TRANSFORMATIONS TO SYMMETRY BASED ON THE PROBABILITY WEIGHTED CHARACTERISTIC FUNCTION

SIMOS G. MEINTANIS AND GILLES STUPFLER

We suggest a nonparametric version of the probability weighted empirical characteristic function (PWECF) introduced by Meintanis et al. [10] and use this PWECF in order to estimate the parameters of arbitrary transformations to symmetry. The almost sure consistency of the resulting estimators is shown. Finite–sample results for i.i.d. data are presented and are subsequently extended to the regression setting. A real data illustration is also included.

Keywords: characteristic function, empirical characteristic function, probability weighted moments, symmetry transformation

Classification: 62G10, 62G20

1. INTRODUCTION

Transformations are applied on given data sets in order to facilitate statistical inference. These transformations are often used so as to induce finite moments and light tails and/or symmetry. This is important as it is common knowledge that certain statistical procedures are applicable or perform well only under such assumptions. Apart from that, symmetry has definite advantages for identification and consistency of location estimators with i.i.d. data, as well as in the context of regression where Bickel [1] and Newey [11] study the existence of adaptive and efficient regression estimators under symmetric errors. The reader is referred to Chapter 6 of Horowitz [7] for a nice review of transformations in regression and other related models. Lately the symmetry assumption has also been invoked for the consistency and efficiency of the quasi maximum likelihood estimator (QMLE) in GARCH models; see González–Rivera and Drost [6] and Newey and Steigerwald [12]. Finally, we mention that power transformations have recently been used by Savchuk and Schick [16] in order to improve the rate of convergence of the classical Parzen-Rosenblatt (Parzen [13]; Rosenblatt [15]) estimator of the probability density function.

The purpose of this paper is to suggest a procedure by means of which a sample from an unknown distribution is reduced to a sample from a symmetric distribution. To this end we employ the notion of the probability weighted empirical characteristic function (PWECF), introduced recently in Meintanis et al. [10]. However, the PWECF used

in Meintanis et al. [10] is defined in an entirely parametric context and it is therefore not appropriate when pursuing nonparametric inference. In what follows we suggest a nonparametric version of the PWECF and use this quantity in order to estimate the parameters of a transformation to symmetry. The remainder of this work is outlined as follows. In Section 2 we recall some properties of the PWCF and the nonparametric PWECF is introduced. In Section 3 we introduce the new estimation procedure which is based on an appropriate functional of this PWECF; the method is related to those in Yeo and Johnson [21] and Yeo et al. [22]. The strong consistency of our estimator is given in Section 4, while in Section 5 the finite–sample properties of the method are investigated by means of a simulation study. A real data example is included in Section 6 while some auxiliary results and their proofs are deferred to the Appendix.

2. THE NONPARAMETRIC PWECF

Let X denote an arbitrary random variable with an absolutely continuous distribution function $F(x) = \mathbb{P}(X \leq x)$. For $\gamma \geq 0$, the probability weighted characteristic function (PWCF) of X is defined by

$$\varphi(t; \gamma) := \mathbb{E}\left[W(X; \gamma t)e^{itX}\right] = \int_{-\infty}^{\infty} W(x; \gamma t)e^{itx}\mathrm{d}F(x), \ t \in \mathbb{R}, \tag{2.1}$$

where $W(x; s) := [F(x)(1 - F(x))]^{|s|}$. It is noteworthy that the PWCF of X has various useful properties similar to those of the characteristic function (CF) of X, see Meintanis et al. [10]; in particular, a distribution function which is symmetric around zero must yield a real-valued PWCF, see property P5 there, and this will be the basis of our transformation procedure in Section 3. The fact that for $\gamma > 0$ the PWCF is no longer a Fourier transform, however, makes it difficult to prove strong distributional results such as a one-to-one correspondence between PWCFs and probability distributions. Interestingly though, in the context of location-scale families, which was the original framework of Meintanis et al. [10], we may state and prove such a result:

Proposition 2.1. Assume that F_1 and F_2 belong to some location-scale family, namely

$$\forall x \in \mathbb{R}, \ F_1(\sigma_1 x + \mu_1) = F_2(\sigma_2 x + \mu_2) = G(x)$$

where G is an absolutely continuous distribution function and $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$. Then, for any $\gamma > 0$, F_1 and F_2 yield the same PWCF if and only if $F_1 = F_2$.

P r o o f. Let $\varphi_{\mu,\sigma}$ be the PWCF related to $F_{\mu,\sigma}(x) := G((x - \mu)/\sigma)$. Since

$$\varphi_{\mu,\sigma}(t; \gamma) = \int_{-\infty}^{\infty} [F_{\mu,\sigma}(x)(1 - F_{\mu,\sigma}(x))]^{\gamma|t|} e^{itx} \mathrm{d}F_{\mu,\sigma}(x),$$

we get by the change of variables $x = \sigma y + \mu$:

$$\varphi_{\mu,\sigma}(t; \gamma) = \int_{-\infty}^{\infty} [G(y)(1 - G(y))]^{\gamma|t|} e^{i(\sigma t)y + \mu} \mathrm{d}G(y) = e^{it\mu} \varphi_{0,1}(\sigma t; \gamma/\sigma).$$

Assume now that F_1 and F_2 yield the same PWCF, with $\sigma_1 \neq \sigma_2$. Then

$$e^{it\mu_1}\varphi_{0,1}(\sigma_1 t; \gamma/\sigma_1) = e^{it\mu_2}\varphi_{0,1}(\sigma_2 t; \gamma/\sigma_2), \ t \in \mathbb{R}, \tag{2.2}$$

which up to reparametrization is equivalent to

$$\varphi_{0,1}(T; \Gamma) = e^{itM}\varphi_{0,1}(cT; \Gamma/c), \ T \in \mathbb{R},$$

for some $M \in \mathbb{R}$, $c \neq 1$ and $\Gamma > 0$. Without loss of generality, we assume in what follows that $c > 1$; in this case, a straightforward proof by induction shows that for any positive integer m:

$$|\varphi_{0,1}(T; \Gamma)| = |\varphi_{0,1}(c^m T; \Gamma/c^m)|, \ T \in \mathbb{R}.$$

Observe now that $\varphi_{0,1}(0; \Gamma) = 1$ and for any $T > 0$,

$$\begin{aligned}
\varphi_{0,1}(c^m T; \Gamma/c^m) &= \int_{-\infty}^{\infty} [G(y)(1 - G(y))]^{\Gamma|T|} e^{i(c^m T)y} g(y) \, \mathrm{d}y \\
&= \frac{1}{T} \int_{-\infty}^{\infty} [G(z/T)(1 - G(z/T))]^{\Gamma|T|} g(z/T) e^{ic^m z} \, \mathrm{d}z
\end{aligned}$$

where g is the probability density function related to G. The right-hand side is, up to a constant, the Fourier transform of the integrable function

$$z \mapsto [G(z/T)(1 - G(z/T))]^{\Gamma|T|} g(z/T),$$

evaluated at the point c^m. Since $c^m \to \infty$ as $m \to \infty$, the Riemann-Lebesgue lemma states that this expression must converge to 0 as $m \to \infty$. As a conclusion,

$$\varphi_{0,1}(0; \Gamma) = 1 \ \text{ and } \ \varphi_{0,1}(T; \Gamma) = 0, \ T > 0.$$

