$$

completing the proof. □

Now we can prove consistency of $\hat{\boldsymbol{\lambda}}$.

THEOREM 4.1. *If the integrable, nonnegative weight function w in (1.6) is strictly positive in a neighborhood of the origin and zero outside a compact set, then under Assumptions 4.1 to 4.3, $\hat{\boldsymbol{\lambda}} \to \boldsymbol{\lambda}_0$, a.s., as $n \to \infty$.*

PROOF. Let $\boldsymbol{e}_i(\boldsymbol{\lambda})$, $i = 1, \ldots, n$ be the $\boldsymbol{\lambda}$-residual as defined in Section 1. Let

$$\begin{aligned} \boldsymbol{e}_i(\boldsymbol{\lambda})_\infty &= \Sigma_{\boldsymbol{\lambda}}^{-1/2}(\boldsymbol{X}_i^{(\boldsymbol{\lambda})} - \boldsymbol{\mu}_{\boldsymbol{\lambda}}) \quad \text{and} \\ U_n(\boldsymbol{\lambda}) &= \int \left(\frac{1}{n} \sum_{i=1}^{n} \sin(u^t \boldsymbol{e}_i(\boldsymbol{\lambda})_\infty) \right)^2 w(u)du. \end{aligned} \tag{4.4}$$

Our proof of consistency relies on proving the following three facts, from which consistency of $\hat{\boldsymbol{\lambda}}$ follows by a standard argument using Proposition 4.2:

$$\text{(I)} \qquad \sup_{\Lambda} |W_n(\boldsymbol{\lambda}) - U_n(\boldsymbol{\lambda})| \to 0, \qquad \text{a.s.}$$

$$\text{(II)} \qquad \sup_{\Lambda} |U_n(\boldsymbol{\lambda}) - EU_n(\boldsymbol{\lambda})| \to 0, \qquad \text{a.s.,} \qquad \text{and}$$

(III) $EU_n(\boldsymbol{\lambda})$ converges, uniformly in $\boldsymbol{\lambda}$, to a continuous limit $E_{\boldsymbol{\lambda}}$ which is zero only at $\boldsymbol{\lambda}_0$.

Note first that $W_n(\boldsymbol{\lambda})$ can be written as $\int(\frac{1}{n}\sum_{i=1}^n \sin(u^t \boldsymbol{e}_i(\boldsymbol{\lambda})))^2 w(u)du$. Now

$$\begin{aligned} &\left| \frac{1}{n} \sum_{i=1}^{n} \sin(u^t \boldsymbol{e}_i(\boldsymbol{\lambda})_\infty) - \frac{1}{n} \sum_{i=1}^{n} \sin(u^t \boldsymbol{e}_i(\boldsymbol{\lambda})) \right| \\ &\quad \le \frac{1}{n} \sum_{i=1}^{n} |u^t (S_{\boldsymbol{\lambda}}^{-1/2} - \Sigma_{\boldsymbol{\lambda}}^{-1/2})(\boldsymbol{X}_i^{(\boldsymbol{\lambda})} - \boldsymbol{\mu}_{\boldsymbol{\lambda}})| \\ &\quad\quad + \frac{1}{n} \sum_{i=1}^{n} |u^t S_{\boldsymbol{\lambda}}^{-1/2}(\boldsymbol{\mu}_{\boldsymbol{\lambda}} - \bar{\boldsymbol{X}}_{\boldsymbol{\lambda}})|. \end{aligned} \tag{4.5}$$

The second term in (4.5) goes to zero uniformly in $\boldsymbol{\lambda} \in \Lambda$ and $u \in \{u : w(u) > 0\}$ by Propositions 4.1 and 4.2, and the fact that the entries of $\Sigma_{\boldsymbol{\lambda}}^{-1/2}$ being continuous functions of $\boldsymbol{\lambda}$, are uniformly bounded. For a matrix M let $\|M\| = \sup_{\|x\|=1} \|Mx\|$. The first term in (4.5) is bounded above by

