Springer

Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.springerlink.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Threshold selection for extremes under a semiparametric model	
Article Sub-Title		
Article CopyRight	Springer-Verlag Berlin Heidelberg (This will be the copyright line in the final PDF)	
Journal Name	Statistical Methods & Applications	
Corresponding Author	Family Name	**Sued**
	Particle	
	Given Name	**Mariela**
	Suffix	
	Division	
	Organization	Universidad de Buenos Aires, Ciudad Universitaria
	Address	Intendente Giraldes 2160, Buenos Aires, C1428EGA , Argentina
	Division	
	Organization	CONICET
	Address	Buenos Aires, Argentina
	Email	msued@dm.uba.ar
Author	Family Name	**Gonzaleza**
	Particle	
	Given Name	**Juan**
	Suffix	
	Division	
	Organization	Universidad de Buenos Aires, Ciudad Universitaria
	Address	Intendente Giraldes 2160, Buenos Aires, C1428EGA , Argentina
	Email	
Author	Family Name	**Rodrigueza**
	Particle	
	Given Name	**Daniela**
	Suffix	
	Division	
	Organization	Universidad de Buenos Aires, Ciudad Universitaria
	Address	Intendente Giraldes 2160, Buenos Aires, C1428EGA , Argentina
	Division	
	Organization	CONICET
	Address	Buenos Aires, Argentina
	Email	
Schedule	Received	
	Revised	
	Accepted	26 June 2013

Abstract In this work we propose a semiparametric likelihood procedure for the threshold selection for extreme values. This is achieved under a semiparametric model, which assumes there is a threshold above which the excess distribution belongs to the generalized Pareto family. The motivation of our proposal lays on a particular

characterization of the threshold under the aforementioned model. A simulation study is performed to show empirically the properties of the proposal and we also compare it with other estimators.

Keywords (separated by '-')	Extreme values theory - Fisher consistency - Semiparametric models - Threshold selection
Footnote Information	

Journal: 10260
Article: 234

Author Query Form

Springer
the language of science

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details required	Author's response
1.	Please check and confirm that the authors and their respective affiliations have been correctly identified and amend if necessary.	
2.	Please provide caption to Figure 1.	
3.	Please check and confirm the inserted citation of Fig. 5 and Table 12 is correct. If not, please suggest an alternative citation. Please note that figures and tables should be cited in sequential order in the text.	

Stat Methods Appl
DOI 10.1007/s10260-013-0234-7

Threshold selection for extremes under a semiparametric model

Juan Gonzaleza · Daniela Rodrigueza · Mariela Sued

Accepted: 26 June 2013

Abstract In this work we propose a semiparametric likelihood procedure for the threshold selection for extreme values. This is achieved under a semiparametric model, which assumes there is a threshold above which the excess distribution belongs to the generalized Pareto family. The motivation of our proposal lays on a particular characterization of the threshold under the aforementioned model. A simulation study is performed to show empirically the properties of the proposal and we also compare it with other estimators.

Keywords Extreme values theory · Fisher consistency · Semiparametric models · Threshold selection

1 Introduction

In the context of extreme values theory, the concept of threshold selection has different meanings. For instance, to estimate the extreme value index γ for a Pareto type distribution, in his seminar paper Hill (1975) proposed to use observations above *a threshold* u_0. Ever since, many others such as Guillou and Hall (2001), also refer to the threshold as the number k of largest order statistics that should be used to estimate γ to control bias and variance. These two problems, are among those known as threshold selection.

J. Gonzaleza · D. Rodrigueza · M. Sued (✉)
Universidad de Buenos Aires, Ciudad Universitaria, Intendente Giraldes 2160,
C1428EGA Buenos Aires, Argentina
e-mail: msued@dm.uba.ar

D. Rodrigueza · M. Sued
CONICET, Buenos Aires, Argentina

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

On the other hand, Pickands (1975) proved that if the underlying distribution F satisfies the extremal limit principle, then the excess distribution above u can be approximated by a Generalized Pareto distribution, when u increases to infinity.

In the context of data treatment, there is an interest in determining the threshold value u for which the approximation becomes reliable. Several graphical methods have been developed to solve this problem. One way of choosing the threshold is through the Mean Residual Plot, see for example Embrechts et al. (1997). Another possibility simply consists in choosing a high percentile of the distribution as in DuMouchel (1983). Behrens et al. (2004) proposed a Bayesian method for threshold selections, while Cabras and Morales (2007) introduced the idea that extreme observations can be considered outliers of a specified parametric model. Recently, MacDonald et al. (2011) considered the same mixture model we will introduce here, but their inference is done under a Bayesian framework. Finally, Wong and Li (2010) also considered a mixture model where both densities (above and below the threshold) are modeled in a parametrical way.

Once the threshold is chosen, the data above its value are used to estimate the parameters of the tail distribution, the extreme value index being one of them. With this problem on mind, Beirlant et al. (1996) suggested to consider a threshold which minimizes the bias and the variance of the model. Methods based on resample or bootstrap were developed by Hall (1990), and Gomes and Oliveira (2001). Drees and Kaufmann (1998) presented a sequential procedure based on the law of the iterated logarithm. An exhaustive summary touching upon these and other methods can be found in Coles (2001) and Beirlant et al. (2004a,b).

As pointed out by Davison and Smith (1990) and Coles and Tawn (1994), to properly estimate the threshold is not an easy task. If the threshold chosen is too high, only a few observations will be used to estimate the tail of the distribution, increasing the variance of the estimator. On the other hand, small values for the threshold will lead to biased estimators. Therefore, an adequate threshold must achieve a proper balance between the variance and the bias of tail estimators. Several authors have studied the influence of the choice of the threshold on the parameter estimation of the tail of the distribution (see for example, Smith 1987; Frigessi et al. 2003; Coles and Tawn 1996; Coles and Powell 1996).

Most of these sample methods are designed to determine a value that, in general, is not even well defined for the entire population, since there is not a proper definition of the threshold for the distribution F generating the data. To overcome this limitation, in this work we introduce a model where F_u, the excess distribution above u, belongs to the generalized Pareto family for u big enough. The stability property of this family allows to define the threshold as the smallest value for which the excess distribution is a generalized Pareto distribution. Thus, the threshold is defined for each distribution in the model. Once this has been established, we provide a characterization of the threshold which is used to estimate it from a random sample.

The paper is organized as follows. Section 2 summarizes the main properties related to the excess distribution and the generalized Pareto family. In Sect. 3, we introduce the model while some theoretical properties are studied in Sect. 4. The estimation procedures are introduced in Sect. 5, followed by an improved version which captures the real nature of the threshold. Section 6 includes a Monte Carlo study designed to

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

evaluate the behavior of the proposed estimators. In Sect. 7, we compare our estimators with other methods developed in the literature.

2 Preliminaries

Given a random variable X distributed according to F, for each $u \in \mathbb{R}$, the excess distribution $F_u(y)$ is defined by

$$F_u(y) = P(X \leq y + u | X > u), \quad \text{for } y \geq 0. \tag{1}$$

In this way, we get that F_u is the conditional distribution of $X - u$ given that X is bigger than u: $X - u | X > u \sim F_u$. In extreme values theory it is known that, under certain conditions, as u increases to infinity, F_u can be approximated by a generalized Pareto distribution. The generalized Pareto family (G.P.F), denoted from now on by $\mathcal{H}$, is a parametric model indexed by $\theta = (\sigma, \gamma)$, with $\sigma > 0$ and $\gamma \in \mathbb{R}$, while for each parameter the distribution function $H_{\sigma,\gamma}$ is given by

$$H_{\sigma,\gamma}(z) = 1 - \left(1 + \frac{z\gamma}{\sigma}\right)_+^{-1/\gamma}, \quad z > 0$$

for $\gamma \neq 0$, while $H_{\sigma,0}(z) = 1 - e^{-\frac{z}{\sigma}}$, for $z > 0$; i.e. the exponential family is included in the G.P.F.

