QUANTIFICATION OF ESTIMATION INSTABILITY AND ITS APPLICATION TO THRESHOLD SELECTION IN EXTREMES

Thomas L. Berning
Stellenbosch University
e-mail: *tberning@sun.ac.za*

Key words: estimation instability, extreme value index, threshold selection, bias-reduced estimator.

Summary: The main goal of this paper is to propose a measure which quantifies the instability of estimates over a range of chosen values of some other parameter. The measure is used to identify a region where the estimates are considered "stable". Methods for identifying stable regions are then developed. These methods are applied to threshold selection in an extreme value analysis context, where the perturbed Pareto distribution is fitted to observed relative excesses. As a result a more accurate estimator of the extreme value index is obtained. As a further application the instability measure is employed in second order parameter estimation, which leads to an adjustment of an existing estimator, having desirable properties as far as its role in threshold selection is concerned, as well as improved estimation of the second order parameter for large samples.

1. Introduction

In extreme value theory (EVT) the emphasis is on describing the tail of the underlying distribution. In this paper the focus will be on the right tail of the underlying distribution, which is assumed to be heavy-tailed.

A distribution is heavy-tailed if and only if its right tail function is regularly varying (de Haan and Ferreira, 2006). The definition is given below. See for instance Geluk, de Haan, Resnick and Stărică (1997).

Definition 1 Let X *be a random variable with distribution function* F *concentrated on* $[0,\infty)$ *and tail function* $\overline{F} = 1-F$. $\overline{F}$ *is said to be regularly varying with index* $-1/\gamma$, $\gamma > 0$, *if*

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\frac{1}{\gamma}} \tag{1}$$

for all $x > 0$.

The right hand side of 1 can be recognised as the tail function of the Pareto distribution. (It is defined for $x \geq 1$ in case of the Pareto distribution.)

AMS: 60G70, 62F10

The parameter γ is called the extreme value index (EVI). A positive EVI indicates a heavy right tail. The EVI is the crucial parameter, since accurate estimation of the EVI leads to accurate inferences concerning extreme quantiles. In this paper attention will only be given to EVI estimation. Once the methodology concerning the estimation of the EVI has been established, the extension to inferences concerning extreme quantiles is standard for maximum likelihood estimation (MLE). Refer for instance to Coles (2001).

Assume a sequence of independent, identically distributed random variables $X_1, X_2, ..., X_n$ with common heavy-tailed distribution F, concentrated on $[0, \infty)$. Denote the corresponding order statistics by $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. Let k denote the number of observations exceeding a positive threshold t. The relation between t and k will be taken as $t = X_{n-k,n}$. If the random variable X has distribution F, and Z is defined as the relative (or multiplicative) excess above the threshold t, then $Z = X/t$ given $X \geq t$. For $j = 1, 2, ..., k$, define $Z_j = X_i/t$ conditional on $X_i > t$, where i is the index of the j-th excess, with $Z_j \geq 1$.

Let $\overline{F}_t$ be the tail function of Z. Then, applying Definition 1, it follows that

$$\overline{F}_t(z) = P(X/t > z | X \geq t) = P(X > tz | X \geq t) = P(X > tz)/P(X \geq t) = \overline{F}(tz)/\overline{F}(t) \to z^{-1/\gamma}$$

as $t \to \infty$ for all $z > 0$ (and in particular for $z \geq 1$), where $\gamma > 0$.

In words this result states that the distribution of the relative excesses from a heavy-tailed distribution is approximately Pareto, if the threshold is large.

The most natural way to estimate the EVI for relative excesses $Z_1, Z_2, ..., Z_k$ is to fit the Pareto distribution by means of maximum likelihood estimation. It is easy to show that this yields $H_{k,n} := \frac{1}{k}\sum_{i=1}^{k} log Z_i$, which is the famous Hill estimator of the EVI (Hill, 1975). Note that the formula for the Hill estimator is derived as one would an MLE, but strictly speaking the variables $Z_1, Z_2, ..., Z_k$ are not independent.

The Hill estimator performs poorly as far as accuracy of estimation is concerned, since the estimate is extremely sensitive to the choice of k. For small k the estimates have low bias, but large variance. This is because the threshold is large enough for the first order approximation (Pareto distribution) to be valid, but the number of observations k on which the estimate is based, is small. Conversely large values of k lead to low variance, but large bias.

A variety of methods for choosing the threshold (or k) has been proposed in the literature, for instance the method by Drees and Kaufmann (1998), the method by Guillou and Hall (2001), and the method by Beirlant, Goegebeur, Segers and Teugels (2004), the last of which is based on the minimisation of the estimated asymptotic mean square error. Despite the improvements in estimation of the EVI brought about by these threshold selection methods, the Hill estimator still performs unsatisfactorily.

A proposed solution to the bias problem associated with the Hill estimator is to consider a bias-reduced estimator, obtained in this instance by fitting the perturbed Pareto distribution (PPD) to the excesses, instead of the (strict) Pareto as in the case of the Hill estimator. The motivation behind this approach is to obtain a larger range of values of k for which the bias is relatively low.

In the above setting, being able to *objectively* determine such a range of values of k for which the estimator is "stable" will not only be more scientific, but will also enable comparison of these methods by means of simulation studies. A method for quantifying the stability (or rather instability) of estimates over a range of values of k is developed in this paper. The *stable region* is then defined as the region for which the instability measure yields the lowest value.

Even though the concept is only illustrated in the context of EVT, the same measure can be used in any application where the stability of estimates needs to be quantified over a range of some other parameter. Examples of such situations are mentioned in the concluding remarks at the end of the paper.

The purpose of this paper is to define a measure which quantifies the instability of estimates and applying this measure to develop a method of threshold selection when fitting the PPD to relative excesses. Note that the purpose is not an exhaustive comparative study of all estimators of the first and second order parameters or of all threshold selection methods.

The next section reviews some results from the literature. The necessity of bias-reduced estimators is demonstrated by considering the properties of the Hill estimator for a specific sample. Fitting the PPD to the excesses instead of the Pareto yields a second order generalisation of the Hill estimator. Some computational aspects are also highlighted.

In Section 3 the proposed instability measure is defined. It is then described how the instability measure can be used to locate *stable regions* (sets of consecutive estimates which yield a low value of the instability measure). Once a stable region has been identified, the EVI is estimated as the mean of the estimates over the stable region, leading to the concept of an implied threshold. Section 4 discusses six candidate external estimators of the second order parameter. The choice of this estimator crucially affects the accuracy of the resulting EVI estimation. In Section 5 a simulation study design is proposed which is applicable to any study of estimators of a positive EVI. This design is then applied to assess the performance of the second order parameter estimators, reducing the number of candidate external estimators to three. In Section 6 it is shown how the proposed EVI estimators are constructed. This section also includes the results of a simulation study comparing the performance of these estimators, including some results illustrating the effect of adaptive threshold selection for the Hill estimator. Some concluding remarks appear at the end of the paper.

2. Second order estimation

Consider a second order regular variation condition on the right tail function. Refer for instance to Geluk et al. (1997).

Definition 2 *A regularly varying tail function $\overline{F}$ is said to be second order regularly varying with first order index $-1/\gamma$, $\gamma > 0$, and second order index ρ/γ, $\rho \leq 0$, if a function $A(t)$ exists which tends to* 0 *and is ultimately of constant sign as $t \to \infty$, such that*

$$\lim_{t\to\infty} \left(\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-1/\gamma} \right) (A(t))^{-1} = H(x)$$

for all $x > 0$, where $H(x) = dx^{-1/\gamma} logx$ if $\rho = 0$ and $H(x) = dx^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\rho/\gamma}$ if $\rho < 0$, with $-\infty < d < \infty$.

In this definition, γ is still the EVI (also called the *first order parameter*) and ρ is called the *second order parameter*.

Note that Geluk et al. (1997) specify that $d \neq 0$. de Haan and Ferreira (2006) also specify $d \neq 0$. This restriction is not applied in Beirlant et al. (2004).

Whether or not d may be zero is of theoretical rather than practical significance. Theoretically we should have $d \neq 0$, since if $d = 0$, ρ is undefined, and consequently the second order index is also undefined. The advantage of allowing d to be zero is that the Pareto distribution can be included as a special case. For the Pareto distribution $\overline{F}(tx)/\overline{F}(t) = x^{-1/\gamma}$, and therefore $d = 0$.

For all practical purposes one can assume that all heavy-tailed distributions of interest are first and second order regularly varying. Following limiting arguments similar to those shown above for the first order case, one can show that the second order condition implies that the relative excesses approximately follow the perturbed Pareto distribution (PPD), if the threshold is large. The definition of the PPD follows.

Definition 3 *A random variable X is said to be distributed perturbed Pareto with parameters γ, ρ* and *c, denoted X ~ PPD(γ,ρ,c), when*

$$\overline{F}(x) = (1-c)x^{-1/\gamma} + cx^{-(1-\rho)/\gamma},$$

where $x \geq 1$, $\gamma > 0$, $\rho < 0$ and $\rho^{-1} \leq c \leq 1$.

The restriction $\rho^{-1} \leq c \leq 1$ ensures that the density function $f(x) \geq 0$ for all $x \geq 1$. The Pareto distribution is a special case of the PPD when $c = 0$ (in which case ρ is not zero, but undefined).

The EVI γ is a parameter of the PPD which can be estimated using MLE, resulting in a second order generalisation of the Hill estimator.

Figure 1 below shows results pertaining to a sample of size $n = 200$ observations simulated from a *Burr*$(\beta = 1, \gamma = 0.5, \rho = -0.7)$ distribution. (Refer to Section 5 for more detail concerning the Burr distribution.) In the first graph the Hill estimates of the EVI were calculated for $k = 1, 2, \ldots, 150$. In the second graph the PPD was fitted to k multiplicative excesses by means of MLE. The resulting estimates of the parameter γ (the EVI) is shown for $k = 1, 2, \ldots, 150$.