This is a contradiction since $T \mapsto \varphi_{0,1}(T; \Gamma)$ is continuous, see property P7 in Meintanis et al. [10]. Hence $\sigma_1 = \sigma_2$, and thus $e^{it\mu_1} = e^{it\mu_2}$ for all $t \in \mathbb{R}$ by (2.2), which entails $\mu_1 = \mu_2$. The proof is complete. $\qquad\square$

Remark 2.2. The location–scale context may actually be dropped under additional moment hypotheses, such as the existence of the moment-generating function of F_1 and F_2 in a neighborhood of 0, by using analytic continuation. In any case, if the PWCF is unique, it can be used to assess symmetry around zero: It is indeed clear that for any t and γ, the PWCF of $-X$ is equal to $\varphi(-t; \gamma)$, and that $\varphi(-t; \gamma) = \overline{\varphi}(t; \gamma)$, where $\overline{z}$ denotes the complex conjugate of z. Now if the PWCF of X is real-valued, this entails $\varphi(-t; \gamma) = \varphi(t; \gamma)$ and thus X and $-X$ have the same PWCF, whence the fact that the distribution function of X is symmetric around zero.

While Meintanis et al. [10] estimated the PWCF in a parametric way, it is interesting to consider the case where F is completely unknown. In this context, it is a natural idea to

define an estimator of the PWCF in an entirely *nonparametric* way. To this end notice
that the PWCF in (2.1) may be written as

$$\varphi(t;\gamma) = \int_0^1 [x(1-x)]^{\gamma|t|} e^{itQ(x)} \mathrm{d}x, \tag{2.3}$$

where $Q(x) = \inf\{t \in \mathbb{R} | F(t) \geq x\}$ denotes the quantile function of X.

In view of (2.3) we suggest the following nonparametric estimator of the PWCF:

$$\widehat{\varphi}_n(t;\gamma) = \int_0^1 [x(1-x)]^{\gamma|t|} e^{it\widehat{Q}_n(x)} \mathrm{d}x, \tag{2.4}$$

with $\widehat{Q}_n(x)$ denoting the empirical quantile function. We shall call $\widehat{\varphi}_n(t;\gamma)$ the prob-
ability weighted empirical characteristic function (PWECF), and for the purpose of
estimation we will use

$$\forall k \in \{1, ..., n\}, \ \forall x \in \left(\frac{k-1}{n}, \frac{k}{n}\right], \ \widehat{Q}_n(x) = X_{k:n},$$

where $X_{1:n} \leq \cdots \leq X_{n:n}$ denote the order statistics corresponding to independent copies
$X_1, \ldots, X_n$ of the random variable X.

3. L2–TYPE PROCEDURES FOR SYMMETRY TRANSFORMATION

The problem we shall consider is to estimate the parameters of a given transformation
which, if applied on the original nonsymmetrically distributed observations $X_1, \ldots, X_n$,
yields transformed observations that are approximately symmetrically distributed with
location zero. To this end, write $\vartheta = (\delta, \lambda) \in \Theta \subset \mathbb{R} \times \mathbf{\Lambda}$ for the transformation
parameter–vector, where δ denotes location and λ denotes the shape parameter which
is assumed to lie in a subset $\mathbf{\Lambda}$ of the real line. For $\vartheta = (\delta, \lambda) \in \Theta$, we let $Q_Z(\cdot; \vartheta)$ be
the quantile function of the transformed random variable $Z(\vartheta) = \psi(X; \lambda) - \delta$, where ψ
is a specific transformation family, and we define

$$\mathcal{S}(t;\gamma;\vartheta) = \int_0^1 [x(1-x)]^{\gamma|t|} \sin(tQ_Z(x;\vartheta)) \, \mathrm{d}x,$$

the imaginary part of the PWCF of $Z(\vartheta)$. It is thus a consequence of Remark 1 that
if the transformed random variable Z has a symmetric distribution around zero then
$\mathcal{S}(t;\gamma;\vartheta) = 0$ for all $t \in \mathbb{R}$, or equivalently $\int_{-\infty}^{\infty} \mathcal{S}^2(t;\gamma;\vartheta) \, \mathrm{d}t = 0$.

This observation is the basic idea we need to build our estimator: we introduce $Z_k(\vartheta) = \psi(X_k; \lambda) - \delta$, we let $\widehat{Q}_{Z,n}(x;\vartheta)$ be the empirical quantile function related to $Z_1(\vartheta), \ldots, Z_n(\vartheta)$
and we define

$$\widehat{\mathcal{S}}_n(t;\gamma;\vartheta) = \int_0^1 [x(1-x)]^{\gamma|t|} \sin(t\widehat{Q}_{Z,n}(x;\vartheta)) \, \mathrm{d}x,$$

the imaginary part of the PWECF of $Z_1(\vartheta), \ldots, Z_n(\vartheta)$. Then $\widehat{\mathcal{S}}_n(t;\gamma;\vartheta)$ is the empirical
counterpart of $\mathcal{S}(t;\gamma;\vartheta)$. We suggest to estimate the true value $\vartheta_0 = (\delta_0, \lambda_0)$ (see

Section 4 for a discussion of the uniqueness of this parameter) by $\widehat{\vartheta}_n$, where

$$\widehat{\vartheta}_n = \underset{\vartheta \in \Theta}{\arg\min}\, \Delta_n(\gamma; \vartheta), \quad \text{with}\quad \Delta_n(\gamma; \theta) = \int_{-\infty}^{\infty} \widehat{\mathcal{S}}_n^2(t; \gamma; \vartheta)\, \mathrm{d}t. \tag{3.1}$$

Remark 3.1. The PWCF $\varphi(t; \gamma)$ and PWECF $\widehat{\varphi}_n(t; \gamma)$ of a random variable X are such that $|\varphi(t; \gamma)| \leq (1/4)^{\gamma|t|}$ and $|\widehat{\varphi}_n(t; \gamma)| \leq (1/4)^{\gamma|t|}$ for every $(t, \gamma) \in \mathbb{R} \times \mathbb{R}^+$. As a consequence, for any ϑ, the integral $\Delta_n(\vartheta)$ is positive and finite.

Remark 3.2. Notice that while we write $\widehat{\vartheta}_n$, the estimator implicitly depends on the value of γ and therefore we have essentially a family of estimators $\{\widehat{\vartheta}_n(\gamma),\, 0 < \gamma < \infty\}$ indexed by γ.

Remark 3.3. Possible choices for the transformation family ψ are the Box-Cox transformation (see Box and Cox [3]), a family introduced by Burbidge et al. [4] as well as the recently introduced method of Yeo and Johnson [19]. Note that while the popular Box-Cox transformation,

$$\psi(x; \lambda) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\[2mm] \log x & \text{if } \lambda = 0, \end{cases}$$

applies only to positive random variables (if λ is not a nonzero integer), its modifications suggested by Manly [9], John and Draper [8] and Bickel and Doksum [2] were designed to allow negative values as well.

