Ann Inst Stat Math (2017) 69:461–487
DOI 10.1007/s10463-015-0548-3

A class of new tail index estimators

Vygantas Paulauskas[1] · Marijus Vaičiulis[2]

Received: 3 November 2014 / Revised: 6 October 2015 / Published online: 14 December 2015

Abstract In the paper, we propose a new class of functions which is used to construct tail index estimators. Functions from this new class are non-monotone in general, but they are the product of two monotone functions: the power function and the logarithmic function, which play essential role in the classical Hill estimator. The newly introduced generalized moment ratio estimator and generalized Hill estimator have a better asymptotic performance compared with the corresponding classical estimators over the whole range of the parameters that appear in the second-order regular variation condition. Asymptotic normality of the introduced estimators is proved, and comparison (using asymptotic mean square error) with other estimators of the tail index is provided. Some preliminary simulation results are presented.

Keywords Tail index estimation · Hill-type estimators · Heavy tails

1 Introduction

From the first papers of Hill and Pickands (see Hill 1975; Pickands 1975), devoted to the estimation of the tail index (or, more generally, the extreme value index), most

The research was supported by Research Council of Lithuania, Grant No. MIP-076/2013.

✉ Vygantas Paulauskas
vygantas.paulauskas@mif.vu.lt

Marijus Vaičiulis
marijus.vaiciulis@mii.vu.lt

1 Department of Mathematics and Informatics, Vilnius University, Naugarduko st. 24, 03225 Vilnius, Lithuania

2 Institute of Mathematics and Informatics, Vilnius University, Akademijos st. 4, 08663 Vilnius, Lithuania

of statistics constructed for this aim were based on order statistics and logarithmic function. Suppose we have a sample $X_1, X_2, \ldots, X_n$, considered as independent, identically distributed (i.i.d.) random variables with a distribution function (d.f.) F satisfying the following relation for large x:

$$\bar{F}(x) := 1 - F(x) = x^{-1/\gamma} L(x). \tag{1}$$

The parameter γ is usually called extreme value index (EVI) and $\alpha = 1/\gamma > 0$ is called the tail index, $L(x) > 0$, for all $x > 0$, and L is a slowly varying at infinity function:

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1.$$

In the paper we consider the case $\gamma > 0$ only. Denote $U(t) = F^{\leftarrow}(1 - (1/t)), t \geq 1$, where $W^{\leftarrow} : I \to \mathbb{R}$ is the left continuous inverse function of a monotone function W, defined by $W^{\leftarrow}(t) := \sup\{x : \ W(x) \leq t\}, t \in I$. It is well-known that in the case $\gamma > 0$ assumption (1) is equivalent to the following one: for all $x > 0$,

$$\lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^{\gamma}, \tag{2}$$

i.e., the quantile function $U(t)$ varies regularly with the index γ. Let $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$ denote the order statistics of $X_1, \ldots, X_n$. Taking some part of the largest values from the sample and the logarithmic function a statistician can form various statistics. In this way one can get Hill and Pickands estimators, moment and moment ratio estimators which are well-known and deeply investigated. The heuristic behind this approach (based on the peaks-over-threshold (POT) phenomenon and the maximum likelihood) is also given in many papers and monographs, therefore we do not provide it here.

There are estimators, based on a different idea: the sample is divided into blocks and in each block the ratio of two largest values is taken. Then the linear function $f(x) = x$ instead of logarithmic one is applied to these ratios. Estimators, based on this idea were constructed in Paulauskas (2003) and Paulauskas and Vaičiulis (2010). The next step was to include the linear and logarithmic functions into some parametric family of functions, and, considering estimators based on block scheme, this was done in Paulauskas and Vaičiulis (2011), taking the family of functions, defined for $x \geq 1$

$$f_r(x) = \begin{cases} \frac{1}{r}\left(x^r - 1\right), & r \neq 0, \\ \ln x, & r = 0. \end{cases} \tag{3}$$

In Paulauskas and Vaičiulis (2013) this family of functions was applied to order statistic. This was done by introducing the statistics

$$H_n^{(j)}(k, r) = \frac{1}{k} \sum_{i=0}^{k-1} f_r^j \left(\frac{X_{n-i,n}}{X_{n-k,n}} \right), \quad j = 1, 2,$$

and some combinations, formed from these statistics. Since in the paper we shall use mainly this statistic with $j = 1$, we use the abbreviation $H_n(k, r) := H_n^{(1)}(k, r)$. Here k is some quantity satisfying $1 \le k < n$, and in the EVI estimation k is chosen as a function of n [thus, strictly speaking, we should denote it by $k(n)$]. In this way the generalizations of the Hill, the moment, and the moment ratio estimators (we shall write the expressions of these estimators later in this section; to denote these estimators we shall use the letters H, M, MR and for generalizations we add the letter G) were obtained, for example, the generalized Hill estimator (GH) is defined as

$$\hat{\gamma}_n^{\mathrm{GH}}(k, r) = \frac{H_n(k, r)}{1 + r \cdot H_n(k, r)} \tag{4}$$

and the Hill estimator is obtained by taking $r = 0$, $\hat{\gamma}_n^H(k) := \hat{\gamma}_n^{\mathrm{GH}}(k, 0)$.

Another estimator of the parameter γ, which can be written as a function of the statistic $H_n(k, r)$, was very recently introduced in Beran et al. (2014) (see also Fabian and Stehlik 2009; Henry III 2009; Stehlik et al. 2010). It is named as the harmonic moment (HM) tail index estimator , and for $\beta > 0$, $\beta \ne 1$, it is defined by formula

$$\gamma_n^{\mathrm{HM}}(k, \beta) = \frac{1}{\beta - 1}\left(\left[k^{-1}\sum_{i=0}^{k-1}\left(\frac{X_{n-k,n}}{X_{n-i,n}}\right)^{\beta-1}\right]^{-1} - 1\right), \tag{5}$$

while for $\beta = 1$ it is defined as a limit as $\beta \to 1$. It is easy to see that by very simple transformation, denoting $\beta = 1 - r$, we have $\gamma_n^{\mathrm{HM}}(k, 1 - r) = \gamma_n^{\mathrm{GH}}(k, r)$, and this means, that the HM is exactly the GH. As it was pointed to us by one of the referees, at the same time as Paulauskas and Vaičiulis (2013), there appeared the paper Brilhante et al. (2013), where the same generalization of the Hill estimator was introduced and investigated. Namely, in Brilhante et al. (2013) the mean of order p (MOP) estimator was introduced, but it is easy to see that this estimator, defined in Brilhante et al. (2013) by formula (8), coincides with GH estimator defined in (4).

Thus, it is possible to say that in three papers Paulauskas and Vaičiulis (2013), Beran et al. (2014), Brilhante et al. (2013), written independently, the same generalization of the Hill estimator was introduced and investigated, and this generalization turned out to be quite successful, since in Paulauskas and Vaičiulis (2013) and Brilhante et al. (2013) it was shown that the GH estimator $\hat{\gamma}_n^{(\mathrm{GH})}(k, r)$ with an optimal r [in the sense of minimal asymptotic mean square error (AMSE)] dominates the Hill estimator in all region $\{\gamma > 0, \rho < 0\}$ of the parameters γ and ρ which are present in the second-order condition, see (12). Also, let us note that in Paulauskas and Vaičiulis (2013) we have considered statistics $H_n^{(j)}(k, r)$ with both $j = 1, 2$, and by means of these statistics the generalizations of the moment and the moment ratio estimators were obtained.

The main goal of this paper is to introduce another parametric family of functions, which has the same property that includes logarithmic function, and to construct new estimators using this family. For $x \ge 1$, let us consider functions

$$g_{r,u}(x) = x^r \ln^u(x),$$

where parameters r and u can be arbitrary real numbers, but for our purposes, connected with consistency, we shall require $\gamma r < 1$ and $u > -1$. Moreover, mainly we shall consider only integer values of parameter u. The family $\{g_{r,u}\}$, similarly to $\{f_r\}$, contains logarithmic function (with $r = 0$), but, contrary to $\{f_r\}$, contains logarithmic function for any value of parameter r (if $u \neq 0$). Also let us note that for $r \geq 0$ the function $g_{r,u}$ is monotone for all values of u, while for $r < 0$ and $u > 0$ there is no monotonicity.

Using these functions we can form statistics, similar to $H_n^{(j)}(k, r)$:

$$G_n(k, r, u) = \frac{1}{k} \sum_{i=0}^{k-1} g_{r,u} \left(\frac{X_{n-i,n}}{X_{n-k,n}} \right). \tag{6}$$

The above-mentioned Hill estimator, the M estimator [introduced in Dekkers et al. (1989)] and the MR estimator [introduced in Danielsson et al. (1996)] can be expressed via statistics $G_n(k, 0, u)$, $u = 1, 2$ as follows:

$$\begin{aligned}
\hat{\gamma}_n^H(k) &= G_n(k, 0, 1), \\
\hat{\gamma}_n^M(k) &= G_n(k, 0, 1) + \frac{1}{2} \left\{ 1 - \left(\frac{G_n(k, 0, 2)}{G_n^2(k, 0, 1)} - 1 \right)^{-1} \right\}, \\
\hat{\gamma}_n^{\mathrm{MR}}(k) &= \frac{G_n(k, 0, 2)}{2G_n(k, 0, 1)}.
\end{aligned}$$

Many estimators, considered earlier, can be expressed in terms of statistics (6), for example, in Gomes and Martins (2001) the following two estimators were considered:

$$\frac{G_n(k, 0, u)}{\Gamma(1 + u) G_n^{u-1}(k, 0, 1)} \quad \text{and} \quad \left(\frac{G_n(k, 0, u)}{\Gamma(1 + u)} \right)^{1/u}.$$

Let us note, that, due to the expressions of functions f_r and $g_{r,u}$, we can express the statistic $H_n(k, r)$ via the statistics $G_n(k, r, u)$:

$$H_n(k, r) = \begin{cases} \left(G_n(k, r, 0) - 1 \right) / r, & r \neq 0, \\ G_n(k, 0, 1), & r = 0, \end{cases} \tag{7}$$

and there is continuity with respect to r in this relation, since it is easy to see that $\lim_{r \to 0} \left(g_{r,0}(x) - 1 \right) / r = g_{0,1}(x)$. Taking into account that $H_n^{(2)}(k, r)$ can be expressed via $H_n(k, r)$ [see (3.2) in Paulauskas and Vaičiulis (2013)], all estimators, which were introduced in Paulauskas and Vaičiulis (2013), can be written by means of statistics $G_n(k, r, u)$ only.

In the paper we provide general method to prove limit theorems for estimators, constructed by means of statistics $G_n(k, r, u)$. We prove the weak consistency for these statistics for general values of parameters $r < 1/\gamma$ and $u > -1$. Then (see Theorem 3) we prove the asymptotic normality for the pair of statistics $(G_n(k, r, 0), G_n(k, r, 1))$. Taking the simplest choice of parameter u, $u = 0$ and $u = 1$, we made the first step, although the final goal would be to prove the asymptotic normality for general pair $(G_n(k, r_1, u_1), G_n(k, r_2, u_2))$.