The generalized Pareto family satisfies the so called stability property. Namely, it states that if for some u_0 the excess distribution F_{u_0} belongs to $\mathcal{H}$, then the same holds for any $u > u_0$: if $F_{u_0} = H_{\sigma_0,\gamma}$, for some $\sigma_0 > 0$ and $\gamma \in \mathbb{R}$, then for all $u \geq u_0$ we get that $F_u = H_{\sigma(u),\gamma}$, where $\sigma(u) = \sigma_0 + \gamma(u - u_0)$.

Pickands (1975) proved that if the underlying distribution F satisfies the extremal limit principle, then the excess distribution can be approximated by a Pareto distribution. More precisely, let $(X_i)_{i \geq 1}$ be a sequence of independent and identically distributed (i.i.d.) variables with common distribution F. Assume that there exist normalizing constants $b_n > 0$ and $a_n \in \mathbb{R}$ such that the sequence $b_n^{-1}(\max\{X_1, \ldots X_n\} - a_n)$ converges in distribution to an extreme value distribution G_γ with the extreme index γ, defined by

$$G_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\} & \text{for } \gamma \neq 0,\ 1 + \gamma x > 0, \\ \exp\{-\exp(-x)\} & \text{for } \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

Thus the excess distribution can be approximated by a Pareto distribution, in the following way: there exists $\sigma(u)$ such that

$$\lim_{u \to x^*} \sup_{0 \leq x < \infty} \left| F_u(x) - H_{\sigma(u),\gamma}(x) \right| = 0,$$

where γ is the extreme value index associated to F and x^* is its right boundary: $x^* := \sup\{x : F(x) < 1\}$.

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

As we mentioned in the Introduction, we will propose here a model defined by those distributions for which there exists a threshold value u_0, above which the excess distribution belongs to the generalized Pareto family.

3 The model

In this work, all distributions to be considered have a density function, and we use f_u to denote the density associated to F_u. We denote by $\mathcal{M}$ the set of distribution functions on $\mathbb{R}$ having density f, for which there exists u such that the excess distribution F_u belongs to the generalized Pareto family. More precisely, there exist $\sigma(u)$ and γ such that $F_u = H_{\sigma(u),\gamma}$. The stability property of this family guarantees that for any $\tilde{u} > u$, $F_{\tilde{u}}$ is also in the G.P.F. Then, if $F \in \mathcal{M}$ we have that $F_u \in \mathcal{H}$ for all u big enough. For $F \in \mathcal{M}$, its threshold will be defined as the smallest value u for which F_u belongs to $\mathcal{H}$. This definition requires the following consideration.

Lemma 1 *Take $F \in \mathcal{M}$. Then,*

$$\inf\{u : F_u \in \mathcal{H}\} \;=\; \min\{u : F_u \in \mathcal{H}\}.$$

Proof Let $u^* = \inf\{u : F_u \in \mathcal{H}\}$. To prove that the infimum is attained, consider $u_n \downarrow u^*$, with $F_{u_n} = H_{\sigma(u_n),\gamma}$. Note that the excess distributions are continuous, meaning that $F_{u_n}(y) \to F_{u^*}(y)$ for all $y \geq 0$. Now, if $\gamma = 0$, we get that F_{u_n} is an exponential distribution with the same parameter λ for any u_n and so, F_{u^*} is also an exponential distribution with parameter λ, implying that F_{u^*} belongs to $\mathcal{H}$, as we wanted to prove. For $\gamma > 0$, the stability property of the G.P.F. also implies that $\sigma(u_n) = \sigma(u_{n+1}) + \gamma(u_n - u_{n+1})$, which guarantees that $\sigma(u_n)$ is a decreasing sequence of positive numbers. Let σ^* denote its limit. If $\sigma^* = 0$, we get that F_{u^*} is the distribution corresponding to the measure concentrated at zero, and so F does not have a density. If $\sigma^* > 0$ we get that $F_{u^*} = H_{\sigma^*,\gamma}$. When $\gamma < 0$ we get that $\sigma(u_n)$ is an increasing function that can not diverge because F has a density function, and so we also get that $F_{u^*} = H_{\sigma^*,\gamma}$, with $\sigma^* = \lim \sigma(u_n)$. □

Now, we are ready to define the threshold, for any distribution in the model $\mathcal{M}$, as follows:

Definition 2 For $F \in \mathcal{M}$, the threshold $u_0(F)$ is defined by

$$u_0(F) \;=\; \min\{u : F_u \in \mathcal{H}\}\,. \tag{2}$$

We use $\theta_0 = (\sigma_0, \gamma_0)$ to denote the parameter associated with the excess distribution at the threshold, meaning that

$$F_{u_0(F)} = H_{\theta_0}\,. \tag{3}$$

The aim of this work is obtain an estimator of the threshold $u_0(F)$ given in the previous definition, based on a sample $X_1, \ldots, X_n$ i.i.d. distributed according to F.

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

4 Some theoretical considerations

From now on, we use $E_F[l(X)]$ and $E_f[l(X)]$ to denote the expected value of $l(X)$ when X is distributed according to F or has density f, respectively. Before presenting our proposal for the estimation of the threshold, we need to introduce some objects. For $\theta = (\sigma, \gamma)$, let h_θ denote the density function of H_θ. Now, for any distribution G (not necessarily in $\mathcal{M}$) with density g and for each u, we choose $\theta_0(u, G)$ in such a way that the distribution $H_{\theta_0(u,G)}$ is the closest element in $\mathcal{H}$ to the excess distribution G_u, minimizing the Kullback–Leibler (KL) divergence. More precisely, if g_u denotes the excedent density, $\theta_0(u, G)$ satisfies

$$K(h_{\theta_0(u,G)}, g_u) \leq K(h_\theta, g_u) \ \forall\, \theta,$$

where for any pair of densities g and f, the Kullback–Leibler divergence between g and f is given by

$$K(g, f) = \int_{\mathbb{R}} \ln\left(\frac{f(x)}{g(x)}\right) f(x)dx = E_f\left[\ln\left(f(X)\right)\right] - E_f\left[\ln\left(g(X)\right)\right]. \qquad (4)$$

This formula guarantees that $\theta_0(u, G)$ can also be characterized as

$$\theta_0(u, G) \ = \underset{\theta}{\operatorname{argmax}} \ E_G\left[\ln h_\theta(W - u) | W > u\right]. \qquad (5)$$

A second tool to be considered is the application of $T(u, \theta, g)$ that is defined as follows. Given a density g, $u \geq 0$ and θ, $T(u, \theta, g)$ is a new density which is equal to g below u, but above u it has tail density in the G.P.F. with parameter θ, i.e.

$$T(u, \theta, g)(x) \ = \ g(x) I_{\{x \leq u\}} + \{1 - c_u(g)\} h_\theta(x - u) I_{\{x > u\}}\ , \qquad (6)$$

where

$$c_u(g) \ = \int_{-\infty}^{u} g(x)\ dx.$$

Note that if W is a random variable with density g, $T(u, \theta, g)$ can also be considered as a convex combination of two density functions with disjoint support, in the following way:

$$T(u, \theta, g)(x) = P(W \leq u)\frac{1}{c_u(g)} g(x) I_{\{x \leq u\}} + \{1 - P(W \leq u)\} h_\theta(x - u) I_{\{x > u\}}. \qquad (7)$$

We have introduced all these objects to get the following Lemma, which proof is straightforward.