The horizontal dashed line indicates the true value of the EVI, namely 0.5.

The first graph of Figure 1 clearly illustrates the sensitivity of the Hill estimate with respect to the choice of k. It is evident from the second graph that a larger value of k (and larger range of k values) can be used when fitting the PPD. This illustrates the typical (desired) effect of a bias-reduced estimator.

Some computational remarks are in order with respect to the MLEs of the PPD parameters.

The first concerns the estimation of ρ. One option is not to estimate ρ at all, but to fix it at some value, usually -1. See for instance Beirlant et al. (2004) and Dierckx (2000). The second option is to estimate ρ simultaneously with the other two parameters, γ and c. The third option, called *external estimation of the second order parameter*, entails using some existing estimator of ρ, substituting the value of ρ in Definition 3 by its estimated value $\widehat{\rho}$, and estimating the remaining two parameters using MLE.

The first option was popular before estimators of ρ existed which perform well. In the simulation studies which were carried out, the estimator of the EVI which performed best was obtained by external estimation of ρ. This is in line with theoretical results which state that external estimation of ρ leads to a smaller asymptotic variance in reduced bias estimators. See for instance Gomes, Figueiredo and Neves (2012). The choice of estimator of ρ will receive attention in later sections.

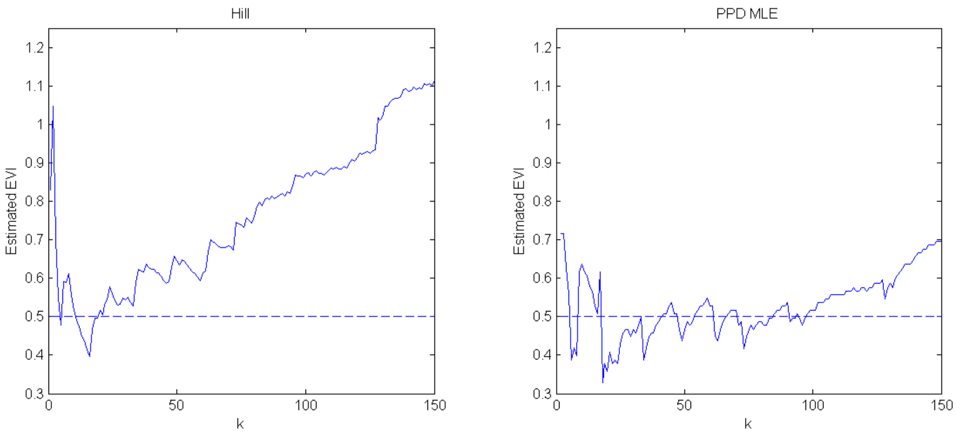

Figure 1: Comparison of first and second order estimates.

The second issue is the computational restrictions on the range of the parameter c. From extensive simulations it was found that applying the restriction $c \leq 0.5$ substantially improves the EVI estimation accuracy. This restriction also makes sense from a theoretical point of view. For $0 \leq c \leq 1$ the PPD survivor function is the weighted sum of a first order component $x^{-1/\gamma}$, and a second order component $x^{-(1-\rho)/\gamma}$. When $c > 0.5$ there is a greater weight on the second order component than on the first, which is not desirable.

3. An instability measure to determine regions of stable EVI estimates

Considering the second graph of Figure 1, one can see that values of k between 40 and 80 (roughly) would yield the most accurate estimates of the EVI. This is the region where the estimates are the most stable. What is meant by "stable" in this context is also clear: one needs to identify a region where the estimates do not vary excessively (low variance), and where the slope is more or less zero (no systematic change in bias). In summary, a region which optimises the variance-bias trade-off needs to be identified.

The proposed measure quantifies the instability of a quantity y with respect to chosen values of another quantity x.

Definition 4 *For $m \geq 2$, let $y_1, y_2, \ldots, y_m$ denote the observed values of a quantity y corresponding to chosen values $x_1, x_2, \ldots, x_m$ of another quantity x, respectively.*

The instability of y with respect to x is defined as

$$\vartheta^2 = \sigma^2 + b^2,$$

where σ^2 is the sample variance of the y values, and b is the slope of the simple least squares regression line of y on x.

In the context of threshold selection in EVT, $\{y_i\}$ is the set of estimated EVIs, and $\{x_i\}$ a set of values corresponding to the respective choices of k. The interpretation of ϑ^2 is similar to that of an MSE, namely that it is, in the sense described above, the sum of the variance and the square of the (change in) bias.

Applying the instability measure for the purpose of threshold selection, the assumption is made that the x values are equally spaced. Prior to calculating the instability measure, the x and y values are also normalised in such a way that the measure ensures location and scale invariance with respect to both the original x values and the original y values.

The chosen x values $x_1 < x_2 < ... < x_m$ can be normalised by calculating $(x_i - x_1)/(x_m - x_1)$ for $i = 1, 2, \ldots, m$. If the original x values are equally spaced and m is large, these values are close to $x_i^* = i/m$. The latter is not only used for the sake of simplicity, but performs better in practice. The reason is that it penalises small values of m. For instance, if $m = 2$ the formula $(x_i - x_1)/(x_m - x_1)$ yields adjusted values of 0 and 1, whereas the formula i/m yields $x_1^* = 0.5$ and $x_2^* = 1$, effectively doubling the value of the slope b, leading to larger value of ϑ^2.

The observed y values are normalised by calculating $y_i^* = y_i/\bar{y}$, where $\bar{y}$ denotes the mean of the y values.

Attention will now be given to procedures for identifying stable regions in the context of threshold selection in EVT. Consider again the second graph of Figure 1. To determine the most stable region in the set of estimates, a smoother plot than that of the right side of Figure 1 is required.

One method which works particularly well in practice is not to calculate the EVI estimates at all possible values of k (namely $1, 2, ..., n-1$), but rather only at $k = 5\%, 10\%, \ldots, 95\%$ of n (rounded to the nearest integer). This leads to 19 estimates of the EVI, denoted in this context by $y_1, ..., y_{19}$. The term *region* refers to a set of consecutive values of y. For instance, the set $\{y_3, y_4, ..., y_9\}$ is a region *of length* 7.

The accuracy of EVI estimation is significantly improved by applying two procedures before applying methods of stable region selection. The first procedure rounds the values $y_1, ..., y_{19}$ to the nearest 5% of their mean to avoid insignificant fluctuations from having an effect when applying Algorithm 3.2 below. Specifically, the first procedure is the following:

Algorithm 3.1

factor := round(100 $\times$ (0.05 $\times$ $\bar{y}$))/100 *Remark*: 5% of mean rounded to two decimals

if factor = 0 then factor := 0.01

$\{y_1, ..., y_{19}\}$:= factor $\times$ round($\{y_1, ..., y_{19}\}$/factor)

The idea behind the second procedure is to remove completely from consideration the region of estimates where the bias becomes significant. This entails reducing the set of $\{y_1, ..., y_{19}\}$ to $\{y_1, ..., y_m\}$, where $m < 19$ if the original set contains non-decreasing values from some index onwards. Specifically, the algorithm is the following:

Algorithm 3.2

$y_{20} := 2 \times y_{19}$ *Remark*: Definition of y_{20} required in final while loop
$m := 19$
while $(y_m \geq y_{m-1})$ and $(y_{m-1} \geq y_{m-2})$ and $(m > 2)$
    $m := m - 1$
end
while $(y_m = y_{m-1})$ and $(m < 19)$ *Remark*: One increase is retained
    $m := m + 1$
end
while $(y_m = y_{m+1})$ and $(m < 19)$
    $m := m + 1$
end

An example of the application of Algorithm 3.2 will be given below. The methods for selecting an optimal (stable) region follow.

Method 0 only applies algorithms 3.1 and 3.2, regarding the remaining m values $\{y_1, ..., y_m\}$ as the optimal region. No use is made of the instability measure.

For the sake of simplicity it will be assumed in the explanations below that m is still 19 after the application of Algorithm 3.2.

Method 1 entails fixing the region length. For a region length of 12, for example, ϑ^2 is calculated for regions $\{y_1, ..., y_{12}\}$, $\{y_2, ..., y_{13}\}$, up to $\{y_8, ..., y_{19}\}$. The region for which ϑ^2 is a minimum, is regarded as the optimal region. If Algorithm 3.2 yields a value of m which is smaller than the region length, the region length is set to m. In other words, if the set of estimates which remains after applying Algorithm 3.2 has fewer elements than the specified region length, the stable region is simply taken as $\{y_1, ..., y_m\}$.

Method 2 involves trimming the region. First ϑ^2 is calculated over the entire region $\{y_1, ..., y_{19}\}$. Then either the first or the last value in the region is omitted, depending on which reduction of the region decreases ϑ^2 by the largest amount. This process is repeated until neither top nor bottom trimming of the region decreases the instability ϑ^2.

Method 3 fixes the upper limit of the region and trims the region from the left. For an upper limit of 6, for example, ϑ^2 is calculated for $\{y_1, ..., y_6\}$, $\{y_2, ..., y_6\}$, up to $\{y_5, y_6\}$, and the region yielding the lowest ϑ^2 is regarded as the optimal region.

Simulation studies were carried out on all three methods (and for all possible parameters of each), across all sample sizes and distributions. The results indicated that algorithms 3.1 and 3.2 should always be applied (as mentioned earlier), and that two methods outperform the others on all counts, namely methods 0 and 3.

Simulation studies were carried out to obtain rules to decide between the methods, and to determine what the choice of upper limit should be in the case of Method 3. These and other issues are addressed in Section 6.

Once the stable region has been identified, the estimate of the EVI is taken as the mean of the *original* EVI estimates over the stable region, i.e. before applying the rounding of Algorithm 3.1.

As an example, consider again the Burr sample used earlier.