The last remark in Sect. 1 concerns the class of the reduced-bias or Hill-corrected estimators (see Brilhante et al. 2013 and references therein). We do not consider these estimators in our paper, since they are not expressed by means of our functions $g_{r,u}$, they are based on a different idea and usually require additional assumptions, like the third-order asymptotic condition [the rate of convergence in the SORV condition (12)].

The rest of the paper is organized as follows. In the next section, we formulate the main results of the paper. In Sect. 3, we investigate asymptotic mean square error of the introduced estimators, and compare these estimators with the H and GH estimators, using the same methodology as in De Haan and Peng (1998), and provide some simulation results. Then there are formulated conclusions, and the last Sect. 5 is devoted to the proofs of the results. At the end of the proof of Theorem 3 we discuss the alternative proof of this result based on the paper Drees (1998).

2 Formulation of results

Our first result shows what quantities are estimated by the introduced statistics $G_n(k, r, u)$.

Theorem 1 *Suppose that $X_1, \ldots, X_n$ are i.i.d. nonnegative random variables whose quantile function U satisfies condition (2). Let $\gamma r < 1$ and $u > -1$. Let us suppose that a sequence $k = k(n)$ satisfies conditions*

$$k(n) \to \infty, \quad n/k(n) \to \infty, \quad \text{as } n \to \infty. \tag{8}$$

Then for statistics, introduced in (6), we have

$$G_n(k, r, u) \xrightarrow{\mathrm{P}} \xi(r, u) := \frac{\gamma^u \Gamma(1+u)}{(1-\gamma r)^{1+u}}, \quad \text{as } n \to \infty. \tag{9}$$

Here $\xrightarrow{\mathrm{P}}$ stands for the convergence in probability and $\Gamma(u)$ denotes the Euler's gamma function.

The following corollary allows to proof the consistency of an estimator, expressed as a function of statistics $G_n(k, r, u)$ with different r and u.

Corollary 1 *Suppose that $X_1, \ldots, X_n$ are i.i.d. nonnegative random variables whose quantile function U satisfies condition (2). Let $\gamma r_j < 1$ and $u_j > -1$, $j = 1, 2, \ldots, m$. Let us suppose that a sequence $k = k(n)$ satisfies (8). Let $\chi(t_1, \ldots, t_m) : (0, \infty)^m \to (0, \infty)$ be a continuous function. Then*

$$\chi\left(G_n(k, r_1, u_1), \ldots, G_n(k, r_m, u_m)\right) \xrightarrow{\mathrm{P}} \chi\left(\xi(r_1, u_1), \ldots, \xi(r_m, u_m)\right), \tag{10}$$

as $n \to \infty$.

The relation (10) gives us many possibilities to form consistent estimators of γ using statistics $G_n(k, r, u)$ with different r, u. Since it is not clear which combinations are good ones, we decided to restrict ourselves with the two most simple statistics $G_n(k, r, 0)$, $G_n(k, r, 1)$ (that is, $u = 0$ and $u = 1$) and to consider the following three estimators of $\gamma > 0$

$$\begin{aligned}
\hat{\gamma}_n^{(1)}(k, r) &= \begin{cases} (G_n(k, r, 0) - 1)(rG_n(k, r, 0))^{-1}, & r \neq 0, \\ G_n(k, 0, 1), & r = 0, \end{cases} \\
\hat{\gamma}_n^{(2)}(k, r) &= \frac{2G_n(k, r, 1)}{2rG_n(k, r, 1) + 1 + \sqrt{4rG_n(k, r, 1) + 1}}, \\
\hat{\gamma}_n^{(3)}(k, r) &= \begin{cases} (rG_n(k, r, 1) - G_n(k, r, 0) + 1)(r^2 G_n(k, r, 1))^{-1}, & r \neq 0, \\ \hat{\gamma}_n^{\mathrm{MR}}(k), & r = 0. \end{cases}
\end{aligned} \tag{11}$$

One can note, that the estimator $\hat{\gamma}_n^{(1)}(k, r)$ is exactly the GH estimator, given in (4), only expressed via statistics $G_n(k, r, u)$. For us it will be convenient to use this notation for GH estimator, since we shall compare these two new estimators from (11) with the H and GH estimators. Since $\hat{\gamma}_n^{(2)}(k, 0) = \hat{\gamma}_n^{(1)}(k, 0)$, the second estimator presents another generalization of the Hill estimator, while the third estimator gives us the generalized moment ratio estimator.

The main step in proving asymptotic normality of the introduced estimators $\hat{\gamma}_n^{(j)}(k, r)$, $j = 1, 2, 3$, (and other estimators, expressed via statistics $G_n(k, r, u)$ with $u = 0$ and $u = 1$) is to prove two-dimensional asymptotic normality for statistics $G_n(k, r, 0)$, $G_n(k, r, 1)$. As usual, in order to get asymptotic normality for estimators the so-called second-order regular variation (SORV) condition, in one or another form, is assumed. In this paper we shall use the SORV condition formulated by means of the function U. We assume that there exists a measurable function $A(t)$ with the constant sign near infinity, which is not identically zero, and $A(t) \to 0$ as $t \to \infty$, such that

$$\lim_{t \to \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \tag{12}$$

for all $x > 0$. Here $\rho < 0$ is the so-called second-order parameter. It is known that (12) implies that the function $|A(t)|$ varies regularly with index ρ.

Let us denote $d_r(b) = 1 - b\gamma r$.

Theorem 2 *Suppose that $X_1, \ldots, X_n$ are i.i.d. nonnegative random variables whose quantile function U satisfies condition (12). Suppose that $\gamma r < 1/2$ and that the sequence $k = k(n)$ satisfies (8) and*

$$\lim_{n \to \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \mu \in (-\infty, +\infty). \tag{13}$$

Then, as $n \to \infty$,

$$\begin{aligned}
&\sqrt{k}\,(G_n(k, r, 0) - \xi(r, 0), G_n(k, r, 1) - \xi(r, 1)) \\
&\qquad \xrightarrow{\mathrm{d}} \mu\left(\nu^{(1)}(r), \nu^{(2)}(r)\right) + \left(W^{(1)}, W^{(2)}\right),
\end{aligned} \tag{14}$$

where $\xrightarrow{\mathrm{d}}$ *stands for the convergence in distribution and quantities* $\nu^{(j)}(r),\ j=1,2$ *are as follows*

$$\nu^{(1)}(r)=\frac{r}{d_r(1)(d_r(1)-\rho)},\quad \nu^{(2)}(r)=\frac{1-\rho-\gamma^2r^2}{(d_r(1))^2(d_r(1)-\rho)^2}. \tag{15}$$

$\left(W^{(1)},W^{(2)}\right)$ *is zero mean Gaussian random vector with the variances*

$$\mathrm{E}\left(W^{(j)}\right)^2=s_j^2(r),\quad j=1,2$$

and the covariance $\mathrm{E}\left(W^{(1)}W^{(2)}\right)=s_{12}(r)$*, where*

$$\begin{aligned}
s_1^2(r)&=\frac{\gamma^2r^2}{d_r(2)d_r^2(1)},\\
s_2^2(r)&=\frac{\gamma^2(d_r(2)+2\gamma^4r^4)}{d_r^3(2)d_r^4(1)},\\
s_{12}(r)&=\frac{\gamma^2r(d_r(1)-\gamma^2r^2)}{d_r^2(2)d_r^3(1)}.
\end{aligned}$$

From Theorem 2 we derive the main result of the paper.

Theorem 3 *Under assumptions of Theorem 2, for the estimators, introduced in (11), we have*

$$\sqrt{k}\left(\hat{\gamma}_n^{(j)}(k,r)-\gamma\right)\xrightarrow{\mathrm{d}}\mathcal{N}\left(\mu\nu_j(r),\sigma_j^2(r)\right),\quad j=1,2,3, \tag{16}$$

where

$$\begin{aligned}
\nu_1(r)&=\frac{d_r(1)}{d_r(1)-\rho},\quad \sigma_1^2(r)=\frac{\gamma^2d_r^2(1)}{d_r(2)},\\
\nu_2(r)&=\frac{d_r(1)(1-\rho-\gamma^2r^2)}{(1+\gamma r)(d_r(1)-\rho)^2},\quad \sigma_2^2(r)=\frac{\gamma^2d_r^2(1)(d_r(2)+2\gamma^4r^4)}{(1+\gamma r)^2d_r^3(2)},\\
\nu_3(r)&=\frac{d_r^2(1)}{(d_r(1)-\rho)^2},\quad \sigma_3^2(r)=\frac{2\gamma^2d_r^4(1)}{d_r^3(2)}.
\end{aligned}$$

Having the asymptotic normality of the introduced estimators in Sect. 3 we compare the asymptotic mean square error (AMSE) of these estimators. As in Paulauskas and Vaičiulis (2013), where we have shown that the GH estimator (with the optimal value of r) dominates the Hill estimator in all region of the parameters $\{\gamma>0,\rho<0\}$, now the same dominance is demonstrated when comparing the GMR (again, with the optimal value of r) and MR estimators, more over, from Fig. 2, the left graph we can see that the GMR estimator outperforms MR estimator not only theoretically, but also empirically (simulation results are given as points on the same graph). In Fig. 3, the right graph presents comparison of the GMR estimator with the GH estimator, and asymptotic result (solid line) shows that no one estimator dominates in all region $-\infty < \rho < 0$ (the ratio of AMSE does not depend on γ), while simulation results (points) demonstrate the domination of GMR for all values of ρ, for which the simulation was performed.

Finally, we formulate some results concerning the robustness of the introduced estimators. We follow the paper Beran et al. (2014), where robustness was considered for the HM estimator (or in our notation, for $\hat{\gamma}_n^{(1)}(k,r)$). To define the robustness measure for the estimator $\hat{\gamma}_n^{(j)}(k,r)$, instead of $\hat{\gamma}_n^{(j)}(k,r)$ we will use the notation $\hat{\gamma}_n^{(j)}(k,r;X_1,\ldots,X_n)$. Let us define

$$\Delta\hat{\gamma}_n^{(j)}(k,r,x) = \hat{\gamma}_n^{(j)}(k,r;X_1,\ldots,X_{n-1},x) - \hat{\gamma}_{n-1}^{(j)}(k-1,r;X_1,\ldots,X_{n-1}).$$

Then, following Beran et al. (2014), for fixed n and k, we take the quantity

$$B_n^{(j)}(k,r) = \lim_{x\to\infty} \Delta\hat{\gamma}_n^{(j)}(k,r,x), \tag{17}$$

which measures the worst effect of one arbitrarily large contamination on the estimator $\hat{\gamma}_n^{(j)}(k,r)$. For fixed n and k these quantities are random variables, but it turns out that asymptotically they become constants, depending on γ and r (here it is appropriate to note, that results on robustness are based on Theorem 1, thus there is no dependence on ρ).