A favorable feature of the specific definition of the nonparametric PWECF in (2.4) is that it leads to a criterion in (3.1) which is convenient from the computational point of view. To see this notice that from (2.4) it is straightforward to compute the imaginary part of the PWECF of $Z_1(\vartheta), \dots, Z_n(\vartheta)$ as

$$\widehat{\mathcal{S}}_n(t; \gamma; \vartheta) = \sum_{k=1}^{n} \upsilon_{k,n}(t; \gamma) \sin(t Z_{k:n}(\vartheta)) \quad \text{with}\quad \upsilon_{k,n}(t; \gamma) = \int_{(k-1)/n}^{k/n} [x(1-x)]^{\gamma|t|}\, \mathrm{d}x.$$

Then the criterion statistic in (3.1) follows by direct calculation as

$$\Delta_n(\gamma; \vartheta) = \frac{1}{2} \sum_{j,k=1}^{n} \left(I_{jk}^-(\gamma; \vartheta) - I_{jk}^+(\gamma; \vartheta) \right)$$

where $I_{jk}^-(\gamma; \vartheta) := I(j, k; \gamma; Z_{j:n}(\vartheta) - Z_{k:n}(\vartheta))$ and $I_{jk}^+(\gamma; \vartheta) := I(j, k; \gamma; Z_{j:n}(\vartheta) + Z_{k:n}(\vartheta))$ with

$$I(j, k; \gamma; x) = \int_{-\infty}^{\infty} \upsilon_j(t; \gamma)\upsilon_k(t; \gamma) \cos(tx)\, \mathrm{d}t.$$

4. STRONG CONSISTENCY OF THE ESTIMATOR

Here, we assume that $\gamma > 0$ and that the following hold:

(A_1) The support $\mathcal{D}$ of the distribution of X is an open interval and F is continuous and strictly increasing on $\mathcal{D}$.

(A_2) The transformation family ψ is such that $(x, \lambda) \mapsto \psi(x; \lambda)$ is continuous on $\mathcal{D} \times \boldsymbol{\Lambda}$.

(A_3) For all $\lambda \in \boldsymbol{\Lambda}$, $x \mapsto \psi(x; \lambda)$ is strictly increasing.

Assumption (A_2) is also used in Yeo and Johnson [20], while (A_3) means that the family of transformations preserves ordering: if two observations X_1 and X_2 are such that $X_1 < X_2$, then the transformed observations $\psi(X_1; \lambda)$ and $\psi(X_2; \lambda)$ are such that $\psi(X_1; \lambda) < \psi(X_2; \lambda)$. In particular, in this setting, it is straightforward to show that

$$Q_Z(x; \vartheta) = \psi(Q(x); \lambda) - \delta \quad \text{and} \quad \widehat{Q}_{Z,n}(x; \vartheta) = \psi(\widehat{Q}_n(x); \lambda) - \delta. \qquad (4.1)$$

Under these assumptions, we may state a strong consistency result for our estimator:

Theorem 4.1. Assume that (A_1), (A_2) and (A_3) hold. Let Θ be a compact subset of $\mathbb{R}^2$ contained in $\mathbb{R} \times \boldsymbol{\Lambda}$. If, over Θ, there exists a unique global minimum ϑ_0 of the function

$$\vartheta \mapsto \int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta) \, dt$$

then $\widehat{\vartheta}_n \to \vartheta_0$ almost surely.

P r o o f . By Lemma 6.2 in the Appendix,

$$H_n(\vartheta) := \int_{-\infty}^{\infty} \widehat{\mathcal{S}}_n^2(t; \gamma; \vartheta) \, dt \to H(\vartheta) := \int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta) \, dt$$

almost surely, uniformly in $\vartheta \in \Theta$. Recall that

$$\mathcal{S}(t; \gamma; \vartheta) = \int_0^1 [x(1-x)]^{\gamma|t|} \sin(tQ_Z(x; \vartheta)) \, dx.$$

Because for any x the function $\vartheta \mapsto Q_Z(x; \vartheta)$ is continuous and the integrand in $\mathcal{S}(t; \gamma; \vartheta)$ is dominated by the constant 1, the dominated convergence theorem entails that for any t, the function $\vartheta \mapsto \mathcal{S}(t; \gamma; \vartheta)$ is continuous. Furthermore, since for any ϑ, $|\mathcal{S}(t; \gamma; \vartheta)| \leq (1/4)^{\gamma|t|}$ by Remark 3.1, it is again a corollary of the dominated convergence theorem that the function H is continuous as well. The same arguments apply to show that (H_n) is a random sequence of continuous functions. Using Lemma 6.3 concludes the proof. □

The existence of a global minimum of the function $\vartheta \mapsto \int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta) \, dt$ is for instance guaranteed if there exists ϑ_0 such that the distribution of $Z(\vartheta_0)$ is symmetric around 0, in which case $\mathcal{S}(t; \gamma; \vartheta_0) = 0$ for each t and therefore

$$\forall \vartheta \in \Theta, \quad \int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta) \, dt \geq 0 = \int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta_0) \, dt.$$

The uniqueness of one such ϑ_0 is a more challenging problem. The following proposition is a step towards solving this question for a large class of transformations, including those mentioned in Remark 3.3.

Proposition 4.2. Assume that (A_1) holds and that X has a positive median. Let ψ be a family of transformations, satisfying (A_2) and (A_3), such that

$$\forall x > 0, \ \forall \lambda > 0, \ \psi(x; \lambda) = \frac{[f(x)]^\lambda - 1}{\lambda}$$

where f is a positive, continuous and strictly increasing function on $(0, \infty)$. If there exists a pair $(\delta, \lambda) \in \mathbb{R} \times (0, \infty)$ such that $\psi(X; \lambda) - \delta$ is symmetrically distributed around zero, then (δ, λ) is the unique such pair.

P r o o f. Since (A_1) holds and X has a positive median, we have $Q(x) > 0$ for all x in an open neighborhood U of $1/2$. Define $\vartheta = (\delta, \lambda)$; the monotonicity of f then yields $Q_Z(x; \vartheta) = \psi(Q(x); \lambda) - \delta$ for all $x \in U$. In particular, the median of $Z(\vartheta)$, which is symmetrically distributed around zero, has to be 0 and thus $0 = [f \circ Q(1/2)]^\lambda - c(\vartheta)$, where $c(\vartheta) = 1 + \delta\lambda$. In particular, $c(\vartheta)$ is positive and $f \circ Q(1/2) = [c(\vartheta)]^{1/\lambda}$. Besides, it must hold that $Q_Z(1/2 - s; \vartheta) = -Q_Z(1/2 + s; \vartheta)$ for any $s \in (0, 1/2)$ which entails for all $\varepsilon > 0$ small enough:

$$\frac{[f \circ Q(1/2 - \varepsilon)]^\lambda - 1}{\lambda} - \delta = -\left[\frac{[f \circ Q(1/2 + \varepsilon)]^\lambda - 1}{\lambda} - \delta \right]$$

or equivalently:

$$f \circ Q(1/2 - \varepsilon) = \left(2c(\vartheta) - [f \circ Q(1/2 + \varepsilon)]^\lambda \right)^{1/\lambda}. \tag{4.2}$$

