



HAL
open science

Transformations to symmetry based on the probability weighted characteristic function

Simos Meintanis, Gilles Stupfler

► **To cite this version:**

Simos Meintanis, Gilles Stupfler. Transformations to symmetry based on the probability weighted characteristic function. 2014. hal-01018575v1

HAL Id: hal-01018575

<https://hal.science/hal-01018575v1>

Preprint submitted on 4 Jul 2014 (v1), last revised 14 Apr 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TRANSFORMATIONS TO SYMMETRY BASED ON THE PROBABILITY WEIGHTED CHARACTERISTIC FUNCTION

SIMOS G. MEINTANIS^{a,b}, GILLES STUPFLER^c

^a*Department of Economics, National and Kapodistrian University of Athens, Athens, Greece,*

^b*Unit for Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa*
and

^c*Aix Marseille Université, CERGAM, EA 4225, 13540 Puyricard, France*

Abstract. We suggest a nonparametric version of the probability weighted empirical characteristic function (PWECEF) introduced by Meintanis *et al.* (2014) and use this PWECEF in order to estimate the parameters of arbitrary transformations to symmetry. The almost sure consistency of the resulting estimators is shown. Results with real and simulated data are also included.

Keywords. Characteristic function; Empirical characteristic function; Probability weighted moments; Symmetry transformation

AMS 2000 classification numbers: 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. For instance it is well known that the naïve bootstrap for the sample mean fails with an infinite-variance underlying distribution possessing asymmetric tails; see Abadir and Cornea (2014) for a recent account. 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 (1982) and Newey (1988) study the existence of adaptive and efficient regression estimators under symmetric errors. The reader is referred to Chapter 6 of Horowitz (2009) 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 (1999) and Newey and Steigerwald (1997). Finally, we mention that power transformations have recently been used by Savchuk and Schick (2013) in order to improve the rate of convergence of the classical Parzen-Rosenblatt (Parzen, 1962; Rosenblatt, 1956) estimator of the probability density function.

The purpose of this paper is to suggest a procedure by means of which a sample from an arbitrary distribution is reduced to a sample from a symmetric distribution. To this end we employ the notion of the probability weighted empirical characteristic function (PWECEF), introduced recently in Meintanis *et al.* (2014). The idea of the PWECEF draws from earlier works in which probability weighted moments are employed in the context of estimation of parameters; see for instance, Diebolt *et al.* (2004), Diebolt *et al.* (2007), Diebolt *et al.* (2008), Furrer

and Naveau (2007), and Caeiro and Ivette Gomes (2011). However, the PWEFC used in Meintanis *et al.* (2014) 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 PWEFC 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 the nonparametric PWEFC is introduced and we give some of its properties. In Section 3 we introduce the new estimation procedures which is based on an appropriate functional of this PWEFC. 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. Three real data examples are included in Section 6. Auxiliary results and their proofs are deferred to the Appendix.

2 The nonparametric PWEFC

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

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

where $W(x; s) := [F(x)(1 - F(x))]^{|s|}$. It is well-known that the ordinary characteristic function (CF) of X , which results from (2.1) by setting $\gamma = 0$, is real if and only if F is symmetric around zero. Interestingly the PWCF is also related to symmetry via the following proposition:

Proposition 1. *If F is symmetric around zero, then $\varphi(t; \gamma)$ is real for each $(t, \gamma) \in \mathbb{R} \times \mathbb{R}^+$. Conversely, if for some $\gamma \in \mathbb{R}^+$ the function $\varphi(t; \gamma)$ is real for each t , then F is symmetric around zero.*

Proof of Proposition 1. The first part of this result is a consequence of the fact that $\varphi(-t; \gamma) = \overline{\varphi(t; \gamma)}$ where \overline{z} denotes the conjugate of a complex number z : see Meintanis *et al.* (2014). To prove the converse part, we note that

$$\varphi(t; \gamma) = \int_{-\infty}^{\infty} e^{itx} dF^{[\gamma|t|]}(x), \quad \text{with} \quad F^{[s]}(x) = \int_{-\infty}^x [F(u)(1 - F(u))]^s dF(u) = \int_0^{F(x)} [v(1 - v)]^s dv.$$

Note that for any $s \geq 0$, the function $F^{[s]}$ has a finite positive limit $F^{[s]}(\infty)$ at infinity, and the function $G^{[s]} := F^{[s]}/F^{[s]}(\infty)$ is a cumulative distribution function of a random variable, say Y . Consequently, we have for all $t \in \mathbb{R}$:

$$\varphi(t; \gamma) \in \mathbb{R} \Rightarrow \varphi(t; \gamma) = \varphi(-t; \gamma) \Rightarrow \int_{-\infty}^{\infty} e^{itx} dG^{[\gamma|t|]}(x) = \int_{-\infty}^{\infty} e^{-itx} dG^{[\gamma|t|]}(x)$$

and the random variable Y has an absolutely continuous symmetric distribution around zero. In particular, $G^{[\gamma|t|]}(x) = 1 - G^{[\gamma|t|]}(-x)$ for any x , which entails:

$$\forall x \in \mathbb{R}, \quad \int_0^{F(x)} [v(1 - v)]^{\gamma|t|} dv = \int_{F(-x)}^1 [v(1 - v)]^{\gamma|t|} dv = \int_0^{1 - F(-x)} [v(1 - v)]^{\gamma|t|} dv.$$

Since the function $y \mapsto \int_0^y [v(1 - v)]^{\gamma|t|} dv$ is strictly increasing, we obtain $F(x) = 1 - F(-x)$ for all x , which means that F is symmetric around zero. \square

While Meintanis *et al.* (2014) estimated the PWCF in a parametric way, it would be 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

$$\varphi(t; \gamma) = \int_0^1 [x(1 - x)]^{\gamma|t|} e^{itQ(x)} dx, \quad (2.2)$$

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

In view of (2.2) we suggest the following nonparametric estimator of the PWCF:

$$\widehat{\varphi}_n(t; \gamma) = \int_0^1 [x(1-x)]^{\gamma|t|} e^{it\widehat{Q}_n(x)} dx, \quad (2.3)$$

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

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

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

3 L2-type procedures for symmetry transformation

The problem which we shall consider is to estimate the parameters of a given transformation which, if applied on the original nonsymmetrically distributed observations X_1, \dots, X_n , yields transformed observations that are approximately symmetrically distributed with location zero. To this end, write $\vartheta = (\delta, \lambda) \in \Theta \subset \mathbb{R} \times \mathbf{A}$ for the transformation parameter-vector, where δ denotes location and λ denotes the shape parameter which is assumed to lie in a subset \mathbf{A} 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

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

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

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

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

the imaginary part of the PWECF of $Z_1(\vartheta), \dots, Z_n(\vartheta)$. Then $\widehat{\mathcal{S}}_n(t; \gamma; \vartheta)$ is the empirical counterpart of $\mathcal{S}(t; \gamma; \vartheta)$. We suggest to estimate $\vartheta = (\delta, \lambda)$ by $\widehat{\vartheta}_n$, where

$$\widehat{\vartheta}_n = \arg \min_{\vartheta \in \Theta} \Delta_n(\gamma; \vartheta), \quad \text{with} \quad \Delta_n(\gamma; \vartheta) = \int_{-\infty}^{\infty} \widehat{\mathcal{S}}_n^2(t; \gamma; \vartheta) dt. \quad (3.1)$$

Remark 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 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. Possible choices for the transformation family ψ are the Box-Cox transformation (1964), a family introduced by Burbidge *et al.* (1988) as well as the recently introduced method of Yeo and Johnson (2000). Note

that while the popular Box-Cox transformation,

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

applies only to positive random variables (if λ is not a nonzero integer), its modifications suggested by Manly (1976), John and Draper (1980) and Bickel and Doksum (1981) were designed to allow negative values as well.

