Discussiones Mathematicae
Probability and Statistics 36 (2016) 5–23
doi:10.7151/dmps.1185

ON NON-EXISTENCE OF MOMENT ESTIMATORS OF THE GED POWER PARAMETER

Bartosz Stawiarski [1]

Cracow University of Technology
Institute of Mathematics

e-mail: bstawiarski@pk.edu.pl

Abstract

We reconsider the problem of the power (also called shape) parameter estimation within symmetric, zero-mean, unit-variance one-parameter Generalized Error Distribution family. Focusing on moment estimators for the parameter in question, through extensive Monte Carlo simulations we analyze the probability of non-existence of moment estimators for small and moderate samples, depending on the shape parameter value and the sample size. We consider a nonparametric bootstrap approach and prove its consistency. However, despite its established asymptotics, bootstrap does not substantially improve the statistical inference based on moment estimators once they fall into the *non-existence area* in case of small and moderate sample sizes.

Keywords: Generalized Error Distribution, nonparametric bootstrap, bootstrap consistency, moment estimator, power parameter.

2010 Mathematics Subject Classification: 60E05, 62F12, 62F40, 60F99.

The paper is dedicated to Prof. Roman Zmyślony on his 70-th Birthday

1. Introduction

Generalized Error Distribution, in brief GED, called alternatively Exponential Power Distribution or also Generalized Gaussian allows for flexible modelling the tail behavior with a power (shape) parameter $\alpha > 0$, encompassing Gaussian

[1] The research is supported by scientific grant No. UMO-2013/10/M/ST1/00096, National Centre of Science, Poland.

($\alpha = 2$) and Laplace ($\alpha = 1$) laws as special cases. Peculiarities spotted in real data (skewness, asymmetry) have translated into vast extensions of the pioneer, two-parameter, scale-shape family, see eg. [1, 2, 7, 20]. Parameter estimation within the GED class has been quite extensively addressed in literature since late 1980s. Some long established, classic results can be found e.g. in Varanasi and Aazhang (1989) [19] who considered three-parameter, location-scale-shape GED family. They provided k-th ($k \geq 4$, even) moment (MM) and maximum likelihood (ML) estimators of the shape α under known location μ and scale $\sigma > 0$, and proved the asymptotic efficiency of MLE $\hat{\alpha}$.

Here we focus on the power parameter within a zero-mean, unit-variance symmetric class useful e.g. as a noise in conditionally heteroscedastic GARCH-type time series applied in econometrics. Proper estimation and statistical inference about α has quite recently emerged as the challenging and not-so-obvious problem, because only quite recently González-Farías *et al.* (2009) in [8] pointed out that MM estimators in some cases may not exist at all. In our paper, we will address this challenging problem extensively.

In some papers the power parameter estimation has been carried out in the two-parametric, scale-shape GED setup. Chen and Beaulieu (2009) [5] treated the scale as a nuisance parameter and proposed several concurrent shape estimators based either on fractional moments, or proportions of various order moments. These proportional matchings extend the idea of Mallat (1989) [11], who originally considered the ratio $E|X|/DX$. Due to GED moment properties, analytically cumbersome ratios of gamma functions appear there, which calls for numerical derivation of MM estimators. MLE can also be obtained numerically from still more nonlinear equation based on the accompanying score function. Detailed performance analysis of moment and ML estimators was provided in [8]. They argue that in order to ensure the existence of MM estimators sample size n has to be increased, especially for larger α (i.e., closer to the Gaussian case). This is consistent with results of Chen and Beaulieu (2009) [5], who reported faster growth of RMSE for $\alpha > 1.2$. Separately, Krupiński and Purczyński (2006) considered in [9] the shape estimation by approximating the moments proportion by suitable, elementary functions, which was shown to work well for $\alpha \in [0.3; 3]$. These marginal amendments are achieved by both approximations of non-elementary, highly nonlinear functions describing GED theoretical moments and fine-tunings of asymptotic variances of the associated estimators. [15], in turn, provided a globally convergent estimation technique, circumventing the interpolations of polygamma functions appearing in equations used for the estimator derivation. As argued therein, the whole estimation procedures, even if performed using painstaking iterative methods, may be vulnerable to numerical inaccuracy whenever initial guess for α is poor. Indeed, the form of Fisher information $I(\alpha)$ provided in [19] contains trigamma functions, compare also [13].

In this paper, we further explore the problematic issues concerning the shape parameter estimation. By extensive simulations we estimate the non-existence probability of MM for small and moderate sample size. We also examine whether bootstrap approach is helpful in enhancing the statistical inference quality whenever important applicational problems are addressed.

The paper is organized as follows. In Subsection 1.1 we briefly discuss the "classic" (location-scale-shape) GED parametrizations and show how to obtain the standardized one-parameter class, focal in our research. Section 2 deals first with analysis of theoretical moments functions and presents detailed results concerning derivation of MM estimators, together with their asymptotics (using delta method). We also estimate the probability of non-existence of MM for $k = 4$ and 6, depending on and the sample size n. In Section 3 we turn to a qualitatively different approach, namely nonparametric bootstrap technique for *iid* data. We check whether bootstrap approach can improve the discussed estimators performance. Using the suitable delta method theorems for bootstrap in [18], we prove consistency of bootstrap in case of MM estimators. Section 4 provides an extensive Monte Carlo and bootstrap simulation study, with the aim of providing some more in-depth insight into the properties of MM estimators and their bootstrapped versions. Next, after brief conclusions, proofs of theorems collected in Appendix for convenience conclude the paper.

1.1. Concurrent parametrizations. Normalized, symmetric, unit-variance GED(α) class

There are several concurrent GED parametrizations appearing in vast literature. Focusing on up to a three-parameter location-scale-shape family, let us briefly discuss relations between selected scale-shape representations and then proceed to the standardized model indexed solely by a shape parameter.