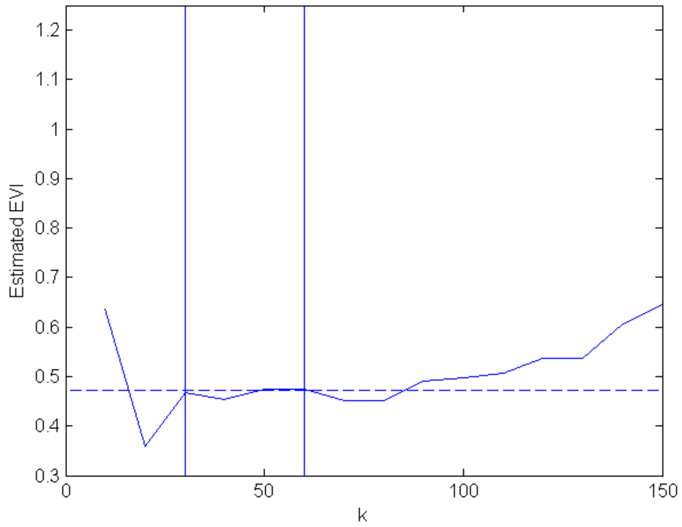

Figure 2: Stable region and EVI estimate for the Burr sample.

Figure 2 is the same as the right side of Figure 1, except that the estimates are calculated at larger intervals of k, namely $k = 10, 20, \ldots, 150$. The plot therefore shows the values $y_1, ..., y_{15}$. The vertical lines in Figure 2 indicate the optimal region which was found to be $k = 30, \ldots, 60$. The dashed line indicates the final EVI estimate of 0.4727.

After rounding the values by using Algorithm 3.1, Algorithm 3.2 was applied. The first *while* loop in the algorithm yielded $m = 8$ (corresponding to $k = 80$). The second *while* loop yielded $m = 9$. Even though bias reduced estimation is applied, there is often a point from where the estimates increase steadily or, more specifically, are non-decreasing as a function of k. This point is at $k = 100$. The values corresponding to $k = 100, 110, \ldots, 190$ are excluded before applying stable region methods 1, 2 or 3. The increase from $k = 80$ to $k = 90$ is regarded as variation, and not taken as the start of the region where the bias is regarded as significant.

Method 3 with an upper limit of 6 was applied to the remaining region $\{y_1, ..., y_9\}$, yielding $\{y_3, ..., y_6\}$, corresponding to $k = 30, \ldots, 60$.

From this point onwards the original values $y_3, ..., y_6$ (without the rounding of Algorithm 3.1) are used. The estimated EVI for this sample is calculated as the mean of $y_3, ..., y_6$, which is 0.4727.

The EVI in the optimal region which is closest to the final estimate of the EVI indicates the choice of threshold. In the example above, the single estimate in the set $\{y_3, ..., y_6\}$ closest to 0.4727 is $y_5 = 0.4731$. This indicates a threshold of $k = 50$. The term *implied threshold* will be used to refer to the threshold selected in this manner. The resulting 50 excesses constitute the observations which will be used to construct confidence intervals, etc. In the concluding remarks more will be said of how to proceed from this point onwards, in the context of a specific financial application.

4. External estimation of the second order parameter

Attention will now be given to the exact method of obtaining the set of EVI estimates at a specified range of values of k. In Section 2 the need for external estimation of ρ was pointed out. In this section six estimators of ρ will be investigated as candidates for the external estimation of ρ when fitting the PPD to relative excesses using MLE. The first four estimators are those considered by Gomes and Martins (2002), all of which are easy to implement, both in terms of programming and computation time. The fifth estimator is one proposed by Gomes and Martins (2001). This estimator is chosen since it lends itself to the application of the instability measure, which yields an adjusted version of the estimator.

The first two estimators are

$$\widehat{\rho}_1 := -\left|\log\left|\frac{1/M_n^{(1)}\left(\left[n^{0.9}\right]\right)-1/M_n^{(1)}\left(\left[n^{0.5}\right]\right)}{1/M_n^{(1)}\left(\left[n^{0.95}\right]\right)-1/M_n^{(1)}\left(\left[n^{0.5}\right]\right)}\right|/\log\frac{\left[n^{0.9}\right]}{\left[n^{0.95}\right]}\right|$$

by Hall and Welsh (1985), and

$$\widehat{\rho}_2 := -\frac{1}{\log 2}\left|\log\left|\frac{M_n^{(2)}\left(\left[\frac{n}{2\log n}\right]\right)-2\left(M_n^{(1)}\left(\left[\frac{n}{2\log n}\right]\right)\right)^2}{M_n^{(2)}\left(\left[\frac{n}{\log n}\right]\right)-2\left(M_n^{(1)}\left(\left[\frac{n}{\log n}\right]\right)\right)^2}\right|\right|$$

by Peng (1998), where

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)^j, j \geq 1.$$

Note that $M_n^{(1)}(k)$ is the Hill estimator. $M_n^{(2)}(k)$ was introduced by Dekkers, Einmahl and de Haan (1989).

The following two estimators of ρ are special cases of a class of estimators proposed by Fraga Alves, Gomes and de Haan (2003). They are

$$\widehat{\rho}_3 := -\left|\frac{3\left(T_n^{(0)}(k_1)-1\right)}{T_n^{(0)}(k_1)-3}\right|$$

and

$$\widehat{\rho}_4 := -\left|\frac{3\left(T_n^{(1)}(k_1)-1\right)}{T_n^{(1)}(k_1)-3}\right|,$$

where $k_1 = min\left(n-1, [2n/\log\log n]\right)$, and

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^{\tau}-\left(M_n^{(1)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2}-\left(M_n^{(3)}(k)/6\right)^{\tau/3}} & \text{if } \tau>0 \\ \dfrac{\log\left(M_n^{(1)}(k)\right)-\frac{1}{2}\log\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\log\left(M_n^{(2)}(k)/2\right)-\frac{1}{3}\log\left(M_n^{(3)}(k)/6\right)} & \text{if } \tau=0 \end{cases}.$$

The estimator proposed by Gomes and Martins (2001) will be denoted by $\widehat{\rho}_5$. The instability measure will be incorporated, leading to an adjusted estimator $\widehat{\rho}_5^*$. Only the procedure to calculate $\widehat{\rho}_5$ will be stated here, together with some computational issues. For a full discussion of the estimator, refer to Gomes and Martins (2001).

Given a set of n positive, ordered observations $x_{1,n} \leq x_{2,n} \leq \ldots \leq x_{n,n}$ assumed to be from a second order regularly varying distribution, define the following functions:

- For given choices of α and k, let $\widehat{\gamma}_n^{(\alpha)}(k) = \frac{m_n^{(\alpha)}(k)}{\Gamma(\alpha+1)(H_{k,n})^{\alpha-1}}$, where $H_{k,n}$ denotes the Hill estimator, based on k excesses.
- For given choices of α, i_n and j_n, let $\chi(\alpha, i_n, j_n)$ be the sample median of $\widehat{\gamma}_n^{(\alpha)}(k)$ for $i_n \leq k \leq j_n$.
- For given choices of α, i_n and j_n, let $\Sigma(\alpha, i_n, j_n) = \sum_{k=i_n}^{j_n} \left(\widehat{\gamma}_n^{(\alpha)}(k) - \chi(\alpha, i_n, j_n) \right)^2$.

The procedure for estimating ρ is as follows:

1. Choose i_n and j_n which specify a range of k such that the estimates $\widehat{\gamma}_n^{(\alpha)}(k)$ are "stable" for $i_n \leq k \leq j_n$.
2. Obtain an estimate of α as $\widehat{\alpha} = \arg\min_\alpha \Sigma(\alpha, i_n, j_n)$.
3. Estimate ρ by $\widehat{\rho}$ as the solution of $(1-\widehat{\rho})^{\widehat{\alpha}-1}(1+\widehat{\rho}(\widehat{\alpha}-2)) = 1$. The restriction $\widehat{\alpha} \geq 2$ applies. In particular, the value of $\widehat{\alpha} = 2.09272$ corresponds to $\widehat{\rho} = -10$, and $\widehat{\alpha} = 15.02746$ to $\widehat{\rho} = -0.01$. Computationally, let $\widehat{\rho} = -10$, if $2 \leq \widehat{\alpha} \leq 2.09272$, and $\widehat{\rho} = -0.01$, if $\widehat{\alpha} \geq 15.02746$. For $2.09272 < \widehat{\alpha} < 15.02746$, $\widehat{\rho}$ is solved numerically.

Simulations showed that $\Sigma(\alpha, i_n, j_n)$ as a function of α will typically have one of three shapes. The first shape (not shown) sees $\Sigma(\alpha, i_n, j_n)$ decreasing as α increases. In such a case take $\widehat{\alpha} = 15$, yielding $\widehat{\rho} = -0.01$.

The second and third typical shapes are shown in the following graphs:

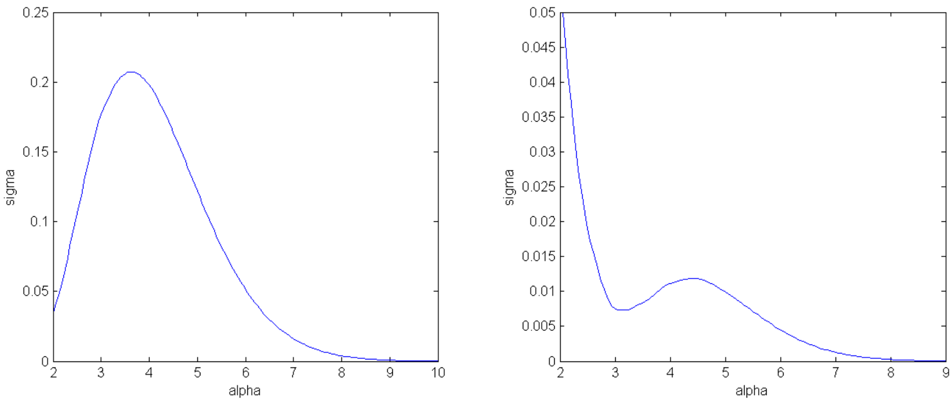

Figure 3: Typical behaviours of $\Sigma(\alpha, i_n, j_n)$ as a function of α.