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

Lemma 3 *Consider $F \in \mathcal{M}$, with density f. Let $u_0(F)$ be its threshold and $\theta_0 = (\sigma_0, \gamma_0)$ the threshold parameter, both introduced at Definition 2. Then,*

1. *$F_u = H_{\theta_0(u,F)}$, for $u \geq u_0$, with $\theta_0(u, F) = (\sigma_0 + \gamma_0(u - u_0), \gamma_0)$*
2. *$T(u, \theta_0(u, F), f) = f$ for all $u \geq u_0(F)$.*
3. *Moreover, by definition of $u_0(F)$, if for some θ we get that $T(u, \theta, f) = f$, then $u \geq u_0$ and $\theta = \theta_0(u, F)$.*

Now, given a random variable X with density f, consider the function

$$\mathcal{F}(u) = E_f \left[\ln T(u, \theta_0(u, F), f)(X)\right] \ . \tag{8}$$

As we mentioned, for $F \in \mathcal{M}$, $T(u, \theta_0(u, F), f) = f$ for all $u \geq u_0$ and so $\mathcal{F}(u) = \mathcal{F}(u_0)$. We will see that $\mathcal{F}(u) < \mathcal{F}(u_0)$ for $u < u_0(F)$, meaning that the behavior of $\mathcal{F}$ changes drastically at $u_0(F)$.

Lemma 4 *If $F \in \mathcal{M}$, then $\mathcal{F}(u) = \mathcal{F}(u_0)$ for $u \geq u_0(F)$ and $\mathcal{F}(u) < \mathcal{F}(u_0)$ for $u < u_0(F)$.*

Proof We already proved that $\mathcal{F}(u) = \mathcal{F}(u_0)$ for $u \geq u_0(F)$. In order to study the behavior of $\mathcal{F}(u)$ for $u < u_0$, note that

$$\mathcal{F}(u_0) - \mathcal{F}(u) = E_f \left[\ln (f(X))\right] - \mathcal{F}(u) = K\left(T(u, \theta_0(u, F), f), f\right)$$

and so

$$\mathcal{F}(u_0) - \mathcal{F}(u) \geq 0,$$

since $K\left(T(u_0, \theta_0(u_0(F), F), f), f\right) \geq 0$. If $\mathcal{F}(u_0) = \mathcal{F}(u)$ for some $u < u_0$, we conclude that the Kullback–Leibler divergence between $T(u, \theta_0(u, F), f)$ and f is zero and so, $T(u, \theta_0(u, F), f), f) = f$, contradicting the definition of $u_0(F)$. □

From the previous result, we get that the functional $M(F)$ defined by

$$M(F) = \min \left\{u : \mathcal{F}(u) = \max_s \mathcal{F}(s)\right\} \ , \tag{9}$$

is Fisher consistent under the model $\mathcal{M}$ for $u_0(F)$, meaning that $M(F) = u_0(F)$, for all $F \in \mathcal{M}$. This fact suggests that $u_0(F)$ can be estimated through an empirical version of $M(F)$. Therefore, if $\mathcal{F}_n$ is an empirical version of $\mathcal{F}$, we can consider the smallest value that maximizes $\mathcal{F}_n$ as an estimator $u_0(F)$.

5 Proposal for the estimation of the threshold

We can now undertake our first attempt to estimate $u_0(F)$ under the model $\mathcal{M}$, based on the results and ideas studied in the previous section. We start by constructing $\mathcal{F}_n(u)$, an empirical version of $\mathcal{F}$, using plug-in estimators and replacing expected values by

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

averages. More precisely, we will consider a nonparametric estimator of f and the maximum likelihood estimator (MLE) of $\theta_0(u, F)$. This estimation procedure can be described in the following steps.

Step 1: Denote by $\widehat{f}$ a non parametric density estimator of f, based on $X_1, \ldots, X_n$.
Step 2: For each u, consider

(a) $\widehat{f}_{1u}(x) = \frac{\widehat{f}(x)1_{\{x\le u\}}}{\widehat{c_u}}$ where $\widehat{c_u} = c_u(\widehat{f}) = \int_{-\infty}^{u} \widehat{f}(x)dx$.

(b) $\widehat{p}_u = \frac{\sum_{i=1}^{n} 1_{\{X_i \le u\}}}{n}$.

(c) $\widehat{\theta}_u$ the maximum likelihood estimator under the G.P.F., based on $X_i - u$, with $X_i > u$, which is a consistent estimator of $\theta_0(u, F)$.

Step 3: For each u, we estimate $T(u, \theta_0(u, F), f)$ with $\widehat{T}_u$, given by

$$\widehat{T}_u(x) \;=\; \widehat{p}_u \widehat{f}_{1u}(x) + (1 - \widehat{p}_u) h_{\widehat{\theta}_u}(x - u) 1_{\{x>u\}}.$$

Step 4: Let $\widehat{\mathcal{F}}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \ln \widehat{T}_u(X_i)$.
Step 5: In the first attempt to estimate the threshold we consider the first point where $\widehat{\mathcal{F}}_n$ attains its maximum, i.e.

$$\tilde{u}_0 = \inf\left\{u : \widehat{\mathcal{F}}_n(u) = \max_s \widehat{\mathcal{F}}_n(s)\right\}. \tag{10}$$

Remark 5 In Step 1, we can consider any consistent nonparametric density estimator, such as a kernel type estimator, splines, orthogonal series, among others. Both in the simulation study as in the comparisons with other methods, we will consider kernel type estimates introduced by Rosenblatt (1956) and Parzen (1962). It is well known that the performance of kernel density estimators depends crucially on the value of the smoothing parameter, commonly referred to as the bandwidth. There are many methods used to select the bandwidth, such as the plug-in and cross validation procedures. Both, the paper of Sheater (2004) and the book of Givens and Hoeting (2005) provide a review and a practical description on these methods.

Note that in Step 4, we are evaluating $\widehat{T}_u$ at X_i. We may also consider $\widehat{T}_{u,-i}(X_i)$, with $\widehat{T}_{u,-i}(\cdot)$ constructed using the sample but leaving out the observation X_i, both for the non parametric and for the parametric steps.

At this point, we wish to highlight some computational aspects involving the calculation of $\tilde{u}_0$. First, note that we are looking for the threshold above which the excess distribution follows a generalized Pareto law. The model $\mathcal{M}$ is inspired in the fact that the excess distribution F_u can be approximated by a G.P.D. for large values of u. This means that we can expect to see very few observations below $u_0(F)$. That is why we will only consider the function $\widehat{\mathcal{F}}_n$ for those values of u between the order statistics $X^{([0.75n])}$ and $X^{(n)}$. We evaluate the function $\widehat{\mathcal{F}}_n$ along an equally spaced grid $\boldsymbol{U} = \{X^{([0.75n])} = u_1 < u_2 < \cdots < u_{t-1} < u_t = X^{([n])}\}$, obtaining the values

$$\boldsymbol{L} = (L_1, \ldots, L_t) = \left(\widehat{\mathcal{F}}_n(u_1), \ldots, \widehat{\mathcal{F}}_n(u_t)\right).$$

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

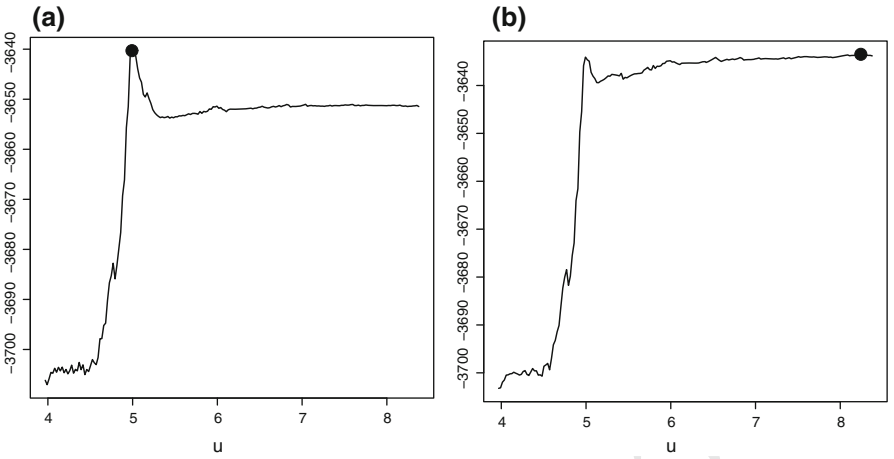

Fig. 1 .