A favorable feature of the specific definition of the nonparametric PWEFC in (2.3) 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 (2.3) it is straightforward to compute the imaginary part of the PWEFC of $Z_1(\vartheta), \dots, Z_n(\vartheta)$ as

$$\widehat{S}_n(t; \gamma) = \sum_{k=1}^n v_{k,n}(t; \gamma) \sin(tZ_{k:n}(\vartheta)) \quad \text{with} \quad v_{k,n}(t; \gamma) = \int_{(k-1)/n}^{k/n} (x(1-x))^{\gamma|t|} dx.$$

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

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

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

$$I(j, k; \gamma; x) = \int_{-\infty}^{\infty} v_j(t; \gamma) v_k(t; \gamma) \cos(tx) dt.$$

We close this section by noting that the new procedures are in the same spirit as the symmetry procedures suggested by Quiroz *et al.* (1996) and Yeo and Johnson (2001). However while in the aforementioned works the ordinary ECF is employed, we expect that the introduction of the PWEFC will yield more efficient estimators, particularly by proper choice of the tuning parameters involved. These expectations will be put to test with simulated as well as with real data.

4 Strong consistency of the estimator

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

(A₁) The support \mathcal{D} of the distribution of X contains a nonempty open interval and F is continuous and strictly increasing on \mathcal{D} .

(A₂) The transformation family ψ is such that $(x, \lambda) \mapsto \psi(x; \lambda)$ is continuous on $E \times \mathbf{\Lambda}$, where E is the support of F .

(A₃) For all $\lambda \in \mathbf{\Lambda}$, $x \mapsto \psi(x; \lambda)$ is strictly increasing.

Assumption (A₂) is also used in Yeo and Johnson (2001), while (A₃) 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

$$Q_Z(x; \vartheta) = \psi(Q(x); \lambda) - \delta \quad \text{and} \quad \widehat{Q}_Z(x; \vartheta) = \psi(\widehat{Q}(x); \lambda) - \delta. \quad (4.1)$$

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

Theorem 1. Assume that (A_1) , (A_2) and (A_3) hold. Let Θ be a compact subset of \mathbb{R}^2 contained in $\mathbb{R} \times \Lambda$. 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 $\widehat{\vartheta}_n \rightarrow \vartheta_0$ almost surely.

Proof of Theorem 1. By Lemma 2,

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

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 1, it is again a corollary of the dominated convergence theorem that the function H is continuous as well. Applying Lemma 3 concludes the proof. \square

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

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

Conversely, if one can find ϑ_0 such that $\int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta_0) dt = 0$, then the distribution of $Z(\vartheta_0)$ must be symmetric around 0 by Proposition 1. The uniqueness of one such ϑ_0 is a more challenging problem: the last result of this section examines what can be said for the Bickel-Doksum family (1981), for which

$$\forall \lambda > 0, \forall x \in \mathbb{R}, \psi(x; \lambda) = \frac{\operatorname{sgn}(x)|x|^\lambda - 1}{\lambda}, \quad \text{with} \quad \operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0. \end{cases} \quad (4.2)$$

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 below.

Proposition 2. 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 with location zero, then (δ, λ) is the unique such pair.

Proof of Proposition 2. 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 (4.2), the random variable $\operatorname{sgn}(X)|X|^\lambda$ would be symmetric. Since (A_1) holds, this would imply, for any $x \leq 0$, that

$$\mathbb{P}(X \leq x) = \mathbb{P}(\operatorname{sgn}(X)|X|^\lambda \leq -(-x)^\lambda) = 1 - \mathbb{P}(\operatorname{sgn}(X)|X|^\lambda \geq (-x)^\lambda) = 1 - \mathbb{P}(X \geq -x).$$

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. Furthermore, since (A_1) holds and (A_2) and (A_3) are satisfied for the Bickel-Doksum family, we may use (4.1) to write that $Q_Z(x; \vartheta) = \psi(Q(x); \lambda) - \delta$. Since $Z(\vartheta)$ is symmetrically distributed around zero, its median is 0 and thus $0 = Q(1/2)^\lambda - c(\vartheta)$, where $c(\vartheta) = 1 + \delta\lambda$. In particular, $c(\vartheta)$ is positive and $Q(1/2) = [c(\vartheta)]^{1/\lambda}$. Finally, it must hold that $Q_Z(1/2 - x; \vartheta) = -Q_Z(1/2 + x; \vartheta)$ for any $x \in (0, 1/2)$, which, since Q is continuous and $Q(1/2) > 0$, entails for all $\varepsilon > 0$ small enough:

$$\frac{[Q(1/2 - \varepsilon)]^\lambda - 1}{\lambda} - \delta = - \left[\frac{[Q(1/2 + \varepsilon)]^\lambda - 1}{\lambda} - \delta \right] \Leftrightarrow Q(1/2 - \varepsilon) = \left(2c(\vartheta) - [Q(1/2 + \varepsilon)]^\lambda \right)^{1/\lambda}. \quad (4.3)$$

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 (4.3), we obtain for all $\varepsilon > 0$ sufficiently small:

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

Since $Q(1/2) = [c(\vartheta_1)]^{1/\lambda_1} = [c(\vartheta_2)]^{1/\lambda_2}$ and the quantile function of X 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 in a neighborhood of 0, we obtain for all h in an open neighborhood of 0:

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

Setting $h = 0$ entails $\lambda_1 = \lambda_2$, which completes the proof. \square

5 A Monte-Carlo simulation study

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 (4.2). The following estimators are compared:

- our estimator (3.1), denoted by M_γ , with $\gamma \in \{1, 2\}$;
- our estimator (3.1), with the weight $[x(1-x)]^{\gamma|t|}$ replaced by $[1/4 - x(1-x)]^{\gamma|t|}$. In other words, the mass is shifted from the median of the distribution to its tails; this estimator is denoted by T_γ and we consider $\gamma \in \{1/2, 1\}$;
- the estimator

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

which corresponds to using the ECF with an exponential weighting function (see Yeo and Johnson, 2001), 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

$$\begin{aligned}\widehat{\lambda} &= \arg \max_{\lambda \in \Lambda} \left\{ -\frac{n}{2} \log(\widehat{\sigma}^2(\lambda)) - \frac{1}{2} \sum_{k=1}^n \frac{(\psi(X_k; \lambda) - \widehat{\delta}(\lambda))^2}{\widehat{\sigma}^2(\lambda)} + (\lambda - 1) \sum_{k=1}^n \log X_k \right\} \\ &= \arg \max_{\lambda \in \Lambda} \left\{ -\frac{n}{2} \log(\widehat{\sigma}^2(\lambda)) + (\lambda - 1) \sum_{k=1}^n \log X_k \right\} \\ \text{with } \widehat{\delta}(\lambda) &= \frac{1}{n} \sum_{k=1}^n \psi(X_k; \lambda) \\ \text{and } \widehat{\sigma}^2(\lambda) &= \frac{1}{n} \sum_{k=1}^n (\psi(X_k; \lambda) - \widehat{\delta}(\lambda))^2;\end{aligned}$$

- a minimum-distance estimator by Nakamura and Ruppert (1990; see also Quiroz *et al.*, 1996), denoted by NR, which is obtained by minimizing

$$\vartheta \mapsto \int_{-\infty}^{\infty} \left(\widehat{F}_n(x; \vartheta) + \widehat{F}_n(-x; \vartheta) - 1 \right)^2 x^2 dx$$

where $\widehat{F}_n(x; \vartheta)$ is the empirical cumulative distribution function of the $Z_k(\vartheta)$.