In the case of the first graph of Figure 3 the value of $\Sigma(\alpha, i_n, j_n)$ at $\alpha = 2$ is not regarded as a local minimum. Take $\widehat{\alpha} = 15$, yielding $\widehat{\rho} = -0.01$. The second graph of Figure 3 shows the situation where the local minimum of $\Sigma(\alpha, i_n, j_n)$ needs to be determined numerically. Note that $\widehat{\gamma}_n^{(\alpha)}(k) \to 0$ (and consequently $\Sigma(\alpha, i_n, j_n) \to 0$) as $\alpha \to \infty$, and that $\widehat{\rho} = 0$ is always a (trivial) solution of $(1-\widehat{\rho})^{\widehat{\alpha}-1}(1+\widehat{\rho}(\widehat{\alpha}-2)) = 1$.

Gomes and Martins (2001) chose $i_n = 500$ and $j_n = 900$ for a Fréchet sample of size $n = 1000$ in their simulation study. Simulations performed in this study confirm that $i_n = 50\%$ of n and $j_n = 90\%$ of n is a good choice, robust over a wide variety of sample sizes and distributions as far as performance is concerned. The estimator $\widehat{\rho}_5$ will imply this choice of i_n and j_n.

The adjusted estimator $\widehat{\rho}_5^*$ chooses adaptively between three possible ranges which also covers 40% of the data, namely

1. $i_n = 10\%$ of n and $j_n = 50\%$ of n,
2. $i_n = 30\%$ of n and $j_n = 70\%$ of n, and
3. $i_n = 50\%$ of n and $j_n = 90\%$ of n.

The procedure for the adjusted estimator is the following:

1. Determine $\widehat{\alpha}_1$, $\widehat{\alpha}_2$ and $\widehat{\alpha}_3$, corresponding to ranges 1, 2 and 3 above, respectively.
2. Determine three sets of estimates: a set of $\widehat{\gamma}_n^{(\widehat{\alpha}_1)}(k)$ values, a set of $\widehat{\gamma}_n^{(\widehat{\alpha}_2)}(k)$ values, and a set of $\widehat{\gamma}_n^{(\widehat{\alpha}_3)}(k)$ values, where k assumes all integer values in ranges 1, 2 and 3, respectively.
3. Calculate the instability measure for each of the three sets of estimates.
4. Let $\widehat{\alpha}$ denote the estimated value of α which corresponds to the range associated with the lowest instability measure.
5. Solve for $\widehat{\rho}_5^*$ from $\left(1-\widehat{\rho}_5^*\right)^{\widehat{\alpha}-1}\left(1+\widehat{\rho}_5^*(\widehat{\alpha}-2)\right) = 1$.

The performance of these six estimators will be evaluated in the next section.

5. Simulation study

In this section the aim is to provide an objective and comprehensive study design, which can be used for all simulation studies concerning positive EVI estimation.

The families of distributions should be those for which the expressions for their respective EVIs and second order parameters are known. The simulation study did not include observations simulated from the limiting distributions, for example the Pareto distribution, the generalised Pareto distribution, the PPD and the generalised extreme value distribution.

A reasonable range for the EVI is from 0 to 1. The choice of the lower bound of ρ is not straightforward. Seventeen papers focussing on the estimation of ρ were reviewed, of which only seven show estimates of ρ for real (case study) data. Fraga Alves (2002) calculated estimates of ρ of -1.9973, -0.4989 and -1.0609 for fire insurance claim data. Fraga Alves (2002) also shows graphs of various estimators over a wide range of values of k. From those graphs it seems as though -2 is the lower limit for these estimates for that specific data set. Goegebeur and de Wet (2012) considered fire insurance claim data, and arrived at a median estimate of ρ of approximately -1.2. Gomes, Caeiro and Figueiredo (2004), estimated the value of ρ as -0.69 for exchange rate data. Gomes, Martins and Neves (2007) obtained an estimate of -0.65, also for exchange rate data. Gomes, Henriques-Rodrigues, Pereira and Pestana (2010) considered daily log-returns of indices and obtained $\widehat{\rho} = -0.72$ for the Dow Jones Industrial, and $\widehat{\rho} = -0.73$ for the NASDAQ. Gomes et al. (2010) studied males diagnosed with cancer of the tongue. Their data set yielded $\widehat{\rho} = -0.654$. Gomes et al. (2012) examined a data set consisting of number of hectares burnt during wildfires, and obtained an estimate $\widehat{\rho} = -0.39$.

For all the cases mentioned above, the estimates exceeded -2. This does not mean that it is impossible to have $\rho < -2$ for real data, but -2 seems to be a reasonable lower bound. As far as the upper limit is concerned, $\rho = 0$ should be included, since for some heavy-tailed distributions $\rho = 0$.

The resulting set of distributions is as follows:

- The Burr distribution with distribution function $F(x) = 1 - \left(\frac{\beta}{\beta + x^{\tau}}\right)^{\lambda}$, defined for $x > 0$, and with β, τ and λ positive. This Burr type XII distribution is also called the Singh-Maddala distribution (Singh and Maddala, 1976). Since the EVI $\gamma = 1/\lambda\tau$ and the second order parameter $\rho = -1/\lambda$, the Burr can be reparameterised in terms of β, γ and ρ. The Burr distributions considered are the *Burr*($\beta = 1, \gamma = 0.25, \rho = -2$), the *Burr*($\beta = 1, \gamma = 0.5, \rho = -2$), the *Burr*($\beta = 1, \gamma = 1, \rho = -2$), the *Burr*($\beta = 1, \gamma = 0.25, \rho = -0.5$), the *Burr*($\beta = 1, \gamma = 0.5, \rho = -0.5$) and the *Burr*($\beta = 1, \gamma = 1, \rho = -0.5$).

- The Fréchet distribution with distribution function $F(x) = \exp(-x^{-\alpha})$, defined for $x > 0$, and with $\alpha > 0$. For the Fréchet $\gamma = 1/\alpha$ and $\rho = -1$. Three Fréchet distributions are considered in this study, namely those with $\alpha = 1$, $\alpha = 2$ and $\alpha = 4$, respectively.

- The student t distribution. Since a variable following the t distribution can assume negative values, the distribution is inflected on the positive half-line. The resulting distribution function is $F(x) = \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \int_0^x \left(1 + \frac{w^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} dw$, defined for $x \geq 0$, and with $n > 0$. For the $|t|$ distribution $\gamma = 1/n$ and $\rho = -2/n$. The t distributions considered in this study are the $|t_1|$, $|t_2|$ and $|t_4|$ distributions.

- The loggamma distribution with distribution function $F(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_1^x w^{-\lambda-1} (\log w)^{\alpha-1} dw$, defined for $x > 1$, and with λ and α positive. For the loggamma distribution $\gamma = 1/\lambda$ and $\rho = 0$. Since $\alpha = 1$ yields the Pareto distribution, we take $\alpha = 2$. Three loggamma distributions will be considered, namely $\log\Gamma(\lambda = 1, \alpha = 2)$, $\log\Gamma(\lambda = 2, \alpha = 2)$ and $\log\Gamma(\lambda = 4, \alpha = 2)$.

The above yields fifteen distributions from four families: five distributions with $\gamma = 0.25$, five with $\gamma = 0.5$, five with $\gamma = 1$, four with $\rho = -2$, four with $\rho = -1$, four with $\rho = -0.5$, and three with $\rho = 0$.

Sample sizes of $n = 100$, 200, 500, 1000, 2000 and 5000 will be considered. As measure of error, the mean square error (MSE) will be used. In the tables which show results pertaining to estimators of the EVI, the values of the MSEs will be multiplied by a factor of 1000 to ease comparison.

All results shown in the following sections were based on 1000 samples from each combination of distribution and sample size. The standard errors of the MSEs were obtained by calculating 10000 bootstrap repetitions of the MSE.

The simulation study design described above was applied to the estimators of the second order parameter described in the previous section. Consistent with the results of Gomes and Martins (2002) it was found that (1) $\widehat{\rho}_1$ and $\widehat{\rho}_2$ do not perform well, (2) $\widehat{\rho}_4$ generally performs better than $\widehat{\rho}_3$ for small values of ρ, and (3) $\widehat{\rho}_3$ has the best performance overall (of the first four estimators).

The estimators $\widehat{\rho}_5$ and $\widehat{\rho}_5^*$ yield in general more accurate estimates than $\widehat{\rho}_3$ for large values of ρ, and outperform $\widehat{\rho}_3$ overall for $n > 100$. The best overall performer for $200 \leq n \leq 2000$ is $\widehat{\rho}_5$, and $\widehat{\rho}_5^*$ for $n = 5000$. More extensive simulations showed $\widehat{\rho}_5^*$ to be the best overall performer for sample sizes $n = 10000$, $n = 20000$ and $n = 50000$ as well.

Based on these results, only the estimators $\widehat{\rho}_3$, $\widehat{\rho}_5$ and $\widehat{\rho}_5^*$ will be considered as candidates for the external estimation of ρ.

6. The PPD maximum likelihood estimators

The simulation results referred to in this section were obtained by estimating the EVI by fitting the PPD to relative excesses using MLE, after externally estimating the second order parameter by using $\widehat{\rho}_3$, $\widehat{\rho}_5$ or $\widehat{\rho}_5^*$. The final result of this section is the construction of two estimators: one by applying $\widehat{\rho}_5$ for EVI estimation and $\widehat{\rho}_5^*$ for threshold selection, and one by applying $\widehat{\rho}_3$ for both EVI estimation and threshold selection. The second estimator is included to provide an alternative estimator which is far easier to implement, and still yields good results.