Subbotin (1923) in [16] proposed a pioneer, two-parameter GED model:

$$f(x,h,m) = \frac{mh}{2\Gamma(1/m)} \exp\{-h^m|x|^m\}, \quad x \in R, \tag{1}$$

with $h > 0, m \geq 1$ being the scale and shape parameters, respectively, and Γ denoting the Euler gamma function. Therefore, this "error" model did not allow for tails heavier than those of Laplace distribution, see also [4]. Coin (2011) in [6] pointed at cumbersome instability of the shape estimation procedure whenever the parameter approaches unity. Lunetta (1963) considered an alternative representation for the three-parameter GED class, namely

$$f(x;\mu,\sigma,\alpha) = \frac{1}{2\sigma\alpha^{1/\alpha}\Gamma(1+1/\alpha)} \exp\left\{-\frac{1}{\alpha}\left|\frac{x-\mu}{\sigma}\right|^{\alpha}\right\} \tag{2}$$

with location $\mu \in R$, scale $\sigma > 0$ and shape $\alpha > 0$, see [10]. In this approach tails heavier than Laplace's are allowed, thus widening the scope of potential applications. Obviously, m in (1) equals α in (2) but $h = (\alpha^{1/\alpha}\sigma)^{-1}$, while by $r\Gamma(r) = \Gamma(r+1)$ the equations (1) and (2) are equivalent as long as $\mu = 0$. The class (2) was recently reconsidered in [20] as the starting point for their work concerning an asymmetric, skew generalization. For a r.v. X with the density f given by (2) it holds $\sigma = \{E|X-\mu|^\alpha\}^{1/\alpha}$, which gives the normal case for $\alpha = 2$, see [12] for some complementary discussion. In the above approach, whenever $\alpha \neq 2$, σ must not be confused with standard deviation of X. However, in the following parametrization of [19]:

$$f(x;\sigma,\alpha) = (2\Gamma(1+1/\alpha)A(\sigma,\alpha))^{-1} \exp\left\{-\left|\frac{x-\mu}{A(\sigma,\alpha)}\right|^\alpha\right\}, \tag{3}$$

where $A(\sigma,\alpha) = \sigma\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}$, σ indeed stands for dispersion.

Setting $\mu = 0$ and $\sigma = 1$ in (3), we obtain the one-parameter, unit-variance family which we will exploit in the sequel:

$$f(x,\alpha) = \frac{\alpha}{2A(\alpha)\Gamma(1/\alpha)} \exp\left\{-\left|\frac{x}{A(\alpha)}\right|^\alpha\right\}, \tag{4}$$

where $A(\alpha) = \sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}, \alpha > 0, x \in R$; compare II.C in [19]. For succinctness, we will denote $X \sim GED(\alpha)$ whenever X obeys the law (4).

Numerous concurrent parametrizations (at various generalization stages) that can be found in literature, see e.g. [1], accordingly call for adequate calculus regarding broadly understood statistical inference and modelling.

As mentioned in Introduction, reliable estimation of α poses a challenging task due to highly nonlinear moment structure of the GED class. Varanasi and Aazhang (1989), [19], considered: moment (MM), maximum likelihood (ML) and hybrid, moment/Newton-step (MMNS) estimators for various combinations of the three parameters in (3), including estimation of α alone under given μ and σ. In this specific case they provide (without derivation) a formula for the inverse asymptotic variance of the efficient ML estimator $\hat{\alpha}_{ML}$, i.e., $(nI(\alpha))^{-1}$, containing elaborate di- and trigamma functions.

2. MM estimators and their non-existence probability

Let $\underline{X}^{(n)} = \{X_1, \ldots, X_n\}$ be a random sample from the GED(α) family given by (4). We focus on k-th, $k \geq 4$, even, moment estimators of the shape parameter α. As $E(X^2) = 1$, direct and well-known calculation gives, for any even $k \geq 4$

$$\mu_k(\alpha) = E(X^k) = \frac{\Gamma((k+1)/\alpha)}{\Gamma(1/\alpha)}[C_0(\alpha)]^{-k} \tag{5}$$

with $C_0(\alpha) = 1/A(\alpha) = \sqrt{\Gamma(3/\alpha)/\Gamma(1/\alpha)}$.

On the other hand, the corresponding empirical moment $\hat{m}_k^{(n)} = n^{-1}\sum_{i=1}^n X_i^k$ is by the laws of large numbers, both weakly and strongly consistent estimator of $\mu_k(\alpha)$:

$$\hat{m}_k^{(n)} = n^{-1}\sum_{i=1}^n X_i^k \overset{n\to\infty}{\longrightarrow} \mu_k(\alpha). \tag{6}$$

Hence, quite intuitively, numerical inversion of μ_k is carried out, leading to an "estimator" $\hat{\alpha}_{MM,k}^{(n)}$ proposed long ago in [19]

$$\hat{\alpha}_{MM,k}^{(n)} = \mu_k^{-1}(\hat{m}_k^{(n)}); k = 4, 6, \ldots \tag{7}$$

provided μ_k are indeed invertible, which we will address soon.

Until quite recently, the existence of MM estimators had been taken for granted, so the solution (7) would have been found along the above lines for some predetermined even, k-th order $k \geq 4$. However, the authors of [8] and [5] questioned the existence of this solution in some cases, hence the quotation marks referring to $\hat{\alpha}_{MM,k}^{(n)}$. Let us examine this problem more closely and define a set

$$N = N(n, k, \alpha) = \{\omega \in \Omega : \neg\exists \hat{\alpha}_{MM,k}^{(n)}(\omega)\} \tag{8}$$

on which the MM estimator does not exist. Clearly, on the set N the random variable $\hat{\alpha}_{MM,k}^{(n)} = \hat{\alpha}_{MM,k}$ is not a valid moment estimator, which may severely impair a statistical inference based on it. We will show more precisely, that at least for small or moderate n (7) can be solved only with some non-unit probability, depending on: the shape parameter α itself, sample size n and moment order k, but not depending on shift μ or scale σ, whenever a three-parameter family is considered, see also [8]. However, asymptotically, for large n, $\hat{\alpha}_{MM,k}$ is an estimator for which standard limit results hold.

Before assessing the probability $P(N)$ (in practice: the risk of nonexistence of $\hat{\alpha}_{MM,k}$), let us first explore analytical properties of $\mu_k(\alpha)$. In our study and simulations below we will consider $k = 4$ and 6. Smoothness of gamma function ensures that $\mu_k(\alpha)$ is smooth on R_+ for any even k. By (5),

$$\begin{aligned} \mu_4(\alpha) &= \frac{\Gamma(5/\alpha)}{\Gamma(1/\alpha)}[C_0(\alpha)]^{-4} = \frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma^2(3/\alpha)}, \\ \mu_6(\alpha) &= \frac{\Gamma(7/\alpha)}{\Gamma(1/\alpha)}[C_0(\alpha)]^{-6} = \frac{\Gamma(7/\alpha)\Gamma^2(1/\alpha)}{\Gamma^3(3/\alpha)}, \end{aligned} \tag{9}$$

which can be consistently estimated by $\hat{m}_4$ and $\hat{m}_6$, respectively. The theorem and conjecture below deal with monotonicity of μ_4 and μ_6 in our GED(α) family (4).