Theorem 4 *Suppose that $X_1,\ldots,X_n$ are i.i.d. nonnegative random variables whose quantile function U satisfies condition (2). Let $\gamma r < 1$ and let $B^{(j)} := B^{(j)}(\gamma,r)$ be the limit in probability of $B_n^{(j)}(k,r)$, $j = 1,2,3$, defined in (17), as $n \to \infty$ and (8) holds. Then we have*

$$B^{(j)} = \begin{cases} 0, & r < 0, \\ \infty, & r = 0, \\ (1-\gamma r)/r, & 0 < r < 1/\gamma. \end{cases} \qquad j = 1,2,3.$$

Assuming the SORV condition (12) we were able to find optimal values of r for $\hat{\gamma}_n^{(j)}(k,r)$, $j = 1,3$ [see formulas (24) and (25) in Sect. 3], therefore, for the generalized Hill estimator we get $B^{(1)}(\gamma,r_1^*) = \gamma(1-\rho+\sqrt{(2-\rho)^2-2})$. For the generalized moment ratio estimator the situation is even better, since optimal value $r_3^* < 0$, therefore $B^{(3)}(\gamma,r_3^*) = 0$, while $B^{(j)}(\gamma,0) = \infty$. At first it seemed for us a little bit strange, that for all three estimators with $r = 0$ we get the same infinite value of $B^{(j)}$, but looking more carefully to the construction of this measure of robustness, we realized that it is quite natural and even the proof is almost trivial. Since the second term in the expression of $\Delta\hat{\gamma}_n^{(j)}(k,r,x)$ is independent of x, moreover, if $n\to\infty$, it tends to γ, we have

$$B_n^{(j)}(k,r) = \lim_{x\to\infty} \hat{\gamma}_n^{(j)}(k,r;X_1,\ldots,X_{n-1},x) - \hat{\gamma}_{n-1}^{(j)}(k-1,r;X_1,\ldots,X_{n-1}). \tag{18}$$

Thus, robustness of the given estimator depends essentially on this first limit, which can be zero, infinity, or finite, depending on the term $g_{r,u}\left(x/X_{n-k,n}\right)$. For all classical estimators $\hat{\gamma}^H$, $\hat{\gamma}^M$, $\hat{\gamma}^{\mathrm{MR}}$ $(r = 0)$ this term tends to infinity, therefore we get infinite value for robustness, while for all three generalizations there appear ratios of such terms, and we are getting that this first limit in (18), as $x \to \infty$, is $1/r$, if $r > 0$ and is γ, if $r < 0$, thus we are getting result of Theorem 4. Moreover, the proof of this theorem shows that we can contaminate the sample not by one large value, but by several, and the asymptotic result will be the same—generalized estimators will remain (asymptotically) robust.

3 Theoretical comparison of the estimators and Monte Carlo simulations

In this section, we assume that the assumptions of Theorem 3 are satisfied, therefore the condition (12) holds with $\rho < 0$. We excluded the case $\rho = 0$, since in this case one faces principal difficulties in finding the optimal sequence $k(n)$, see p. 81–83 in De Haan and Ferreira (2006) for details.

For theoretical comparison of estimators under consideration, as a first step we find optimal values (in the sense of minimal AMSE) of both tuning parameters k and r. As in De Haan and Peng (1998) we can write the following relation for the asymptotic mean squared error of the estimator $\hat{\gamma}_n^{(j)}(k, r)$

$$\mathrm{AMSE}\left(\hat{\gamma}_n^{(j)}(k, r)\right) \sim \nu_j^2(r) A^2\left(\frac{n}{k}\right) + \frac{\sigma_j^2(r)}{k}, \quad j = 1, 2, 3, \tag{19}$$

where a sequence $k = k(n)$ satisfies (13). Here and below we write $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$. As in Paulauskas and Vaičiulis (2013), assuming that r is fixed, we perform the minimization of AMSE $\left(\hat{\gamma}_n^{(j)}(k, r)\right)$ with respect to k . We will obtain the optimal value of $k(n)$ for the estimator $\hat{\gamma}_n^{(j)}(k, r)$, which will be denoted by $k_j^*(r)$. Then we minimize the asymptotic mean squared error AMSE $\left(\hat{\gamma}_n^{(j)}\left(k_j^*(r), r\right)\right)$ with respect to r. Since the optimization procedure, based on De Haan and Peng (1998), is used in many papers dealing with EVI estimation, we do not provide all calculations and formulate only results of this optimization.

We define the function a by the following relation:

$$A^2(t) \sim \int_t^\infty a(u)\,\mathrm{d}u, \quad t \to \infty, \tag{20}$$

and, assuming that r is fixed, we get

$$k_j^*(r) = \left(\frac{\sigma_j^2(r)}{\nu_j^2(r)}\right)^{1/(1-2\rho)} \frac{n}{a^{\leftarrow}(1/n)}. \tag{21}$$

Having (21), from (19) we obtain

$$\mathrm{AMSE}\left(\hat{\gamma}_n^{(j)}\left(k_j^*(r), r\right)\right) \sim \frac{1-2\rho}{(-2\rho)}\left(\nu_j^2(r)\left(\sigma_j^2(r)\right)^{-2\rho}\right)^{1/(1-2\rho)} \frac{a^{\leftarrow}(1/n)}{n}. \tag{22}$$

It is easy to see that to minimize the right-hand side of (22) with respect to r it is sufficient to minimize the product $\nu_j^2(r)\left(\sigma_j^2(r)\right)^{-2\rho}$ with respect to r. Let us note that asymptotic parameters $\nu_j(r)$, $1 \le j \le 3$ depend on parameters ρ and γr, while quantities $\sigma_j^2(r)/\gamma^2$, $1 \le j \le 3$ depend on the product γr only. Therefore, it is convenient to introduce notation $R = \gamma r$ and to consider minimization of the function

$$\eta_j(R) = \gamma^{4\rho}\nu_j^2(R/\gamma)\left(\sigma_j^2(R/\gamma)\right)^{-2\rho},$$

with respect to R satisfying inequality $R < 1/2$. Equating the derivative of this function to zero, we get

$$\sigma_j^2(R/\gamma)\frac{\mathrm{d}\nu_j(R/\gamma)}{\mathrm{d}R} - \rho\nu_j(R/\gamma)\frac{\mathrm{d}\sigma_j^2(R/\gamma)}{\mathrm{d}R} = 0. \tag{23}$$

By substituting the values of $\nu_3(R/\gamma)$ and $\sigma_3^2(R/\gamma)$ into Eq. (23) we get the equation $R^2 - R(2-\rho)+\rho = 0$. Whence it follows that the optimal value of R for the estimator $\hat{\gamma}_n^{(3)}\left(k_3^*(r), r\right)$ is

$$R_3^* = \frac{(2-\rho) - \sqrt{(2-\rho)^2 - 4\rho}}{2}, \tag{24}$$

since the second root of the quadratic equation does not satisfy the relation $R < 1/2$. As for the estimator $\hat{\gamma}_n^{(1)}\left(k_1^*(r), r\right)$, the optimal value of the parameter r was found in Paulauskas and Vaičiulis (2013) (see also Brilhante et al. 2013), and in our notation [it is necessary to note, that SORV condition in Paulauskas and Vaičiulis (2013) was used with a different parametrization, see (1.3) therein] the optimal value of R is

$$R_1^* = \frac{(2-\rho) - \sqrt{(2-\rho)^2 - 2}}{2}. \tag{25}$$

Unfortunately, the situation with the estimator $\hat{\gamma}_n^{(2)}\left(k_2^*(r), r\right)$ is more complicated. By substituting the expressions of $\nu_2(R/\gamma)$ and $\sigma_2^2(R/\gamma)$ into (23) we get the equation

$$\begin{aligned}&2R^9 - 2R^8(1-\rho) - 2R^7(5-3\rho) + 2R^6(\rho^2 - 3\rho + 6) - 2\rho R^5(5-2\rho)\\ &-6R^4(1-\rho)^2 + R^3(8\rho^2 - 22\rho + 15) - 2R^2(5\rho^2 - 14\rho + 9)\\ &+4R(\rho^2 - 3\rho + 2) - (1-\rho) = 0.\end{aligned}$$

For a fixed given value of ρ this equation of the 9th order was solved with "Wolfram Mathematica 6.0". It turns out that depending on the parameter ρ it has 3 real roots and three pairs of conjugate roots or 5 real roots and two pairs of conjugate roots. All real roots were substituted into the function $\eta_2(R)$ and optimal value was found in this way. Numerical values of optimal value R_2^* as a function of ρ are provided in Fig. 1. Although we got explicit and very simple expressions for the optimal values R_1^* and R_3^*, we provide these two functions (as functions of ρ) in the same Fig. 1.

Let $r_j^* = R_j^*/\gamma$ denote the optimal value of the parameter r. From this picture we see that the first two functions R_i^*, $i = 1, 2$ has comparatively small range of values

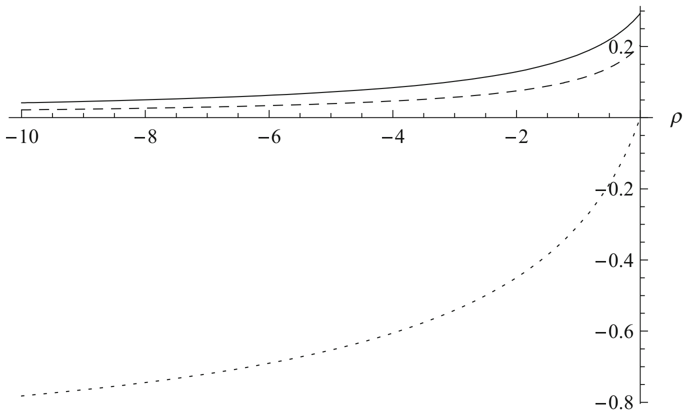

Fig. 1 Graph of the functions $R_1^*(\rho)$ (*solid line*), $R_2^*(\rho)$ (*dashed line*) and $R_3^*(\rho)$ (*dotted line*), $-10 < \rho < 0$

[for $R_3^*(\rho)$ the range is $(-1, 0)$], this means that optimal value of parameters r_1^* and r_2^* are not sensitive to the parameter ρ, but more sensitive to γ.