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, \dots, 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, \dots, X_n such that $X_k = \tau(Y_k + \delta; \lambda)$ where

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

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;
- Y follows a Laplace(0, 1) distribution;
- $Y|V = v$ 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 $\Gamma(k, 1)$;
- $Y|V = v$ is Gaussian centered with variance v , where V is Inverse-Gaussian distributed with location parameter μ and shape parameter σ . This distribution is denoted by Variance $\text{IG}(\mu, \sigma)$;
- Y follows a stable distribution with shape parameter α , location parameter zero and unit scale. This distribution is denoted by $\text{Stable}(\alpha, 0, 1)$.

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 on Tables 1–3 the mean L^1 -error for λ and δ and on Tables 4–6 the sum of the mean L^1 -errors.

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 $\text{Stable}(1, 0, 1)$ distribution. Furthermore, while the GMLE and NR methods appear superior at estimating λ when the tail is light or when the distribution is leptokurtic, our technique is comparable to and sometimes better than these two methods 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). Note further that for light-tailed or leptokurtic distributions, the estimator T_γ is most often comparable to or better than the M_γ estimator. Our method is however superior in most cases when estimating δ , in particular for heavy-tailed and platykurtic distributions. Finally, according to Tables 4–6, it appears that the M_2 and GMLE methods are the most competitive overall in these examples.

6 Real data examples

In this section, we showcase our method on three sets of real data. For the first one, we consider the daily closing values (p_t) of the DAX index from October 1, 2007 to April 1, 2009, and our data is either the daily percentage of return $r_{t,1} = 100(p_t/p_{t-1} - 1)$ of size $n = 378$ or the 2-day percentage of return $r_{t,2} = 100(p_{2t+1}/p_{2t-1} - 1)$ of size $n = 190$. 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. The second data set is the Florida data set from Risebrough (1972). The variable of interest is the concentration in parts per million of the pollutant PCB (polychlorinated biphenyl) in $n = 87$ brown pelican eggs. The third data set consists of the hourly wages of $n = 533$ people selected at random from the 1985 U.S. Current Population Survey; this data set was also considered by Savchuk and Schick (2013). Observe that the last two data sets consist of positive values, and thus in these cases the Bickel-Doksum transformation coincides with a Box-Cox transformation.

We compare the results found with the M_1 and M_2 methods with what we find when using the GMLE method. In Table 7, 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 on Figures 1–4.

All in all, the data sets transformed using the M_1 and M_2 methods look somewhat “more” symmetric than the original data sets, particularly in the case of the PCB-Florida and wage data sets. This is confirmed by Table 7, in which 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 at this point 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. Finally, we mention that although the values of $\hat{\lambda}$ and $\hat{\delta}$ are almost identical for all methods when considering the PCB-Florida and wage data sets and similar for the daily DAX data set, they are actually quite different for the 2-day DAX data set. In the latter case, the M_γ method seems to detect a bimodal distribution, whereas the GMLE method points to a unimodal distribution.

Appendix: auxiliary results and their proofs

The first lemma is a useful result of real analysis:

Lemma 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*

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

Proof of Lemma 1. 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 $|y_n - x_n| \rightarrow 0$ with

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

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

The second lemma is the cornerstone to prove Theorem 1.

Lemma 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*

$$\int_{-\infty}^{\infty} \widehat{\mathcal{S}}_n^2(t; \gamma; \vartheta) dt \rightarrow \int_{-\infty}^{\infty} \mathcal{S}^2(t; \gamma; \vartheta) dt$$

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

Proof of Lemma 2. 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 ,

$$\int_{-\infty}^{\infty} \left| \widehat{\mathcal{S}}_n^2(t; \gamma; \vartheta) dt - \mathcal{S}^2(t; \gamma; \vartheta) \right| dt \leq 2 \int_{-\infty}^{\infty} |\widehat{\varphi}_Z(t; \gamma; \vartheta) dt - \varphi_Z(t; \gamma; \vartheta)| dt$$

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 1 thus makes it possible to choose $M > 0$ such that for any ϑ :

$$\begin{aligned} \int_{-\infty}^{\infty} \left| \widehat{\mathcal{S}}_n^2(t; \gamma; \vartheta) dt - \mathcal{S}^2(t; \gamma; \vartheta) \right| dt &\leq \frac{\varepsilon}{4} + 2 \int_{-M}^M |\widehat{\varphi}_{Z,n}(t; \gamma; \vartheta) dt - \varphi_Z(t; \gamma; \vartheta)| dt \\ &\leq \frac{\varepsilon}{4} + 4M \sup_{-M \leq t \leq M} |\widehat{\varphi}_{Z,n}(t; \gamma; \vartheta) dt - \varphi_Z(t; \gamma; \vartheta)|. \end{aligned} \quad (6.1)$$

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

$$\begin{aligned} |\widehat{\varphi}_{Z,n}(t; \gamma; \vartheta) - \varphi_Z(t; \gamma; \vartheta)| &= \left| \int_0^1 [x(1-x)]^{\gamma|t|} \left\{ e^{it\widehat{Q}_{Z,n}(x; \vartheta)} - e^{itQ_Z(x; \vartheta)} \right\} dx \right| \\ &\leq \frac{\varepsilon}{16M} + \int_{\varepsilon'}^{1-\varepsilon'} [x(1-x)]^{\gamma|t|} \left| e^{it\widehat{Q}_{Z,n}(x; \vartheta)} - e^{itQ_Z(x; \vartheta)} \right| dx \\ &\leq \frac{\varepsilon}{16M} + \sup_{\varepsilon' \leq x \leq 1-\varepsilon'} \left| e^{it\widehat{Q}_{Z,n}(x; \vartheta)} - e^{itQ_Z(x; \vartheta)} \right|. \end{aligned} \quad (6.2)$$

Moreover

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

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

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

Furthermore, since K is a compact subset of \mathbb{R} contained in $\mathbf{\Lambda}$, we 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 $I = [a_{\min}, a_{\max}]$. 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, 1998) the convergence must be uniform on compact intervals contained in $(0, 1)$; in particular

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

which entails that with probability 1, we have $\widehat{Q}_n(x) \in J = [a_{\min} - 1, a_{\max} + 1]$ 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

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

for n large enough. By Lemma 1, the right-hand side can be made arbitrarily small as $N \rightarrow \infty$, which concludes the proof. \square

The last lemma is a classical result (see Lemma 2 in Yeo and Johnson, 2001) 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 3. *Assume that (H_n) is a sequence of random 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 $x_n \rightarrow x^$ almost surely.*

Proof of Lemma 3. 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. \square

References

- [1] Abadir, K.M., Cornea, A. (2014). Approximating moments by nonlinear transformations with an application to resampling from fat-tailed distributions. Presented in the 2nd Conference of the International Association of NonParametric Statistics, Cádiz, June 12–16, 2014. Submitted to *Ann. Statist.*.
- [2] Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.*, 10, 647–671.
- [3] Bickel, P.J., Doksum, K.A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.*, 76, 296–311.
- [4] Box, G.E.P., Cox, D.R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. B*, 26, 211–243.
- [5] Burbidge, J.B., Magee, L., Robb, A.L. (1988). Alternative transformations to handle extreme values of the dependent variable. *J. Amer. Statist. Assoc.*, 83, 123–127.
- [6] Caeiro, F., Ivette Gomes, M. (2011). Semi-parametric tail inference through probability-weighted moments. *J. Statist. Plann. Inference*, 141, 937–950.