In order to illustrate some of the computational details, we consider the following example. We generate a random sample $X_1, \ldots, X_n$ of size $n = 2{,}000$, according to the following density function

$$f(x) = 0.95\,\frac{1}{5}1_{\{0\le x\le 5\}} + (1-0.95)\,0.5e^{-0.5(x-5)}1_{\{x>5\}}.$$

If we consider the proposed estimator employing a nonparametric kernel density estimator $\widehat{f}(x)$, based on the Epanechnicov kernel with $h = 0.5$, $\widehat{\mathcal{F}}_n$ attains its maximum at 4.971, which is close to $u_0(F) = 5$, the true value of the threshold that we want to estimate. Figure 1a shows the values of $\boldsymbol{L}$ for the mentioned value h of the smoothing parameter. However, if we consider another smoothing parameter to compute the nonparametric density estimator, $\widehat{\mathcal{F}}_n$ may continue to increase, attaining its maximum at the final point of the interval that we are considering. This is the case shown in Fig. 1b. Nevertheless, in both cases we observe that there is a value $\bar{u}$, above which the function $\widehat{\mathcal{F}}_n$ increases very slowly with respect to its behavior below $\bar{u}$. Moreover, we can consider that the function $\widehat{\mathcal{F}}_n$ increases until $\bar{u}$, and is almost constant above $\bar{u}$, except for stochastic and empirical fluctuations. The analysis of the behavior of $\mathcal{F}$ presented at Lemma 4 suggests that $\bar{u}$ should be the estimator of $u_0(F)$ that we are looking for.

Given the need to distinguish the value of $\bar{u}$, we used an heuristic strategy which consists in recognizing the value of u from which the function $\widehat{\mathcal{F}}_n$ is approximately constant by comparing its value with its partial average value $\boldsymbol{S}$.

To be more precise, given $\mathbf{L} = (L_1, \ldots, L_t)$, we consider their partial average values given by $\mathbf{S} = \{S_j : 1 \le j \le t\}$, where S_j is defined as

$$S_j = \frac{1}{j}\sum_{i=1}^{j} L_i.$$

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

The differences $\mathbf{D} = \{D_j : 1 \leq j \leq t-1\}$ are given by

$$D_j = L_{j+1} - S_j, \quad \text{for } 1 \leq j \leq t-1.$$

The main idea behind this construction, is that the difference between $\boldsymbol{L}$ and $\boldsymbol{S}$ attains its maximum in those positions where $\widehat{\mathcal{F}}_n$ is maximized, or its growth rate changes drastically, like in the case shown in Fig. 1b.

After all these considerations, in order to estimate $u_0(F)$ we will compute both $\widehat{u}_{0,A} = \tilde{u}_0$, the first attempt to estimate the threshold presented at (10), and its modified version $\widehat{u}_{0,B} = \bar{u}$ using the heuristic strategy presented. Both strategies can be summarized as follow:

1. Concerning the computation of $\widehat{u}_{0,A} = \tilde{u}_0$, consider

$$J_{0A} = \min\{j : L_j = \max_{1 \leq k \leq t} L_k\} \quad \text{and} \quad \widehat{u}_{0A} = u_{J_{0A}}\,. \tag{11}$$

2. On the other hand, in order to compute $\widehat{u}_{0,B} = \bar{u}$, we consider

$$J_{0B} = \min\{j : D_i \leq D_j, \quad 1 \leq i \leq t-1\} \quad \text{and} \quad \widehat{u}_{0B} = u_{J_{0B}}\,. \tag{12}$$

In Fig. 2 we can see that the heuristic method proposed allows to identify the value where the growth of $\widehat{\mathcal{F}}_n$ changes, both in case a) and b) from Fig. 1.

Even if the extreme index γ plays a relevant role in the field of extreme values, in this work we focus on the threshold itself. However, once the threshold is chosen, we can estimate γ considering the MLE under a generalized Pareto family. More specifically, recalling that $\widehat{\theta}_u = (\widehat{\sigma}_u, \widehat{\gamma}_u)$ is the maximum likelihood estimator under the G.P.F., based on $X_i - u$ with $X_i > u$, we estimate the extreme value index with

$$\widehat{\gamma}_{0A} = \widehat{\gamma}_{\widehat{u}_{0,A}} \quad \text{and} \quad \widehat{\gamma}_{0B} = \widehat{\gamma}_{\widehat{u}_{0,B}}\,. \tag{13}$$

6 Some simulation studies

In this section we perform two different simulations to evaluate the behavior of both our original proposal and its modified version.

The uniform—exponential model: The first case to be considered assumes that the density of the observations is given by

$$f(x) = 0.95\,\frac{1}{5}\,1_{\{0 \leq x \leq 5\}} + 0.05\; 0.5e^{-0.5(x-5)}\,1_{\{x>5\}},$$

as in the example introduced in the previous section. This density implies that (i) the threshold is $u_0 = 5$, (ii) with probability 0.95 the observations are below the threshold uniformly distributed, (iii) the excess distribution above the threshold is exponential with parameter $\lambda_0 = 0.5$.

We perform 1,000 replications of sample size n, with $n = 500, 1{,}000, 1{,}500, 2{,}000$. For each sample, we compute $\widehat{\mathcal{F}}_n(u)$ following the steps described in the previous

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

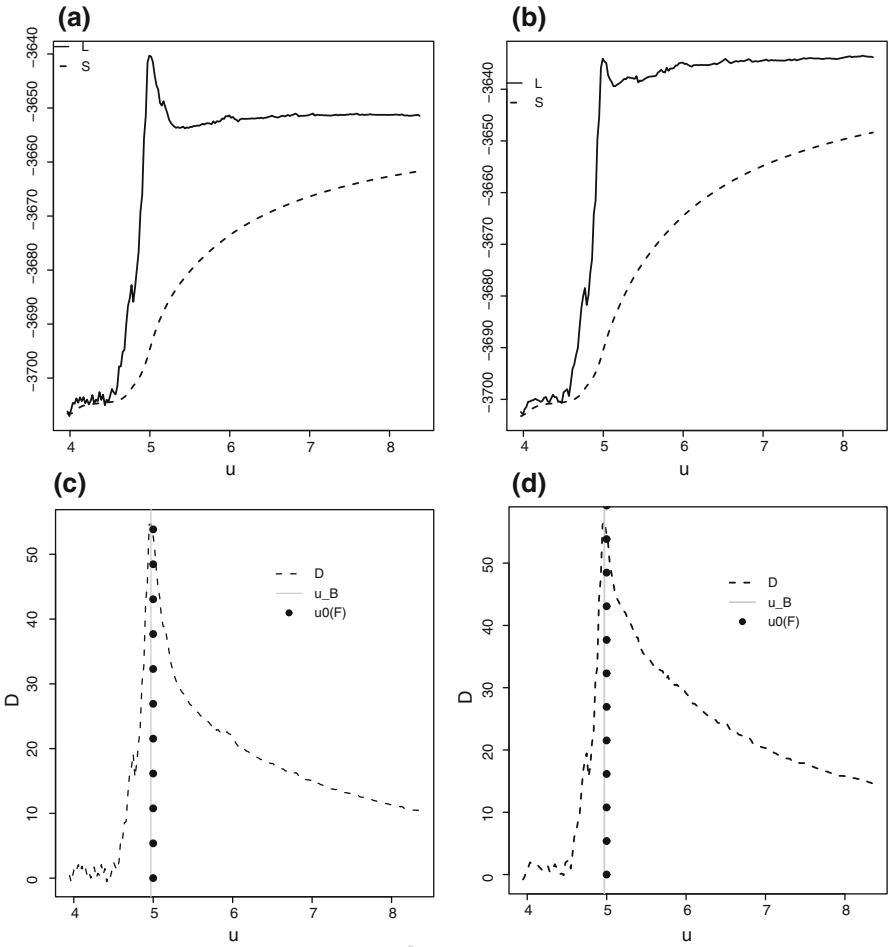

Fig. 2 *Solid* and *dashed lines* represent $\boldsymbol{L}$ and $\boldsymbol{S}$, respectively: **a** using bandwidth $h = 0.5$, **b** using bandwidth $h = 0.25$. **c**, **d** $\boldsymbol{D}$ using bandwidth $h = 0.5$ and $h = 0.25$, respectively