Assume now that there exist two pairs $\vartheta_1 = (\delta_1, \lambda_1)$ and $\vartheta_2 = (\delta_2, \lambda_2)$ such that $Z(\vartheta_1)$ and $Z(\vartheta_2)$ are symmetrically distributed around zero. Note that it is enough to show that $\lambda_1 = \lambda_2$. Using (4.2), we obtain for all $\varepsilon > 0$ sufficiently small:

$$\left(2c(\vartheta_1) - [f \circ Q(1/2 + \varepsilon)]^{\lambda_1} \right)^{1/\lambda_1} = \left(2c(\vartheta_2) - [f \circ Q(1/2 + \varepsilon)]^{\lambda_2} \right)^{1/\lambda_2}.$$

Since $f \circ Q(1/2) = [c(\vartheta_1)]^{1/\lambda_1} = [c(\vartheta_2)]^{1/\lambda_2}$ and the function $f \circ Q$ is continuous and strictly increasing, this entails for all $h > 0$ small enough:

$$\left(2c(\vartheta_1) - \left[[c(\vartheta_1)]^{1/\lambda_1} + h \right]^{\lambda_1} \right)^{1/\lambda_1} = \left(2c(\vartheta_2) - \left[[c(\vartheta_2)]^{1/\lambda_2} + h \right]^{\lambda_2} \right)^{1/\lambda_2}.$$

Noting that $[c(\vartheta_1)]^{1/\lambda_1} = [c(\vartheta_2)]^{1/\lambda_2} > 0$, we get that for all $h > 0$ small enough:

$$\left(2 - [1 + h]^{\lambda_1} \right)^{1/\lambda_1} = \left(2 - [1 + h]^{\lambda_2} \right)^{1/\lambda_2}.$$

Taking logarithms and differentiating twice, we obtain for $h > 0$ sufficiently small:

$$\frac{(1 + h)^{\lambda_1 - 2} \left[2(\lambda_1 - 1) + (1 + h)^{\lambda_1} \right]}{[2 - (1 + h)^{\lambda_1}]^2} = \frac{(1 + h)^{\lambda_2 - 2} \left[2(\lambda_2 - 1) + (1 + h)^{\lambda_2} \right]}{[2 - (1 + h)^{\lambda_2}]^2}.$$

Letting $h \downarrow 0$ entails $\lambda_1 = \lambda_2$, which completes the proof.   □

We note that this result requires the median of X to be positive. For some families such as the Bickel–Doksum family (see Bickel and Doksum [2]), also called the "signed power" transformation family:

$$\forall x \in \mathbb{R}, \ \forall \lambda > 0, \ \psi(x; \lambda) = \frac{\text{sgn}(x)|x|^\lambda - 1}{\lambda}, \quad \text{with} \quad \text{sgn}(x) = \left\{ \begin{array}{lll} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \end{array} \right. \quad (4.3)$$

this assumption may actually be dropped, as shown by Corollary 4.3 below. This particular family of transformations, which coincides with the Box-Cox family of transformations for positive values of x and λ, is the one we shall consider in our simulation study.

Corollary 4.3. Let ψ be the Bickel–Doksum family of transformations. Assume that (A_1) holds and that the distribution of X is not symmetric around zero. If there exists a pair $(\delta, \lambda) \in \mathbb{R} \times (0, \infty)$ such that $\psi(X; \lambda) - \delta$ is symmetrically distributed around zero, then (δ, λ) is the unique such pair.

P r o o f. We first note that for any such pair $\vartheta = (\delta, \lambda)$, then $\delta \neq -1/\lambda$. If indeed we had that $\delta = -1/\lambda$, then using (4.3), the random variable $\text{sgn}(X)|X|^\lambda$ would be symmetric. This would imply, for any $x \leq 0$, that

$$\mathbb{P}(X \leq x) = \mathbb{P}(\text{sgn}(X)|X|^\lambda \leq -(-x)^\lambda) = \mathbb{P}(\text{sgn}(X)|X|^\lambda \geq (-x)^\lambda) = \mathbb{P}(X \geq -x).$$

Then X would be symmetrically distributed around zero, which is a contradiction. Moreover, we may assume without loss of generality that the median $Q(1/2)$ of X is nonnegative: if indeed this is not the case then $-X$ has a nonnegative median and, letting $\delta' = -(\delta + 2/\lambda) \neq -1/\lambda$, the random variable

$$\psi(-X; \lambda) - \delta' = -[\psi(X; \lambda) - \delta]$$

is symmetrically distributed around zero. Finally, since (A_1) holds and (A_2) and (A_3) are satisfied for the Bickel–Doksum family, we have $Q_Z(x; \vartheta) = \psi(Q(x); \lambda) - \delta$ by (4.1). Since $Z(\vartheta)$ is symmetrically distributed around zero, we must have $0 = Q(1/2)^\lambda - (1 + \delta\lambda)$. Especially, the median $Q(1/2) = (1 + \delta\lambda)^{1/\lambda}$ of X is positive. Applying Proposition 4.2 concludes the proof.   □

5. A MONTE-CARLO SIMULATION STUDY

5.1. Finite sample performance of the presented technique

In this section, we present the results of a Monte-Carlo study conducted to assess the performance of our method. In what follows, the transformation family considered is the Bickel–Doksum family (4.3). The following estimators are compared:

- our estimator (3.1), denoted by M_γ, with $\gamma \in \{1, 2\}$;

- the estimator

$$\arg\min_{\vartheta\in\Theta} \int_{-\infty}^{\infty} \left[\frac{1}{n}\sum_{k=1}^{n}\sin(tZ_k(\vartheta))\right]^2 e^{-|t|}\mathrm{d}t$$

which corresponds to using the ECF with an exponential weighting function (see Yeo and Johnson [20]), and will be denoted by EECF;

- the Gaussian maximum likelihood estimator (GMLE), assuming that the target symmetric distribution is Gaussian. While this estimator actually attempts to transform to normality, we include it for comparative reasons. The shape estimator is $\widehat{\lambda}$ and the location estimator is $\widehat{\delta}(\widehat{\lambda})$ where

$$\widehat{\lambda} = \arg\max_{\lambda\in\Lambda}\left\{-\frac{n}{2}\log(\widehat{\sigma^2}(\lambda)) - \frac{1}{2}\sum_{k=1}^{n}\frac{(\psi(X_k;\lambda)-\widehat{\delta}(\lambda))^2}{\widehat{\sigma^2}(\lambda)} + (\lambda-1)\sum_{k=1}^{n}\log|X_k|\right\}$$

$$= \arg\max_{\lambda\in\Lambda}\left\{-\frac{n}{2}\log(\widehat{\sigma^2}(\lambda)) + (\lambda-1)\sum_{k=1}^{n}\log|X_k|\right\}$$

with

$$\widehat{\delta}(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\psi(X_k;\lambda) \ \text{ and } \ \widehat{\sigma^2}(\lambda) = \frac{1}{n}\sum_{k=1}^{n}(\psi(X_k;\lambda)-\widehat{\delta}(\lambda))^2.$$