Theorem 2.1. *The function $\mu_4(\alpha)$ is strictly decreasing on $(0,\infty)$.*

Theorem 2.2 (Conjecture)**.** *The function $\mu_6(\alpha)$ is strictly decreasing on $(0,\infty)$.*

In Appendix we provide an analytical proof of Theorem 2.1 whereas 2.2 is shown numerically, but with an arbitrary degree of accuracy. In some papers these results are just stated with no proof at all, see e.g. p. 4216 in [8]. Theorem 2.1 and Conjecture 2.2 state that both $\mu_4(\alpha)$ and $\mu_6(\alpha)$ are invertible, which ensures the existence of MM estimators on $N^C = (N(n,k,\alpha))^C$ for respective k's, for small samples, with N defined in (8). Their asymptotic behavior will be addressed later, and now let us deal with precise estimation of $P(N)$.

The GED random number generator in Matlab is based on appropriate power transformation of gamma distribution taken with a random sign, see e.g. [17]. For our estimation purposes we will work with the interval $[\alpha_-;\alpha^+] = [0.2; 3.5]$ which encompasses the crucial set $[1,2]$ with an ample margin. From (9) we calculate numerically: $\mu_4(\alpha^+) = 2.282, \mu_4(\alpha_-) = 1959.297$; $\mu_6(\alpha^+) = 7.335, \mu_6(\alpha_-) \cong 2.567 \cdot 10^8$. Gradients of μ_4 and μ_6 are very steep for small α and substantially flatten out when α crosses ca. 1 for $k = 4$ and ca. 1.5 for $k = 6$, see Figure 6 in Appendix. Both functions have above-zero horizontal asymptotes in $+\infty$, namely for $k = 4$ and 6, numerical approximations are: 1.8 and 3.8571, respectively. Therefore we can anticipate poorer estimation results for α rather closer to 2 than 1. The event N will occur whenever $\hat{m}_k^{(n)} \notin [\mu_k(\alpha^+); \mu_k(\alpha_-)]$. In practice, only $\hat{m}_k < \mu_k(\alpha^+)$ may happen. Below we estimate $P(N)$ by Monte Carlo simulations based on $M = 50000$ loops. Namely, for each n-element GED sample $\underline{X}^{(n,i)}$ we obtain $\hat{m}_{k,i}^{(n)}$ and approximate

$$\begin{aligned} P(N) &\cong \hat{P}(N(n,k,\alpha)) \\ &= \frac{1}{M} \cdot card\left\{ i \in \{1,\dots,M\} : \hat{m}_{k,i}^{(n)} \notin [\mu_k(\alpha^+); \mu_k(\alpha_-)] \right\}. \end{aligned} \tag{10}$$

The probabilities (10) have been calculated on the following grid of points:

$\alpha \in \{1; 1.1; \dots; 1.9; 2\}; n \in \{50; 100; 200; 300; 400; 500; 1000\}; k \in \{4; 6\}$.

Figures below illustrate the results rounded to the third decimal point. (For clarity and due to capacity constraints, the tables with specific values of $\hat{P}(N)$ were not quoted here, but can be provided at an individual request.)

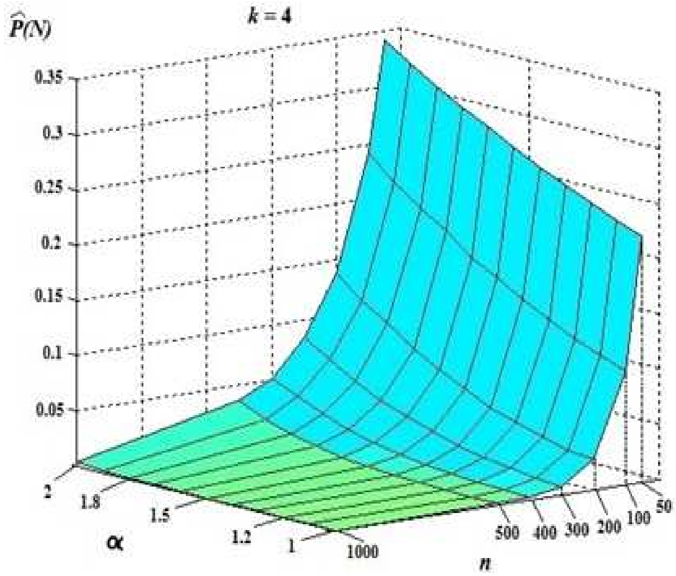

Figure 1. The Monte Carlo-estimated values of $\hat{P}(N(n,4,\alpha))$.

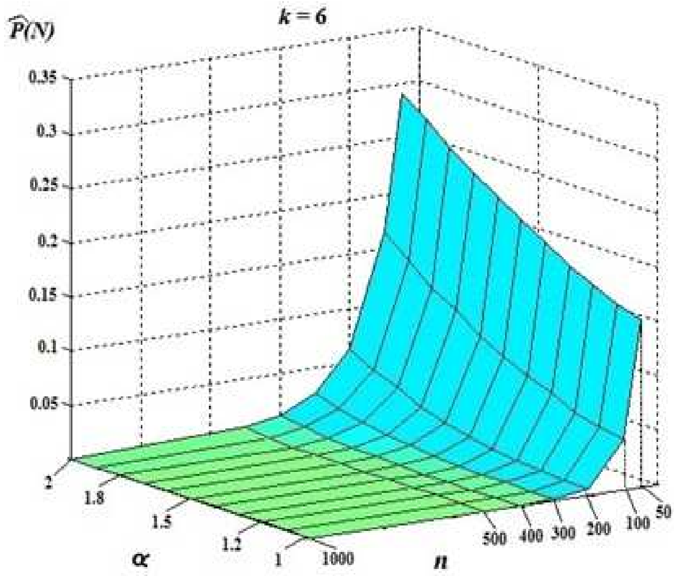

Figure 2. The Monte Carlo-estimated values of $\hat{P}(N(n,6,\alpha))$.