Now we are able to compare the generalized Hill estimator $\hat{\gamma}_n^{(1)}\left(k_1^*(r_1^*), r_1^*\right)$ with the another generalization of the Hill estimator $\hat{\gamma}_n^{(2)}\left(k_2^*(r_2^*), r_2^*\right)$ and generalized moment ratio estimator $\hat{\gamma}_n^{(3)}\left(k_3^*(r_3^*), r_3^*\right)$. But before performing comparison of these estimators, at first we demonstrate that GMR estimator $\hat{\gamma}_n^{(3)}\left(k_3^*(r_3^*), r_3^*\right)$ outperforms the initial MR estimator $\hat{\gamma}_n^{(3)}\left(k_3^*(0), 0\right)$ in the whole area $\{(\gamma, \rho) : \gamma > 0, \rho < 0\}$. Denoting

$$\psi_{\mathrm{MR}}(\rho) = \lim_{n\to\infty} \frac{\mathrm{AMSE}\left(\hat{\gamma}_n^{(3)}\left(k_3^*(0), 0\right)\right)}{\mathrm{AMSE}\left(\hat{\gamma}_n^{(3)}\left(k_3^*(r_3^*), r_3^*\right)\right)},$$

it is not difficult to get that

$$\psi_{\mathrm{MR}}(\rho) = \left(\frac{2^{-8\rho}\,(v(\rho)-\rho)^4\,(v(\rho)-1+\rho)^{-6\rho}}{(1-\rho)^4\,(v(\rho)+\rho)^{4-8\rho}}\right)^{1/(1-2\rho)},$$

where $v(\rho) = \left((2-\rho)^2 - 4\rho\right)^{1/2}$. Since we must investigate this function on negative half-line $\{\rho < 0\}$, it is convenient to denote $-\rho = x$ and to write

$$\tilde{\psi}_{\mathrm{MR}}(x) = \psi_{\mathrm{MR}}(-x) = (f(x))^{g(x)},$$

with

$$f(x) = \frac{2^{-8x}\,(\tilde{v}(x)+x)^4\,(\tilde{v}(x)-1-x)^{6x}}{(1+x)^4\,(\tilde{v}(x)-x)^{4+8x}}, \quad g(x) = \frac{1}{1+2x},$$

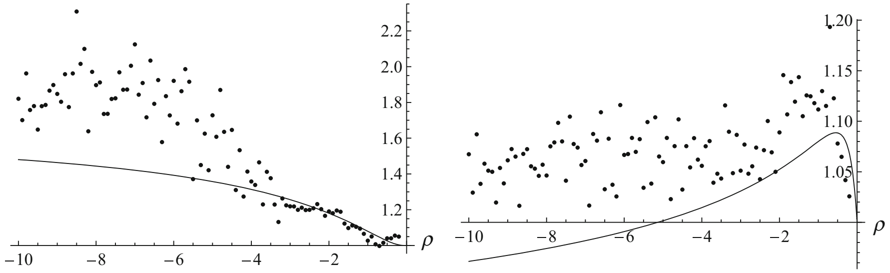

Fig. 2 Graph of the functions $\psi_{\mathrm{MR}}(\rho)$ (*solid line*, on the *left*), $\psi_H(\rho)$ (*solid line*, on the *right*) and results of Monte Carlo simulations (*points*) with the Burr distribution and $\gamma = 1$

where $\tilde{v}(x) = v(-x)$. Taking logarithm of $\tilde{\psi}_{\mathrm{MR}}(x)$, using the fact that f is the product of several elementary functions, and using the simple relation $v(x) - x = 4 + O(x^{-1})$, one can get

$$\lim_{x\to\infty} \ln(\tilde{\psi}_{\mathrm{MR}}(x)) = 3\ln 3 - 4\ln 2, \quad \text{or} \quad \lim_{x\to\infty} \tilde{\psi}_{\mathrm{MR}}(x) = \frac{27}{16} = 1.6875.$$

In a similar way one can get

$$\lim_{x\to 0} \ln(\tilde{\psi}_{\mathrm{MR}}(x)) = 0, \quad \text{or} \quad \lim_{x\to 0} \tilde{\psi}_{\mathrm{MR}}(x) = 1.$$

As a matter of fact, $\psi_{\mathrm{MR}}(0) = 1$, but considering the asymptotic normality and AMSE of the estimators under consideration we excluded the case $\rho = 0$, therefore we calculate this last limit. More difficult is to show that the function $\tilde{\psi}_{\mathrm{MR}}(x)$ is monotone and increasing (or $\psi_{\mathrm{MR}}(\rho)$ is decreasing), we skip these considerations, only we mention that we use the fact that the logarithmic derivative of a product of functions is a sum of logarithmic derivatives of these functions. The graph of the function $\psi_{\mathrm{MR}}(\rho)$, $\rho < 0$ is provided in Fig. 2 in left. In the same picture we gave also results (in form of separate points) of simulations, namely, we calculated the ratio of MSE of these estimators taking samples of size $n = 1000$ from Burr distribution (details on simulation will be explained below). Surprisingly, simulation results are even better than theoretical asymptotical result—most points are above the graph of $\psi_{\mathrm{MR}}(\rho)$.

As it was recommended by the referee, we also compare the GMR estimator with the Hill estimator. If we denote

$$\psi_H(\rho) = \lim_{n\to\infty} \frac{\mathrm{AMSE}\left(\hat{\gamma}_n^{(1)}\left(k_1^*(0), 0\right)\right)}{\mathrm{AMSE}\left(\hat{\gamma}_n^{(3)}\left(k_3^*(r_3^*), r_3^*\right)\right)},$$

then we have

$$\psi_H(\rho) = \left(\frac{(1-\rho-R_3^*)^4(1-2R_3^*)^{-6\rho}}{2^{-2\rho}(1-\rho)^2(1-R_3^*)^{4-8\rho}}\right)^{1/(1-2\rho)}.$$

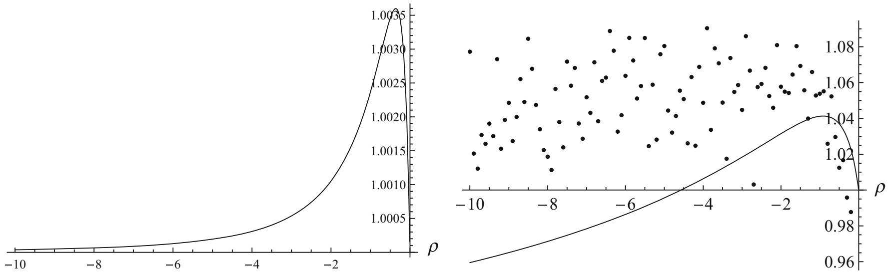

Fig. 3 Graph of the functions $\varphi_2(\rho)$ (on the *left*), $\varphi_3(\rho)$ (*solid line*, on the *right*) and results of Monte Carlo simulations (*points*) with the Burr distribution and $\gamma = 1$

The graph of the function $\psi_H(\rho)$, $\rho < 0$ is provided in Fig. 2 on the right, and, as on the left picture, there are simulation results, which are even more surprising, since the empirical results are not only above the theoretical curve, but they show that empirical MSE of the GMR estimator is smaller than MSE of the Hill estimator for all considered values of ρ.

Both graphs in Fig. 2 shows good performance of the GMR estimator, especially comparing with the MR estimator, since in a big range of the parameter ρ the function ψ_{MR} is bigger than 1.3, and the maximal value is close to 1.7.

Now we return to the comparison of the estimators $\hat{\gamma}_n^{(2)}\left(k_2^*(r_2^*), r_2^*\right)$ and $\hat{\gamma}_n^{(3)}\left(k_3^*(r_3^*), r_3^*\right)$ with the GH estimator $\hat{\gamma}_n^{(1)}\left(k_1^*(r_1^*), r_1^*\right)$, and we must investigate the following two functions:

$$\varphi_j(\rho) := \lim_{n\to\infty} \frac{\mathrm{AMSE}\left(\hat{\gamma}_n^{(1)}\left(k_1^*(r_1^*), r_1^*\right)\right)}{\mathrm{AMSE}\left(\hat{\gamma}_n^{(j)}\left(k_j^*(r_j^*), r_j^*\right)\right)}, \quad j = 2, 3.$$

It is important to note that both functions are independent of γ and depend only on ρ. In view of (22) we have

$$\varphi_j(\rho) = \left(\frac{\nu_1^2\left(r_1^*\right)\left(\sigma_1^2\left(r_1^*\right)\right)^{-2\rho}}{\nu_j^2\left(r_j^*\right)\left(\sigma_j^2\left(r_j^*\right)\right)^{-2\rho}}\right)^{1/(1-2\rho)}, \quad j = 2, 3.$$

Since we were able to obtain the optimal value of R_2^* only numerically, we can provide only a numerically obtained graph of the function $\varphi_2(\rho)$, see Fig. 3 on the left; therefore, in this case we did not provide simulation results.

Although the graph of $\varphi_2(\rho)$ allows to believe that the new generalization of the Hill estimator dominates the GH estimator [which is the same as HM estimator from Beran et al. (2014) or MOP estimator from Brilhante et al. (2013)] in all region of parameters $\{(\gamma, \rho) :\ \gamma > 0,\ \rho < 0\}$, but without explicit expression of the function $\varphi_2(\rho)$ we cannot prove this.

Finally, comparing GH and GMR estimators (these two estimators, in our opinion, are the most successful, since for both of them we have quite simple expression for optimal value of R_j^*, $j = 1, 3$) we have

$$\varphi_3(\rho) = \left(\frac{3^{-6\rho}\,(v(\rho)-\rho)^{8-8\rho}\,(w(\rho)+(1-\rho))^{-2\rho}}{4^{3-5\rho}(1-2\rho)^2\,(v(\rho)+(1-\rho))^{-6\rho}\,(w(\rho)-\rho)^{4-4\rho}} \right)^{1/(1-2\rho)},$$

where $w(\rho) = \left((2-\rho)^2 - 2\right)^{1/2}$. As can be seen from Fig. 3 on right, GMR estimator $\hat{\gamma}_n^{(3)}\left(k_3^*(r_3^*), r_3^*\right)$ dominates the GH estimator $\hat{\gamma}_n^{(1)}\left(k_1^*(r_1^*), r_1^*\right)$ for $\rho \in (\tilde{\rho}, 0)$, where $\tilde{\rho} \approx -4.57018$. It is not difficult to show that

$$\lim_{\rho \to -\infty} \varphi_3(\rho) = \frac{3^3}{2^5} = 0.84375.$$

The empirical results, as in the case of function $\psi_H(x)$ in Fig. 2 (on the right), show the same picture—almost all empirical points in the figure are not only above the graph of $\varphi_3(x)$, but they are bigger than 1. This means that the empirical MSE of GMR estimator is smaller than the empirical MSE of the GH estimator for all values of ρ in the interval $-10 < \rho < 0$, not only for $\rho \in (\tilde{\rho}, 0)$, as gives asymptotic theoretical result. The similarity of the graphs on the right of both Figs. 2 and 3 is not coincidental, since in Paulauskas and Vaičiulis (2013) we had compared GH and H estimators (see Fig. 2 therein) and saw that the improvement of the GH estimator over the Hill estimator is important only theoretically, since the maximal value of the ratio is only 1.05.

Now we shall provide some results of Monte Carlo simulations (part of these results are given in Figs. 2 and 3 together with theoretical results). We must admit that these results are very preliminary, since we had based our simulation only on two families of heavy-tailed distributions. Also it is necessary to investigate the stability of estimators under consideration with respect to r. Figure 1 shows that stability for different estimators is different. We intend to return to this problem in a separate paper.