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \|u\| \|S_{\boldsymbol{\lambda}}^{-1/2} - \Sigma_{\boldsymbol{\lambda}}^{-1/2}\| \|\boldsymbol{X}_i^{(\boldsymbol{\lambda})} - \boldsymbol{\mu}_{\boldsymbol{\lambda}}\| \\ &\quad \le \frac{1}{n} \sum_{i=1}^{n} \|u\| \|S_{\boldsymbol{\lambda}}^{-1/2} - \Sigma_{\boldsymbol{\lambda}}^{-1/2}\| \left(F_1(\boldsymbol{X}_i) + \sup_{\Lambda} \|\boldsymbol{\mu}_{\boldsymbol{\lambda}}\| \right), \end{aligned}$$

and this approaches zero uniformly in $\boldsymbol{\lambda}$ by Proposition 4.2 and the strong law of large numbers applied to the function F_1. Therefore (I) is proved.

For $(\boldsymbol{X}_i, \boldsymbol{X}_j) \in R^d \times R^d$, and $\boldsymbol{\lambda} \in \Lambda$ let

$$g_{\boldsymbol{\lambda}}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \int \sin(u^t e_i(\boldsymbol{\lambda})_\infty) \sin(u^t e_j(\boldsymbol{\lambda})_\infty) w(u) du, \tag{4.6}$$

with $\boldsymbol{e}_i(\boldsymbol{\lambda})_\infty$ and $\boldsymbol{e}_j(\boldsymbol{\lambda})_\infty$ as given in (4.4). Let

$$U'_n(\boldsymbol{\lambda}) = \frac{1}{n(n-1)} \sum_{i \neq j} g_{\boldsymbol{\lambda}}(\boldsymbol{X}_i, \boldsymbol{X}_j). \tag{4.7}$$

Since $U_n(\boldsymbol{\lambda})$ and $U'_n(\boldsymbol{\lambda})$ differ only in the *diagonal* terms, we have

$$\sup_{\Lambda} |U_n(\boldsymbol{\lambda}) - U'_n(\boldsymbol{\lambda})| = O\left(\frac{1}{n}\right). \tag{4.8}$$

The stochastic process $\{U'_n(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \Lambda\}$ is a U-process, as those considered in Arcones and Giné (1993). It is indexed on the class of symmetric kernels

$$\boldsymbol{G} = \{g_{\boldsymbol{\lambda}} : R^d \times R^d \to R; \boldsymbol{\lambda} \in \Lambda\},$$

with $g_{\boldsymbol{\lambda}}$ as defined in (4.6). By (4.8) and Corollary 3.5 of Arcones and Giné (1993), in order to prove (II), it suffices to show that, for every $\epsilon > 0$,

$$N^{(1)}_{[\ ]}(\epsilon, \boldsymbol{G}, P \otimes P) < \infty, \tag{4.9}$$

where $N^{(1)}_{[\ ]}(\epsilon, \boldsymbol{G}, P \otimes P)$ is the bracketing metric entropy of the class $\boldsymbol{G}$, as defined in page 1512 of Arcones and Giné (1993).

Let $K = \int w(u)du$. Given $\epsilon > 0$, take $M > 0$ such that $P(F_1(\boldsymbol{X}_i) > M) < \epsilon/(8K)$. As noted above, the components of $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$ and $\boldsymbol{X}^{(\boldsymbol{\lambda})}$, and the entries of $\Sigma_{\boldsymbol{\lambda}}^{-1/2}$ are differentiable with respect to the components of $\boldsymbol{\lambda}$. It follows that on the set $\{(\boldsymbol{X}_i, \boldsymbol{X}_j) : F_1(\boldsymbol{X}_i) \leq M \text{ and } F_1(\boldsymbol{X}_j) \leq M\}$, $g_{\boldsymbol{\lambda}}(\boldsymbol{X}_i, \boldsymbol{X}_j)$ has partial derivatives, with respect to the components of $\boldsymbol{\lambda}$, uniformly bounded by a positive constant C. Cover Λ with a regular grid of points at distance $\epsilon/(2C)$. It follows that for each $\boldsymbol{\lambda} \in \Lambda$ there is a $\boldsymbol{\beta}$ in the grid, such that

$$|g_{\boldsymbol{\lambda}}(\boldsymbol{X}_i, \boldsymbol{X}_j) - g_{\boldsymbol{\beta}}(\boldsymbol{X}_i, \boldsymbol{X}_j)| \leq \delta(\boldsymbol{X}_i, \boldsymbol{X}_j),$$

with

$$\delta(\boldsymbol{X}_i, \boldsymbol{X}_j) = \begin{cases} \epsilon/2, & \text{if } F_1(\boldsymbol{X}_i) \leq M \text{ and } F_1(\boldsymbol{X}_j) \leq M, \\ 2K, & \text{otherwise.} \end{cases}$$