The above results show that the quality of MM estimators for smaller sample sizes (up to 200) proves to be at least questionable. When approaching the Gaussian case the event N occurs roughly in one third of all Monte Carlo loops, whereas around the Laplace case the risk of MM nonexistence drops nearly by half but is still intolerably large. In both cases, $k = 4$ and 6, $\hat{P}(N)$ drops below 0.05 for n at least ca. 400 and 250, respectively. Such a behavior practically disqualifies the MM estimation technique for small sample sizes within our one-parameter GED framework, which will be further confirmed in Section 4. The occurrence of N translates into either $\hat{\alpha}_{MM,k}$ hitting the upper boundary α^+ or being fairly close to it. Increasing α^+ does not improve the overall performance of $\hat{\alpha}_{MM,k}$ for small n, again due to the gradient of μ_k quickly approaching zero for $\alpha > 3.5$. Hence, $\hat{P}(N)$ reported in Figures 1 and 2 drops hardly noticeably. Besides, as ample research shows, the GED class is suitable for modeling heavy-tailed instead of ultra-light-tailed phenomena.

The above structural deficiency is the main driver of overall upward bias of MM evident in simulations presented later in Section 4. Fortunately, however, extensive numerical experiments suggest that $P(N(n,k,\alpha)) \overset{n\to\infty}{\longrightarrow} 0$ uniformly at least for $[1,2]$ and $k \in \{4;6\}$, which is crucial for the asymptotic results for $\hat{\alpha}^{(n)}_{MM,k}$.

First, notice that $\hat{m}_k$ from (6) is an unbiased and consistent estimator of $\mu_k(\alpha)$. Consider a zero-mean random sequence $\sqrt{n}(\hat{m}_k - \mu_k(\alpha))$. Direct derivation gives, by X_i's being *iid* copies of $X \sim GED(\alpha)$,

$$
\begin{aligned}
E\left\{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i^k - \mu_k(\alpha))\right)\right\}^2 &= \frac{1}{n}\sum_{i,j=1}^{n} E(X_i^k - \mu_k(\alpha))(X_j^k - \mu_k(\alpha)) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i^k - \mu_k(\alpha))^2 = D^2(X^k) = \mu_{2k}(\alpha) - \mu_k^2(\alpha).
\end{aligned} \tag{11}
$$

Hence and by the central limit theorem,

$$
\sqrt{n}(\hat{m}_k^{(n)} - \mu_k(\alpha)) \overset{D}{\longrightarrow} N(0, \varsigma_k^2) \tag{12}
$$

as $n \to \infty$ with $\varsigma_k^2 = D^2(X^k)$. Equivalently, we can write $\hat{m}_k^{(n)} \sim AN(\mu_k(\alpha), \varsigma_{n,k}^2)$ with $\varsigma_{n,k}^2 = n^{-1}\varsigma_k^2 \overset{n\to\infty}{\longrightarrow} 0$. At this stage we can use the delta-method technique, cf. [18], to obtain the asymptotic normality of the k-th (even) order MM estimator $\hat{\alpha}^{(n)}_{MM,k}$.

Theorem 2.3. *Let* $\hat{m}_k^{(n)}$ *be an estimator defined by* (6)*, taking values in* R_+*, such that* $\hat{m}_k^{(n)} \sim AN(\mu_k(\alpha), \varsigma_{n,k}^2)$ *with* $\varsigma_{n,k}^2 \overset{n\to\infty}{\longrightarrow} 0$*. Then, for any continuous function* $g : R_+ \to R$ *with non-vanishing derivative* g' *it holds*

$$
g(\hat{m}_k^{(n)}) \sim AN(g(\alpha), [g'(\alpha)]^2 \varsigma_{n,k}^2). \tag{13}
$$

In our context, for $k = 4$ we identify $g(\alpha) = \mu_4^{-1}(\alpha)$ and $g(\alpha) = \mu_6^{-1}(\alpha)$ for $k = 6$. Accordingly, $g(\hat{m}_4^{(n)}) = \hat{\alpha}_{MM,4}^{(n)}$ and $g(\hat{m}_6^{(n)}) = \hat{\alpha}_{MM,6}^{(n)}$. For small n they are valid estimators on the sets $(N(n,4,\alpha))^C$ and $(N(n,6,\alpha))^C$, respectively, which does not invalidate their asymptotics stated in Theorem 2.3, again thanks to $P(N^C) \to 1$ with $n \to \infty$. To be more precise here, let $L_k = \lim_{\alpha\to\infty} \mu_k(\alpha)$ be the horizontal asymptote of μ_k in infinity. These values can be accurately approximated numerically, cf. p. 5, namely $L_4 \cong 1.8$ and $L_6 \cong 3.8571$. Moreover, $\lim_{\alpha\to 0^+} \mu_k(\alpha) = U = +\infty$. By consistency of the moment estimate $\hat{\mu}_k(\alpha)$ based on the *iid* $GED(\alpha)$ sample, for any fixed α it holds $P(\hat{\mu}_k(\alpha) \in (L_k, U)) \to 1$ with $n \to \infty$. Hence indeed $P(N^C(n,k,\alpha)) \to 0$, and the thesis results from Slutzky lemma and continuous invertibility of $\mu_k(\alpha)$ at least for $k = 4$ and 6 we are focused on.

In both cases $g'(\alpha)$ are non-elementary functions which can be found by the inverse function derivative formula applied to (9), but we will not need them in the sequel. Thus (13) states the asymptotic normality of the MM estimator $\hat{\alpha}_{MM,k}^{(n)}$.

As far as the parameter estimation is concerned, the researcher faces the problem of dealing with just one random sample, $\underline{X}^{(n)}$. Both cost constraints accompanying the process of sampling (e.g. very expensive measurements) and real non-replicability of the observed stochastic phenomena (e.g. once-in-a-lifetime recorded econometric data) are serious hindrances disabling precise assessment of the estimators performance. Intolerably large estimated standard error of $\hat{\alpha}_{MM,k}^{(n)}$ may be the main source of this instability. Therefore, in order to gain a better insight into the estimation performance within our GED(α) framework, we consider the nonparametric bootstrap approach. More importantly, this technique still seems to be widely unexplored in the GED context. Sampling with replacement from the original n-element data set in our rather regular parametric family case (the functions involved are smooth) enables us to use some well-established general theorems and obtain the desired asymptotic validity of bootstrap methods.