To get a grasp of how these estimators behave in practice, we use the following generating algorithm: for a given $n-$independent sample $Y_1,\ldots,Y_n$ of random copies of a symmetric random variable Y, we pick (known) values of λ and δ and we consider the $n-$independent sample $X_1,\ldots,X_n$ such that $X_k = \tau(Y_k+\delta;\lambda)$ where

$$\tau(y;\lambda) = \mathrm{sgn}(\lambda y+1)|\lambda y+1|^{1/\lambda}$$

is the inverse of the Bickel–Doksum transformation. With this notation, we thus have $\psi(X_k;\lambda)-\delta = Y_k$ which are symmetric random variables and we may apply our various procedures to assess the quality of the estimation of λ and δ in each case. In what follows, λ is picked in the set $\{1/4,1/2,3/4\}$, $\delta=1$ and the symmetric distributions considered are the following:

- $Y = W\exp(hW^2/2)$ with W standard normal, namely Y follows a Tukey$(0,h)$ distribution. The higher is h, the higher is the kurtosis of Y. When $h=0$, Y is standard Gaussian, denoted by N$(0,1)$;

- $Y|V=v$ is Gaussian centered with variance v, where V is Gamma distributed with shape parameter $k>0$ and unit scale. This distribution is denoted by Variance $\Gamma(k,1)$;

- Y follows a symmetric stable distribution with shape parameter α, location parameter zero and unit scale. This distribution is denoted by Stable$(\alpha,0,1)$.

In each case, the estimation is carried out on 1000 samples of size $n = 100$ and we compute the mean L^1-error (i. e. the mean absolute deviation) related to $\widehat{\lambda}$ and $\widehat{\delta}$. We

display in Table 1 the mean L^1−error for λ and δ as well as the standard deviation of the estimates.

It appears from these tables that our M_γ estimator performs fairly well in all cases for both values of γ. In particular, it performs better than the EECF method at estimating λ, and equally well at estimating δ except when the tail is very heavy as is the case for the Stable$(1, 0, 1)$ distribution. Furthermore, while the GMLE method appears superior at estimating λ when the tail is light or when the distribution is leptokurtic, our technique is comparable to and sometimes better than this method when $\lambda \geq 1/2$ and the tail is heavy (for instance, the stable distribution) or if the distribution is platykurtic (as is the case for the Tukey$(0, 3/4)$ distribution). Finally, it can be seen by computing the sum of the mean L^1−errors that overall, our technique competes well with the GMLE method and outperforms the EECF technique. In this connection we would like to stress that our method does not involve the choice of a weighting function unlike what must be done when using the conventional ECF.

We conclude this section by highlighting how our technique may be used prior to a statistical analysis of a data set. The context is the following: We assume that we observe a sample of independent copies $(X_1, Z_1), \ldots, (X_n, Z_n)$ of a random pair (X, Z) such that for some (λ, δ):

$$\psi(X; \lambda) - \delta = m_0 + m_1 Z + \varepsilon$$

where ψ is a given family of transformations, m_0, $m_1 \in \mathbb{R}$ and (Z, ε) are such that Z and ε are two independent random variables which both possess symmetric around zero distributions. The goal is to estimate the parameters m_0 and m_1. In the framework of linear regression, one can think of m_0 as the intercept and m_1 as the slope, Z is the regressor and ε is the random error. For a nice account of transformations in the context of regression the reader is referred to Chen et al. [5]. Of course, a first, crucial task is to estimate (λ, δ) as accurately as possible so as to recover enough information on the hidden regression setting. Note that

$$\psi(X; \lambda) - (\delta + m_0) = m_1 Z + \varepsilon$$

so that without loss of generality, we may assume that the intercept m_0 is zero. Observe then that the right-hand side is a symmetric random variable, which makes it possible to implement our method in order to estimate (λ, δ). A possible procedure is as follows:

1. estimate (λ, δ) by a symmetry procedure, such as our PWECF–based technique or the GMLE;

2. if $(\widehat{\lambda}, \widehat{\delta})$ is the estimate, compute the transformed observations $\widehat{Y}_k = \psi(X_k; \widehat{\lambda}) - \widehat{\delta}$;

3. choose an estimation procedure for the regression parameters (m_0, m_1), such as ordinary least squares (OLS) and use the random pairs $(Z_k, \widehat{Y}_k)$ for the estimation.

In fact, a robust method such as the Theil–Sen estimator (Theil [18]; Sen [17]), may be preferred to the basic OLS estimator at the final step because nothing is known regarding the moments of ε. In this connection, a small simulation study which we do not report here tends to indicate that the Theil–Sen estimator combined with our technique works better than the classical GMLE–OLS method under a heavy–tailed error distribution.

6. REAL DATA EXAMPLES

In this section, we showcase our method on a set of real data. We consider the daily closing values (p_t) of the DAX index from October 1, 2007 to April 1, 2009, and our data is the daily percentage of return $r_t = 100(p_t/p_{t-1} - 1)$ of size $n = 378$. During this period of time, European markets generally followed a downward trend, so that we can expect these percentages to have a left-skewed distribution.

We compare the results found with the M_1 and M_2 methods with what we find when using the GMLE method. In Table 2, we summarize the results, along with the mean, variance, skewness and kurtosis of the transformed data set (using the Bickel–Doksum family) with the estimated parameters given by each method. Histograms of the raw and transformed data sets are given in Figure 1.
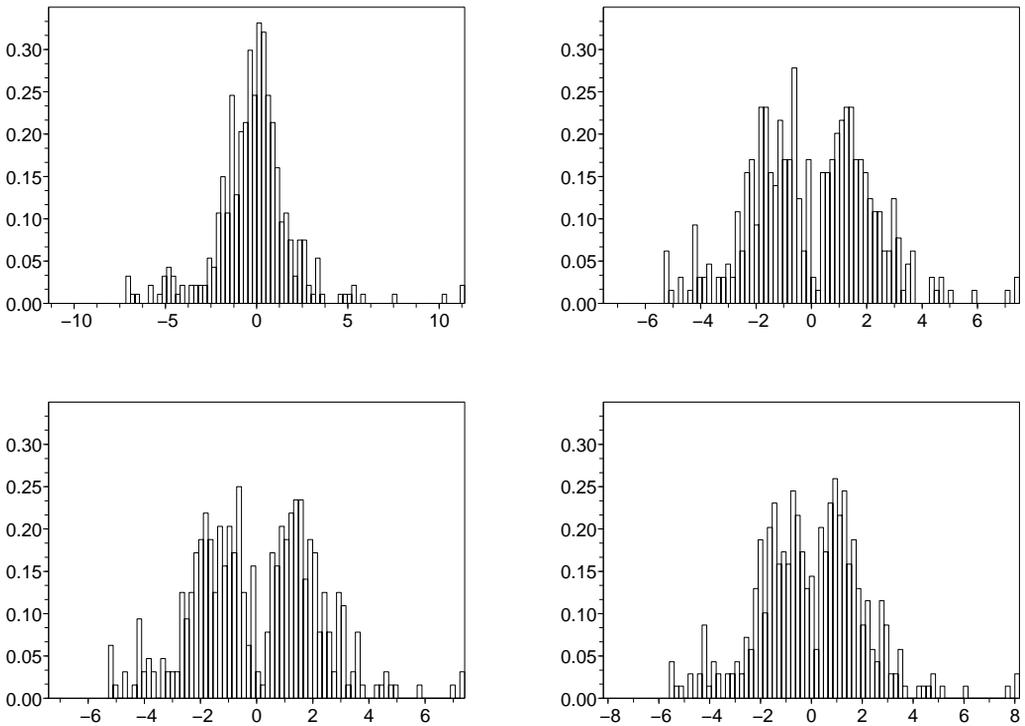
Fig. 1. DAX daily data set, top left: original data, top right: data transformed with the parameters obtained by the M_1 technique, bottom left: data transformed with the parameters obtained by the M_2 technique, bottom right: data transformed with the parameters obtained by the GMLE technique.