Since Λ is compact, the grid is finite, and by the choice of M, $(P \otimes P)(\delta) \leq \epsilon$. This suffices for (4.9) to hold, proving (II).

To prove (III) observe that $E(U_n(\boldsymbol{\lambda}) - U'_n(\boldsymbol{\lambda})) = O(\frac{1}{n})$, uniformly in $\boldsymbol{\lambda}$ and

$$E(U'_n(\boldsymbol{\lambda})) = E(g_{\boldsymbol{\lambda}}) = \int g_{\boldsymbol{\lambda}}(\boldsymbol{X}_i, \boldsymbol{X}_j) dP(\boldsymbol{X}_i) dP(\boldsymbol{X}_j), \qquad \text{for all} \quad n,$$

where P is the probability law of $\boldsymbol{X}$. It is clear that $E(g_{\boldsymbol{\lambda}})$ is continuous in $\boldsymbol{\lambda}$ and, by the strong law of large numbers and Proposition 2.4 it reaches its only minimum, namely zero, at $\boldsymbol{\lambda}_0$, finishing the proof. □

Once consistency has been shown, the differentiability of the statistic considered allows us to obtain $\sqrt{n}$-consistency, at little extra cost: For $\boldsymbol{\lambda} \in \Lambda$, $c \in R^d$ and a real, positive definite symmetric, $d \times d$ matrix M, let $\theta = (\boldsymbol{\lambda}, c, M)$ (we will think of θ as a p-dimensional parameter for $p = 2d + d(d+1)/2)$ and let $h_\theta : R^d \times R^d \to R$, be given by

$$h_\theta(x,y) = \int_{R^d} \sin(u^t M^{-1/2}(x^{(\boldsymbol{\lambda})} - c)) \sin(u^t M^{-1/2}(y^{(\boldsymbol{\lambda})} - c)) w(u) du.$$

Put $\theta_0 = (\boldsymbol{\lambda}_0, \boldsymbol{\mu}_0, \Sigma_0)$ and $\hat{\theta} = (\hat{\boldsymbol{\lambda}}, \bar{\boldsymbol{X}}_{\hat{\boldsymbol{\lambda}}}, S_{\hat{\boldsymbol{\lambda}}})$. Let $R_n(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} h_\theta(\boldsymbol{X}_i, \boldsymbol{X}_j)$ and

$$R(\theta) = ER_n(\theta) = \int_{R^d} h_\theta(x,y) dP(x) dP(y).$$

For a function $h : R^d \times R^d \to R$, not in the collection $\{h_\theta, \theta \in \mathcal{N}\}$, let

$$R_n(h) = \frac{1}{n(n-1)} \sum_{i \neq j} h(\boldsymbol{X}_i, \boldsymbol{X}_j), \qquad \text{and} \qquad R(h) = ER_n(h).$$

Applying Proposition 2.4 we have that, in a neighborhood $\mathcal{N}$ of θ_0, $R(\theta)$ has a unique minimum, namely zero, attained at $\theta = \theta_0$. From the previous theorem we can restrict our attention to $\theta \in \mathcal{N}$. We can assume as well, that the approximations used in the proof of the following proposition are valid in $\mathcal{N}$. $R(\theta)$ is clearly twice differentiable at θ_0. Call V_R the matrix of second derivatives of $R(\theta)$ with respect to the coordinates of θ. We need the following

ASSUMPTION 4.4. The probability law P is such that V_R is positive definite.

Then, we have

PROPOSITION 4.3. *Under the assumptions of Theorem* 4.1 *and Assumption* 4.4, $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 = O_p(1/\sqrt{n})$.