3. Bootstrap approach

Suppose we draw B samples of size n (each time with replacement) from our original random sample $\underline{X}^{(n)} = \{X_1, \ldots, X_n\} \sim GED(\alpha)$, obtaining $\underline{X}^{(n),*b} = \{X_1^{*b}, \ldots, X_n^{*b}\}, b = 1, \ldots, B$. Applying (7) and (9) we get B bootstrap replications

$$\hat{m}_k^{(n),*b} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^{*b}\right)^k, \quad \hat{\alpha}_{MM,k}^{(n),*b} = \mu_k^{-1}\left(\hat{m}_k^{(n),*b}\right) \tag{14}$$

with the earlier discussion concerning the existence of the MM estimator $\hat{\alpha}_{MM,k}^{(n),*b}$ valid here, too.

For each b-th bootstrap sample $\underline{X}^{(n),*b}$ we will successively show the consistency of the corresponding bootstrap estimators (14). For notational convenience, denote $Y_{i,k} = X_i^k, Y_{i,k}^{*b} = (X_i^{*b})^k$, $1 \le i \le n, 1 \le b \le B$. Recall that $Y_{i,k}$ are *iid* with common mean $\mu_k(\alpha)$ given by (9) and variance ς_k^2 by (11). Then, using the notation above, we have the following theorem.

Theorem 3.1. *For any fixed $b \in \{1, \dots, B\}$ and at least $k = 4; 6$ it holds as $n \to \infty$:*

(i)

$$\sqrt{n}\left(\hat{m}_k^{(n),*b} - \hat{m}_k^{(n)}\right) \xrightarrow{D} N\left(0, \varsigma_k^2\right) \tag{15}$$

conditionally on $\underline{X}^{(n)}$;

(ii)

$$\left(s_k^{(n),*b}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_{i,k}^{*b} - \hat{m}_k^{(n),*b}\right)^2 \longrightarrow \varsigma_k^2 \tag{16}$$

in probability conditionally on $\underline{X}^{(n)}$.

Hence we immediately get, for any fixed $b \in \{1, \dots, B\}$ and $k = 4; 6$,

$$\frac{\sqrt{n}\left(\hat{m}_k^{(n),*b} - \hat{m}_k^{(n)}\right)}{s_k^{(n),*b}} \xrightarrow{D} N(0, 1) \tag{17}$$

conditionally on $\underline{X}^{(n)}$ as $n \to \infty$.

To prove Theorem 3.1 and the convergence (17) it suffices to apply directly Theorem 2.1 in [3]. The above results also hold true if we consider subsampling, i.e., drawing l-element (not necessarily equal to n) bootstrap samples, as long as both l and n tend to infinity. We can expect that the above result holds for any even $k \ge 4$, but this thesis would call for formal proof.

Now, we turn our attention to the bootstrap version $\hat{\alpha}_{MM,k}^{(n),*b}$ of the original MM estimator $\hat{\alpha}_{MM,k}^{(n)}$. The asymptotic result (17) has to be interpreted to hold *conditionally almost surely*, i.e., along almost all sample sequences $Y_{1,k}, Y_{2,k}, \dots$ given $\{Y_{1,k}, \dots, Y_{n,k}\}$ or equivalently, given the original input sample $\underline{X}^{(n)}$. We will establish the bootstrap analogue of (13) from Theorem 2.3.

Theorem 3.2. *Let $g : R \to R$ be a measurable map, defined and continuously differentiable in a neighborhood of θ. Assume that $\hat{\theta}_n$ is a random sequence, based on the sample $\underline{X}^{(n)}$, taking its values in the domain of g and converging almost* *surely to* θ *as* $n \to \infty$. *Then, if for some r.v.* T *it holds* $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} T$ *and for each b-th bootstrap version* $\sqrt{n}(\hat{\theta}_n^{*b} - \hat{\theta}_n) \xrightarrow{D} T$ *conditionally almost surely with* $n \to \infty$, *then*

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta)\right) \xrightarrow{D} (g)'_\theta(T) \tag{18}$$

and

$$\sqrt{n}\left(g(\hat{\theta}_n^{*b}) - g(\hat{\theta}_n)\right) \xrightarrow{D} (g)'_\theta(T) \tag{19}$$

as $n \to \infty$.

The above result states the bootstrap consistency.

4. Simulation study

In this section we compare the performance of MM-4, MM-6 (4^{th} and 6^{th} moment, respectively) estimators of α by extensive Monte Carlo and bootstrap simulation study. This work is conducted completely independently of the estimation studies using alternative estimators considered in papers mentioned in Introduction.

The simulation experiment is designed as follows. For generated zero-mean, unit-variance *iid* random samples $\{X_1, \ldots, X_n\} \sim GED(\alpha)$ of various lengths $n \in \{100; 200; 500; 1000\}$ and gradually changing shape $\alpha \in \{1; 1.1; \ldots, 2\}$, $MC = 2000$ Monte Carlo iterations are carried out to estimate sample means and standard errors of $\hat{\alpha}_4$ and $\hat{\alpha}_6$. The first collection of results is provided in Tables 1–4. The following columns contain: the true parameter value, sample means of MM estimators searched on the 0.001 grid within the [0.2; 3.5] interval; their estimated SE's.

Evident upward bias results mainly from the aforementioned small gradient of μ_k, which translates into either invalid estimator, or substantial shift to the right. This effect tapers off with growing n.

Below, by simulations we explore the quality of inference about α by bootstrapping all the considered estimators. We can expect that bootstrapping the possibly non-existent MM estimators (i.e., when the N event occurs) especially for small samples will not improve the statistical inference, independent of the bootstrap consistency proved in Section 3.

For fixed n we generate a random sample of size n from $GED(\alpha)$ which will be called an original sample. Tables 5–8 contain an extensive bootstrap simulation study, with fixed $B = 2000$, retaining the results for the original sample subject to bootstrapping.

Whenever $\hat{P}(N)$ is unacceptably high, we can be suspicious that the event N occurred for the original sample itself with quite large probability. Accordingly, the originally calculated MM might have not been a valid moment estimator. In that case it is easily seen that bootstrap is incapable of amending this structural deficiency. We can say that the original flaw inherent in original samples percolates into bootstrapped samples. Bias-corrected version of bootstrap would not be a remedy here, either.

Table 1. MM-4, MM-6 estimators performance Monte Carlo study, $n = 100$.