The simulation study design of Section 5 was applied. For each of the estimators $\widehat{\rho}_3$, $\widehat{\rho}_5$ or $\widehat{\rho}_5^*$ the EVI was estimated at 19 values of k (5%, 10%, …, 95% of n) for 1000 samples from each combination of sample size and distribution.

All possible combinations of threshold selection techniques were applied to the resulting estimates: Method 1 with region lengths 2 to 19, Method 2, Method 3 with upper limits 2 to 19, each with and without applying Algorithm 3.2. Note that Method 1 with region length 19 yields Method 0.

It was found that $\widehat{\rho}_5$ consistently delivers the lowest MSE for the estimated EVI, even for sample sizes $n = 100$ and $n = 5000$. This implies that even though an estimator of ρ yields the lowest MSE in terms of estimation of ρ, it does not necessarily perform the best in terms of estimation of the EVI when used as external estimator of ρ.

It was found that all methods across all combinations of sample size and distribution benefited from the application of Algorithm 3.2. Furthermore, Method 3 consistently yielded the best results, except for a couple of cases where Method 0 performed better.

The question of how to choose the optimal technique for a given sample size will now be addressed. The term *technique* denotes one of the 19 possibilities: Method 0 and Method 3 with each of the upper limits $2, ..., 19$. Initially it was assumed that the choice of technique will depend on the family of the underlying distribution, and specifically the respective EVI. The results showed however that the optimal technique is for all practical purposes independent of the family of the underlying distribution and the underlying EVI, and only a function of the underlying value of ρ.

This information is used to develop an estimator which adaptively chooses the appropriate technique. The idea is to determine $\widehat{\rho}$, and if $\widehat{\rho} < -1.5$ to use the optimal technique for distributions for which the underlying value of ρ is -2. If $-1.5 \leq \widehat{\rho} < -0.75$ use the optimal technique for $\rho = -1$. If $-0.75 \leq \widehat{\rho} < -0.25$ use the optimal technique for $\rho = -0.5$. If $\widehat{\rho} > -0.25$ use the optimal technique for $\rho = 0$.

Another surprising result was that, even though $\widehat{\rho}_5$ yields the lowest MSEs for EVI estimation, $\widehat{\rho}_5^*$ consistently performs better than $\widehat{\rho}_5$ when choosing between the four categories mentioned in the previous paragraph. The "best" estimator is achieved by using the value of $\widehat{\rho}_5^*$ to decide which technique to apply, and then applying the technique to the EVI estimates with $\widehat{\rho}_5$ as external estimator. The results in the table which follow are also shown for the estimator resulting from using $\widehat{\rho}_3$ to choose between the four categories, and applying the appropriate technique to the EVI estimates with $\widehat{\rho}_3$ as external estimator.

Table 1: Optimal techniques for threshold selection.

ρ	**n**	$\widehat{\rho}_5$,$\widehat{\rho}_5^*$	$\widehat{\rho}_3$	ρ	**n**	$\widehat{\rho}_5$,$\widehat{\rho}_5^*$	$\widehat{\rho}_3$
-2	100	19	19	-0.5	100	5	5
-2	200	19	19	-0.5	200	5	5
-2	500	17	17	-0.5	500	5	5
-2	1000	17	17	-0.5	1000	5	5
-2	2000	15	15	-0.5	2000	4	4
-2	5000	14	14	-0.5	5000	$\underline{3}$	$\underline{2}$
-1	100	19	19	0	100	0	0
-1	200	18	18	0	200	0	0
-1	500	16	16	0	500	$\underline{0}$	$\underline{9}$
-1	1000	14	14	0	1000	9	9
-1	2000	12	12	0	2000	6	6
-1	5000	10	10	0	5000	3	3

The numbers in the $\widehat{\rho}_5$,$\widehat{\rho}_5^*$ and $\widehat{\rho}_3$ columns refer to the upper limit for Method 3. A zero indicates Method 0. As an example, suppose $n = 5000$ and $\widehat{\rho}_5^* < -1.5$. The appropriate technique is therefore to be found in the section of the table where $\rho = -2$. Method 3 with an upper limit of 14 (corresponding to $k = 70\%$ of n) is applied to the EVI estimates obtained with $\widehat{\rho}_5$.

The upper limits were chosen in a way that it is as robust as possible with respect to the choice of external estimator, in this case only $\widehat{\rho}_3$ and $\widehat{\rho}_5$. No group MSE (average MSE for the group of distributions with the same value of ρ) obtained using the table differed by more than two standard errors from the actual observed minimum MSE. The two exceptions are underlined in Table 1. The first exception is for $\rho = -0.5$ and $n = 5000$. For such a large sample size one must bear in mind that there is a huge difference between $k = 10\%$ and $k = 15\%$ of n. The second exception sees $\widehat{\rho}_3$ making a quicker transition than $\widehat{\rho}_5$ from Method 0 to Method 3 (with an upper limit of 9) as the sample size increases.

In summary, the procedure for the proposed estimator, which will be denoted by $\widehat{\gamma}_5$, is as follows:

1. Calculate $\widehat{\rho}_5$ and $\widehat{\rho}_5^*$ as described in Section 4.

2. Calculate 19 estimates of the EVI (at $k = 5\%, 10\%, \ldots, 95\%$ of n) by fitting the PPD to the relative excesses using MLE, keeping ρ fixed at $\widehat{\rho}_5$ and maximising the log likelihood over $\gamma > 0$ (EVI) and $1/\widehat{\rho}_5 \leq c \leq 0.5$.

3. Choose the category for ρ. For $\widehat{\rho}_5^*$ in $(-\infty, -1.5)$, choose $\rho = -2$, in $[-1.5, -0.75)$, choose $\rho = -1$, in $[-0.75, -0.25)$, choose $\rho = -0.5$, and in $[-0.25, 0)$, choose $\rho = 0$.

4. Use the $\widehat{\rho}_5, \widehat{\rho}_5^*$ column of Table 1 to determine which technique to use.

5. Apply algorithms 3.1 and 3.2 to the 19 estimates, followed by the technique resulting from Step 4 to determine the optimal region.

6. The estimate is calculated as the mean of the original estimates (not rounded) in the optimal region.

The estimator $\widehat{\gamma}_3$ is defined in the same way, except that both $\widehat{\rho}_5$ and $\widehat{\rho}_5^*$ are replaced by $\widehat{\rho}_3$. This estimator does not perform as well as $\widehat{\gamma}_5$, but is considerably easier to program and computationally much less intensive.

Below are tables containing simulation results comparing the performance of EVI estimators in terms of MSE.

The effect of adaptive threshold selection was also investigated for the Hill estimator. The three threshold selection methods considered were those mentioned in the introduction, namely the method by Drees and Kaufmann (1998), the method by Guillou and Hall (2001), and the method by Beirlant et al. (2004). The method by Drees and Kaufmann (1998), yielding the estimator denoted by $\widehat{\gamma}_{DK}^{Hill}$, performed best overall and is the only method for which the results are shown. For a discussion of these three adaptive threshold selection methods, please refer to the respective papers.

The alternative to adaptive threshold selection is to simply fix k at some value prior to estimation. In this study the value of k was fixed at $k_0 = 2\sqrt{n}$ for the Hill estimator. This is the choice made by Drees and Kaufmann (1998) to obtain their initial estimate $\widehat{\gamma}_0^{Hill} = H_{2\sqrt{n},n}$. The performance of $\widehat{\gamma}_0^{Hill}$ is compared to that of $\widehat{\gamma}_{DK}^{Hill}$.

The value of $k_0 = 2\sqrt{n}$ performs well as a default fixed choice of k for the Hill estimator, as can be seen from the results below. A similar formula might be useful when fixing the value of k when fitting the PPD. In the case of the PPD k/n should tend to zero slower than in the first order case. A reasonable choice would be $k_0 = 2n^{2/3}$. It turns out that this choice closely models that which was observed in the simulation studies. The estimator $\widehat{\gamma}_0$ is defined as the estimate of the MLE of the EVI obtained by fitting the PPD to $k_0 = 2n^{2/3}$ relative excesses, with $\widehat{\rho}_5$ as external estimator of the second order parameter.

The simulation study results are shown in the tables below. In each row the minimum value of the MSE is underlined.

Table 2: 1000× MSEs for estimates of the EVI for sample size $n = 100$.