section. Furthermore, at Step 1, we consider a nonparametric kernel density estimator $\widehat{f}(x)$, based on different kernels and different bandwidths: (i) h fixed taking the values 0.1, 0.25, 0.5, 0.75, 1 and the Epanechnikov kernel; (ii) h_{ucv} and h_{bcv} bandwidths chosen automatically according to the unbiased and biased cross-validation criteria, respectively (see Härdle 1991) and a Gaussian kernel. At Step 2, we take advantage of the fact that our data have been generated using an exponential model for the excess distribution ($\gamma = 0$) and so, for each u, we only estimate the exponential parameter with the MLE based on those observations above u, considering

$$\widehat{\lambda}(u) = \frac{\sum_{i=1}^{n} 1_{\{X_i > u\}}}{\sum_{i=1}^{n} (X_i - u) 1_{\{X_i > u\}}}.$$

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

Table 1 Mean squared error and Bias of $\widehat{u}_{0A}$ over the 1,000 replications for the uniform-exponential model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	1.152	0.974	5.684	2.341	10.077	3.145	14.090	3.727
0.25	0.694	0.568	4.338	1.826	7.776	2.489	11.245	3.053
0.50	0.167	0.145	0.651	0.280	0.817	0.278	0.617	0.194
0.75	0.041	0.053	0.047	0.042	0.087	0.050	0.002	0.025
1.00	0.040	−0.071	0.022	−0.041	0.010	−0.023	0.008	−0.018
hucv	0.009	0.025	0.055	0.039	0.086	0.048	0.163	0.065
hbcv	0.003	0.009	0.002	0.015	0.001	0.017	0.001	0.020

Table 2 Mean squared error and Bias of $\widehat{u}_{0B}$ over the 1,000 replications for the uniform-exponential model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	0.084	0.045	0.027	−0.035	0.004	−0.058	0.005	−0.064
0.25	0.020	−0.023	0.006	−0.049	0.004	−0.060	0.005	−0.067
0.50	0.007	−0.034	0.003	−0.050	0.004	−0.061	0.005	−0.068
0.75	0.006	−0.032	0.003	−0.050	0.005	−0.061	0.005	−0.068
1.00	0.055	−0.124	0.052	−0.126	0.042	−0.121	0.039	−0.122
hucv	0.003	−0.033	0.003	−0.050	0.004	−0.061	0.005	−0.068
hbcv	0.005	−0.044	0.004	−0.053	0.005	−0.063	0.006	−0.070

In this way, we get that $\widehat{\mathcal{F}}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \ln \widehat{T}_u(X_i)$, where

$$\widehat{T}_u(x) \;=\; \widehat{p_u}\,\widehat{f}_{1u}(x) + (1-\widehat{p_u})\widehat{\lambda}(u)e^{-\widehat{\lambda}(u)(x-u)}1_{\{x>u\}}.$$

To summarize the simulation results, we report the mean squared error (MSE) and Bias corresponding to both $\widehat{u}_{0,A}$ (Table 1) and $\widehat{u}_{0,B}$ (Table 2), along the 1,000 replications. In Tables 3 and 4 we present the results corresponding to $\widehat{\lambda}_{0,J} = \widehat{\lambda}(\widehat{u}_{0,J})$, for $J = A, B$, respectively.

As we can see in Table 1, the performance of $\widehat{u}_{0A}$ is deficient for small values of h. This can be explained because, in most of the 1,000 replications, for small values of h the estimator $\widehat{u}_{0,A}$ behaves as we show in Fig. 1b. This fact can not be corrected increasing the sample size. For large values of h, for example $h = 1$, we can conclude by comparing Tables 1 and 2, that the MSE for both estimators $\widehat{u}_{0,A}$ and $\widehat{u}_{0,B}$ are small and have the same magnitude. In Table 2, we can observe that for small values of h, $\widehat{u}_{0,B}$ has a good behavior. $\widehat{\lambda}_{0A}$ seems to perform better when h is chosen automatically, while $\widehat{\lambda}_{0B}$ has a good behavior for any sample size and does not seem to be very sensitive to the bandwidth selection criterion.

Even though we started by considering $\widehat{u}_{0A}$ as our first attempt to estimate u_0, the simulations show that a correction is required to achieve a consistent procedure. We

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

Table 3 Mean squared error and Bias of $\widehat{\lambda}_{0A}$ over the 1,000 replications for the uniform-exponential model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	0.018	0.046	0.019	0.046	0.017	0.045	0.017	0.042
0.25	0.016	0.035	0.017	0.039	0.015	0.037	0.015	0.035
0.50	0.012	0.015	0.008	0.014	0.005	0.010	0.004	0.007
0.75	0.010	0.007	0.004	0.005	0.003	0.006	0.002	0.003
1.00	0.009	0.005	0.004	0.005	0.002	0.005	0.002	0.003
hucv	0.009	0.006	0.004	0.006	0.003	0.006	0.002	0.004
hbcv	0.009	0.009	0.004	0.007	0.002	0.007	0.002	0.004

Table 4 Mean squared error and Bias of $\widehat{\lambda}_{0B}$ over the 1,000 replications for the uniform-exponential model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	0.016	0.060	0.009	0.060	0.007	0.066	0.007	0.067
0.25	0.015	0.057	0.009	0.061	0.007	0.067	0.008	0.070
0.50	0.013	0.053	0.009	0.061	0.007	0.068	0.008	0.071
0.75	0.013	0.052	0.009	0.061	0.007	0.068	0.008	0.071
1.00	0.013	0.051	0.008	0.060	0.007	0.068	0.008	0.071
hucv	0.012	0.051	0.009	0.060	0.007	0.068	0.008	0.071
hbcv	0.014	0.057	0.009	0.063	0.008	0.070	0.008	0.073

see that $\widehat{u}_{0B}$ overcomes the limitations displayed by $\widehat{u}_{0A}$, and so we believe that this estimator should be consistent for u_0.

The Gaussian–Pareto model: Now, we consider a truncated normal density below u_0 and a generalized Pareto distribution above u_0. Namely, if $g(x, a, b)$ denotes a density function corresponding to a normal distribution $\mathcal{N}(a, b^2)$, data are generated according to the following density function:

$$f(x) = 0.93 \frac{g(x, a, b)}{\int_{-\infty}^{u_0} g(s, a, b)ds} 1_{\{x \le u_0\}} + (1 - 0.93) h_{\sigma_0, \gamma_0}(x - u_0) 1_{\{x > u_0\}}, \quad (14)$$

with $u_0 = 3.480633$, $a = 2.3$, $b = 0.8$, $\sigma_0 = 1.1$ and $\gamma_0 = 0.3$.

As in the previous case, we generate 1,000 replications of sample size n, with $n = 500, 1{,}000, 1{,}500, 2{,}000$. The nonparametric density estimator was constructed as in the previous simulation, using the Epanechnicov kernel for fixed bandwidths and a Gaussian kernel for bandwidths automatically selected. At Step 2 c), the maximum likelihood estimator $\widehat{\theta}_u$, based on $X_i - u$ with $X_i > u$, was computed using the package *envir*, developed by Mc Neil (2011), in R.