$n = 100$				
α	$\overline{\hat{\alpha}_4}$	$\overline{\hat{\alpha}_6}$	$\hat{SE}(\hat{\alpha}_4)$	$\hat{SE}(\hat{\alpha}_6)$
1	1.526	1.512	0.888	0.713
1.1	1.638	1.612	0.886	0.709
1.2	1.680	1.673	0.871	0.700
1.3	1.756	1.746	0.874	0.702
1.4	1.853	1.851	0.874	0.720
1.5	1.948	1.951	0.870	0.722
1.6	2.034	2.031	0.869	0.721
1.7	2.105	2.117	0.862	0.718
1.8	2.189	2.210	0.850	0.711
1.9	2.268	2.297	0.854	0.725
2	2.345	2.377	0.834	0.699

Table 2. MM-4, MM-6 estimators performance Monte Carlo study, $n = 200$.

$n = 200$				
α	$\overline{\hat{\alpha}_4}$	$\overline{\hat{\alpha}_6}$	$\hat{SE}(\hat{\alpha}_4)$	$\hat{SE}(\hat{\alpha}_6)$
1	1.306	1.296	0.632	0.457
1.1	1.378	1.369	0.623	0.449
1.2	1.488	1.472	0.665	0.488
1.3	1.588	1.569	0.654	0.482
1.4	1.674	1.657	0.668	0.506
1.5	1.803	1.774	0.701	0.532
1.6	1.889	1.864	0.703	0.541
1.7	2.003	1.977	0.729	0.568
1.8	2.060	2.047	0.713	0.567
1.9	2.144	2.138	0.702	0.566
2	2.249	2.232	0.709	0.574

Table 3. MM-4, MM-6 estimators performance Monte Carlo study, $n = 500$.

$n = 500$				
α	$\overline{\hat{\alpha}_4}$	$\overline{\hat{\alpha}_6}$	$\hat{SE}(\hat{\alpha}_4)$	$\hat{SE}(\hat{\alpha}_6)$
1	1.097	1.125	0.288	0.231
1.1	1.204	1.223	0.315	0.243
1.2	1.300	1.315	0.341	0.265
1.3	1.403	1.412	0.351	0.267
1.4	1.501	1.506	0.369	0.279
1.5	1.621	1.615	0.400	0.297
1.6	1.746	1.725	0.443	0.321
1.7	1.836	1.815	0.456	0.330
1.8	1.922	1.902	0.461	0.324
1.9	2.022	1.997	0.484	0.352
2	2.138	2.113	0.500	0.363

Table 4. MM-4, MM-6 estimators performance Monte Carlo study, $n = 1000$.

$n = 1000$				
α	$\overline{\hat{\alpha}_4}$	$\overline{\hat{\alpha}_6}$	$\hat{SE}(\hat{\alpha}_4)$	$\hat{SE}(\hat{\alpha}_6)$
1	1.047	1.074	0.194	0.172
1.1	1.156	1.176	0.196	0.174
1.2	1.249	1.265	0.212	0.183
1.3	1.354	1.367	0.225	0.190
1.4	1.456	1.463	0.242	0.198
1.5	1.553	1.559	0.251	0.203
1.6	1.660	1.660	0.281	0.221
1.7	1.765	1.758	0.299	0.232
1.8	1.870	1.860	0.317	0.234
1.9	1.979	1.961	0.343	0.246
2	2.082	2.063	0.361	0.259

It is worth mentioning, that results for MM-6 are somewhat better than for MM-4, mainly thanks to $P(N(n, 6, \alpha)) < P(N(n, 4, \alpha))$. The consistency is visible, albeit rather slow. Poorer quality is more pronounced for larger shape values, which stands in accordance with results reported in [5].

Table 5. Bootstrapped MM-4 and MM-6 estimators performance, $n = 100$.

$n = 100$											
original sample											
	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
$\hat{\alpha}_4$	0.775	1.132	1.739	1.606	0.797	1.703	2.921	1.655	2.077	0.982	2.903
$\hat{\alpha}_6$	0.919	1.135	1.958	1.628	0.894	2.070	2.575	1.929	2.253	0.959	3.427
bootstrap											
MM-4											
$\overline{\hat{\alpha}_4^*}$	1.021	1.652	2.023	1.969	1.192	1.931	2.826	1.900	2.326	1.815	2.878
$\hat{SE}(\hat{\alpha}_4^*)$	0.659	1.008	0.750	0.856	0.851	0.644	0.734	0.671	0.754	1.282	0.609
$\hat{P}^*(N)$	0.065	0.249	0.294	0.280	0.119	0.227	0.713	0.227	0.444	0.365	0.778
MM-6											
$\overline{\hat{\alpha}_6^*}$	1.098	1.572	2.148	1.850	1.245	2.242	2.737	2.107	2.450	1.840	3.244
$\hat{SE}(\hat{\alpha}_6^*)$	0.500	0.845	0.539	0.585	0.751	0.508	0.597	0.512	0.587	1.259	0.335
$\hat{P}^*(N)$	0	0	0	0	0	0	0	0	0	0	0

Table 6. Bootstrapped MM-4 and MM-6 estimators performance, $n = 200$.

$n = 200$											
original sample											
α	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
$\hat{\alpha}_4$	1.120	2.093	0.792	1.210	1.539	1.107	0.982	1.600	1.650	1.850	1.902
$\hat{\alpha}_6$	1.358	1.807	0.857	1.281	1.782	1.185	1.220	1.845	1.671	2.141	1.964
bootstrap											
MM-4											
$\overline{\hat{\alpha}_4^*}$	1.187	2.345	1.030	1.399	1.684	1.285	1.033	1.741	1.904	2.001	2.144
$\hat{SE}(\hat{\alpha}_4^*)$	0.285	0.805	0.502	0.548	0.466	0.514	0.220	0.494	0.696	0.552	0.656
$\hat{P}^*(N)$	0.009	0.442	0.030	0.072	0.096	0.054	0.002	0.117	0.225	0.238	0.337
MM-6											
$\overline{\hat{\alpha}_6^*}$	1.402	2.008	1.105	1.420	1.906	1.341	1.256	1.953	1.851	2.240	2.137
$\hat{SE}(\hat{\alpha}_6^*)$	0.191	0.559	0.420	0.375	0.380	0.384	0.169	0.381	0.497	0.388	0.480
$\hat{P}^*(N)$	0	0	0	0	0	0	0	0	0	0	0

Table 7. Bootstrapped MM-4 and MM-6 estimators performance, $n = 500$.