For simulations we use a slightly more restrictive condition than (12), namely, we assume that the distribution function $F(x)$ under consideration belongs to the Hall's class of Pareto type distributions (Hall 1982; Hall and Welsh 1985), i.e.,

$$1 - F(x) = \left(\frac{x}{C}\right)^{-1/\gamma} \left(1 + \frac{\beta}{\rho}\left(\frac{x}{C}\right)^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right), \quad x \to \infty, \tag{26}$$

where $C > 0$, $\beta \in \mathbf{R}\backslash\{0\}$ and $\rho < 0$. This assumption is assumed in many papers dealing with simulations for the following reason. Taking the ratio of AMSE of two estimators we do not need to know the function $a^{\leftarrow}$, but for simulations, having given sample size n, we must calculate the value of $k_j^*(r)$ in (21) and the empirical MSE of estimators, and for this we must have the function $a^{\leftarrow}$. Assuming (26), we have that the second-order condition (12) holds with $A(t) = \gamma\beta t^{\rho}$ and from (20) it follows that $a^{\leftarrow}(t) = \left(-2\rho\gamma^2\beta^2\right)^{1/(1-2\rho)} t^{1/(2\rho-1)}$. Now we can rewrite (21) as follows:

$$k_j^*(r, \beta, \rho) = \left(\frac{\sigma_j^2(r)}{-2\rho\beta^2\gamma^2\nu_j^2(r)} \right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}.$$

In fact, quantities $k_j^*(0, \beta, \rho)$, $j = 1, 3$ depend on β, ρ and n only, thus replacing β and ρ by some estimators $\hat{\beta}_n$ and $\hat{\rho}_n$, we obtain the empirical values of the parameter $k(n)$ for the Hill and the moment ratio estimators:

$$\hat{k}_{1,n} = \left(\frac{(1-\hat{\rho}_n)^2}{-2\hat{\rho}_n \hat{\beta}_n^2} \right)^{1/(1-2\hat{\rho}_n)} n^{-2\hat{\rho}_n/(1-2\hat{\rho}_n)}, \tag{27}$$

$$\hat{k}_{3,n} = \left(\frac{(1-\hat{\rho}_n)^4}{-\hat{\rho}_n \hat{\beta}_n^2} \right)^{1/(1-2\hat{\rho}_n)} n^{-2\hat{\rho}_n/(1-2\hat{\rho}_n)}. \tag{28}$$

For corresponding generalized estimators we have additionally to take estimators of optimal parameter R, therefore we have

$$\hat{K}_{1,n} = \left(\frac{(1-\hat{\rho}_n - R_1^*(\hat{\rho}_n))^2}{-2\hat{\rho}_n \hat{\beta}_n^2 (1-2R_1^*(\hat{\rho}_n))} \right)^{1/(1-2\hat{\rho}_n)} n^{-2\hat{\rho}_n/(1-2\hat{\rho}_n)}, \tag{29}$$

$$\hat{K}_{3,n} = \left(\frac{(1-\hat{\rho}_n - R_3^*(\hat{\rho}_n))^4}{-\hat{\rho}_n \hat{\beta}_n^2 (1-2R_3^*(\hat{\rho}_n))^3} \right)^{1/(1-2\hat{\rho}_n)} n^{-2\hat{\rho}_n/(1-2\hat{\rho}_n)}. \tag{30}$$

Thus, in our simulations the comparison is made between the Hill estimator $\hat{\gamma}_n^{(1)}\left(\hat{k}_{1,n}, 0\right)$, the GH estimator $\hat{\gamma}_n^{(1)}\left(\hat{K}_{1,n}, r_1^*\left(\hat{\gamma}_n^{(1)}\left(\hat{k}_{1,n}, 0\right), \hat{\rho}_n\right)\right)$, the MR estimator $\hat{\gamma}_n^{(3)}\left(\hat{k}_{3,n}, 0\right)$ and the GMR estimator

$$\hat{\gamma}_n^{(3)}\left(\hat{K}_{3,n}, r_3^*\left(\hat{\gamma}_n^{(3)}\left(\hat{k}_{3,n}, 0\right), \hat{\rho}_n\right)\right),$$

with parameters given in (27)–(30).

We generated 1000 times samples $X_1, \ldots, X_n$ of i.i.d. random variables of size $n = 1000$ with the following two d. f. with the extreme value index γ, parameter ρ and satisfying (26):

(i) the Burr d.f. $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$;
(ii) the Kumaraswamy generalized exponential d.f.

$$F(x) = 1 - \left(1 - \exp\left\{-x^{\rho/\gamma}\right\}\right)^{-1/\rho}, \quad x \geq 0.$$

The parameter β which is present in (26) for the Burr distribution is 1 and for the Kumaraswamy distribution $-1/2$, and $C = 1$ for both distributions. To calculate the H and the GH estimators we used the following algorithm:

1. Estimate the parameter ρ by the following estimator proposed in Fraga Alves et al. (2009):

$$\hat{\rho}_n(k, \tau) = -\left| 3\left(T_n^{(\tau)}(k) - 1\right)\left(T_n^{(\tau)}(k) - 3\right)^{-1} \right|,$$

where

$$T_n^{(\tau)}(k) = \frac{(G_n(k,0,1))^\tau - (G_n(k,0,2)/2)^{\tau/2}}{(G_n(k,0,2)/2)^{\tau/2} - (G_n(k,0,3)/6)^{\tau/3}}$$

with $\tau > 0$, and

$$T_n^{(0)}(k) = \frac{\ln\left(G_n(k,0,1)\right) - (1/2)\ln\left(G_n(k,0,2)/2\right)}{(1/2)\ln\left(G_n(k,0,2)/2\right) - (1/3)\ln\left(G_n(k,0,3)/6\right)}.$$

To decide which values of parameters τ (0 or 1) and k to take in the above written estimator $\hat{\rho}_n(k,\tau)$, we realized the algorithm provided in Gomes et al. (2009).

2. To estimate the parameter β use the estimator $\hat{\beta}_n\left(k, \hat{\rho}_n(k,\tau)\right)$, where

$$\hat{\beta}_n(k,\rho) = \frac{\left(\frac{k}{n}\right)^{\rho}\left\{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum_{i=1}^{k}W_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}W_i\right)\right\}}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}W_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\rho}W_i\right)}$$

and $W_i = i\ln\left(X_{n+i-1,n}/X_{n+i,n}\right)$, $1 \le i \le k < n$. This estimator was introduced in Gomes and Martins (2002). Again, as in the step 1, to choose the parameter k we applied the algorithm from Gomes et al. (2009).

3. By using (27) estimate parameter k for the Hill estimator and then obtain $\hat{\gamma}_n^{(1)}\left(\hat{k}_{1,n}, 0\right)$;
4. Estimate $R_1^*\left(\hat{\rho}_n\right)$ and $r_1^*\left(\hat{\gamma}_n^{(1)}\left(\hat{k}_{1,n}, 0\right), \hat{\rho}_n\right)$;
5. By using (29) estimate the parameter k for the GH estimator and find the estimate $\hat{\gamma}_n^{(1)}\left(\hat{K}_{1,n}, r_1^*\left(\hat{\gamma}_n^{(1)}\left(\hat{k}_{1,n}, 0\right), \hat{\rho}_n\right)\right)$.

We used the similar algorithm (with obvious changes) for the MR estimator $\hat{\gamma}_n^{(3)}\left(\hat{k}_{3,n}, 0\right)$ and the GMR estimator $\hat{\gamma}_n^{(3)}\left(\hat{K}_{3,n}, r_3^*\left(\hat{\gamma}_n^{(3)}\left(\hat{k}_{3,n}, 0\right), \hat{\rho}_n\right)\right)$. Having values of the estimators we calculate MSE and bias of these estimators and these results of simulations are summarized in Fig. 4 (for the Burr distribution) and in Fig. 5 (for the Kumaraswamy distribution). Also these calculated MSE were used in Figs. 2 and 3, providing empirical results for comparison of some pairs of estimators, for example in Fig. 2, on the left, points are obtained calculating the ratio of MSE of MR and GMR estimators.

For the Burr distribution we took parameters γ and ρ in the intervals $(0, 4]$ and $(-5, -0.2]$, respectively. In Fig. 4 (left) we divided this rectangle $(0, 4] \times (-5, -0.2]$ into squares $\Delta_{i,j} = (i/10, (i+1)/10] \times (-(j+1)/10, -j/10]$, $i = 0, 1, \ldots, 40$, $j = 2, 3, \ldots, 50$. By taking true values of parameters γ and ρ as coordinates of the center of the rectangle $\Delta_{i,j}$, we performed Monte Carlo simulations. We colored the square $\Delta_{i,j}$ in black if empirical MSE of the GH estimator $\hat{\gamma}_n^{(1)}\left(\hat{K}_{1,n}, r_1^*\left(\hat{\gamma}_n^{(1)}\left(\hat{k}_{1,n}, 0\right), \hat{\rho}_n\right)\right)$ is the smallest among all four estimators under consideration, while areas of domination of the estimators $\hat{\gamma}_n^{(1)}\left(\hat{k}_{1,n}, 0\right)$, (H) $\hat{\gamma}_n^{(3)}\left(\hat{k}_{3,n}, 0\right)$ (MR), and $\hat{\gamma}_n^{(3)}\left(\hat{K}_{3,n}, r_3^*\left(\hat{\gamma}_n^{(3)}\left(\hat{k}_{3,n}, 0\right), \hat{\rho}_n\right)\right)$ (GMR) are in dark grey, grey, and in white, respectively. In Fig. 4(right) there are given the areas of domination with respect to absolute value of the bias (domination means that absolute value of the bias is the smallest), using the same colors as in the left picture.

The results of the simulations based on Kumaraswamy distribution (areas of domination of MSE and bias for this distribution) are performed in the same way and are presented in Fig. 5.

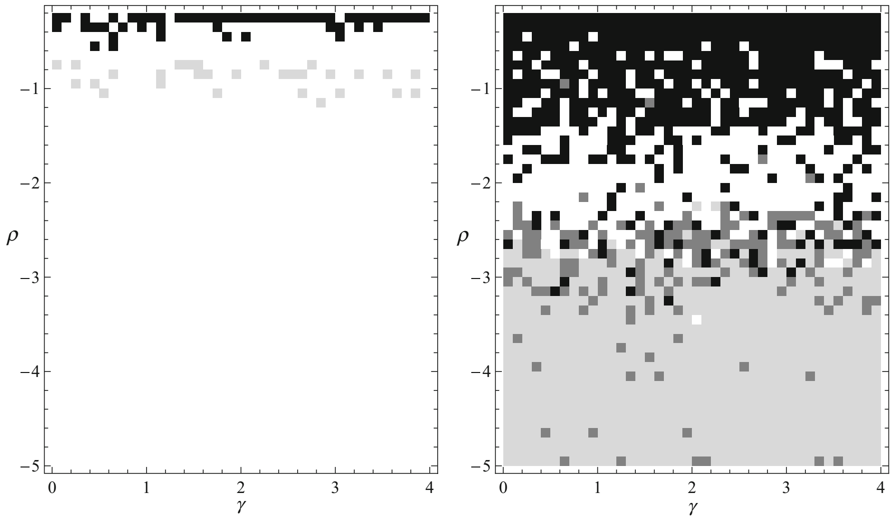

Fig. 4 Empirical comparison of the estimators by using Burr distribution; MSE on the *left*, BIAS on the *right*; *black color*—if at this point GH is the smallest, *dark grey*, *grey*, and *white*—if *H*, MR, and GMR, respectively, are the smallest