- [7] Diebolt, J., Guillou, A., Naveau, P., Ribereau, P. (2008). Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT*, 6, 33–50.
- [8] Diebolt, J., Guillou, A., Rached, I. (2004). A new look at probability-weighted moments estimators. *C. R. Acad. Sci. Paris, Ser. I 338*, 629–634.
- [9] Diebolt, J., Guillou, A., Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method. *J. Statist. Plann. Infer.*, 137, 841–857.
- [10] Furrer, R., Naveau, P. (2007). Probability weighted moments properties for small samples. *Statist. Probab. Lett.*, 77, 190–195.
- [11] González-Rivera, G., Drost, F.C. (1999). Efficiency comparisons of maximum-likelihood-based estimators in GARCH models. *J. Econometr.*, 93, 93–111.
- [12] Horowitz, J.L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*, Springer.
- [13] John, J.A., Draper, N.R. (1980). An alternative family of transformations. *J. Roy. Statist. Soc. C*, 29, 190–197.
- [14] Manly, B.F.J. (1976). Exponential data transformations. *J. Roy. Statist. Soc. D*, 25, 37–42.
- [15] Meintanis, S.G., Swanepoel, J., Allison, J. (2013). The probability weighted characteristic function and goodness-of-fit testing. *J. Statist. Plann. Infer.*, 146, 122–132.
- [16] Nakamura, M., Ruppert, D. (1990). Semi-parametric estimation of symmetrizing transformations with application to the shifted power transformation. Unpublished manuscript.
- [17] Newey, W.K. (1988). Adaptive estimation of regression models via moment restrictions. *J. Econometr.*, 38, 301–339.
- [18] Newey, W.K., Steigerwald, D.G. (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica*, 65, 587–599.
- [19] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065–1076.
- [20] Pólya, G., Szegő, G. (1998). *Problems and Theorems in Analysis, Vol. I.*, Springer.
- [21] Quiroz, A.J., Nakamura, M., Pérez, F.J. (1996). Estimation of a multivariate Box-Cox transformation to elliptical symmetry via the empirical characteristic function. *Ann. Instit. Statist. Math.*, 48, 687–709.
- [22] Risebrough, R. (1972). Effects of environmental pollutants upon animals other than man. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, 6, 443–463.
- [23] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832–837.
- [24] Savchuk, O.Y., Schick, A. (2013). Density estimation for power transformations. *J. Nonparametr. Statist.*, 25, 545–559.
- [25] Yeo, I.-K., Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87, 954–959.
- [26] Yeo, I.-K., Johnson, R.A. (2001). A uniform law of large numbers for U -statistics with application to transforming to near symmetry. *Statist. Probab. Lett.*, 51, 63–69.

Case	M_1	M_2	$T_{1/2}$	T_1	EECF	GMLE	NR
Standard Gaussian	[0.112, 0.124]	[0.104, 0.120]	[0.0867, 0.111]	[0.0844, 0.112]	[0.116, 0.119]	[0.0774, 0.105]	[0.0875, 0.111]
Tukey(0, 1/4)	[0.107, 0.0987]	[0.0984, 0.0995]	[0.116, 0.101]	[0.103, 0.110]	[0.148, 0.0984]	[0.0905, 0.119]	[0.0888, 0.136]
Tukey(0, 1/2)	[0.126, 0.104]	[0.114, 0.109]	[0.112, 0.108]	[0.0970, 0.117]	[0.148, 0.0876]	[0.0664, 0.215]	[0.0745, 0.273]
Tukey(0, 3/4)	[0.144, 0.105]	[0.135, 0.123]	[0.123, 0.119]	[0.0995, 0.138]	[0.151, 0.0885]	[0.0847, 0.507]	[0.0859, 0.660]
Laplace(0, 1)	[0.0929, 0.0946]	[0.0821, 0.0933]	[0.0924, 0.112]	[0.0783, 0.120]	[0.131, 0.0937]	[0.0722, 0.113]	[0.0734, 0.138]
Variance $\Gamma(1, 1)$	[0.126, 0.0803]	[0.119, 0.0786]	[0.120, 0.0919]	[0.110, 0.101]	[0.152, 0.0850]	[0.108, 0.0891]	[0.107, 0.107]
Variance $\Gamma(2, 1)$	[0.0888, 0.118]	[0.0783, 0.115]	[0.0807, 0.122]	[0.0692, 0.126]	[0.117, 0.117]	[0.0630, 0.116]	[0.0683, 0.135]
Variance $\Gamma(3, 1)$	[0.0702, 0.150]	[0.0612, 0.144]	[0.0596, 0.147]	[0.0522, 0.149]	[0.0939, 0.146]	[0.0460, 0.144]	[0.0508, 0.159]
Variance $\Gamma(4, 1)$	[0.0669, 0.166]	[0.0557, 0.161]	[0.0537, 0.166]	[0.0450, 0.165]	[0.0877, 0.164]	[0.0378, 0.158]	[0.0417, 0.177]
Variance IG(1, 1)	[0.125, 0.0937]	[0.118, 0.0914]	[0.123, 0.0950]	[0.114, 0.0998]	[0.153, 0.0983]	[0.111, 0.0924]	[0.114, 0.103]
Variance IG(3, 1)	[0.100, 0.0986]	[0.0837, 0.0963]	[0.107, 0.108]	[0.0869, 0.116]	[0.151, 0.0954]	[0.0727, 0.126]	[0.0682, 0.148]
Variance IG(6, 1)	[0.101, 0.101]	[0.0870, 0.106]	[0.105, 0.121]	[0.0828, 0.130]	[0.150, 0.101]	[0.0592, 0.155]	[0.0623, 0.195]
Variance IG(10, 1)	[0.104, 0.104]	[0.0956, 0.113]	[0.109, 0.130]	[0.0851, 0.144]	[0.146, 0.0996]	[0.0476, 0.199]	[0.0564, 0.261]
Variance IG(1, 10)	[0.113, 0.116]	[0.106, 0.113]	[0.0960, 0.108]	[0.0916, 0.110]	[0.122, 0.113]	[0.0847, 0.103]	[0.0957, 0.109]
Variance IG(10, 10)	[0.0589, 0.215]	[0.0503, 0.209]	[0.0441, 0.241]	[0.0348, 0.244]	[0.0760, 0.231]	[0.0276, 0.246]	[0.0325, 0.269]
Stable(7/4, 0, 1)	[0.129, 0.115]	[0.122, 0.111]	[0.121, 0.104]	[0.130, 0.115]	[0.149, 0.113]	[0.127, 0.122]	[0.127, 0.133]
Stable(3/2, 0, 1)	[0.145, 0.104]	[0.146, 0.114]	[0.146, 0.0917]	[0.139, 0.100]	[0.167, 0.0961]	[0.115, 0.146]	[0.107, 0.164]
Stable(5/4, 0, 1)	[0.165, 0.104]	[0.192, 0.140]	[0.146, 0.0861]	[0.134, 0.0987]	[0.163, 0.0724]	[0.0939, 0.232]	[0.0991, 0.279]
Stable(1, 0, 1)	[0.213, 0.142]	[0.202, 0.174]	[0.203, 0.180]	[0.188, 0.196]	[0.197, 0.133]	[0.114, 0.323]	[0.118, 0.385]

Table 1: Case $\lambda = 1/4$ and $\delta = 1$: mean L^1 -errors. First number: error related to the estimator of λ , second number: error related to the estimator of δ .