Dist	γ	ρ	$\widehat{\gamma}_0^{Hill}$	$\widehat{\gamma}_{DK}^{Hill}$	$\widehat{\gamma}_0$	$\widehat{\gamma}_3$	$\widehat{\gamma}_5$
Burr	0.25	-2	3.086	2.626	3.253	2.893	3.669
			(0.144)	(0.111)	(0.124)	(0.150)	(0.157)
Burr	0.25	-0.5	19.493	19.356	10.039	14.340	8.349
			(0.647)	(0.744)	(0.469)	(0.614)	(0.437)
Fréchet (4)	0.25	-1	3.299	3.376	3.196	2.787	3.746
			(0.168)	(0.356)	(0.127)	(0.158)	(0.165)
$\|t_4\|$	0.25	-0.5	36.255	30.310	20.352	15.796	10.818
			(0.977)	(1.053)	(0.589)	(0.769)	(0.549)
$\log\Gamma(4,2)$	0.25	0	8.890	10.348	4.726	6.129	4.319
			(0.432)	(1.342)	(0.210)	(0.282)	(0.176)
Mean	**0.25**		**14.205**	**13.203**	**8.313**	**8.389**	**6.180**
			(0.254)	(0.380)	(0.160)	(0.209)	(0.152)
Burr	0.5	-2	11.572	9.455	12.135	10.663	13.973
			(0.560)	(0.465)	(0.473)	(0.733)	(0.614)
Burr	0.5	-0.5	73.088	74.419	36.679	54.358	31.727
			(2.434)	(2.647)	(1.604)	(2.316)	(1.523)
Fréchet (2)	0.5	-1	14.676	13.405	14.912	13.479	16.562
			(0.810)	(0.622)	(0.534)	(0.692)	(0.656)
$\|t_2\|$	0.5	-1	24.513	27.493	12.968	24.968	25.179
			(1.108)	(1.163)	(0.768)	(1.106)	(1.147)
$\log\Gamma(2,2)$	0.5	0	36.270	38.995	21.113	27.481	19.343
			(1.589)	(2.234)	(0.906)	(1.255)	(0.779)
Mean	**0.5**		**32.024**	**32.754**	**19.562**	**26.190**	**21.357**
			(0.653)	(0.747)	(0.424)	(0.606)	(0.449)
Burr	1	-2	46.160	40.643	50.740	43.836	58.115
			(2.220)	(2.805)	(1.936)	(2.304)	(2.485)
Burr	1	-0.5	296.088	293.840	138.417	192.518	118.216
			(10.427)	(10.573)	(5.762)	(7.391)	(5.374)
Fréchet (1)	1	-1	58.165	65.676	51.091	49.477	59.510
			(3.393)	(6.556)	(1.972)	(2.589)	(2.497)
$\|t_1\|$	1	-2	53.228	47.448	57.310	54.341	72.719
			(2.447)	(2.142)	(2.240)	(2.690)	(2.858)
$\log\Gamma(1,2)$	1	0	138.476	161.175	79.031	97.364	67.987
			(6.442)	(9.671)	(3.318)	(4.003)	(2.696)
Mean	**1**		**118.423**	**121.756**	**75.318**	**87.507**	**75.309**
			(2.628)	(3.230)	(1.508)	(1.896)	(1.506)
Overall mean			**54.884**	**55.904**	**34.397**	**40.695**	**34.282**
			(0.907)	(1.112)	(0.525)	(0.667)	(0.526)

Table 3: 1000× MSEs for estimates of the EVI for sample size $n = 200$.

Dist	γ	ρ	$\widehat{\gamma}_0^{Hill}$	$\widehat{\gamma}_{DK}^{Hill}$	$\widehat{\gamma}_0$	$\widehat{\gamma}_3$	$\widehat{\gamma}_5$
Burr	0.25	-2	2.339	1.429	2.212	1.530	1.973
			(0.111)	(0.064)	(0.092)	(0.083)	(0.101)
Burr	0.25	-0.5	11.028	12.476	5.763	8.909	5.906
			(0.394)	(0.419)	(0.274)	(0.358)	(0.293)
Fréchet (4)	0.25	-1	2.183	1.739	2.374	1.560	2.367
			(0.101)	(0.080)	(0.099)	(0.086)	(0.105)
$\|t_4\|$	0.25	-0.5	20.427	20.038	10.414	9.918	7.008
			(0.570)	(0.630)	(0.315)	(0.414)	(0.291)
$\log\Gamma(4,2)$	0.25	0	6.297	7.190	4.375	5.199	3.683
			(0.268)	(0.549)	(0.168)	(0.180)	(0.128)
Mean	**0.25**		**8.455**	**8.574**	**5.028**	**5.423**	**4.187**
			(0.152)	(0.188)	(0.094)	(0.118)	(0.091)
Burr	0.5	-2	8.745	5.270	8.617	6.072	8.549
			(0.409)	(0.247)	(0.372)	(0.327)	(0.409)
Burr	0.5	-0.5	46.050	50.958	23.165	35.136	22.839
			(1.671)	(1.703)	(1.072)	(1.390)	(1.071)
Fréchet (2)	0.5	-1	10.503	8.011	9.785	7.502	10.075
			(0.509)	(0.383)	(0.383)	(0.384)	(0.440)
$\|t_2\|$	0.5	-1	13.974	15.850	9.710	15.646	15.575
			(0.708)	(0.714)	(0.738)	(0.740)	(0.738)
$\log\Gamma(2,2)$	0.5	0	26.677	27.398	18.727	21.408	16.279
			(1.093)	(0.822)	(0.716)	(0.727)	(0.604)
Mean	**0.5**		**21.190**	**21.497**	**14.001**	**17.153**	**14.663**
			(0.443)	(0.414)	(0.316)	(0.361)	(0.311)
Burr	1	-2	36.532	22.108	33.994	24.887	33.816
			(1.812)	(1.023)	(1.394)	(1.355)	(1.619)
Burr	1	-0.5	181.968	202.763	93.315	143.099	87.062
			(6.034)	(7.185)	(3.992)	(5.736)	(3.972)
Fréchet (1)	1	-1	39.442	32.050	39.533	28.668	39.296
			(2.016)	(1.368)	(1.588)	(1.398)	(1.759)
$\|t_1\|$	1	-2	34.808	26.771	42.175	30.064	44.497
			(2.058)	(1.289)	(1.909)	(1.634)	(2.002)
$\log\Gamma(1,2)$	1	0	98.580	107.525	69.386	79.382	59.001
			(4.368)	(3.606)	(2.628)	(2.731)	(2.130)
Mean	**1**		**78.266**	**78.243**	**55.681**	**61.220**	**52.735**
			(1.638)	(1.664)	(1.113)	(1.368)	(1.096)
Overall mean			**35.970**	**36.105**	**24.903**	**27.932**	**23.862**
			(0.568)	(0.575)	(0.387)	(0.473)	(0.381)

Table 4: 1000× MSEs for estimates of the EVI for sample size $n = 500$.

Dist	γ	ρ	$\widehat{\gamma}_0^{Hill}$	$\widehat{\gamma}_{DK}^{Hill}$	$\widehat{\gamma}_0$	$\widehat{\gamma}_3$	$\widehat{\gamma}_5$
Burr	0.25	-2	1.355	0.581	1.299	0.764	0.912
			(0.061)	(0.026)	(0.057)	(0.045)	(0.053)
Burr	0.25	-0.5	5.926	7.746	2.795	4.411	3.050
			(0.200)	(0.232)	(0.137)	(0.176)	(0.136)
Fréchet (4)	0.25	-1	1.433	0.854	1.522	0.872	1.232
			(0.071)	(0.037)	(0.069)	(0.049)	(0.062)
$\|t_4\|$	0.25	-0.5	10.815	11.716	5.005	6.064	3.918
			(0.309)	(0.357)	(0.203)	(0.250)	(0.179)
$\log\Gamma(4,2)$	0.25	0	4.530	5.406	3.551	4.278	3.077
			(0.191)	(0.134)	(0.141)	(0.116)	(0.105)
Mean	**0.25**		**4.812**	**5.260**	**2.835**	**3.278**	**2.438**
			(0.085)	(0.090)	(0.059)	(0.067)	(0.052)
Burr	0.5	-2	5.427	2.455	5.490	3.160	3.949
			(0.260)	(0.117)	(0.266)	(0.206)	(0.234)
Burr	0.5	-0.5	24.800	30.077	12.066	18.161	12.854
			(0.878)	(0.969)	(0.635)	(0.726)	(0.573)
Fréchet (2)	0.5	-1	6.002	3.704	6.362	3.769	4.888
			(0.267)	(0.149)	(0.267)	(0.199)	(0.234)
$\|t_2\|$	0.5	-1	7.675	8.255	6.591	8.505	8.604
			(0.346)	(0.314)	(0.428)	(0.432)	(0.405)
$\log\Gamma(2,2)$	0.5	0	18.454	22.668	15.466	17.604	13.109
			(0.697)	(0.522)	(0.564)	(0.452)	(0.409)
Mean	**0.5**		**12.472**	**13.432**	**9.195**	**10.240**	**8.681**
			(0.246)	(0.232)	(0.205)	(0.200)	(0.176)
Burr	1	-2	23.276	9.907	21.099	14.252	15.177
			(1.049)	(0.533)	(0.901)	(0.835)	(0.896)
Burr	1	-0.5	103.445	116.008	49.345	74.587	49.197
			(3.634)	(3.886)	(2.928)	(2.950)	(2.309)
Fréchet (1)	1	-1	22.443	17.245	24.628	15.934	20.271
			(0.992)	(0.760)	(1.159)	(0.831)	(0.994)
$\|t_1\|$	1	-2	22.916	20.853	30.080	15.524	20.996
			(1.111)	(9.036)	(1.775)	(0.842)	(1.098)
$\log\Gamma(1,2)$	1	0	73.957	87.196	59.608	68.296	51.872
			(2.972)	(2.149)	(2.194)	(1.818)	(1.633)
Mean	**1**		**49.207**	**50.242**	**36.952**	**37.719**	**31.503**
			(1.007)	(2.022)	(0.865)	(0.751)	(0.663)
Overall mean			**22.164**	**22.978**	**16.327**	**17.079**	**14.207**
			(0.347)	(0.679)	(0.297)	(0.260)	(0.229)

Table 5: 1000× MSEs for estimates of the EVI for sample size $n = 1000$.