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

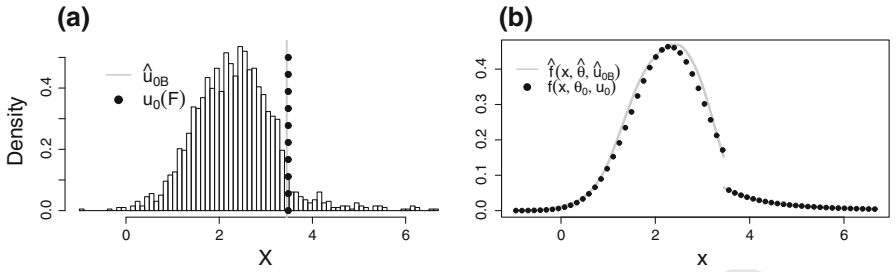

Fig. 3 **a** Histogram of a random sample. The *vertical dotted line* indicates the true value $u_0(F) = 3.480633$, while the *vertical gray line* highlights the estimator $\widehat{u}_{0B} = 3.44862$. **b** The *dotted* and *solid lines* represent the true density and the semiparametric estimator $\widehat{f}(x, \widehat{\theta}, \widehat{u}_{0B}) = \widehat{T}_{\widehat{u}_{0B}}(x)$ using $h = 0.5$, respectively

Table 5 Mean square error and Bias of $\widehat{u}_{0A}$ over the 1,000 replications for the Gaussian–Pareto model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	0.38	0.52	2.50	1.54	5.06	2.21	7.89	2.77
0.25	0.20	0.23	1.69	1.07	3.91	1.77	6.28	2.29
0.50	0.06	0.00	0.24	0.15	0.33	0.15	0.42	0.16
0.75	0.04	−0.05	0.02	−0.01	0.01	−0.01	0.01	−0.00
1.00	0.04	−0.07	0.02	−0.04	0.01	−0.02	0.01	−0.02
h_{ucv}	0.06	−0.13	0.05	−0.08	0.03	−0.05	0.07	−0.01
h_{bcv}	0.06	−0.13	0.04	−0.09	0.02	−0.06	0.01	−0.04

Figure 3 shows the true density $f(x)$ defined at (14) and its semiparametric estimator $\widehat{T}_{u_1}(x)$, defined at Step 4 of Sect. 5, with $u_1 = \widehat{u}_{0A}$ and $h = 0.5$, based on one sample of size $n = 2{,}000$.

Tables 5 and 6 show the MSE and Bias of $\widehat{u}_{0A}$ and $\widehat{u}_{0B}$, respectively, under different scenarios combining the sample size n with the smoothing parameters h, both taken fixed or chosen automatically. We see that the MSE of $\widehat{u}_{0A}$ is small for large values of h, while $\widehat{u}_{0B}$ performs pretty well and better than $\widehat{u}_{0A}$ for small values of h. When the bandwidths are chosen automatically, the results are comparable or even better than those obtained using fixed h.

Tables 7 and 8 show the MSE and Bias of the extreme value index estimators presented at (13). We can observe that for all the different scenarios considered, the behavior of the estimators of γ is satisfactory. It is important to note that, when the bandwidth is selected using an automatic procedure, our recipes give rise to completely data driven estimators (Fig. 5).

7 Comparisons with some other methods

To conclude this work, we compare our proposals with some other estimators already considered in the literature.

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

Table 6 Mean squared error and Bias of $\widehat{u}_{0B}$ over the 1,000 replications for the Gaussian–Pareto model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	0.15	0.15	1.04	0.70	1.88	0.89	2.47	0.94
0.25	0.09	−0.02	0.27	0.13	0.30	0.09	0.30	0.05
0.50	0.07	−0.09	0.12	−0.04	0.06	−0.09	0.06	−0.09
0.75	0.05	−0.10	0.06	−0.10	0.04	−0.11	0.04	−0.11
1.00	0.06	−0.12	0.05	−0.13	0.04	−0.12	0.04	−0.12
h_{ucv}	0.09	−0.01	0.31	0.16	0.36	0.14	0.34	0.08
h_{bcv}	0.05	−0.12	0.05	−0.10	0.03	−0.10	0.03	−0.09

Table 7 Mean squared error and Bias of $\hat{\gamma}_A$ over the 1,000 replications for the Gaussian–Pareto model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	0.14	−0.07	0.13	−0.07	0.13	−0.08	0.14	−0.10
0.25	0.13	−0.07	0.12	−0.07	0.13	−0.08	0.13	−0.11
0.50	0.12	−0.09	0.07	−0.06	0.04	−0.04	0.03	−0.04
0.75	0.10	−0.08	0.04	−0.05	0.02	−0.03	0.02	−0.01
1.00	0.09	−0.07	0.04	−0.03	0.02	−0.02	0.02	−0.00
h_{ucv}	0.14	−0.07	0.12	−0.07	0.13	−0.08	0.14	−0.10
h_{bcv}	0.09	−0.07	0.04	−0.04	0.02	−0.03	0.02	−0.02

Table 8 Mean squared error and Bias of $\hat{\gamma}_B$ over the 1,000 replications for the Gaussian–Pareto model

h	$n = 500$		$n = 1{,}000$		$n = 1{,}500$		$n = 2{,}000$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.10	0.10	−0.01	0.07	−0.01	0.07	0.01	0.07	0.02
0.25	0.09	−0.00	0.06	0.02	0.04	0.05	0.04	0.08
0.50	0.09	0.00	0.05	0.04	0.03	0.08	0.03	0.10
0.75	0.08	0.01	0.04	0.05	0.03	0.09	0.03	0.12
1.00	0.08	0.02	0.04	0.07	0.03	0.10	0.03	0.13
h_{ucv}	0.09	−0.00	0.06	0.02	0.04	0.05	0.04	0.07
h_{bcv}	0.08	0.02	0.04	0.06	0.03	0.09	0.03	0.12

Comparison 1 Mean Residual Plot. Figure 4 shows the Mean Residual Plot (MRP) for the same sample considered in Fig. 3.

This graphical method provides a classic estimator for the threshold and is motivated by the following fact: notice that if $Z \sim H_{\sigma,\gamma}$ with $\gamma < 1$, it follows that $E[Z] = \sigma/(1-\gamma)$. Then, if $F_u = H_{\sigma(u),\gamma}$ for $u \geq u_0$, we get that

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

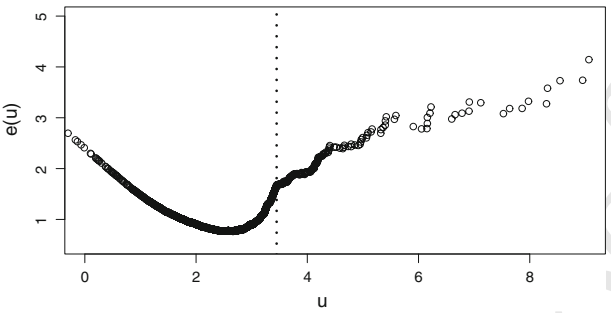

Fig. 4 Mean Residual Plot. The *dotted line* shows a possible choice of the estimator using the MRP

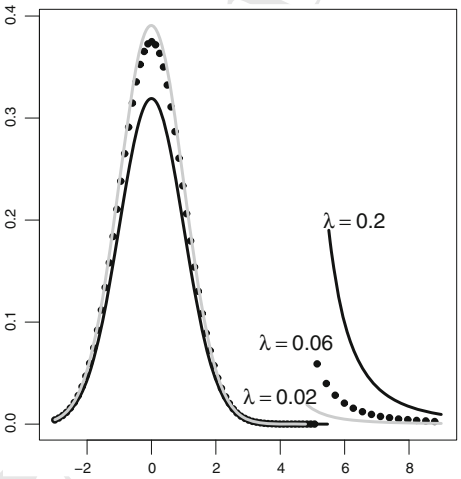

Fig. 5 Densities f_λ for different values of λ: *black solid line* for $\lambda = 0.2$, *dotted line* for $\lambda = 0.06$ and *gray solid line* for $\lambda = 0.02$

$$e(u) = E[X - u | X > u] = \frac{\sigma(u_0) + \gamma u}{1 - \gamma}, \qquad \forall\, u \geq u_0,$$

meaning that mean residual function $e(u)$ is linear, for $u \geq u_0$. Given a random sample $X_1, \ldots, X_n$ the mean residual function can be estimated by

$$\widehat{e}(u) = \frac{\sum_{i=1}^{n} (X_i - u)\, I_{\{X_i > u\}}}{\sum_{i=1}^{n} I_{\{X_i > u\}}}.$$

Therefore, the MRP is defined as the graph of $\widehat{e}(u)$, i.e. $\left\{\left(u, \hat{e}(u)\right) : u < X_{(n)}\right\}$. Thus, once we have this plot, we can choose the threshold as the smallest value of u above which $\widehat{e}(u)$ is almost linear. This method is presented in Coles (2001).