Case	M_1	M_2	$T_{1/2}$	T_1	EECF	GMLE	NR
Standard Gaussian	[0.133, 0.125]	[0.121, 0.122]	[0.102, 0.114]	[0.0967, 0.115]	[0.140, 0.121]	[0.0835, 0.105]	[0.103, 0.116]
Tukey(0, 1/4)	[0.118, 0.104]	[0.109, 0.103]	[0.127, 0.107]	[0.111, 0.116]	[0.195, 0.104]	[0.0962, 0.125]	[0.105, 0.145]
Tukey(0, 1/2)	[0.129, 0.103]	[0.123, 0.106]	[0.142, 0.108]	[0.123, 0.126]	[0.221, 0.0863]	[0.122, 0.259]	[0.131, 0.341]
Tukey(0, 3/4)	[0.146, 0.105]	[0.147, 0.120]	[0.139, 0.122]	[0.122, 0.144]	[0.210, 0.0803]	[0.190, 0.530]	[0.183, 0.724]
Laplace(0, 1)	[0.113, 0.0985]	[0.101, 0.0980]	[0.109, 0.114]	[0.0885, 0.122]	[0.188, 0.0999]	[0.0793, 0.118]	[0.0807, 0.141]
Variance $\Gamma(1, 1)$	[0.150, 0.0855]	[0.139, 0.0836]	[0.142, 0.0982]	[0.124, 0.105]	[0.197, 0.0935]	[0.123, 0.0991]	[0.119, 0.115]
Variance $\Gamma(2, 1)$	[0.0959, 0.113]	[0.0840, 0.113]	[0.0852, 0.119]	[0.0737, 0.123]	[0.145, 0.116]	[0.0687, 0.122]	[0.0692, 0.133]
Variance $\Gamma(3, 1)$	[0.0883, 0.148]	[0.0744, 0.149]	[0.0702, 0.149]	[0.0618, 0.151]	[0.127, 0.148]	[0.0519, 0.148]	[0.0578, 0.161]
Variance $\Gamma(4, 1)$	[0.0770, 0.158]	[0.0592, 0.168]	[0.0551, 0.166]	[0.0510, 0.169]	[0.0946, 0.165]	[0.0437, 0.163]	[0.0503, 0.174]
Variance IG(1, 1)	[0.144, 0.0962]	[0.135, 0.0929]	[0.145, 0.101]	[0.128, 0.106]	[0.198, 0.105]	[0.119, 0.0989]	[0.122, 0.108]
Variance IG(3, 1)	[0.112, 0.102]	[0.0989, 0.101]	[0.123, 0.112]	[0.103, 0.122]	[0.211, 0.0981]	[0.0773, 0.125]	[0.0865, 0.152]
Variance IG(6, 1)	[0.108, 0.107]	[0.0966, 0.110]	[0.124, 0.125]	[0.100, 0.141]	[0.205, 0.0973]	[0.0691, 0.180]	[0.0873, 0.224]
Variance IG(10, 1)	[0.109, 0.108]	[0.100, 0.111]	[0.117, 0.129]	[0.101, 0.145]	[0.196, 0.0939]	[0.0827, 0.231]	[0.0947, 0.297]
Variance IG(1, 10)	[0.132, 0.118]	[0.121, 0.113]	[0.104, 0.104]	[0.101, 0.107]	[0.149, 0.117]	[0.0896, 0.0982]	[0.104, 0.107]
Variance IG(10, 10)	[0.0857, 0.205]	[0.0566, 0.227]	[0.0545, 0.248]	[0.0537, 0.255]	[0.0828, 0.236]	[0.0432, 0.259]	[0.0530, 0.285]
Stable(7/4, 0, 1)	[0.145, 0.114]	[0.142, 0.112]	[0.155, 0.112]	[0.168, 0.127]	[0.186, 0.118]	[0.157, 0.133]	[0.165, 0.147]
Stable(3/2, 0, 1)	[0.158, 0.103]	[0.153, 0.102]	[0.181, 0.101]	[0.171, 0.115]	[0.224, 0.106]	[0.148, 0.155]	[0.161, 0.177]
Stable(5/4, 0, 1)	[0.172, 0.0935]	[0.170, 0.102]	[0.195, 0.0926]	[0.169, 0.109]	[0.252, 0.0901]	[0.157, 0.256]	[0.172, 0.318]
Stable(1, 0, 1)	[0.222, 0.198]	[0.231, 0.264]	[0.363, 0.128]	[0.332, 0.172]	[0.419, 0.0443]	[0.210, 0.281]	[0.218, 0.345]

Table 2: Case $\lambda = 1/2$ and $\delta = 1$: mean L^1 -errors. First number: error related to the estimator of λ , second number: error related to the estimator of δ .

Case	M_1	M_2	$T_{1/2}$	T_1	EECF	GMLE	NR
Standard Gaussian	[0.153, 0.130]	[0.138, 0.125]	[0.108, 0.111]	[0.100, 0.111]	[0.169, 0.126]	[0.0941, 0.102]	[0.105, 0.113]
Tukey(0, 1/4)	[0.132, 0.0996]	[0.128, 0.0995]	[0.137, 0.108]	[0.130, 0.124]	[0.231, 0.104]	[0.106, 0.127]	[0.126, 0.154]
Tukey(0, 1/2)	[0.151, 0.104]	[0.148, 0.107]	[0.158, 0.114]	[0.145, 0.138]	[0.239, 0.0951]	[0.192, 0.290]	[0.198, 0.382]
Tukey(0, 3/4)	[0.155, 0.103]	[0.155, 0.111]	[0.163, 0.124]	[0.156, 0.157]	[0.240, 0.0896]	[0.306, 0.562]	[0.293, 0.779]
Laplace(0, 1)	[0.127, 0.100]	[0.118, 0.0998]	[0.118, 0.114]	[0.107, 0.125]	[0.200, 0.104]	[0.0946, 0.118]	[0.101, 0.135]
Variance $\Gamma(1, 1)$	[0.170, 0.0863]	[0.157, 0.0834]	[0.152, 0.0947]	[0.133, 0.102]	[0.239, 0.0994]	[0.137, 0.101]	[0.130, 0.107]
Variance $\Gamma(2, 1)$	[0.107, 0.120]	[0.0975, 0.118]	[0.0946, 0.118]	[0.0889, 0.124]	[0.158, 0.119]	[0.0798, 0.119]	[0.0876, 0.129]
Variance $\Gamma(3, 1)$	[0.0954, 0.153]	[0.0864, 0.151]	[0.0784, 0.150]	[0.0772, 0.157]	[0.126, 0.147]	[0.0642, 0.148]	[0.0741, 0.161]
Variance $\Gamma(4, 1)$	[0.0796, 0.171]	[0.0699, 0.169]	[0.0646, 0.172]	[0.0652, 0.177]	[0.102, 0.169]	[0.0529, 0.169]	[0.0632, 0.184]
Variance IG(1, 1)	[0.172, 0.106]	[0.159, 0.102]	[0.160, 0.103]	[0.148, 0.111]	[0.241, 0.116]	[0.143, 0.107]	[0.144, 0.110]
Variance IG(3, 1)	[0.128, 0.101]	[0.123, 0.102]	[0.134, 0.113]	[0.123, 0.131]	[0.225, 0.100]	[0.0918, 0.130]	[0.115, 0.158]
Variance IG(6, 1)	[0.132, 0.111]	[0.123, 0.113]	[0.134, 0.129]	[0.119, 0.153]	[0.223, 0.103]	[0.112, 0.194]	[0.122, 0.248]
Variance IG(10, 1)	[0.138, 0.110]	[0.135, 0.114]	[0.141, 0.131]	[0.132, 0.157]	[0.215, 0.0972]	[0.149, 0.279]	[0.151, 0.383]
Variance IG(1, 10)	[0.151, 0.125]	[0.138, 0.120]	[0.117, 0.107]	[0.109, 0.109]	[0.171, 0.124]	[0.104, 0.103]	[0.114, 0.110]
Variance IG(10, 10)	[0.0829, 0.223]	[0.0761, 0.225]	[0.0768, 0.246]	[0.0834, 0.257]	[0.0974, 0.231]	[0.0743, 0.268]	[0.0863, 0.320]
Stable(7/4, 0, 1)	[0.157, 0.117]	[0.153, 0.113]	[0.161, 0.112]	[0.172, 0.126]	[0.213, 0.120]	[0.159, 0.130]	[0.183, 0.152]
Stable(3/2, 0, 1)	[0.171, 0.104]	[0.173, 0.106]	[0.193, 0.111]	[0.185, 0.132]	[0.264, 0.115]	[0.186, 0.180]	[0.212, 0.214]
Stable(5/4, 0, 1)	[0.192, 0.0936]	[0.193, 0.0984]	[0.230, 0.0997]	[0.199, 0.121]	[0.299, 0.101]	[0.247, 0.270]	[0.264, 0.344]
Stable(1, 0, 1)	[0.271, 0.199]	[0.260, 0.210]	[0.461, 0.119]	[0.380, 0.185]	[0.526, 0.0712]	[0.300, 0.233]	[0.320, 0.311]