Case $\lambda = 1/4$	M_1		M_2		EECF		GMLE	
N(0, 1)	0.112	(0.141)	0.104	(0.131)	0.116	(0.145)	0.0774	(0.0972)
	0.124	(0.156)	0.120	(0.151)	0.119	(0.149)	0.105	(0.131)
Tukey(0, 3/4)	0.144	(0.199)	0.135	(0.195)	0.150	(0.183)	0.0853	(0.0635)
	0.105	(0.138)	0.123	(0.162)	0.0873	(0.113)	0.507	(0.567)
Variance $\Gamma(1, 1)$	0.124	(0.151)	0.116	(0.143)	0.151	(0.181)	0.106	(0.128)
	0.0807	(0.102)	0.0792	(0.100)	0.0848	(0.108)	0.0881	(0.112)
Variance $\Gamma(4, 1)$	0.0678	(0.0884)	0.0562	(0.0737)	0.0876	(0.114)	0.0390	(0.0504)
	0.173	(0.220)	0.167	(0.213)	0.173	(0.218)	0.169	(0.212)
Stable(7/4, 0, 1)	0.132	(0.185)	0.125	(0.176)	0.142	(0.173)	0.131	(0.156)
	0.117	(0.162)	0.113	(0.162)	0.111	(0.141)	0.124	(0.157)
Stable(1, 0, 1)	0.213	(0.156)	0.207	(0.203)	0.196	(0.0918)	0.110	(0.111)
	0.140	(0.0979)	0.175	(0.125)	0.133	(0.0645)	0.336	(0.449)
Case $\lambda = 1/2$	M_1		M_2		EECF		GMLE	
N(0, 1)	0.133	(0.169)	0.121	(0.156)	0.140	(0.179)	0.0835	(0.105)
	0.125	(0.160)	0.122	(0.156)	0.121	(0.153)	0.105	(0.129)
Tukey(0, 3/4)	0.146	(0.188)	0.147	(0.194)	0.212	(0.240)	0.191	(0.0891)
	0.105	(0.135)	0.120	(0.161)	0.0778	(0.0976)	0.530	(0.438)
Variance $\Gamma(1, 1)$	0.147	(0.187)	0.138	(0.176)	0.207	(0.245)	0.126	(0.156)
	0.0887	(0.113)	0.0871	(0.110)	0.100	(0.122)	0.0996	(0.122)
Variance $\Gamma(4, 1)$	0.0794	(0.116)	0.0601	(0.0776)	0.0974	(0.139)	0.0418	(0.0531)
	0.159	(0.201)	0.169	(0.215)	0.165	(0.213)	0.166	(0.211)
Stable(7/4, 0, 1)	0.138	(0.175)	0.131	(0.163)	0.174	(0.212)	0.147	(0.184)
	0.110	(0.141)	0.107	(0.136)	0.112	(0.142)	0.131	(0.168)
Stable(1, 0, 1)	0.222	(0.258)	0.231	(0.291)	0.420	(0.128)	0.201	(0.160)
	0.200	(0.152)	0.264	(0.200)	0.0415	(0.0656)	0.266	(0.353)
Case $\lambda = 3/4$	M_1		M_2		EECF		GMLE	
N(0, 1)	0.153	(0.194)	0.138	(0.175)	0.169	(0.220)	0.0940	(0.119)
	0.130	(0.168)	0.125	(0.161)	0.126	(0.161)	0.102	(0.129)
Tukey(0, 3/4)	0.155	(0.197)	0.155	(0.192)	0.239	(0.278)	0.308	(0.124)
	0.103	(0.131)	0.111	(0.143)	0.0862	(0.107)	0.561	(0.389)
Variance $\Gamma(1, 1)$	0.167	(0.215)	0.155	(0.199)	0.243	(0.304)	0.136	(0.165)
	0.0917	(0.115)	0.0885	(0.111)	0.108	(0.130)	0.101	(0.120)
Variance $\Gamma(4, 1)$	0.0824	(0.103)	0.0748	(0.0944)	0.101	(0.143)	0.0560	(0.0709)
	0.184	(0.232)	0.182	(0.229)	0.180	(0.228)	0.178	(0.222)
Stable(7/4, 0, 1)	0.156	(0.201)	0.154	(0.195)	0.212	(0.263)	0.156	(0.195)
	0.119	(0.155)	0.114	(0.146)	0.121	(0.150)	0.133	(0.170)
Stable(1, 0, 1)	0.282	(0.325)	0.273	(0.319)	0.545	(0.213)	0.307	(0.204)
	0.201	(0.177)	0.215	(0.181)	0.0690	(0.0926)	0.233	(0.297)

Tab. 1. Mean L^1-errors for the estimates; in each case, $\delta = 1$, first line: mean L^1-errors for the parameter λ, second line: mean L^1-errors for the parameter δ. Between brackets: sample standard deviation of the estimates.

	$\widehat{\lambda}$	$\widehat{\delta}$	Mean	Std. deviation	Skewness	Kurtosis
Raw data	1	-1	-0.148	2.208	0.641	8.702
M_1	0.629	-1.756	0.00568	2.173	0.152	3.244
M_2	0.611	-1.808	0.00883	2.196	0.142	3.086
GMLE	0.722	-1.541	0	2.101	0.221	4.193

Tab. 2. Estimated values of λ and δ for our real data set.

In Table 2, we see that in each case, the absolute value of the skewness of the transformed data set is smaller than that of the raw data set. Note that while the value of the skewness of the daily DAX data set is positive and thus seems to indicate a right-skewed distribution, the 2% trimmed skewness is actually -0.292, which confirms that we have a left-skewed data set. It is also interesting that the transformations yield transformed data sets having lower kurtosis in all cases.