Dist	γ	ρ	$\widehat{\gamma}_0^{Hill}$	$\widehat{\gamma}_{DK}^{Hill}$	$\widehat{\gamma}_0$	$\widehat{\gamma}_3$	$\widehat{\gamma}_5$
Burr	0.25	-2	1.009	<u>0.341</u>	0.833	0.445	0.437
			(0.045)	(0.015)	(0.034)	(0.024)	(0.023)
Burr	0.25	-0.5	4.176	5.690	<u>1.828</u>	2.908	1.972
			(0.144)	(0.151)	(0.092)	(0.111)	(0.084)
Fréchet (4)	0.25	-1	0.974	0.527	0.994	<u>0.510</u>	0.687
			(0.039)	(0.022)	(0.045)	(0.027)	(0.033)
$\lvert t_4 \rvert$	0.25	-0.5	6.640	8.049	2.534	3.911	<u>2.248</u>
			(0.194)	(0.217)	(0.095)	(0.139)	(0.092)
$\log\Gamma(4,2)$	0.25	0	3.200	4.717	2.886	3.476	<u>2.718</u>
			(0.122)	(0.085)	(0.114)	(0.081)	(0.084)
Mean	**0.25**		**3.200**	**3.865**	**1.815**	**2.250**	<u>**1.612**</u>
			(0.055)	(0.056)	(0.037)	(0.040)	(0.031)
Burr	0.5	-2	3.909	<u>1.317</u>	3.613	1.964	1.966
			(0.175)	(0.062)	(0.191)	(0.117)	(0.114)
Burr	0.5	-0.5	16.367	21.214	<u>8.016</u>	11.766	8.401
			(0.573)	(0.646)	(0.497)	(0.483)	(0.417)
Fréchet (2)	0.5	-1	4.032	2.479	4.111	<u>2.137</u>	2.848
			(0.187)	(0.100)	(0.179)	(0.108)	(0.128)
$\lvert t_2 \rvert$	0.5	-1	<u>4.948</u>	5.330	5.454	5.028	5.093
			(0.237)	(0.206)	(0.319)	(0.268)	(0.244)
$\log\Gamma(2,2)$	0.5	0	14.281	19.911	11.880	14.414	<u>11.461</u>
			(0.552)	(0.378)	(0.438)	(0.344)	(0.347)
Mean	**0.5**		**8.707**	**10.050**	**6.615**	**7.062**	<u>**5.954**</u>
			(0.174)	(0.157)	(0.156)	(0.134)	(0.124)
Burr	1	-2	14.867	<u>5.632</u>	13.485	8.039	8.193
			(0.636)	(0.251)	(0.560)	(0.437)	(0.479)
Burr	1	-0.5	60.773	74.966	<u>28.898</u>	43.199	30.549
			(2.091)	(2.424)	(1.532)	(1.700)	(1.342)
Fréchet (1)	1	-1	16.287	9.833	17.703	<u>9.077</u>	12.514
			(0.752)	(0.378)	(0.836)	(0.485)	(0.627)
$\lvert t_1 \rvert$	1	-2	15.082	<u>5.984</u>	20.733	8.443	9.886
			(0.705)	(0.277)	(1.191)	(0.507)	(0.568)
$\log\Gamma(1,2)$	1	0	54.459	78.858	48.269	57.814	<u>46.241</u>
			(2.094)	(1.538)	(1.789)	(1.368)	(1.420)
Mean	**1**		**32.294**	**35.055**	**25.818**	**25.314**	<u>**21.477**</u>
			(0.639)	(0.584)	(0.565)	(0.467)	(0.436)
Overall mean			**14.734**	**16.323**	**11.416**	**11.542**	<u>**9.681**</u>
			(0.222)	(0.202)	(0.196)	(0.162)	(0.152)

Table 6: 1000× MSEs for estimates of the EVI for sample size $n = 2000$.

Dist	γ	ρ	$\widehat{\gamma}_0^{Hill}$	$\widehat{\gamma}_{DK}^{Hill}$	$\widehat{\gamma}_0$	$\widehat{\gamma}_3$	$\widehat{\gamma}_5$
Burr	0.25	-2	0.672	<u>0.187</u>	0.543	0.250	0.251
			(0.028)	(0.008)	(0.022)	(0.015)	(0.015)
Burr	0.25	-0.5	2.748	4.195	1.294	2.089	<u>1.231</u>
			(0.094)	(0.102)	(0.087)	(0.075)	(0.060)
Fréchet (4)	0.25	-1	0.679	<u>0.333</u>	0.711	0.361	0.424
			(0.031)	(0.013)	(0.035)	(0.019)	(0.022)
$\lvert t_4 \rvert$	0.25	-0.5	4.596	5.877	1.476	2.999	<u>1.471</u>
			(0.137)	(0.146)	(0.068)	(0.097)	(0.064)
$\log\Gamma(4,2)$	0.25	0	2.720	4.399	2.393	2.857	<u>2.361</u>
			(0.092)	(0.062)	(0.088)	(0.061)	(0.064)
Mean	**0.25**		**2.283**	**2.998**	**1.283**	**1.711**	<u>**1.148**</u>
			(0.039)	(0.038)	(0.029)	(0.028)	(0.022)
Burr	0.5	-2	2.753	<u>0.771</u>	2.364	1.025	1.134
			(0.125)	(0.034)	(0.109)	(0.057)	(0.073)
Burr	0.5	-0.5	10.218	14.722	<u>4.498</u>	7.940	4.770
			(0.353)	(0.426)	(0.239)	(0.292)	(0.206)
Fréchet (2)	0.5	-1	3.110	<u>1.505</u>	2.895	1.583	1.814
			(0.137)	(0.056)	(0.148)	(0.081)	(0.095)
$\lvert t_2 \rvert$	0.5	-1	3.432	3.309	4.507	<u>2.832</u>	3.475
			(0.163)	(0.125)	(0.241)	(0.131)	(0.161)
$\log\Gamma(2,2)$	0.5	0	10.146	18.076	9.624	11.292	<u>9.239</u>
			(0.371)	(0.282)	(0.368)	(0.262)	(0.270)
Mean	**0.5**		**5.932**	**7.677**	**4.778**	**4.934**	<u>**4.087**</u>
			(0.114)	(0.106)	(0.107)	(0.085)	(0.079)
Burr	1	-2	11.327	<u>3.277</u>	9.841	4.821	5.127
			(0.511)	(0.155)	(0.465)	(0.306)	(0.361)
Burr	1	-0.5	41.164	50.202	19.393	30.540	<u>19.035</u>
			(1.459)	(1.606)	(1.229)	(1.158)	(0.911)
Fréchet (1)	1	-1	11.635	6.621	11.756	<u>6.362</u>	7.553
			(0.543)	(0.237)	(0.602)	(0.361)	(0.418)
$\lvert t_1 \rvert$	1	-2	12.535	<u>3.900</u>	16.609	6.132	7.106
			(0.538)	(0.201)	(0.917)	(0.374)	(0.432)
$\log\Gamma(1,2)$	1	0	43.365	73.104	<u>38.197</u>	47.759	38.903
			(1.596)	(1.144)	(1.410)	(1.008)	(1.059)
Mean	**1**		**24.005**	**27.421**	**19.159**	**19.123**	<u>**15.545**</u>
			(0.470)	(0.400)	(0.444)	(0.330)	(0.313)
Overall mean			**10.740**	**12.699**	**8.407**	**8.589**	<u>**6.926**</u>
			(0.162)	(0.139)	(0.152)	(0.114)	(0.108)

Table 7: 1000× MSEs for estimates of the EVI for sample size $n = 5000$.

Dist	γ	ρ	$\widehat{\gamma}_0^{Hill}$	$\widehat{\gamma}_{DK}^{Hill}$	$\widehat{\gamma}_0$	$\widehat{\gamma}_3$	$\widehat{\gamma}_5$
Burr	0.25	-2	0.460	<u>0.090</u>	0.410	0.118	0.129
			(0.020)	(0.004)	(0.025)	(0.009)	(0.010)
Burr	0.25	-0.5	1.670	2.808	0.692	1.168	<u>0.627</u>
			(0.056)	(0.061)	(0.043)	(0.046)	(0.030)
Fréchet (4)	0.25	-1	0.441	0.198	0.446	<u>0.182</u>	0.205
			(0.020)	(0.008)	(0.022)	(0.009)	(0.011)
$\lvert t_4 \rvert$	0.25	-0.5	2.444	3.621	<u>0.611</u>	1.532	0.671
			(0.070)	(0.083)	(0.038)	(0.062)	(0.029)
$\log\Gamma(4,2)$	0.25	0	2.260	3.862	2.047	2.337	<u>2.044</u>
			(0.074)	(0.044)	(0.069)	(0.047)	(0.048)
Mean	**0.25**		**1.455**	**2.116**	**0.841**	**1.067**	<u>**0.735**</u>
			(0.024)	(0.023)	(0.019)	(0.018)	(0.013)
Burr	0.5	-2	1.709	<u>0.327</u>	1.405	0.429	0.453
			(0.073)	(0.015)	(0.067)	(0.027)	(0.029)
Burr	0.5	-0.5	5.938	8.981	2.895	4.236	<u>2.561</u>
			(0.217)	(0.252)	(0.237)	(0.193)	(0.161)
Fréchet (2)	0.5	-1	1.808	0.922	1.805	<u>0.790</u>	0.866
			(0.082)	(0.033)	(0.097)	(0.041)	(0.047)
$\lvert t_2 \rvert$	0.5	-1	1.725	1.629	3.190	<u>1.512</u>	2.002
			(0.079)	(0.064)	(0.147)	(0.058)	(0.087)
$\log\Gamma(2,2)$	0.5	0	7.999	15.302	<u>7.392</u>	9.006	7.793
			(0.258)	(0.186)	(0.248)	(0.188)	(0.189)
Mean	**0.5**		**3.836**	**5.432**	**3.337**	**3.195**	<u>**2.735**</u>
			(0.073)	(0.064)	(0.078)	(0.056)	(0.054)
Burr	1	-2	7.560	<u>1.311</u>	6.261	2.140	2.180
			(0.338)	(0.058)	(0.261)	(0.154)	(0.159)
Burr	1	-0.5	24.769	31.584	10.124	16.946	<u>9.408</u>
			(0.851)	(0.914)	(0.565)	(0.682)	(0.446)
Fréchet (1)	1	-1	7.302	3.358	7.388	<u>3.184</u>	3.695
			(0.352)	(0.123)	(0.413)	(0.160)	(0.203)
$\lvert t_1 \rvert$	1	-2	6.657	<u>1.632</u>	9.254	2.341	2.582
			(0.292)	(0.073)	(0.624)	(0.130)	(0.155)
$\log\Gamma(1,2)$	1	0	33.432	59.177	<u>30.625</u>	35.967	31.001
			(1.102)	(0.782)	(1.044)	(0.719)	(0.744)
Mean	**1**		**15.944**	**19.412**	**12.730**	**12.116**	<u>**9.773**</u>
			(0.301)	(0.243)	(0.285)	(0.205)	(0.184)
Overall mean			**7.078**	**8.987**	**5.636**	**5.459**	<u>**4.415**</u>
			(0.103)	(0.084)	(0.099)	(0.071)	(0.064)

The following observations can be made:

1. All methods based on the fitting of the PPD perform significantly better than all methods based on the Hill estimator. This indicates that the choice of first or second order model outweighs all other concerns, at least for this limited number of estimators considered.