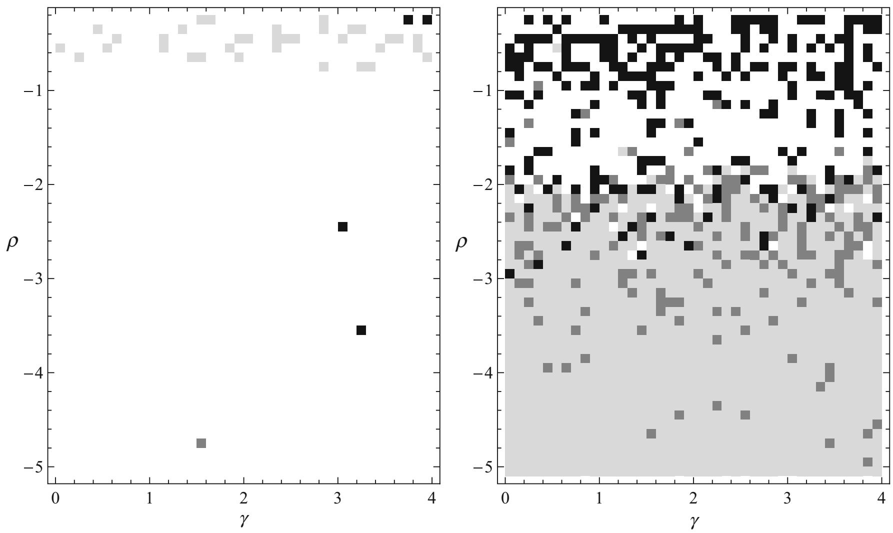

Fig. 5 Empirical comparison of the estimators by using Kumaraswamy distribution; MSE on the *left*, BIAS on the *right*; meaning of *colors* is the same as in Fig. 4

Figures 4 and 5 demonstrate that areas of domination almost do not depend on parameter γ and essentially depend only on ρ. This corresponds well to theoretical results which state that functions $\varphi_j(\rho)$, $j = 2, 3$, $\psi_H(\rho)$, and $\psi_{\mathrm{MR}}(\rho)$ depend on ρ only. Therefore, taking the particular value $\gamma = 1$ and the hundred of values of ρ in the interval $(-10, 0)$ for the Burr distribution we obtained ratios of empirical

MSEs to complement theoretical comparison and included these ratios (as separate points) in Figs. 2 and 3. Slightly unexpected, Figs. 2 and 3 reveal that empirical results differ from theoretical ones. In Fig. 2 (left), where GMR estimator is compared with the MR estimator, in the interval $-4 \le \rho < 0$ empirical points are very close to the theoretical function $\psi_{\mathrm{MR}}(\rho)$, but in the interval $-10 < \rho < -4$ all points are above the theoretical curve, this means that the MR estimator has MSE almost two times bigger that MSE of the GMR estimator. Empirical results in Fig. 3 (right) show that the GMR estimator performs better than GH for all values of ρ in the interval $-10 < \rho < 0$, while theoretical result predict such result only for $\rho \in (\tilde{\rho}, 0)$.

4 Conclusions

We introduced a new parametric class of functions $g_{r,u}$ which allows to construct many new generalizations of the well-known estimators, including such as the Hill, the M, and the MR estimators. We proved the asymptotic normality of all these generalized estimators in a unified way and demonstrate that in the sense of AMSE new estimators are better than the classical ones, especially promising looks GMR estimator. Also we hope that this new parametric class of functions will be useful in the difficult problem of estimating the second-order parameter ρ.

Preliminary simulation results show quite good correspondence with the obtained theoretical results, but we admit that future work on the construction of new estimators by means of statistics $G_n(k, r, u)$ and on studying the behavior of the new estimators for middle size samples is needed.

5 Proofs

Proof of Theorem 1 There is nothing to prove in the trivial case $r = u = 0$. Keeping in mind relation (7), conclusion (9) is the immediate consequence of Theorem 1 in Paulauskas and Vaičiulis (2013) in the case $u = 0$ and $\gamma r < 1$. The case $r = 0,\ u \ge 1$ was investigated in Gomes and Martins (2001).

Consider the case $r \ne 0$ and $u > -1$, $u \ne 0$. Let us recall that the function $U(t),\ t \ge 1$ varies regularly at infinity with the index γ, thus, by applying Potter's bound, we have: for arbitrary $\epsilon > 0$ there exits t_0, such that, for $x \ge 1$ and $t \ge t_0$,

$$((1-\epsilon)x)^{\gamma-\epsilon} < \frac{U(tx)}{U(t)} < ((1+\epsilon)x)^{\gamma+\epsilon}. \tag{31}$$

In order to apply (31) for the function t^r it is convenient to introduce the notation $\epsilon_{\pm}(r)$, where $\epsilon_{\pm}(r) = \epsilon$, if $r > 0$, and $\epsilon_{\pm}(r) = -\epsilon$, if $r < 0$. Then we get

$$\{(1-\epsilon_{\pm}(r))x\}^{r(\gamma-\epsilon_{\pm}(r))} < \left(\frac{U(tx)}{U(t)}\right)^r < \{(1+\epsilon_{\pm}(r))x\}^{r(\gamma+\epsilon_{\pm}(r))}. \tag{32}$$

Similarly we get the following inequalities

$$(\gamma-\epsilon_{\pm}(u))^u \ln^u((1-\epsilon_{\pm}(u))x) < \ln^u\left(\frac{U(tx)}{U(t)}\right) \tag{33}$$

$$\ln^u\left(\frac{U(tx)}{U(t)}\right) < (\gamma+\epsilon_\pm(u))^u \ln^u\left((1+\epsilon_\pm(u))x\right). \tag{34}$$

By multiplying inequalities (32) and (33), (34), we obtain

$$c_1 x^{r(\gamma-\epsilon_\pm(r))}\ln^u\left((1-\epsilon_\pm(u))x\right) < g_{r,u}\left(\frac{U(tx)}{U(t)}\right)$$
$$g_{r,u}\left(\frac{U(tx)}{U(t)}\right) < c_2 x^{r(\gamma+\epsilon_\pm(r))}\ln^u\left((1+\epsilon_\pm(u))x\right),$$

where

$$c_1 = (\gamma-\epsilon_\pm(u))^u(1-\epsilon_\pm(r))^{r(\gamma-\epsilon_\pm(r))},$$
$$c_2 = (\gamma+\epsilon_\pm(u))^u(1+\epsilon_\pm(r))^{r(\gamma+\epsilon_\pm(r))}.$$

Let $Y_1,\ldots,Y_n$ be i.i.d. random variables with distribution function $G(x)=1-(1/x)$, $x\geq 1$. Taking

$$t = Y_{n-k,n}, \quad x = Y_{n-i,n}/Y_{n-k,n}, \tag{35}$$

for $i=0,1,\ldots,k-1$, we get

$$c_1\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{r(\gamma-\epsilon_\pm(r))}\ln^u\left((1-\epsilon_\pm(u))\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) < g_{r,u}\left(\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})}\right), \tag{36}$$
$$c_2\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{r(\gamma+\epsilon_\pm(r))}\ln^u\left((1+\epsilon_\pm(u))\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) > g_{r,u}\left(\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})}\right). \tag{37}$$

Note that $U(Y_i)=X_i$, $i=1,2,\ldots,n$, thus

$$G_n(k,r,u) = \frac{1}{k}\sum_{i=0}^{k-1} g_{r,u}\left(\frac{U(Y_{n-i,n})}{U(Y_{n-k,n})}\right).$$

From this equality, by summing inequalities (36) and (37), we get

$$\frac{c_1}{k}\sum_{i=0}^{k-1}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{r(\gamma-\epsilon_\pm(r))}\ln^u\left((1-\epsilon_\pm(u))\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) < G_n(k,r,u), \tag{38}$$
$$\frac{c_2}{k}\sum_{i=0}^{k-1}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{r(\gamma+\epsilon_\pm(r))}\ln^u\left((1+\epsilon_\pm(u))\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) > G_n(k,r,u). \tag{39}$$

By means of the standard argument (see e.g. De Haan and Ferreira 2006) one can deduce that the left-hand side of (38) equals (in distribution) to the sum

$$\frac{c_1}{k}\sum_{i=1}^{k} Y_i^{r(\gamma-\epsilon_\pm(r))}\left(\ln(Y_i)+\ln(1-\epsilon_\pm(u))\right)^u.$$

The expectation of this quantity equals to $c_1\,(\Delta_1 + \Delta_2)$, where

$$\Delta_1 = \int_1^\infty x^{r(\gamma-\epsilon_\pm(r))-2}\ln^u(x)\mathrm{d}x,$$
$$\Delta_2 = \int_1^\infty x^{r(\gamma-\epsilon_\pm(r))-2}\ln^u(x)\left\{\left(1+\frac{\ln(1-\epsilon_\pm(u))}{\ln(x)}\right)^u - 1\right\}\mathrm{d}x.$$

One can verify that

$$\int_1^\infty x^{a-2}\ln^b(x)\mathrm{d}x = (1-a)^{-1-b}\Gamma(1+b), \quad a<1,\ b>-1. \tag{40}$$

By using the last identity, assumptions $1-\gamma r > 0$, $u > -1$ and the fact that $r\epsilon_\pm > 0$ we get

$$c_1\Delta_1 = \frac{(1-\epsilon_\pm(r))^{r(\gamma-\epsilon_\pm(r))}(\gamma-\epsilon_\pm(u))^u\Gamma(u+1)}{(1-r(\gamma-\epsilon_\pm(r)))^{u+1}},$$

whence we get $c_1\Delta_1 \to \xi(r,u)$, as $\epsilon \to 0$.

Consider the quantity Δ_2 now. If $0 < u \le 1$, then $\epsilon_\pm(u) = \epsilon$ and we use the following inequality which holds for any real numbers a and b: $||a|^u - |b|^u| \le |a-b|^u$. We have

$$|\Delta_2| \le |\ln(1-\epsilon)|^u \int_1^\infty x^{r(\gamma-\epsilon_\pm(r))-2}\mathrm{d}x$$
$$= \frac{|\ln(1-\epsilon)|^u}{1-r(\gamma-\epsilon_\pm(r))}.$$

Now it follows that $c_1\Delta_2 \to 0$, as $\epsilon \to 0$. In the case $-1 < u < 0$ we have $\epsilon_\pm(u) = -\epsilon$, and since $0 < -u < 1$, applying the same inequality as above, we get

$$\left|(\ln(x)+\ln(1+\epsilon))^u - \ln^u(x)\right| = \frac{\left|\ln(x)^{-u} - (\ln(x)+\ln(1+\epsilon))^{-u}\right|}{\ln(x)^{-u}\,(\ln(x)+\ln(1+\epsilon))^{-u}}$$
$$\le \frac{|\ln(1+\epsilon)|^{-u}}{\ln(x)^{-u}\,(\ln(x)+\ln(1+\epsilon))^{-u}}.$$

We take small $\delta > 0$ such that $u-\delta > -1$, for example, one can take $\delta = (u+1)/2$. Keeping in mind that $\ln(1+\epsilon) > 0$, we can estimate

$$\frac{1}{(\ln(x)+\ln(1+\epsilon))^{-u}} \le \frac{1}{(\ln(x))^\delta\,(\ln(1+\epsilon))^{-u-\delta}}.$$

Collecting the last two estimates we get

$$|\Delta_2| \le |\ln(1+\epsilon)|^\delta \int_1^\infty x^{r(\gamma-\epsilon_\pm(r))-2}\ln^{u-\delta}(x)\mathrm{d}x$$
$$= \frac{(\ln(1+\epsilon))^\delta\,\Gamma(1+u-\delta)}{(1-r(\gamma-\epsilon_\pm(r)))^{1+u-\delta}}.$$

This allows to deduce that $c_1\Delta_2 \to 0$, as $\epsilon \to 0$.