APPENDIX: AUXILIARY RESULTS AND THEIR PROOFS

The first lemma is a useful result of real analysis:

Lemma 6.1. Assume that H is a continuous real-valued function on $E \times E'$, where E and E' are two subsets of $\mathbb{R}$. Let K, K' be compact subsets of $\mathbb{R}$ which are contained in E and E' respectively. Then the family of functions $x \mapsto H(x; \lambda)$, $\lambda \in K'$, is uniformly equicontinuous on K, in the sense that

$$\lim_{\substack{h \to 0 \\ h > 0}} \sup_{(x, \lambda) \in K \times K'} \sup_{\substack{y \in K \\ |y - x| \le h}} |H(y; \lambda) - H(x; \lambda)| = 0.$$

P r o o f . If the statement were false then one could find a sequence $(x_n, \lambda_n) \subset K \times K'$ and a sequence $(y_n) \subset K'$ such that $|y_n - x_n| \to 0$ with

$$\liminf_{n \to \infty} |H(y_n; \lambda_n) - H(x_n; \lambda_n)| > 0.$$

Since K and K' are compact subsets of $\mathbb{R}$, we may assume, up to extracting a suitable subsequence, that $(x_n, \lambda_n) \to (x^*, \lambda^*) \in K \times K'$. In particular, $y_n \to x^*$ as well. By the continuity of H, $|H(y_n; \lambda_n) - H(x_n; \lambda_n)| \to 0$, which is a contradiction. $\square$

The second lemma is the cornerstone to prove Theorem 4.1.

Lemma 6.2. Assume that (A_1), (A_2) and (A_3) hold. If K is a compact subset of $\mathbb{R}$ contained in $\mathbf{\Lambda}$ then

$$\int_{-\infty}^{\infty} \widehat{\mathcal{S}}_n^2(t; \gamma; \vartheta) \, \mathrm{d}t \to \int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta) \, \mathrm{d}t$$

almost surely, uniformly in $\vartheta = (\delta, \lambda) \in \mathbb{R} \times K$ as $n \to \infty$.

P r o o f .   Since $|\widehat{\mathcal{S}}_n(t;\gamma;\vartheta)| \leq 1$, $|\mathcal{S}(t;\gamma;\vartheta)| \leq 1$ and the imaginary part of a complex number is less than its modulus, it is clear that for any ϑ,

$$\int_{-\infty}^{\infty} \left|\widehat{\mathcal{S}}_n^2(t;\gamma;\vartheta) - \mathcal{S}^2(t;\gamma;\vartheta)\right| \mathrm{d}t \leq 2 \int_{-\infty}^{\infty} |\widehat{\varphi}_{Z,n}(t;\gamma;\vartheta) - \varphi_Z(t;\gamma;\vartheta)| \, \mathrm{d}t$$

where $\varphi_Z(\cdot;\gamma;\vartheta)$ and $\widehat{\varphi}_{Z,n}(\cdot;\gamma;\vartheta)$ are the PWCF and PWECF related to $Z(\vartheta)$. Pick $\varepsilon > 0$; Remark 3.1 thus makes it possible to choose $M > 0$ such that for any ϑ:

$$\int_{-\infty}^{\infty} \left|\widehat{\mathcal{S}}_n^2(t;\gamma;\vartheta) - \mathcal{S}^2(t;\gamma;\vartheta)\right| \mathrm{d}t \leq \frac{\varepsilon}{4} + 2 \int_{-M}^{M} |\widehat{\varphi}_{Z,n}(t;\gamma;\vartheta) - \varphi_Z(t;\gamma;\vartheta)| \, \mathrm{d}t$$

$$\leq \frac{\varepsilon}{4} + 4M \sup_{-M \leq t \leq M} |\widehat{\varphi}_{Z,n}(t;\gamma;\vartheta) - \varphi_Z(t;\gamma;\vartheta)| \,.$$

$$(6.1)$$

Let $\varepsilon' = \varepsilon/(64M) > 0$ and observe that for any t:

$$|\widehat{\varphi}_{Z,n}(t;\gamma;\vartheta) - \varphi_Z(t;\gamma;\vartheta)| = \left| \int_0^1 [x(1-x)]^{\gamma|t|} \left\{ e^{it\widehat{Q}_{Z,n}(x;\vartheta)} - e^{itQ_Z(x;\vartheta)} \right\} \mathrm{d}x \right|$$

$$\leq \frac{\varepsilon}{16M} + \int_{\varepsilon'}^{1-\varepsilon'} [x(1-x)]^{\gamma|t|} \left| e^{it\widehat{Q}_{Z,n}(x;\vartheta)} - e^{itQ_Z(x;\vartheta)} \right| \mathrm{d}x$$

$$\leq \frac{\varepsilon}{16M} + \sup_{\varepsilon' \leq x \leq 1-\varepsilon'} \left| e^{it\widehat{Q}_{Z,n}(x;\vartheta)} - e^{itQ_Z(x;\vartheta)} \right| . \qquad (6.2)$$

Moreover

$$\left| e^{it\widehat{Q}_{Z,n}(x;\vartheta)} - e^{itQ_Z(x;\vartheta)} \right| = \left| it \int_{Q_Z(x;\vartheta)}^{\widehat{Q}_{Z,n}(x;\vartheta)} e^{itz} \mathrm{d}z \right| \leq |t| \left| \widehat{Q}_{Z,n}(x;\vartheta) - Q_Z(x;\vartheta) \right| . \quad (6.3)$$

Collecting (6.1), (6.2) and (6.3) entails

$$\sup_{\vartheta \in \mathbb{R} \times K} \int_{-\infty}^{\infty} \left|\widehat{\mathcal{S}}_n^2(t;\gamma;\vartheta) - \mathcal{S}^2(t;\gamma;\vartheta)\right| \mathrm{d}t \leq \frac{\varepsilon}{2} + 4M^2 \sup_{\substack{\varepsilon' \leq x \leq 1-\varepsilon' \\ \vartheta \in \mathbb{R} \times K}} \left| \widehat{Q}_{Z,n}(x;\vartheta) - Q_Z(x;\vartheta) \right| .$$

We thus get by using (4.1):

$$\sup_{\substack{\varepsilon' \leq x \leq 1-\varepsilon' \\ \vartheta \in \mathbb{R} \times K}} \left| \widehat{Q}_{Z,n}(x;\vartheta) - Q_Z(x;\vartheta) \right| \leq \sup_{\substack{\varepsilon' \leq x \leq 1-\varepsilon' \\ \lambda \in K}} \left| \psi(\widehat{Q}_n(x);\lambda) - \psi(Q(x);\lambda) \right| .$$