$n = 500$											
original sample											
α	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
$\hat{\alpha}_4$	0.912	1.085	1.114	1.285	1.427	1.442	1.638	1.743	1.632	1.955	1.921
$\hat{\alpha}_6$	0.979	1.102	1.202	1.345	1.399	1.627	1.682	1.856	1.794	1.953	1.847
bootstrap											
MM-4											
$\overline{\hat{\alpha}_4^*}$	0.972	1.193	1.211	1.363	1.550	1.485	1.745	1.834	1.697	2.090	2.068
$\hat{SE}(\hat{\alpha}_4^*)$	0.208	0.336	0.269	0.303	0.417	0.241	0.422	0.379	0.314	0.502	0.536
$\hat{P}^*(N)$	0	0.001	0.001	0.002	0.006	0	0.008	0.009	0.002	0.031	0.034
MM-6											
$\overline{\hat{\alpha}_6^*}$	1.035	1.162	1.245	1.398	1.464	1.660	1.761	1.917	1.843	2.030	1.957
$\hat{SE}(\hat{\alpha}_6^*)$	0.166	0.188	0.175	0.192	0.237	0.186	0.300	0.255	0.237	0.314	0.378
$\hat{P}^*(N)$	0	0	0	0	0	0	0	0	0	0	0

Table 8. Bootstrapped MM-4 and MM-6 estimators performance, $n = 1000$.

$n = 1000$											
original sample											
α	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
$\hat{\alpha}_4$	0.946	1.095	1.282	1.396	1.458	1.526	1.545	1.758	1.827	1.899	2.085
$\hat{\alpha}_6$	1.013	1.163	1.287	1.360	1.510	1.594	1.588	1.762	1.852	2.017	2.099
bootstrap											
MM-4											
$\overline{\hat{\alpha}_4^*}$	0.975	1.127	1.330	1.465	1.502	1.572	1.580	1.820	1.887	1.945	2.162
$\hat{SE}(\hat{\alpha}_4^*)$	0.136	0.155	0.223	0.273	0.221	0.234	0.224	0.298	0.297	0.265	0.361
$\hat{P}^*(N)$	0	0	0.003	0.016	0.006	0.011	0.009	0.072	0.096	0.113	0.303
MM-6											
$\overline{\hat{\alpha}_6^*}$	1.045	1.189	1.318	1.435	1.550	1.630	1.614	1.801	1.886	2.047	2.147
$\hat{SE}(\hat{\alpha}_6^*)$	0.116	0.115	0.139	0.226	0.175	0.171	0.160	0.193	0.188	0.184	0.247
$\hat{P}^*(N)$	0	0	0	0	0	0	0	0	0	0	0

5. Conclusions

We examined the performance of 4^{th} and 6^{th} moment estimators of the shape parameter in zero-mean, unit-variance *iid* GED framework. It may be of special practical interest in the context of conditionally GED-distributed GARCH models, applied in econometric modeling. Provided that the time series model is properly specified, one can deal with the filtered noise. Especially strong emphasis was put on the problems concerning the poor quality of the MM estimators, which is caused by their occasional non-existence for smaller sample sizes. This undesired event, called here N, may considerably impair the statistical inference, therefore we focused on possibly exact assessing its probability. Consistency of bootstrap applied to MM has been established, which is confirmed by our simulations. The results are promising but it must be stated that bootstrap is not a remedy for structural deficiencies of the original estimators when computed out of small data sets.

In opposition to MM, the maximum likelihood estimator can be considered as far more reliable in real modeling problems, which is a subject of current research, also in the context of bootstrap approach. It is also quite natural to analyze the bootstrap performance in case of the more general three-parameter setup, or even admitting skewness and various left and right tail behavior, such as proposed in [1] or [20]. In general, studying the performance of bootstrap procedures in the context of GED class is an attractive and still relatively poorly examined path of research.

6. Appendix

***Proof of Theorem* 2.1.** It suffices to show that derivatives $\mu_4^{'}(\alpha)$ and $\mu_6^{'}(\alpha)$ are strictly negative functions of α.

For any $r, s > 0$ define an auxiliary function on R_+

$$\xi(x) = \frac{\Gamma^2((r+s)x)}{\Gamma(2rx)\Gamma(2sx)}. \tag{20}$$

We will show that ξ is strictly decreasing. It will be enough to show that $(\log \xi(x))^{'} < 0$. By Euler formula it holds

$$\Psi(x) = (\log \Gamma(x))^{'} = -\gamma + \int_0^1 \frac{1-t^{x-1}}{1-t} dt \tag{21}$$

with Euler constant $\gamma \approx 0.5772$. Hence, by (20), (21) and direct calculus $(\log \xi(x))^{'} = 2\Psi((r+s)x) - \Psi(2rx) - \Psi(2sx) = -\int_0^1 \frac{(t^{rx}-t^{sx})(2rt^{rx-1}-2st^{sx-1})}{1-t} dt.$

On the other hand, integration by parts yields

$$\begin{aligned} -\int_0^1 \frac{(t^{rx}-t^{sx})^2}{x(1-t)^2} dt &= \left[\frac{(t^{rx}-t^{sx})^2}{x(1-t)} \right]_{t=0}^{t=1^-} \\ -2\int_0^1 \frac{(t^{rx}-t^{sx})(rt^{rx-1}-st^{sx-1})}{1-t} dt &< 0 \end{aligned} \tag{22}$$

so $(\log \xi(x))^{'} < 0$ indeed, because by de l'Hospital rule, $\lim_{t\to 1^-} \frac{(t^{rx}-t^{sx})^2}{x(1-t)} = 0$.

To complete the proof of Theorem 2.1, set $r = 5/2, s = 1/2$ in (20) and for $x = 1/\alpha$ observe that $\mu_4(\alpha) = 1/\xi(1/\alpha)$ is strictly decreasing as well.

***Proof of Conjecture* 2.2.** This time the former approach does not apply. The function $\mu_6(\alpha)$ is hardly tractable analytically, therefore we will show the monotonicity numerically (this can be done at any level of accuracy). We calculate the function values on a dense grid of α coordinates and plot it. The Figure 3 below presents the graphs restricted to $[0.5; 3.5]$ for μ_4 and $[1.1; 3.5]$ for μ_6 with a grid resolution equal to 0.001. Steepness of the slope is by far larger in case of the 6^{th} moment. Both functions explode to infinity when α approaches 0.