Table 3: Case $\lambda = 3/4$ and $\delta = 1$: mean L^1 -errors. First number: error related to the estimator of λ , second number: error related to the estimator of δ .

Case	M_1	M_2	$T_{1/2}$	T_1	EECF	GMLE	NR
Standard Gaussian	0.236	0.224	0.198	0.196	0.235	0.182	0.199
Tukey(0, 1/4)	0.206	0.198	0.217	0.213	0.246	0.210	0.225
Tukey(0, 1/2)	0.230	0.223	0.220	0.214	0.236	0.281	0.348
Tukey(0, 3/4)	0.249	0.258	0.242	0.238	0.240	0.592	0.746
Laplace(0, 1)	0.188	0.175	0.204	0.198	0.225	0.185	0.211
Variance $\Gamma(1, 1)$	0.206	0.198	0.212	0.211	0.237	0.197	0.214
Variance $\Gamma(2, 1)$	0.207	0.193	0.203	0.195	0.234	0.179	0.203
Variance $\Gamma(3, 1)$	0.220	0.205	0.207	0.201	0.240	0.190	0.210
Variance $\Gamma(4, 1)$	0.233	0.217	0.220	0.210	0.252	0.196	0.219
Variance IG(1, 1)	0.219	0.209	0.218	0.214	0.251	0.203	0.217
Variance IG(3, 1)	0.199	0.180	0.215	0.203	0.246	0.199	0.216
Variance IG(6, 1)	0.202	0.193	0.226	0.213	0.251	0.214	0.257
Variance IG(10, 1)	0.208	0.209	0.239	0.229	0.246	0.247	0.317
Variance IG(1, 10)	0.229	0.219	0.204	0.202	0.235	0.188	0.205
Variance IG(10, 10)	0.274	0.259	0.285	0.279	0.307	0.274	0.302
Stable(7/4, 0, 1)	0.244	0.233	0.225	0.245	0.262	0.249	0.260
Stable(3/2, 0, 1)	0.249	0.260	0.238	0.239	0.263	0.261	0.271
Stable(5/4, 0, 1)	0.269	0.332	0.232	0.233	0.235	0.326	0.378
Stable(1, 0, 1)	0.355	0.376	0.383	0.384	0.330	0.437	0.503

Table 4: Case $\lambda = 1/4$ and $\delta = 1$: sum of the mean L^1 -errors related to the estimators.

Case	M_1	M_2	$T_{1/2}$	T_1	EECF	GMLE	NR
Standard Gaussian	0.258	0.243	0.216	0.212	0.261	0.189	0.219
Tukey(0, 1/4)	0.222	0.212	0.234	0.227	0.299	0.221	0.250
Tukey(0, 1/2)	0.232	0.229	0.250	0.249	0.307	0.381	0.472
Tukey(0, 3/4)	0.251	0.267	0.261	0.266	0.290	0.720	0.907
Laplace(0, 1)	0.212	0.199	0.223	0.211	0.288	0.197	0.222
Variance $\Gamma(1, 1)$	0.236	0.223	0.240	0.229	0.291	0.222	0.234
Variance $\Gamma(2, 1)$	0.209	0.197	0.204	0.197	0.261	0.191	0.202
Variance $\Gamma(3, 1)$	0.236	0.223	0.219	0.213	0.275	0.200	0.219
Variance $\Gamma(4, 1)$	0.235	0.227	0.221	0.220	0.260	0.207	0.224
Variance IG(1, 1)	0.240	0.228	0.246	0.234	0.303	0.218	0.230
Variance IG(3, 1)	0.214	0.200	0.235	0.225	0.309	0.202	0.239
Variance IG(6, 1)	0.215	0.207	0.249	0.241	0.302	0.249	0.311
Variance IG(10, 1)	0.217	0.211	0.246	0.246	0.290	0.314	0.392
Variance IG(1, 10)	0.250	0.234	0.208	0.208	0.266	0.188	0.211
Variance IG(10, 10)	0.291	0.284	0.303	0.309	0.319	0.302	0.338
Stable(7/4, 0, 1)	0.259	0.254	0.267	0.295	0.304	0.290	0.312
Stable(3/2, 0, 1)	0.261	0.255	0.282	0.286	0.330	0.303	0.338
Stable(5/4, 0, 1)	0.266	0.272	0.288	0.278	0.342	0.413	0.490
Stable(1, 0, 1)	0.420	0.495	0.491	0.504	0.463	0.491	0.563

Table 5: Case $\lambda = 1/2$ and $\delta = 1$: sum of the mean L^1 -errors related to the estimators.

Case	M_1	M_2	$T_{1/2}$	T_1	EECF	GMLE	NR
Standard Gaussian	0.283	0.263	0.219	0.211	0.295	0.196	0.218
Tukey(0, 1/4)	0.232	0.228	0.245	0.254	0.335	0.233	0.280
Tukey(0, 1/2)	0.255	0.255	0.272	0.283	0.334	0.482	0.580
Tukey(0, 3/4)	0.258	0.266	0.287	0.313	0.330	0.868	1.072
Laplace(0, 1)	0.227	0.218	0.232	0.232	0.304	0.213	0.236
Variance $\Gamma(1, 1)$	0.256	0.240	0.247	0.235	0.338	0.238	0.237
Variance $\Gamma(2, 1)$	0.227	0.216	0.213	0.213	0.277	0.199	0.217
Variance $\Gamma(3, 1)$	0.248	0.237	0.228	0.234	0.273	0.212	0.235
Variance $\Gamma(4, 1)$	0.251	0.239	0.237	0.242	0.271	0.222	0.247
Variance IG(1, 1)	0.278	0.261	0.263	0.259	0.357	0.250	0.254
Variance IG(3, 1)	0.229	0.225	0.247	0.254	0.325	0.222	0.273
Variance IG(6, 1)	0.243	0.236	0.263	0.272	0.326	0.306	0.370
Variance IG(10, 1)	0.248	0.249	0.272	0.289	0.312	0.428	0.534
Variance IG(1, 10)	0.276	0.258	0.224	0.218	0.295	0.207	0.224
Variance IG(10, 10)	0.306	0.301	0.323	0.340	0.328	0.342	0.406
Stable(7/4, 0, 1)	0.274	0.266	0.273	0.298	0.333	0.289	0.335
Stable(3/2, 0, 1)	0.275	0.279	0.304	0.317	0.379	0.366	0.426
Stable(5/4, 0, 1)	0.286	0.291	0.330	0.320	0.400	0.517	0.608
Stable(1, 0, 1)	0.470	0.470	0.580	0.565	0.597	0.533	0.631

Table 6: Case $\lambda = 3/4$ and $\delta = 1$: sum of the mean L^1 -errors related to the estimators.