Let $u > 1$. By using the inequality $||a|^u - |b|^u| \le u\left(|a|^{u-1} - |b|^{u-1}\right)|a-b|$, which holds for any real numbers a and b, we get

$$\begin{aligned}|\Delta_2| &\le 2u\,|\ln(1-\epsilon_\pm(u))| \int_1^\infty x^{r(\gamma-\epsilon_\pm(r))-2}\ln^{u-1}(x)\mathrm{d}x\\ &= \frac{2u\Gamma(u)\,|\ln(1-\epsilon_\pm(u))|}{(1-r(\gamma-\epsilon_\pm(r)))^u},\end{aligned}$$

and this implies $c_1\Delta_2 \to 0$, as $\epsilon \to 0$. Thus, by applying the Khintchine weak law of large numbers, the left-hand side of (38) converges to zero in probability. In a similar way we can prove that the left-hand side of (39) tends to zero in probability, too. Theorem 1 is proved. □

Proof of Corollary 1 From (9) it follows

$$\left(G_n(k,r_1,u_1),\ldots,G_n(k,r_j,u_j)\right) \xrightarrow{\mathrm{P}} \left(\xi(r_1,u_1),\ldots,\xi(r_j,u_j)\right),$$

as $n \to \infty$. Applying Corollary 2 of Theorem 5.1 in Billingsley (1968) we derive (10). □

Proof of Theorem 2 Let $r = 0$. Theorem 3.2.5 in De Haan and Ferreira (2006) states

$$\sqrt{k}\,(G_n(k,0,1)-\gamma) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\frac{\mu}{1-\rho},\gamma^2\right), \quad n\to\infty. \tag{41}$$

The relation (41), together with $\sqrt{k}\,(G_n(k,0,0)-1) \xrightarrow{\mathrm{P}} 0$ and Theorem 4.4 in Billingsley (1968), give (14) for $r = 0$.

Consider now the case $r \ne 0$. Adjusting Potter's type bounds (3.4) in Paulauskas and Vaičiulis (2013) for our purposes (such adjustment is needed since the second-order condition (12) and the corresponding condition in Paulauskas and Vaičiulis (2013) are slightly different), we get that, for possibly different function $A_0(t)$ with $A_0(t) \sim A(t)$, as $t \to \infty$, and for each $\epsilon > 0$, $\delta > 0$, there exists t_0 such that for $t > t_0$, $x \ge 1$,

$$\left|g_{r,0}\left(\frac{U(tx)}{U(t)}\right) - x^{\gamma r} - r x^{\gamma r} A_0(t) f_\rho(x)\right| \le \epsilon r x^{\gamma r+\rho+\delta}\,|A_0(t)|\,, \tag{42}$$

where $f_\rho(x)$ is defined in (3). It is well-known that similar Potter's type bounds hold for the logarithmic function, see, e.g., inequalities (3.2.7) in De Haan and Ferreira (2006). Namely, for a possibly different function $A_1(t)$ with $A_1(t) \sim A(t)$, as $t \to \infty$, and for each $\epsilon > 0$, $\delta > 0$, there exists t_1 such that for $t > t_1$, $x \ge 1$,

$$\left|\ln\left(\frac{U(tx)}{U(t)}\right) - \gamma\ln(x) - A_1(t) f_\rho(x)\right| \le \epsilon x^{\rho+\delta}\,|A_1(t)|\,. \tag{43}$$

Let $\tilde{t} = \max\{t_0, t_1\}$. By multiplying inequalities (42) and (43) we get that for $t > \tilde{t}$, $x \ge 1$,

$$\Bigg| g_{r,1}\left(\frac{U(tx)}{U(t)}\right) - g_{r,0}\left(\frac{U(tx)}{U(t)}\right)\left\{\gamma \ln(x) + A_1(t) f_\rho(x)\right\}$$
$$- \ln\left(\frac{U(tx)}{U(t)}\right) x^{\gamma r}\left\{1 + rA_0(t) f_\rho(x)\right\}$$
$$+ x^{\gamma r}\left\{1 + rA_0(t) f_\rho(x)\right\}\left\{\gamma \ln(x) + A_1(t) f_\rho(x)\right\}\Bigg|$$
$$\le \epsilon^2 r x^{\gamma r + 2\rho + 2\delta} |A_0(t)| |A_1(t)| .$$

Suppose that $\tilde{t}$ is large enough that, for $t > \tilde{t}$, $x \ge 1$,

$$\gamma \ln(x) + A_1(t) f_\rho(x) > 0, \quad 1 + rA_0(t) f_\rho(x) > 0.$$

Then, by applying inequalities (42) and (43) one more time, we obtain

$$-\epsilon d_1(x,t) \le g_{r,1}\left(\frac{U(tx)}{U(t)}\right) - b_1(x) - c_1(x,t) \le \epsilon d_1(x,t), \tag{44}$$

where

$$\begin{aligned}
b_1(x) &= \gamma x^{\gamma r} \ln(x),\\
c_1(x,t) &= (A_1(t) + \gamma r A_0(t) \ln(x))\, x^{\gamma r} f_\rho(x) + rA_0(t)A_1(t) x^{\gamma r} f_\rho^2(x),\\
d_1(x,t) &= x^{\gamma r + \rho + \delta}\Big(|A_1(t)| \left\{1 + rA_0(t) f_\rho(x)\right\}\\
&\quad + r\,|A_0(t)| \left\{\gamma \ln(x) + A_1(t) f_\rho(x)\right\} + \epsilon r x^{\rho+\delta} |A_0(t)| |A_1(t)| \Big),
\end{aligned}$$

To prove two-dimensional Central Limit Theorem (14) we shall use the well-known Cramer–Wald method. Let $(\theta_0, \theta_1) \in \mathbf{R}^2$. From (42) and (44) we get

$$-\epsilon d(x,t) \le \theta_0 g_{r,0}\left(\frac{U(tx)}{U(t)}\right) + \theta_1 g_{r,1}\left(\frac{U(tx)}{U(t)}\right) - b(x) - c(x,t) \le \epsilon d(x,t), \tag{45}$$

where

$$\begin{aligned}
b(x) &= \theta_0 x^{\gamma r} + \theta_1 b_1(x),\\
c(x,t) &= \theta_0 r A_0(t) x^{\gamma r} f_\rho(x) + \theta_1 c_1(x,t),\\
d(x,t) &= |\theta_0|\, r x^{\gamma r + \rho + \delta} |A_0(t)| + |\theta_1|\, d_1(x,t).
\end{aligned}$$

We claim that

$$\frac{1}{\sqrt{k}} \sum_{i=0}^{k-1} d\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}, Y_{n-k,n}\right) \xrightarrow{\mathrm{P}} \frac{r\,|\theta_0 \mu|}{d_r(1) - \rho - \delta} + \frac{(1-\rho-\delta)\,|\theta_1 \mu|}{(d_r(1) - \rho - \delta)^2}, \quad n \to \infty. \tag{46}$$

From Lemma 1 in Paulauskas and Vaičiulis (2013) we know that if $\nu < 1$, $\nu \neq 0$, then

$$\frac{1}{k}\sum_{i=0}^{k-1} g_{\nu,0}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) \xrightarrow{\mathrm{P}} \frac{1}{1-\nu}, \quad n \to \infty. \tag{47}$$

Similarly one can prove

$$\frac{1}{k}\sum_{i=0}^{k-1} g_{\nu,1}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) \xrightarrow{\mathrm{P}} \frac{1}{(1-\nu)^2}, \quad n \to \infty. \tag{48}$$

The relation

$$\sqrt{k}A\left(Y_{n-k,n}\right) \xrightarrow{\mathrm{P}} \mu, \quad n \to \infty, \tag{49}$$

where μ is the same as in (13), is proved in Paulauskas and Vaičiulis (2013). Now, by combining (47)–(49) one can obtain (46).

Taking into account (46), substituting the values of t and x from (35) into (45) and performing summation over $i = 0, 1, \ldots, k-1$ we get distributional representation

$$\begin{aligned}&\sqrt{k}\left\{\theta_0\left(G_n(k,r,0) - \xi(r,0)\right) + \theta_1\left(G_n(k,r,1) - \xi(r,1)\right)\right\}\\ &\stackrel{\mathrm{d}}{=} \sqrt{k}B_n(k,r) + \sqrt{k}C_n(k,r) + o_p(1),\end{aligned} \tag{50}$$

where

$$B_n(k,r) = \frac{1}{k}\sum_{i=0}^{k-1}\left\{b\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) - \theta_0\xi(r,0) - \theta_1\xi(r,1)\right\},$$

$$C_n(k,r) = \frac{1}{k}\sum_{i=0}^{k-1} c\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}, Y_{n-k,n}\right).$$

By applying relations (47)–(49) one more time, we find

$$\sqrt{k}C_n(k,r) \xrightarrow{\mathrm{P}} \mu\left(\theta_0\nu^{(1)}(r) + \theta_1\nu^{(2)}(r)\right), \quad n \to \infty, \tag{51}$$

where $\nu^{(j)}(r)$, $j = 1, 2$ are defined in (15). By using the well-known Rényi's representation [see e.g., Section 2 in Paulauskas and Vaičiulis (2013) for details] we obtain

$$\sqrt{k}B_n(k,r) \stackrel{\mathrm{d}}{=} \tilde{B}_n(k,r), \tag{52}$$

where

$$\tilde{B}_n(k,r) = \frac{1}{\sqrt{k}}\sum_{i=0}^{k-1}\left\{\theta_0\left(g_{\gamma r,0}(Y_i) - \xi(r,0)\right) + \theta_1\gamma\left(g_{\gamma r,1}(Y_i) - \frac{1}{d_r^2(1)}\right)\right\}.$$

Keeping in mind equality (40) one can deduce that the quantity $\tilde{B}_n(k,r)$ presents normalized sum of i.i.d. zero mean random variables. Moreover, under assumption $\gamma r < 1/2$,

$$\mathrm{E}\left\{\theta_0\left(g_{\gamma r,0}(Y_i)-\xi(r,0)\right)+\theta_1\gamma\left(g_{\gamma r,1}(Y_i)-\frac{1}{d_r^2(1)}\right)\right\}^2$$
$$=\theta_0^2 s_1^2(r)+2\theta_0\theta_1 s_{12}(r)+\theta_1^2 s_1^2(r),$$

where $s_1^2(r)$, $s_2^2(r)$ and $s_{12}(r)$ are defined in Theorem 2. Thus, applying Lindeberg–Lévy central limit theorem, we get the relation $\tilde{B}_n(k,r)\xrightarrow{\mathrm{d}}\theta_0 W^{(1)}+\theta_1 W^{(2)}$. This, together with (52) gives

$$\sqrt{k}B_n(k,r)\xrightarrow{\mathrm{d}}\theta_0 W^{(1)}+\theta_1 W^{(2)}, \tag{53}$$

as $n\to\infty$. Applying Theorem 4.4 in Billingsley (1968), from (51) and (53) we get

$$\sqrt{k}\left(B_n(k,r),C_n(k,r)\right)\xrightarrow{\mathrm{d}}\left(\theta_0 W^{(1)}+\theta_1 W^{(2)},\mu\left(\theta_0\nu^{(1)}(r)+\theta_1\nu^{(2)}(r)\right)\right),$$

as $n\to\infty$. Continuous Mapping Theorem gives us the relation

$$\sqrt{k}\left(B_n(k,r)+C_n(k,r)\right)\xrightarrow{\mathrm{d}}\theta_0\left(W^{(1)}+\mu\nu^{(1)}(r)\right)+\theta_1\left(W^{(2)}+\mu\nu^{(2)}(r)\right),$$

as $n\to\infty$. The last relation together with (50) gives (14). Theorem 2 is proved. □