2. The threshold selection method by Drees and Kaufmann (1998) performs significantly worse than fixing the threshold at $2\sqrt{n}$, despite the fact that it is the best performing method of the three methods considered for the Hill estimator. The only exception is the distributions with an EVI of 0.25, for samples of size $n = 100$. This might be an indication that, if an estimator performs extremely poorly, the very underlying mechanics of an adaptive threshold selection method are undermined by the instability of the estimator. In such a case a reasonable choice of a fixed threshold may perform better on average.

3. The estimator $\widehat{\gamma}_0$ outperforms $\widehat{\gamma}_3$ consistently, except when the sample size and underlying EVI is large. In particular $\widehat{\gamma}_3$ performed better than $\widehat{\gamma}_0$ for $n = 1000$ and $n = 2000$ for the group with an EVI of 1, and for $n = 5000$ for groups with EVI 0.5 and 1. The fact that the estimator $\widehat{\gamma}_0$ outperforms $\widehat{\gamma}_3$ in general shows that threshold selection is of secondary importance relative to the choice of external estimator.

4. The estimator $\widehat{\gamma}_5$ has the best average performance for each combination of EVI group and sample size. The only two exceptions are the groups with EVI 0.5, for sample sizes of both $n = 100$ and $n = 200$, where the estimator of choice was $\widehat{\gamma}_0$.

Overall the results suggest that the most important contributing factors to the accuracy of estimation of the EVI are, in decreasing order of importance, the following:

1. Whether a first or second order model is fitted to the relative excesses.

2. The method of parameter estimation, for a given threshold. This refers to the choice of external estimator of the second order parameter, as well as to limits on the ranges of the parameters.

3. The method of threshold selection.

Concluding remarks

Frequently a situation is encountered where an estimator of the parameter of interest depends on the choice of some nuisance parameter. More often than not suboptimal ranges of the nuisance parameter exist, within which the estimates of the parameter of interest are extremely sensitive to the slightest change in the value of the nuisance parameter. Failure to avoid these problem ranges may have severe adverse effects on estimation accuracy. This paper proposes a new approach to circumnavigate this problem to a large extent by defining a measure which quantifies the instability of an observed quantity over a range of chosen values of another parameter.

Two applications of the proposed approach were considered, both in an extreme value analysis context. One application was to refine the second order parameter estimator proposed by Gomes and Martins (2001), by using the instability measure to distinguish between possible ranges of estimates. An improvement in the accuracy of second order parameter estimation was obtained for large samples, specifically when the sample size exceeds 5000. More importantly, the adjusted estimator performed better in the ability to choose between different threshold selection techniques, based on the range of the underlying value of ρ.

The main application of the instability measure involved threshold selection when estimating the EVI by fitting the PPD to relative excesses. This was accomplished by determining a region where the EVI estimates can be regarded as most stable. It led to the construction of the estimator $\widehat{\gamma}_5$, which shows significant improvement on the corresponding estimator which uses a fixed threshold.

The rest of the discussion will focus on some possible applications of the results presented in this paper. Financial risk managers need to estimate, inter alia, the value at risk (VaR) of portfolios, which is the level below which the future portfolio will drop with a specified small probability. EVT is used to calculate the maximum (extreme) losses that can occur in a given time period. Consider again the results at the end of Section 3, where it was found that for a data set of size $n = 200$ the EVI was estimated as 0.4727, yielding an implied threshold of $k = 50$. The estimated value of ρ (determined externally), as well as the MLEs of γ and c when fitting the PPD to the 50 excesses, yields a fitted model for observations exceeding the threshold. Estimated values of extreme quantiles can be obtained from this model for a given small exceedance probability. If the data represented insurance claims, for example, the corresponding extreme claim amount can be estimated. The accuracy of this estimated claim amount depends heavily on the accuracy of the estimated EVI. The techniques presented in this paper improve on EVI estimation by proposing an improved method of threshold selection.

There are also several possible areas of application of the instability measure outside the realm of EVT. Generally the variance is used as a measure of stability. The main advantage of the proposed instability measure is that it also takes into account the slope of the estimates over the specified region. It therefore does not only detect high variability but also systematic change.

The following are some examples of possible areas of application. Many applications are simply assessing stability of estimates over time, where there is no nuisance parameter which has to be chosen.

- In statistical process control (SPC), both the variance and mean of a production process need to be in-control for the process to run effectively. The instability measure can be applied here as an additional measure which needs to be in-control. Alternatively, some threshold can be decided upon which, if exceeded by the instability measure, can be used to flag the possibility of future out-of-control signals.

- The amount of health data measured on each individual is expected to increase exponentially in the near future with the availability of applications on electronic hand-held devices which record these data automatically. A similar concept to that of the application of the measure in SPC can be applied here to give early warning signals if measurements pertaining to health are becoming unstable.

- Stacy, Guarino, Reckase and Wooldridge (2013) investigate in their working paper the effect of student characteristics on the value-added estimates of teachers. The purpose of these estimates is to evaluate teacher performance, but inter-year instability of estimates is a problem. The problem is partially solved by taking into account in the regression model various variables which explain the variance in the teachers' performance, and also by calculating moving averages. However, the authors warn that care should be taken when applying these measures. In this setting the proposed instability measure can easily be applied to a large data base in order to identify estimates which are unstable over time. The measure also takes systematic change into account. The cases with the most extreme instability can be identified timeously and objectively. Investigating these cases can lead to new insights as to the causes of unstable performance estimates. These insights can be used to assist with policy decisions, or to help identify additional variables which should be included in the model.

Acknowledgements

The author would like to express his sincere gratitude towards the following persons:

- Professor Tertius de Wet, for his invaluable contributions, specifically his advice and suggestions during the research leading up to this paper, as well as for his long-standing guidance and support.

- Professor Yuri Goegebeur, for his constructive criticism and advice which led to a major revision of the content and presentation of the material.

- The reviewer and associate editor, for their recommendations and suggestions which enhanced the quality of this paper.

References

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., AND TEUGELS, J. L. (2004). *Statistics of Extremes*. Chichester: John Wiley & Sons.

COLES, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

DE HAAN, L. AND FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Boston: Springer.

DEKKERS, A. L. M., EINMAHL, J. H. J., AND DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, **17**, 1833–1855.

DIERCKX, G. (2000). *Estimation of the Extreme Value Index*. Phd thesis, Leuven: Katholieke Universiteit Leuven.

DREES, H. AND KAUFMANN, E. (1998). Selecting the optimal sampling fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, **75**, 149–172.

FRAGA ALVES, M. I. (2002). Estimation of first and second order parameters in heavy tails. Technical report, Lisbon: University of Lisbon.

FRAGA ALVES, M. I., GOMES, M. I., AND DE HAAN, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliæ Mathematica*, **60** (2), 193–213.

GELUK, J., DE HAAN, L., RESNICK, S., AND STĂRICĂ, C. (1997). Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications*, **69**, 139–159.

GOEGEBEUR, Y. AND DE WET, T. (2012). Local estimation of the second order parameter in extreme value statistics and local unbiased estimation of the tail index. *Communications in Statistics - Theory and Methods*, **41**, 1–33.

GOMES, M. I., CAEIRO, F., AND FIGUEIREDO, F. (2004). Bias reduction of a tail index estimator through an external estimation of the second-order parameter. *Statistics: A Journal of Theoretical and Applied Statistics*, **38** (6), 497–510.

GOMES, M. I., FIGUEIREDO, F., AND NEVES, M. M. (2012). Adaptive estimation of heavy right tails: resampling-based methods in action. *Extremes*, **15** (4), 463–489.

GOMES, M. I. AND HENRIQUES-RODRIGUES, L. (2010). A heuristic choice of tuning parameters in a location invariant reduced-bias estimation of the extreme value index: application to financial log-returns and simulated data. *In Proceedings of the ITI 2010 32nd International Conference on Information Technology Interfaces*. Cavtat, Croatia, 21–24 June, p. 527.

GOMES, M. I., HENRIQUES-RODRIGUES, L., PEREIRA, H., AND PESTANA, D. (2010). Tail index and second-order parameters' semi-parametric estimation based on the log-excesses. *Journal of Statistical Computation and Simulation*, **80** (6), 653–666.

GOMES, M. I. AND MARTINS, M. J. (2001). Generalizations of the Hill estimator - asymptotic versus finite sample behaviour. *Journal of Statistical Planning and Inference*, **93**, 161–180.

GOMES, M. I. AND MARTINS, M. J. (2002). "asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes*, **5** (1), 5–31.

GOMES, M. I., MARTINS, M. J., AND NEVES, M. M. (2007). Improving second order reduced bias extreme value index estimation. *REVSTAT Statistical Journal*, **5** (2), 117–207.

GUILLOU, A. AND HALL, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society*, **63**, 293–305.

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

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

PENG, L. (1998). Asymptotically unbiased estimator for the extreme-value index. *Statistics and Probability Letters*, **38** (2), 107–115.

SINGH, S. K. AND MADDALA, G. S. (1976). A function for the size distribution of incomes. *Econometrica*, **44**, 963–970.

STACY, B., GUARINO, C. M., RECKASE, M. D., AND WOOLDRIDGE, J. M. (2013). Does the precision and stability of value-added estimates of teacher performance depend on the types of students they serve? *IZA Discussion Paper*, **7676**.

Manuscript received, 2013-11-26, revised, 2014-04-14, accepted, 2014-04-28.