PROOF. Differentiation with respect to θ under the integral sign, can be used repeatedly, in the definition of h_θ, to get that there exist a vector of p functions on R^{2d}, Δ, and a $p \times p$ matrix of functions, $V_{h,\theta}$, (depending on θ), such that for $\theta \in \mathcal{N}$,

$$h_\theta(x,y) = h_{\theta_0}(x,y) + (\theta - \theta_0)^t \Delta(x,y) + \frac{1}{2}(\theta - \theta_0)^t V_{h,\theta^*}(x,y)(\theta - \theta_0), \tag{4.10}$$

where θ^* is a point falling in the segment that joins θ to θ_0. A lengthy but elementary calculation, using our moment assumptions, shows that for θ in $\mathcal{N}$, each entry of $V_{h,\theta^*}(x,y)$ is bounded, in absolute value, by a function $H(x,y)$ in

$L^2(P \otimes P)$. Let $\gamma_n(\theta) = \sqrt{n}(R_n(\theta) - R(\theta))$. Define likewise $\gamma_n(h)$ for functions not of the form h_θ. Using (4.10) and Assumption 4.4, we have

$$
\begin{aligned}
R_n(\theta) &= \frac{\gamma_n(\theta)}{\sqrt{n}} + R(\theta) \\
&= \frac{\gamma_n(\theta_0)}{\sqrt{n}} + (\theta - \theta_0)^t \frac{\gamma_n(\Delta)}{\sqrt{n}} \\
&\quad + \frac{1}{2\sqrt{n}}(\theta - \theta_0)^t \gamma_n(V_{h,\theta^*})(\theta - \theta_0) + R(\theta) \\
&\geq R_n(\theta_0) + O_p(\|\theta - \theta_0\|/\sqrt{n}) \\
&\quad + O_p(\|\theta - \theta_0\|^2/\sqrt{n}) + k\|\theta - \theta_0\|^2,
\end{aligned}
\tag{4.11}
$$

for some positive constant k. γ_n applied to a vector (or matrix) of functions is understood to mean the vector (matrix) obtained by applying γ_n to each coordinate (entry). The second term in the right hand side of last inequality comes from the application of the usual Central Limit Theorem for U-statistics (see Randles and Wolfe (1979), Theorem 3.3.13, for instance) to the fixed functions forming the vector Δ, while the $O_p(\|\theta - \theta_0\|^2/\sqrt{n})$ term corresponds to the application of that same theorem to the function $H(x, y)$ introduced above. Given the differentiability of h_θ and $R(\theta)$ there is no need to use empirical processes results to get these (or the following) inequalities. On the other hand, by definition of $\hat{\boldsymbol{\lambda}}$,

$$
\begin{aligned}
R_n(\hat{\theta}) = W_n(\hat{\boldsymbol{\lambda}}) + O_p(1/n) &\leq W_n(\boldsymbol{\lambda}_0) + O_p(1/n) \\
&\leq R_n(\theta_0) + O_p(1/n).
\end{aligned}
\tag{4.12}
$$

The last inequality comes from the $\sqrt{n}$-consistency of $\bar{\boldsymbol{X}}_{\boldsymbol{\lambda}_0}$ and $S_{\boldsymbol{\lambda}_0}$, which is easy to obtain ($\boldsymbol{\lambda}_0$ is fixed), and an approximation argument like the one given in inequality (4.5) and the lines after it. Assume, by the previous theorem, that $\hat{\theta}$ is in $\mathcal{N}$ and combine (4.11) and (4.12) to get

$$
\begin{aligned}
O_p(1/n) &\geq R_n(\hat{\theta}) - R_n(\theta_0) \\
&\geq k\|\hat{\theta} - \theta_0\|^2 + O_p(\|\hat{\theta} - \theta_0\|/\sqrt{n}) + o_p(\|\hat{\theta} - \theta_0\|^2).
\end{aligned}
$$

With this last inequality, use the *completion of the square* trick in the proof of Theorem 1, page 126, of Sherman (1993) to finish the proof of $\sqrt{n}$-consistency for $\hat{\theta}$, which implies, of course, the $\sqrt{n}$-consistency of $\hat{\boldsymbol{\lambda}}$. □

Technical difficulties, mainly the fact that $R_n(\hat{\theta})$ might not fall within $o_p(1/n)$ of minimizing $R_n(\theta)$ (see the argument in the proof of Theorem 2 of Sherman (1993), or in the proof of Theorem 5, Section VII.1, Pollard (1984)) prevent us, at the present moment, from pushing the argument in order to prove asymptotic normality of $\hat{\boldsymbol{\lambda}}$. We plan to address this topic in future research. Two important, immediate applications of Gaussian asymptotics, would be to provide analytical tools for optimal selection of the constant A, as well as methods for building confidence intervals and tests.