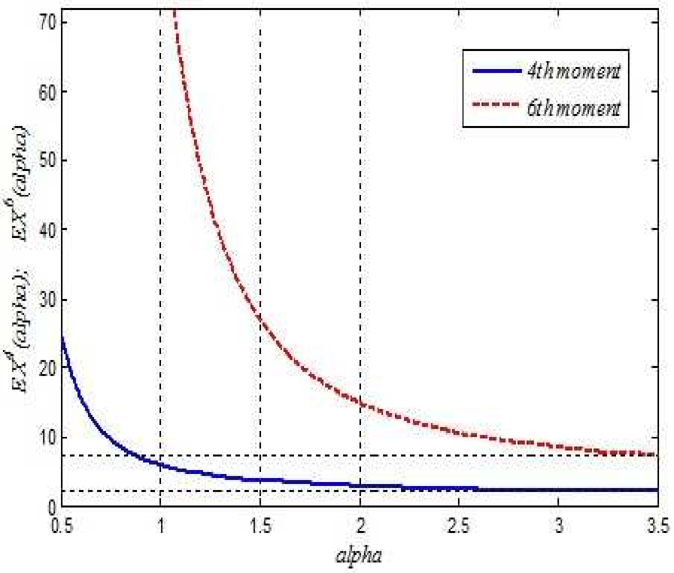

Figure 3. Behavior of $\mu_4(\alpha)$ and $\mu_6(\alpha)$ for applicationally relevant shape values.

In fact, rapidly evanescing gradient of μ_4 for $\alpha > 1.2$ translates into poor statistical quality of the estimator $\hat{m}_4$ provided it exists, which could be noticed in our simulations. Obviously, monotonicity of μ_6 can be shown graphically along the whole domain R_+ for arbitrarily dense grid resolution of α.

***Proof of Theorem* 2.3.** The thesis of this theorem can be directly inferred from the univariate version of Theorem 23.5 in [19], which we now adapt to our specific GED case. Accordingly, we successively identify for $k = 4; 6$: $\theta = \theta_k = \mu_k(\alpha); \alpha > 0$; $\hat{\theta}_n = \hat{m}_k^{(n)}; g(\theta) = \mu_k^{-1}(\theta)$ is for $k = 4$ continuously differentiable by Theorem 2.1 and the inverse function derivative (for $k = 6$ Conjecture 2.2 suggests that μ_6 has non-zero derivative, too). Next, $T \sim N(0, \varsigma_k^2)$, see (12), while derivative $(g)'_\theta = (\mu_k^{-1}(\theta))'_\theta$ is continuous by Theorem 2.1. Finally, $g(\hat{\theta}_n) = \hat{\alpha}_{MM,k}^{(n)}$ and $g(\hat{\theta}_n^{*b}) = \hat{\alpha}_{MM,k}^{(n),*b}$ for $1 \leq b \leq B$.

In fact, (18) is just a more convenient restatement of (13) whereas the crucial result (19) is equivalent to bootstrap consistency of for each fixed $b = 1, \ldots, B$ and even $k \geq 4$.

Acknowledgement

The author would like to thank K. Sheppard for providing a Matlab script with the GED random number generator as well as an anonymous Referee for her/his precious remarks.

References

[1] A. Ayebo and T.J. Kozubowski, *An asymmetric generalization of Gaussian and laplace laws*, J. Probab. Statist. Sci. **1** (2) (2004) 187–210.

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

[3] P.J. Bickel and D.A. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Stat. **9** (6) (1981) 1196–1217. doi:10.1214/aos/1176345637

[4] G.E. Box and G.C. Tiao, Bayesian Inference in Statistical Analysis (Addison Wesley Ed., 1973).

[5] Y. Chen and N.C. Beaulieu, *Novel low-complexity estimators for the shape parameter of the Generalized Gaussian Distribution*, IEEE Transactions on Vehicular Technology **58** (4) (2009) 2067–2071. doi:10.1109/TVT.2008.2003079

[6] D. Coin, *A method to estimate power parameter in exponential power distribution via polynomial regression*, Banca D'Italia 834 (2011), working paper.

[7] C. Fernandez, J. Osiewalski and M.F.J. Steel, *Modeling and inference with v-distributions*, J. Amer. Statist. Association **90** (432) (1995) 1331–1340. doi:10.1080/01621459.1995.10476637

[8] G. González-Farías, J.A. Domnguez-Molina and R.M. Rodrguez-Dagnino, *Efficiency of the approximated shape parameter estimator in the generalized Gaussian distribution*, IEEE Transactions on Vehicular Technology **58** (8) (2009) 4214–4223. doi:10.1109/TVT.2009.2021270

[9] R. Krupiński and J. Purczyński, *Approximated fast estimator for the shape parameter of Generalized Gaussian Distribution*, Signal Processing **86** (2006) 205–211. doi:10.1016/j.sigpro.2005.05.003

[10] G. Lunetta, *Di una generalizzazione dello schema della curva normale*, Annali della Facolta di Economia e Commercio di Palermo **17** (1963) 237–244.

[11] S. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Trans. Pattern Anal. Machine Intell. **11** (1989) 674–693. doi:10.1109/34.192463

[12] A.M. Mineo and M. Ruggieri, *A software tool for the exponential power distribution: The normalp package*, J. Statist. Software **12** (4) (2005) 1–24. doi:10.18637/jss.v012.i04

[13] S. Nadarajah, *A generalized normal distribution*, J. Appl. Stat. **32** (7) (2005) 685–694.

[14] J. Purczyński, *Simplified method of GED distribution parameters estimation*, Folia Oeconomica Stetinensia **10** (2) (2012) 35–49.

[15] K.-S. Song, *A globally convergent and consistent method for estimating the shape parameter of a Generalized Gaussian Distribution*, IEEE Transactions on Information Theory **52** (2) (2006) 510–527. doi:10.1109/TIT.2005.860423

[16] M.T. Subbotin, *On the law of frequency of errors*, Matematicheskii Sbornik **31** (1923) 296–301.

[17] P.R. Tadikamalla, *Random sampling from the exponential power distribution*, J. Amer. Statist. Association **75** (1980) 683–686. doi:10.1080/01621459.1980.10477533

[18] Van der Vaart, Asymptotic Statistics (Cambridge University Press, 1998).

[19] M.K. Varanasi and B. Aazhang, *Parametric generalized Gaussian density estimation*, J. Acoustical Society of America **86** (4) (1989) 1404–1415. doi:10.1121/1.398700

[20] D. Zhu and V. Zinde-Walsh, *Properties and estimation of asymmetric exponential power distribution*, J. Econometrics **148** (2009) 86–99. doi:10.1016/j.jeconom.2008.09.038

Received 7 October 2015
Revised 10 December 2015