Proof of Theorem 3 In the case $j=1$ the proof of the relation (16) can be found in Paulauskas and Vaičiulis (2013) (proof of the Corollary 1) or in Beran et al. (2014) (proof of Theorem 2). But the asymptotic normality of all estimators $\hat{\gamma}_n^{(j)}(k,r)$, $j=1,2,3$ can be obtained in a unified way, expressing these estimators as functions of statistics $G_n(k,r,0)$ and $G_n(k,r,1)$, and then combining Theorems 1, 2 and Continuous mapping Theorem. Namely, it is easy to see that

$$\hat{\gamma}_n^{(1)}(k,r)-\gamma=\frac{(1-\gamma r)\left(G_n(k,r,0)-\xi_{r,0}\right)}{rG_n(k,r,0)}.$$

For the estimator $\hat{\gamma}_n^{(2)}(k,r)$ we have

$$\hat{\gamma}_n^{(2)}(k,r)-\gamma=\frac{2(1-\gamma r)G_n(k,r,1)-\gamma-\gamma\sqrt{4rG_n(k,r,1)+1}}{2rG_n(k,r,1)+1+\sqrt{4rG_n(k,r,1)+1}}.$$

Multiplying the numerator and denominator of the right-hand side by $2(1-\gamma r)G_n(k,r,1)-\gamma+\gamma\sqrt{4rG_n(k,r,1)+1}$, we get

$$\hat{\gamma}_n^{(2)}(k,r)-\gamma=4(1-\gamma r)^2\left(G_n(k,r,1)-\xi_{r,1}\right)$$

$$\times \frac{\left(G_n(k,r,1)+\xi_{r,1}\right)-4\gamma\left(G_n(k,r,1)-\xi_{r,1}\right)}{2rG_n(k,r,1)+1+\sqrt{4rG_n(k,r,1)+1}}$$
$$\times \frac{1}{2(1-\gamma r)G_n(k,r,1)-\gamma+\gamma\sqrt{4rG_n(k,r,1)+1}}.$$

For the third estimator the following representation holds:

$$\hat{\gamma}_n^{(3)}(k,r)-\gamma=\frac{r(1-\gamma r)\left(G_n(k,r,1)-\xi_{r,1}\right)-\left(G_n(k,r,0)-\xi_{r,0}\right)}{r^2G_n(k,r,1)}.$$

As it was said, it remains to combine Theorems 1, 2 and Continuous Mapping Theorem, and we deduce (16) with $1\le j\le 3$. For example, we have

$$\begin{aligned}&\hat{\gamma}_n^{(3)}(k,r)-\gamma\\&=f\left(\sqrt{k}(G_n(k,r,1)-\xi(r,1)),\sqrt{k}(G_n(k,r,0)-\xi(r,0)),G_n(k,r,1)\right)\end{aligned}$$

with

$$f(x,y,z)=\frac{r(1-\gamma r)x-y}{r^2z}.$$

From Theorems 1 and 2 and Theorem 3.9 from Billingsley (1968), we have

$$\left(\sqrt{k}\left(G_n(k,r,i)-\xi(r,i)\right),i=0,1,G_n(k,r,1)\right)$$

$$\xrightarrow{\mathrm{d}}\left(W^{(i)}+\mu\nu^{(i)}(r),i=1,2,\xi(r,1)\right),$$

and now we apply Continuous Mapping Theorem (Theorem 2.7 in Billingsley (1968)).

As it was mentioned at the end of Sect. 1, for the proof of the asymptotic normality of the introduced estimators there is possibility to use general approach, suggested in Drees (1998). Let F_n stand for the empirical d.f. based on the sample $X_1,\ldots,X_n$ and let the empirical tail quantile function is defined as

$$Q_n(t):=F_n^{-1}\left(1-\frac{k_n}{n}t\right)=X_{n-[k_nt],n},\quad t\in[0,1].$$

Then almost all known estimators of the EVI that are based on some part of largest order statistics can be written as some functional T (defined on some functional space), applied to Q_n. Then, having estimator written as $T(Q_n)$, the idea in Drees (1998) is to use invariance principle for the process Q_n in the functional space on which T is defined and then requiring some smoothness of T one can try to derive asymptotic normality of the estimator under consideration. Smoothness of a functional T is required in terms of Hadamard differentiability in linear topological space; usual $D[0,1]$ space with the Skorokhod topology is a metric, but not a linear space, while $D[0, 1]$ with supremum norm is non-separable normed space, therefore usually one deals with the so-called countable normed spaces. For estimators, considered in the paper, it is possible to use this scheme, but the functionals, which appear using this approach, are quite complicated. For example, one can write $\hat{\gamma}_n^{(3)}(k, r) = T_{\mathrm{GMR}}(Q_n)$ with

$$T_{\mathrm{GMR}}(z) = f(T_0(z), T_1(z), T_2(z)),$$

$$f(x, y, v) = \frac{ry - x + 1}{r^2 y}, \quad \text{for} \quad r \neq 0 \quad \text{and} \quad f(x, y, v) = \frac{z}{2y}, \quad \text{for} \quad r = 0,$$

and

$$T_i(z) = \int_0^1 g_{r,i}\left(\frac{z(t)}{z(1)}\right) \mathrm{d}t, \quad i = 0, 1, \quad T_2(z) = \int_0^1 g_{0,2}\left(\frac{z(t)}{z(1)}\right) \mathrm{d}t.$$

For this complicated functional T_{GMR} we must prove Hadamard differentiability on some linear topological space (to choose the appropriate space is also non trivial problem). Thus, it seems that our approach is much more simple, and we do not need more restrictive conditions [such that appears in Drees (1998)], since instead of invariance principle for the tail quantile process we prove two-dimensional CLT for two particular statistics and then apply continuous mapping theorem in R^3. □

Proof of Theorem 4 The expression of $B^{(1)}(r)$ is given in Beran et al. (2014), and at first we followed the pattern of the proof in Beran et al. (2014), but, as it was noticed in Introduction, there is more simple proof.

From the expression (18) we see that, in order to find $B_n^{(j)}(k, r)$, it is sufficient to find the limit $\lim_{x\to\infty} \hat{\gamma}_n^{(j)}(k, r; X_1, \ldots, X_{n-1}, x)$, since $\hat{\gamma}_{n-1}^{(j)}(k-1, r; X_1, \ldots, X_{n-1}) \xrightarrow{\mathrm{P}} \gamma$, as $n \to \infty$ and (8) holds. For all three estimators calculations are simple and similar, therefore we demonstrate the proof for the estimator $\hat{\gamma}_n^{(2)}(k, r, x)$, having the most complicated expression. It is clear that, for sufficiently large value of x, $\hat{\gamma}_n^{(j)}(k, r; X_1, \ldots, X_{n-1}, x)$ can be written as

$$\frac{2\,(h(x) + b)}{2r\,(h(x) + b) + 1 + \sqrt{4r\,(h(x) + b) + 1}}, \tag{54}$$

where $h(x) = g_{r,1}\left(x/X_{n-k,n}\right)$ and b is the sum of the rest summands from statistic $G_n(k, r, 1)$ and does not depend on x. If $r > 0$, then $h(x) \to \infty$ and the limit of the quantity in (54) is $1/r$, while for $r < 0$ $h(x) \to 0$, and the limit in (54) is

$$\frac{2b}{2rb + 1 + \sqrt{4rb + 1}}$$

and this expression almost (this word is used for the reason that in the above written expression the sum is divided by k, while for complete coincidence division should be by $k-1$) coincides with $\hat{\gamma}_{n-1}^{(j)}(k-1, r; X_1, \ldots, X_{n-1})$; therefore, passing to the limit as $n \to \infty$ we get in limit $B^{(2)} = 0$. In the case $r = 0$ only nominator contains function $h(x) = g_{0,1}\left(x/X_{n-k,n}\right)$ which grows unboundedly; therefore, we get infinite value for $B_n^{(j)}(k, r)$.

Acknowledgements The authors are grateful to anonymous referees and the editor whose remarks and suggestions were very helpful in improving the paper.

References

Beran, J., Schell, D., Stehlik, M. (2014). The harmonic moment tail index estimator: Asymptotic distribution and robustness. *Annals of the Institute of Statistical Mathematics*, *66*, 193–220.

Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.

Brilhante, M. F., Gomes, M. I., Pestana, D. (2013). A simple generalization of the Hill estimator. *Computational Statistics and Data Analysis*, *57*, 518–535.

Danielsson, J., Jansen, D. W., De Vries, C. G. (1996). The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics: Theory and Methods*, *25*, 711–720.

De Haan, L., Ferreira, A. (2006). *Extreme value theory: An introduction*. New York: Springer.

De Haan, L., Peng, L. (1998). Comparison of tail index estimators. *Statistica Nederlandica*, *52*, 60–70.

Dekkers, A. L. M., De Haan, L. (1993). Optimal choise of sample fraction in extreme-value estimation. *Journal of Multivariate Analysis*, *47*, 173–195.

Dekkers, A. L. M., Einmahl, J. H. J., De Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, *17*, 1833–1855.

Drees, H. (1998). A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference*, *66*, 95–112.

Fabian, Z., Stehlik, M. (2009). On robust and distribution sensitive Hill like method. IFAS Research report 43. Linz: Department for Applied Statistics, Johannes Kepler university.

Fraga Alves, M. I., Gomes, M. I., De Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, *60*, 193–214.

Gomes, M. I., Martins, M. J. (2001). Alternatives to Hill's estimator: Asymptotic versus finite sample behaviour. *Journal of Statistical Planning and Inference*, *93*, 161–180.

Gomes, M. I., Martins, M. J. (2002). Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter. *Extremes*, *5*, 5–31.

Gomes, M. I., Pestana, D., Caeiro, F. (2009). A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Statistics and Probability Letters*, *79*, 295–303.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society Series B. Statistical Methodology*, *44*, 37–42.

Hall, P., Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Annals of Statistics*, *13*, 331–341.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, *3*, 1163–1174.

Henry, J. B, I. I. I. (2009). A harmonic moment tail index estimator. *Journal of Statistical Theory and Applications*, *8*, 141–162.

Paulauskas, V. (2003). A new estimator for a tail index. *Acta Applicandae Mathematicae*, *79*, 55–67.

Paulauskas, V., Vaičiulis, M. (2010). Once more on comparison of tail index estimators. unpublished manuscript.

Paulauskas, V., Vaičiulis, M. (2011). Several modifications of DPR estimator of the tail index. *Lithuanian Mathematical Journal*, *51*, 36–50.

Paulauskas, V., Vaičiulis, M. (2013). On the improvement of Hill and some others estimators. *Lithuanian Mathematical Journal*, *53*, 336–355.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, *3*, 119–131.

Stehlik, M., Potocký, R., Waldl, H., Fabián, Z. (2010). On the favourable estimation of fitting heavy tailed data. *Computational Statistics*, *25*, 485–503.