		$\hat{\lambda}$	$\hat{\delta}$	Mean	Std. deviation	Skewness	Kurtosis
Daily DAX data set	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
2-day DAX data set	Raw data	1	-1	-0.300	2.902	-0.173	4.580
	M_1	0.460	-2.352	-0.0321	3.025	-0.0227	1.725
	M_2	0.440	-2.462	-0.0164	3.092	-0.0177	1.669
	GMLE	0.805	-1.495	0	2.681	-0.126	3.244
PCB-Florida data set	Raw data	1	-1	71.248	52.47	2.113	9.822
	M_1	0.214	6.511	0.00351	1.704	0.0314	3.480
	M_2	0.210	6.458	0.00218	1.681	0.0220	3.481
	GMLE	0.213	6.495	0	1.696	0.0281	3.480
Wage data set	Raw data	1	-1	8.958	4.908	1.273	4.465
	M_1	0	2.049	0.00662	0.523	0.0521	2.749
	M_2	0	2.050	0.00642	0.523	0.0521	2.749
	GMLE	0	2.056	0	0.523	0.0523	2.749

Table 7: Estimated values of λ and δ for our real data sets

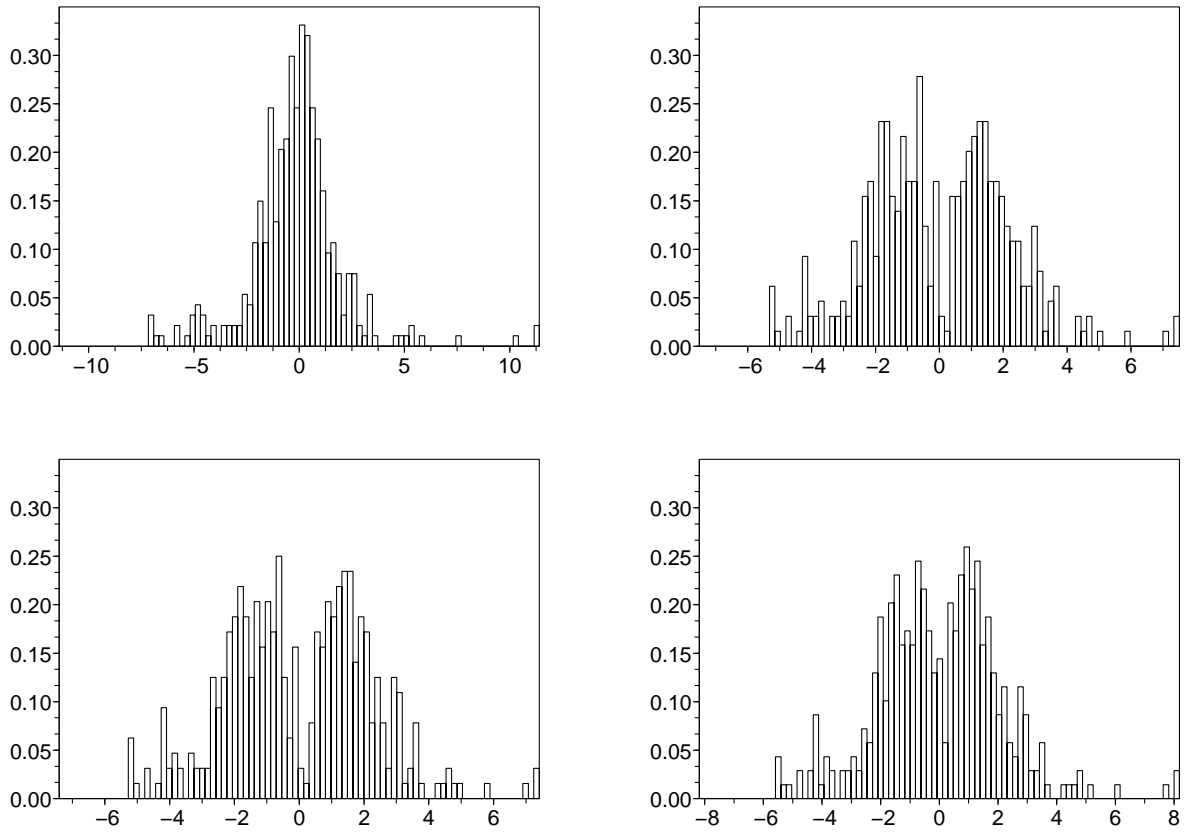


Figure 1: DAX daily data set, top left: original data, top right: data transformed with the parameters obtained by the M_1 technique, bottom left: data transformed with the parameters obtained by the M_2 technique, bottom right: data transformed with the parameters obtained by the GMLE technique.

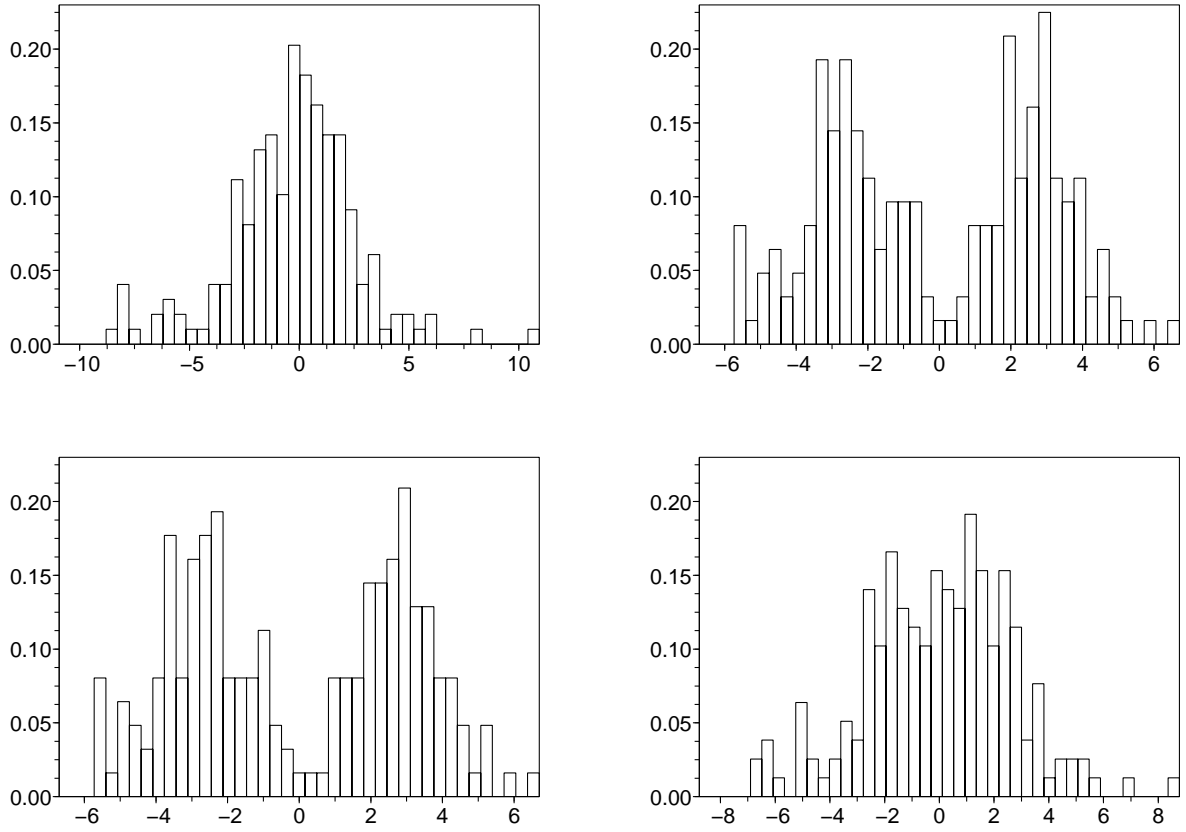


Figure 2: DAX 2-day data set, top left: original data, top right: data transformed with the parameters obtained by the M_1 technique, bottom left: data transformed with the parameters obtained by the M_2 technique, bottom right: data transformed with the parameters obtained by the GMLE technique.