It is worth noting that the mean residual plot leads to a threshold selection method dependent on the particular criterion of the practitioner. Instead, our method provides estimators which are not subject to the experience or the ability of the user when the bandwidth is selected automatically.

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

Table 9 Comparison with competing methods (u_0's estimation)

G	$\widehat{u}_{WL}$		$\widehat{u}_{GH}$		$\widehat{u}_{0A}(h_{ucv})$		$\widehat{u}_{0A}(h_{bcv})$		$\widehat{u}_{0B}(h_{ucv})$		$\widehat{u}_{0B}(h_{bcv})$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
$W(1,0,5)$	0.001	0.00	182.33	10.41	0.033	0.016	0.000	0.007	0.001	−0.021	0.001	−0.022
$\varepsilon(1)$	0.013	−0.03	174.91	9.70	21.703	3.925	1.754	−0.177	0.990	0.183	0.347	−0.117
$\Gamma(1,5)$	0.21	−0.16	149.06	6.99	0.819	−0.294	0.646	−0.459	0.484	−0.351	0.542	−0.436
$\mathcal{N}(10,1)$	0.001	−0.01	144.094	7.03	0.153	0.019	0.002	−0.004	0.003	−0.039	0.003	−0.040
$\mathcal{T}_5$	0.006	−0.02	383.53	16.99	0.343	0.033	0.011	−0.025	0.014	−0.064	0.016	−0.068

Mean squared error and Bias of the estimates for the five models

Comparison 2 As in this work, Wong and Li (2010) also considered a mixture model where, conditionally on $X > u_0$, the density belongs to the G.P.F. (as in our model), while, conditionally on $X \leq u_0$, the density belongs to a parametric family. Then, these authors estimate both the extreme value index γ and the threshold u_0, defined as in this work.

In their numerical simulation data are generated according to the following density function:

$$f(x) = g(x)1_{\{x \leq u_0\}} + 0.1h_{5,0.4}(x - u_0)1_{\{x > u_0\}}, \tag{15}$$

where $g(x) = G'(x)$, $u_0 = \inf\{x : G(x) \geq 0.9\}$ and h_{σ_0,γ_0} is the generalized Pareto density with $\gamma_0 = 0.4$ and $\sigma_0 = 5$. They consider the distribution G in the Weibull, Exponential, Gamma, Normal and Student families, among others. For each case, the value of u_0 is given by 1.18, 2.3, 7.99, 11.28 and 1.48, respectively.

Unfortunately, we did not have access to the code corresponding to their estimators, neither to those corresponding to the methods used in their comparison. For this reason, we present our results combined with those reported in Wong and Li (2010).

In Tables 9 and 10, we present the MSE and Bias of some estimators of u_0 and γ, respectively. We use $\widehat{u}_{WL}$ and $\widehat{\gamma}_{WL}$ to denote the estimators proposed by Wong and Li (2010) and $\widehat{u}_{GH}$ and $\widehat{\gamma}_{GH}$ for the estimators presented in Guillou and Hall (2001). As in the simulation study, our estimators were computed based on a Gaussian kernel, while the bandwidths were selected using biased (h_{bcv}) and unbiased (h_{ucv}) cross-validation criteria. Table 10 includes the Hill estimator of γ given by $\widehat{u}_{GH}$ (denoted by $\widehat{\gamma}_{GH,Hill}$).

Even if the MSE of our estimators are not as small as the MSE of $\widehat{u}_{WL}$, they illustrate that our procedure is also estimating consistently the threshold. We want to remark that our approach is based on a semiparametric model, while the estimator proposed by Wong and Li requires to specify a parametric model for the density below the threshold. This fact could justify why their proposal has a better performance than ours. Related to the extreme index estimation, our results are comparable.

Comparison 3 To examine the robustness of the proposed procedure to estimate the threshold, we decided to perform a simulation study using a family of distributions

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

Table 10 Comparison with competing methods (γ's estimation)

$\widehat{\gamma}_{WL}$		$\widehat{\gamma}_{GH,Hill}$		$\widehat{\gamma}_{GH}$		$\widehat{\gamma}_{0A}(h_{ucv})$		$\widehat{\gamma}_{0B}(h_{ucv})$		$\widehat{\gamma}_{0A}(h_{bcv})$		$\widehat{\gamma}_{0B}(h_{bcv})$	
MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.064	0.080	0.125	0.240	0.004	0.030	0.052	−0.059	0.343	0.297	0.053	−0.056	0.397	0.329
0.067	0.070	0.098	0.200	0.015	0.100	0.145	−0.044	0.063	−0.002	0.157	0.141	0.093	0.072
0.072	0.070	0.028	0.050	0.096	−0.300	0.067	0.006	0.071	0.054	0.066	0.023	0.072	0.068
0.064	0.080	0.020	0.000	0.075	−0.270	0.050	−0.050	0.060	0.053	0.050	−0.050	0.061	0.057
0.069	0.080	0.020	0.000	0.129	−0.350	0.060	−0.040	0.067	0.049	0.063	−0.036	0.070	0.055

Mean squared error and Bias of the estimates for the five models presented in the first column of Table 9

which do not belong to the model $\mathcal{M}$, but have a notion of threshold associated with them.

Cabras and Morales (2007) proposed to select the threshold assuming that most of the observations, though not all, come from a parametric model, and the threshold is chosen as the smallest observation that can be considered as an outlier for the central model. Outliers are detected through a Bayesian procedure. In this framework the threshold is not identified; they proposed how it should be chosen in terms of the size of the sample, the proportion of contamination in the data distribution, and a parameter fixed to decide which observations are outliers.

In their work, they performed a simulation study under different scenarios. To compare our proposal with theirs, we chose one of these scenarios, considering $N = 5{,}000$ replications data sets with $n = 1{,}000$ observations in each of them, with data generated from a standard normal model contaminated (in different proportions) with a generalized Pareto distribution. That is, the density of data (depending on $u_{0,n,\lambda}$) is given by

$$f_\lambda(x) = (1-\lambda)\, g(x) + \lambda h_{\sigma_0,\gamma_0}(x - u_{0,n,\lambda}) \,, \tag{16}$$

where $g(x)$ denotes the density function of a standard normal distribution, h_{σ_0,γ_0} is the generalized Pareto density with $\gamma_0 = 0.5$ and $\sigma_0 = 1$, λ varies between 0.2 and 0.02, while $u_{0,n,\lambda} = 2 + \Phi^{-1}\left((1-0.05)^{1/n\lambda}\right)$ and $\Phi^{-1}(\beta)$ is the quantile of the standard normal distribution at level β.

Note that these densities do not belong to the model $\mathcal{M}$, where we give a precise definition of the threshold u_0. Even so, we will compute MSE and Bias of our estimators assuming that the threshold is given by $u_{0,n,\lambda}$.

As in the previous comparison, we only were able to compute our estimators. Thus, in Table 11 we present the MSE and Bias reported by Cabras and Morales ($\widehat{u}_C$) for different values of λ, together with the MSE and Bias corresponding to our estimators, constructed using automatic bandwidths and a Gaussian kernel.