It is then enough to show that the supremum on the right-hand side of this inequality converges to 0 almost surely. To this end, we note that since the function F is continuous and strictly increasing on $\mathcal{D}$, so is Q on $(0,1)$. Especially, Q maps the interval $[\varepsilon', 1-\varepsilon']$ onto a compact interval $I \subsetneq \mathcal{D}$. Moreover, since with probability 1, $\widehat{Q}_n$ is a nondecreasing sequence of functions which converges pointwise to the continuous function Q on $(0,1)$,

by a well-known result due to Pólya (see e. g. Problem 127 p.270 in Pólya and Szegő [14]) the convergence must be uniform on compact intervals contained in $(0, 1)$; in particular

$$\sup_{\varepsilon' \leq x \leq 1-\varepsilon'} |\widehat{Q}_n(x) - Q(x)| \to 0 \quad \text{almost surely,}$$

which entails that there is a compact interval $J \subsetneq \mathcal{D}$ such that with probability 1, we have $\widehat{Q}_n(x) \in J$ for any $x \in [\varepsilon', 1 - \varepsilon']$ if n is large enough. As a consequence, for any positive integer N, we have with probability 1

$$\sup_{\substack{\varepsilon' \leq x \leq 1-\varepsilon' \\ \lambda \in K}} \left| \psi(\widehat{Q}_n(x); \lambda) - \psi(Q(x); \lambda) \right| \leq \sup_{(z,\lambda) \in J \times K} \sup_{\substack{y \in J \\ |y-z| \leq 1/N}} |\psi(y; \lambda) - \psi(z; \lambda)|$$

for n large enough. By Lemma 6.1, the right-hand side can be made arbitrarily small as $N \to \infty$, which concludes the proof. $\qquad\square$

The last lemma is a classical result (see Lemma 2 in Yeo and Johnson [20]) which essentially states that under some conditions, if a sequence of random functions (H_n) converges to a (nonrandom) function H which has a unique minimum x^*, then the sequence of the minima of the (H_n) converges to x^*.

Lemma 6.3. Assume that (H_n) is a random sequence of continuous functions on a compact metric space E such that

- (H_n) converges uniformly almost surely to a continuous function H on E;

- H has a unique global minimum x^*.

Then if (x_n) is any sequence such that $x_n = \arg\min_{x \in E} H_n(x)$, it holds that $x_n \to x^*$ almost surely.

P r o o f. If the result were false, we could find a set A with positive probability such that on A, (x_n) fails to converge to x^* but (H_n) converges uniformly almost surely to H on E. Choose $\omega \in A$ and define $y_n = x_n(\omega)$, $h_n = H_n(\cdot; \omega)$. The compactness of E would entail that one could find a subsequence of (y_n) which converges to $x_0 \neq x^*$. Since $h_n(y_n) \leq h_n(x^*)$ and

$$|h_n(y_n) - H(x_0)| \leq |h_n(y_n) - H(y_n)| + |H(y_n) - H(x_0)|$$

we would obtain in the limit $H(x_0) \leq H(x^*)$, which is a contradiction. $\qquad\square$

REFERENCES

[1] P. J. Bickel: On adaptive estimation. Ann. Statist. *10* (1982), 647–671. DOI:10.1214/aos/1176345863

[2] P. J. Bickel and K. A. Doksum: An analysis of transformations revisited. J. Amer. Statist. Assoc. *76* (1981), 296–311. DOI:10.1080/01621459.1981.10477649

[3] G. E. P. Box and D. R. Cox: An analysis of transformations. J. Roy. Statist. Soc. B *26* (1964), 211–243.

[4] J. B. Burbidge, L. Magee, and A. L. Robb: Alternative transformations to handle extreme values of the dependent variable. J. Amer. Statist. Assoc. *83* (1988), 123–127.

[5] G. Chen, R. Lockhart, and M. A. Stephens: Box–Cox transformations in linear models: large sample theory and tests for normality (with discussion). Canad. J. Statist. *30* (2002), 1–59.

[6] G. González–Rivera and F. C. Drost: Efficiency comparisons of maximum–likelihood–based estimators in GARCH models. J. Econometr. *93* (1999), 93–111. DOI:10.1016/s0304-4076(99)00005-6

[7] J. L. Horowitz: Semiparametric and Nonparametric Methods in Econometrics. Springer–Verlag, New York 2009. DOI:10.1007/978-0-387-92870-8

[8] J. A. John and N. R. Draper: An alternative family of transformations. J. Roy. Statist. Soc. C *29* (1980), 190–197. DOI:10.2307/2986305

[9] B. F. J. Manly: Exponential data transformations. J. Roy. Statist. Soc. D *25* (1976), 37–42.

[10] S. G. Meintanis, J. Swanepoel, and J. Allison: The probability weighted characteristic function and goodness-of-fit testing. J. Statist. Plann. Infer. *146* (2014), 122–132. DOI:10.1016/j.jspi.2013.09.011

[11] W. K. Newey: Adaptive estimation of regression models via moment restrictions. J. Econometr. *38* (1988), 301–339. DOI:10.1016/0304-4076(88)90048-6

[12] W. K. Newey and D. G. Steigerwald: Asymptotic bias for quasi–maximum–likelihood estimators in conditional heteroskedasticity models. Econometrica *65* (1997), 587–599. DOI:10.2307/2171754

[13] E. Parzen: On estimation of a probability density function and mode. Ann. Math. Statist. *33* (1962), 1065–1076. DOI:10.1214/aoms/1177704472

[14] G. Pólya and G. Szegő: Problems and Theorems in Analysis, Volume I. Springer–Verlag, Berlin 1998.

[15] M. Rosenblatt: Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. *27* (1956), 832–837. DOI:10.1214/aoms/1177728190

[16] O. Y. Savchuk and A. Schick: Density estimation for power transformations. J. Nonparametr. Statist. *25* (2013), 545–559. DOI:10.1080/10485252.2013.811788

[17] P. K. Sen: Estimates of the regression coefficient based on Kendall's tau. J. Amer. Statist. Assoc. *63* (1968), 1379–1389. DOI:10.1080/01621459.1968.10480934

[18] H. Theil: A rank–invariant method of linear and polynomial regression analysis. I, II, III. Nederl. Akad. Wetensch. Proc. *53* (1950), 386–392, 521–525, 1397–1412.

[19] I.-K. Yeo and R. A. Johnson: A new family of power transformations to improve normality or symmetry. Biometrika *87* (2000), 954–959.

[20] I.-K. Yeo and R. A. Johnson: A uniform law of large numbers for U–statistics with application to transforming to near symmetry. Statist. Probab. Lett. *51* (2001), 63–69.

[21] I.-K. Yeo and R. A. Johnson: An empirical characteristic function approach to selecting a transformation to symmetry. In: Contemporary Developments in Statistical Theory (S. Lahiri, A. Schick, A. SenGupta and T. Sriram, eds.), Springer International Publishing 2014, pp. 191–202. DOI:10.1007/978-3-319-02651-0_11

[22] I.-K. Yeo, R. A. Johnson, and X. W. Deng: An empirical characteristic function approach to selecting a transformation to normality. Commun. Stat. Appl. Methods *21* (2014), 213–224. DOI:10.5351/csam.2014.21.3.213

Simos G. Meintanis, Department of Economics, National and Kapodistrian University of Athens, Athens, Greece and Unit for Business Mathematics and Informatics, North-West University, Potchefstroom. South Africa.[1]
    e-mail: simosmei@econ.uoa.gr

Gilles Stupfler, Aix Marseille Université, CNRS, EHESS, Centrale Marseille, GRE-QAM UMR 7316, 13002 Marseille. France.
    e-mail: gilles.stupfler@univ-amu.fr

---

[1]On sabbatical leave from the University of Athens.