Acknowledgements

The authors wish to thank the editor and two referees for careful reading of the manuscript. Anonymous referee "A" was particularly helpful in detecting both mathematical and typographical errors, and bringing our attention to relevant references. His suggestions for clarifying and shortening some of the proofs were incorporated into the present, much improved, form of the manuscript.

References

Andrews, D. F. and Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, Springer, New York.

Andrews, D. F., Gnanadesikan, R. and Warner, J. L. (1971). Transformations of multivariate data, *Biometrics*, **27**, 825–840.

Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes, *Ann. Probab.*, **21**, 1494–1542.

Assouad, P. (1983). Densité et dimension, *Ann. Inst. Fourier Grenoble*, **33**, 233–282.

Berry, D. A. (1987). Logarithmic transformation in anova, *Biometrics*, **3**, 39–52.

Boos, D. D. (1982). A test for asymmetry associated with a Hodges-Lehmann estimator, *J. Amer. Statist. Assoc.*, **77**, 647–651.

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions, *J. Multivariate Anal.*, **11**, 368–385.

Csörgő, S. (1986). Testing for normality in arbitrary dimension, *Ann. Statist.*, **14**, 708–723.

Csörgő, S. and Heathcote, C. R. (1987). Testing for symmetry, *Biometrika*, **74**, 177–184.

Devlin, S. J., Gnanadesikan, R. and Kettenring, J. R. (1976). Some multivariate applications of elliptical distributions, *Essays in Probability and Statistics* (eds. S. Ideka, T. Hayakawa, H. Hudimoto, M. Okamoto, M. Siotani and S. Yamaoto), 365–395, Shink Tsusho Co., Ltd. Tokyo.

Dudley, R. M. (1984). A course on empirical processes, Ecole d'Eté de Probabilités de Saint-Flour, XII-1982, *Lecture Notes in Math.*, **1097**, 1–142. Springer, New York.

Dudley, R. M. (1987). Universal Donsker classes and metric entropy, *Ann. Probab.*, **15**, 1306–1326.

Fang, K. T. and Anderson, T. W. (eds.) (1990). *Statistical Inference in Elliptically Contoured and Related Distributions*, Allerton Press, New York.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). Symmetric multivariate and related distributions, *Monographs Statist. Appl. Probab.*, **36**, Chapman and Hall, London.

Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications, *Ann. Statist.*, **5**, 88–97.

Ghosh, S. and Ruymgaart, F. (1992). Applications of empirical characteristic functions in some multivariate problems, *Canad. J. Statist.*, **20**(4), 429–440.

Hinkley, D. V. (1977). On quick choice of power transformation, *Appl. Statist.*, **26**, 67–68.

Loève, M. M. (1955). *Probability Theory: Foundations, Random Sequences*, Van Nostrand, New York.

Nakamura, M. and Ruppert, D. (1990). Semi-parametric estimation of symmetrizing transformations with application to the shifted power transformation (unpublished manuscript).

Nelson, C. H., Cox, D. D. and Ndjuenga, J. (1989). Mean variance portfolio choice: a test for elliptical symmetry, Tech. Report, No. 41, Department of Statistics, University of Illinois.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*, Wiley, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator, *Econometrica*, **61**, 123–137.

Shohat, J. A. and Tamarkin, J. D. (1943). *The Problem of Moments*, Mathematical Surveys, Number 1, American Mathematical Society, Rhode Island.

Taylor, J. M. G. (1985). Power transformations to symmetry, *Biometrika*, **72**, 145–152.

Velilla, S. (1993). A note on the multivariate Box-Cox transformation to normality, *Statist. Probab. Lett.*, **17**, 259–263.