Despite that our procedure was designed to estimate the threshold under the model $\mathcal{M}$, in Table 11 we can observe that the results are comparable (in magnitude) with those obtained by Cabras and Morales (2007). It seems that, for small values of λ, our procedure can not distinguish between contaminations and observations in the tail

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

of the central distribution. As λ increases, our procedure overcomes this limitation, mainly because, even if the densities are not in the model $\mathcal{M}$, they become closer to it and higher proportion of data came from G.P.D, providing an improvement on estimations of γ (Table 12).

Once the threshold is determined, we computed $\widehat{\gamma}_{0A}$ and $\widehat{\gamma}_{0B}$, defined in (13), obtaining the following results:

Comparison 4 To enclose this work, we compare our estimators for the tail index with the proposal presented by Guillou and Hall (2001). For that purpose,we generate data from a Frechet distribution, which is not in the model $\mathcal{M}$. For this distribution the threshold is not defined but its estimator will be computed at each sample to determine which observations should be used to estimate the extreme value index γ. Once it is chosen, we compute the MLE for γ based on the observations above it, as defined in (13).

We generate 1,000 replications of sample size $n = 1{,}000$, from the distribution function

$$F_\gamma(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \exp\{-x^{-1/\gamma}\} & \text{if} \quad x \geq 0, \end{cases}$$

Table 11 Comparison with competing methods—threshold

λ	$\hat{u}_C$		$\hat{u}_{0A}(h_{ucv})$		$\hat{u}_{0B}(h_{ucv})$		$\hat{u}_{0A}(h_{bcv})$		$\hat{u}_{0B}(h_{bcv})$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.2	59.273	7.626	1.938	0.264	0.137	−0.212	0.020	0.001	0.145	−0.217
0.1	9.632	3.076	0.887	−0.041	0.888	−0.361	0.787	−0.203	1.371	−0.481
0.08	2.582	1.312	3.097	−0.746	2.715	−0.840	4.208	−1.222	3.930	−1.166
0.06	0.087	−0.036	6.284	−1.784	5.544	−1.656	8.977	−2.719	7.372	−2.171
0.04	0.066	−0.035	8.368	−2.457	8.338	−2.470	10.676	−3.209	10.621	−3.122
0.02	0.046	0.007	10.355	−2.992	11.996	−3.293	11.840	−3.400	13.565	−3.661

Table 12 Comparison with competing methods—tail index

λ	$\hat{\gamma}_{0A}(h_{ucv})$		$\hat{\gamma}_{0B}(h_{ucv})$		$\hat{\gamma}_{0A}(h_{bcv})$		$\hat{\gamma}_{0B}(h_{bcv})$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.2	0.0630	−0.0486	0.049	−0.192	0.018	−0.039	0.049	−0.195
0.1	0.121	−0.0792	0.073	−0.194	0.092	−0.086	0.082	−0.204
0.08	0.185	−0.182	0.119	−0.236	0.223	−0.254	0.148	−0.268
0.06	0.295	−0.349	0.182	−0.302	0.358	−0.496	0.225	−0.351
0.04	0.340	−0.343	0.222	−0.259	0.333	−0.407	0.242	−0.281
0.02	0.179	−0.010	0.103	−0.048	0.118	0.0387	0.066	0.001

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

with $\gamma = 0.5$. For each data set, we computed the thresholds $\widehat{u}_{0A}$ and $\widehat{u}_{0B}$, defined in (11) and (12), respectively, using a Gaussian kernel and bandwidth h_{bcv} and bandwidth h_{ucv}. In all cases, the function $\widehat{\mathcal{F}}_n$ behaves as in the case b) of Fig. 1, and thus, we can not compute $\widehat{\gamma}_{0A}$ since there are not data above $\widehat{u}_{0A}$. The MSE and Bias of $\widehat{\gamma}_{0B}$ using bandwidth h_{bcv} are given by 0.124 and -0.068, while for bandwidth h_{ucv} the MSE and Bias are 0.13 and 0.0636, respectively. Also, for each data set we computed the estimators of γ proposed in Guillou and Hall (2001), with $c_{\text{crit}} = 1.25$ and $c_{\text{crit}} = 1.5$, $p = 1$ in both cases, getting for each c_{crit} a MSE equal to 0.005 and 0.006, with Bias 0.056 and 0.064, respectively. As was expected, their method for the estimation of γ performs much better than ours since it was designed for such purpose, while in this work, we are interested in threshold selection, by itself.

Acknowledgments We are grateful to the anonymous referees and the editor for their comments that have largely contributed to improve the original manuscript, shedding light on several issues. We also thank Damian and Lucio for their careful reading of this manuscript. This research was partially supported by Grants PIP 112-200801-00216 and 0592 from CONICET, PICT 0821 and 0883 from ANPCYT and 20020090100208 and 20020100300057 from the Universidad de Buenos Aires at Buenos Aires, Argentina.

References

Behrens C, Lopes H, Gamerman D (2004) Bayesian analysis of extreme events with threshold estimation. Stat Model 4:227–244

Beirlant J, Joossens E, Segers J (2004a) Discussion of "generalized Pareto fit to the society of actuaries' large claims database" by A. Cebrian, M. Denuit and P. Lambert. N Am Actuar J 8:108–111

Beirlant J, Goegebeur Y, Segers J, Teugels J (2004b) Statistics of extremes—theory and applications. Wiley, New York

Beirlant J, Vynckier P, Teugels JL (1996) Excess functions and estimation of the extreme-value index. Bernoulli 2:293–318

Cabras S, Morales J (2007) Extreme value analysis within a parametric outlier detection. Appl Stoch Models Bus Ind 23:157–164

Coles SG, Tawn JA (1994) Statistical methods for multivariate extremes: an application to structural design. Appl Stat 43:1–48

Coles SG, Tawn JA (1996) A Bayesian analysis of extreme rainfall data. Appl Stat 45(4):463–478

Coles SG, Powell EA (1996) Bayesian methods in extreme value modelling: a review and new developments. Int Stat Rev 64(1):119–136

Coles S (2001) An introduction to statistical modeling of extreme values. Springer, Berlin

Davison AC, Smith RL (1990) Models for exceedances over high thresholds. J R Stat Soc B 52(3):393–442

Drees H, Kaufmann E (1998) Selecting the optimal sample fraction in univariate exteme value estimations. Stoch Process Their Appl 75:149–172

DuMouchel WH (1983) Estimating the stable index α in order to measure tail thickness: a critique. Ann Stat 11(4):1019–1031

Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance. Springer, New York

Frigessi A, Haug O, Rue H (2003) Tail estimation with generalized Pareto distribution without threshold selection. Extremes 5(3):219–236

Givens G, Hoeting J (2005) Computational statistics. Wiley, New York

Gomes MI, Oliveira O (2001) The bootstrap methodology in statistics of extremes-choice of the optimal sample fraction. Extremes 4:331–358

Guillou A, Hall P (2001) A diagnostic for selecting the threshold in extreme value analysis. J R Stat Soc Ser B 63:293–305

Hill B (1975) A simple general approach to inference about the tail of a distribution. Ann Stat 3:1163–1174

Hall P (1990) Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. J Multivar Anal 32:177–203

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X

Härdle W (1991) Applied Nonparametric Regression. Econometric Society Monographs
MacDonald A, Scarrott C, Lee D, Darlow B, Reale M, Russell G (2011) A flexible extreme value mixture model. Comput Stat Data Anal 55:2137–2157
Mc Neil A (2011) evir: Extreme values in R. S original (Evis) by Alexander Mc Neil and R Port by Alec Stephenson
Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 33:1065–1076
Pickands J (1975) Statistical inference using extreme order statistics. Ann Stat 3:119–131
Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. Ann Math Stat 27:832–837
Sheater S (2004) Density estimation. Stat Sci 19(4):588–597
Smith R (1987) Estimating tails of probability distributions. Ann Stat 15:1174–1207
Wong T, Li W (2010) A thereshold approach for peaks-over-threshold modeling using maximum product of spacings. Stat Sinica 20:1257–1272

Journal: 10260 Article No.: 0234 TYPESET DISK LE CP Disp.:2013/7/4 Pages: 20 Layout: Small-X