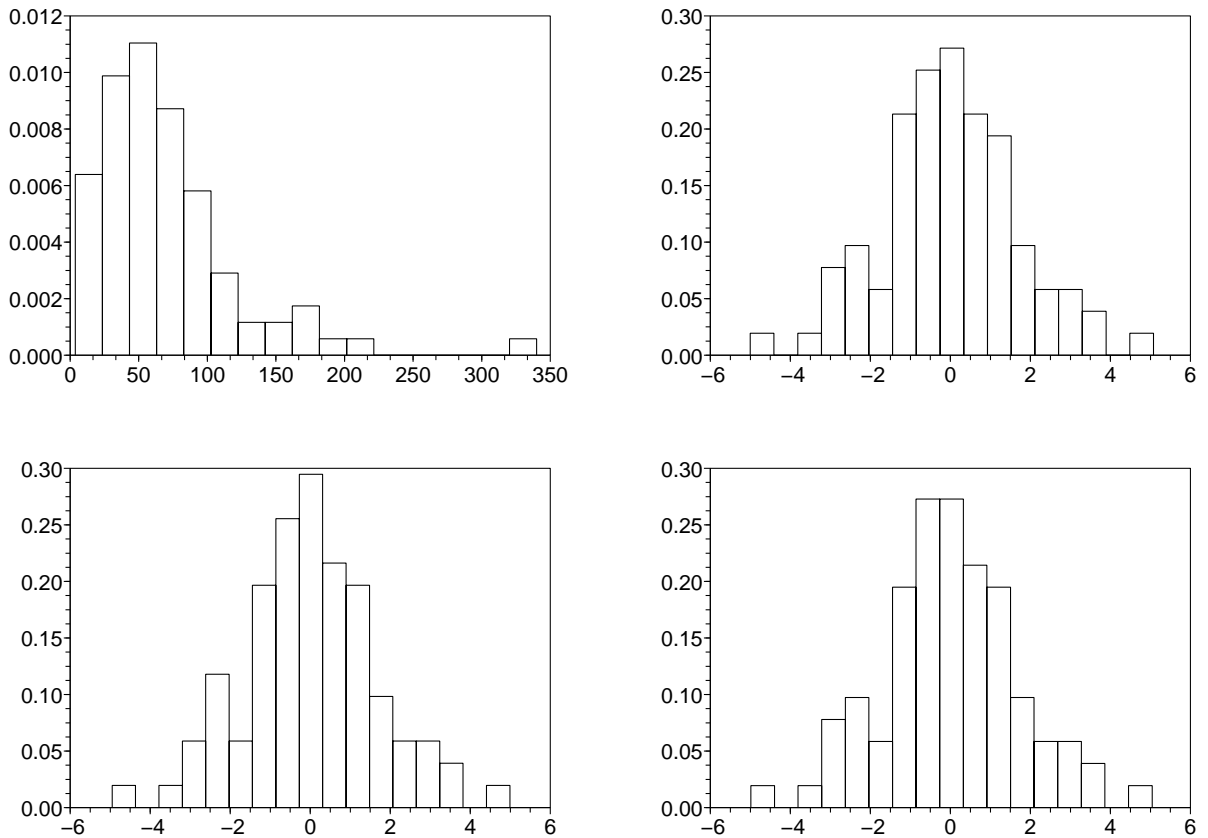


Figure 3: PCB data set, top left: original data, top right: data transformed with the parameters obtained by the M_1 technique, bottom left: data transformed with the parameters obtained by the M_2 technique, bottom right: data transformed with the parameters obtained by the GMLE technique.

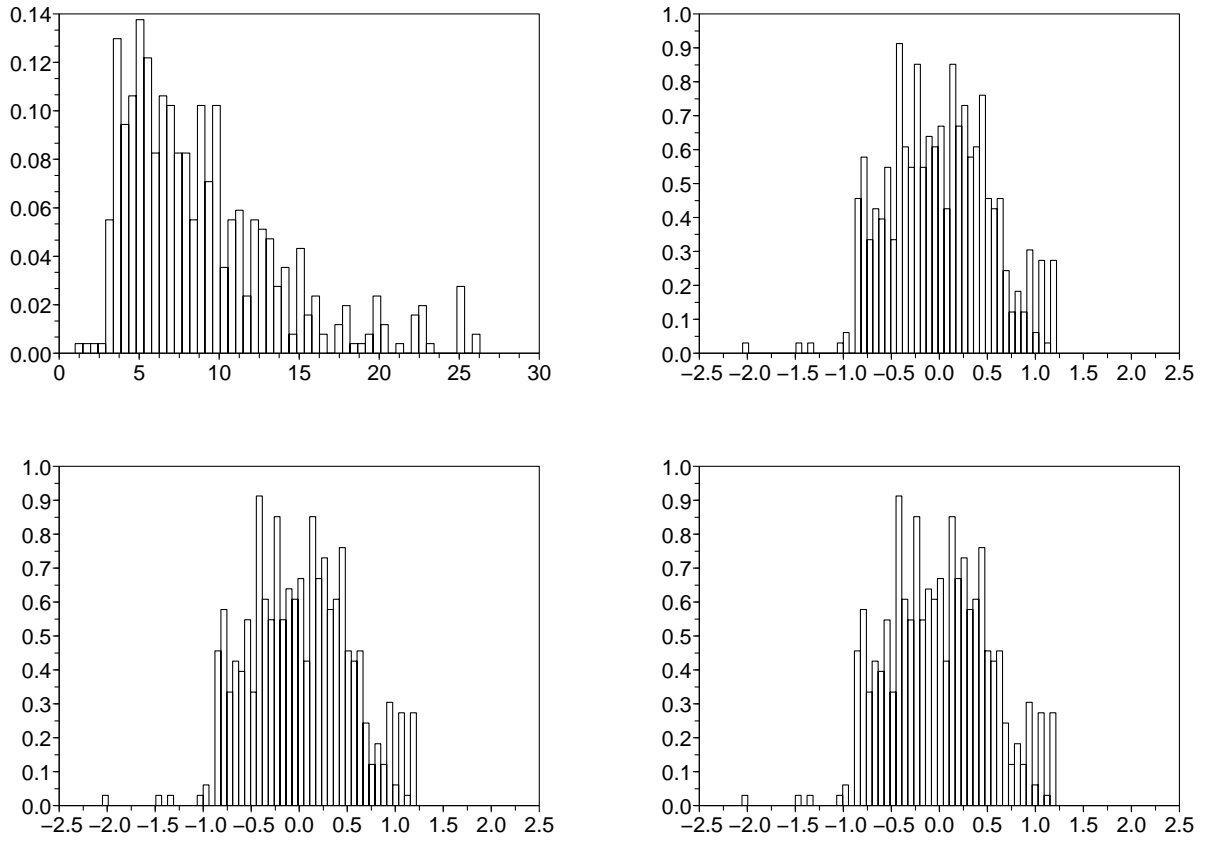


Figure 4: Wage data set, top left: original data, top right: data transformed with the parameters obtained by the M_1 technique, bottom left: data transformed with the parameters obtained by the M_2 technique, bottom right: data transformed with the parameters obtained by the GMLE technique.