ISSN 1645-6726

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT

Statistical Journal

Special Issue on Statistics of Extremes and Related Fields

Guest Editors:
Jan Beirlant
Isabel Fraga Alves
Ross Leadbetter

Volume 6, No. 1
March 2008

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional
de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E.,
2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **EDITOR-IN-CHIEF**
  - *M. Ivette Gomes*
- **CO-EDITOR**
  - *M. Antónia Amaral Turkman*
- **ASSOCIATE EDITORS**
  - *António Pacheco*
  - *Barry Arnold*
  - *Dani Gamerman*
  - *David Cox*
  - *Dinis Pestana*
  - *Edwin Diday*
  - *Gilbert Saporta*
  - *Helena Bacelar Nicolau*
  - *Isaac Meilijson*
  - *Jef Teugels*
  - *João Branco*
  - *Ludger Rüschendorf*
  - *M. Lucília Carvalho*
  - *Marie Husková*
  - *Nazaré Mendes-Lopes*
  - *Stephan Morgenthaler*
  - *Susie Bayarri*
- **EXECUTIVE EDITOR**
  - *Maria José Carrilho*
- **SECRETARY**
  - *Liliana Martins*
- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 426 364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 226 050 748
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *400 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue ………………………………………………… **€ 10**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)… **€ 24**
- Annual subscription (No. 2, No. 3) ………………………… **€ 16**

REVSTAT

Statistical Journal

FOREWORD

Recent years have witnessed a vigorous growth in the use of Extreme Value Theory and Statistics of Extremes, with relevance to applications in a broad spectrum of areas, ranging for example from natural hazards in geophysics and the environment to rare events in financial risk. This special issue of *REVSTAT – Statistical Review* aims at giving the readers a flavor of this exciting area of research, through recent advances in the field.

Statistics of Extremes and Related Fields consists of six articles authored by prominent researchers who participated in the 56th Session of the International Statistical Institute, which was held in PORTUGAL, Lisbon, from August 22–29, 2007. The selection involves some of the original contributions in Invited Paper Meetings (IPMs) and Special Topics Contributed Paper Meetings (STCPMs) in the field of Statistics of Extremes, namely IPM7 (*Bias Reduction in the Estimation of Parameters of Rare Events*), IPM10 (*Extremes, Risk and the Environment*), IPM61 (Are extreme weather events more prevalent now than before?) and STCPM17 (*Extremal methods for action in today's world*). This selection covers a wide range of topics in the field, which are of significant current interest.

The topics discussed include:

Minimum-Variance Reduced-Bias High Quantile Estimation, i.e., estimation of a value that is exceeded with a small pre-specified probability. The semiparametric estimation of this parameter relies essentially on the estimation of the tail index, the primary parameter in statistics of extremes.

An applied article whose main objective is to provide an insight into the geographic distribution of extreme precipitation events in the Southern region of continental Portugal, as a basis for future study of the relationships between extreme rainfall patterns.

The asymptotic properties of the so-called Generalized Probability-Weighted Moments (GPWM), a recent extension of PWM, which address Generalized Extreme Value distributions with large values of the shape parameter.

The problem in a general kernel goodness-of-fit test statistic for assessing whether a sample is consistent with the Pareto-type model. Therein the relation between goodness-of-fit testing and the optimal selection of the sample fraction for tail estimation is examined.

A proposal for use of a stationary max-stable process as a model of the dependence structure in two-dimensional spatial problems, employing a representation of simple max-stable processes.

A brief overview of several tests published in the context of statistical choice of extreme value domains and for assessing extreme value conditions, also illustrated with a teletraffic data set.

It is our hope that this edition also strengthens the ties and encourages collaboration between researchers in Statistics of Extremes and related fields all over the world. We thank the authors for their prompt support with their interesting contributions. We are also most grateful to the referees for careful review of the papers. Finally, we must record our deep appreciation for the encouragement and support of the editor-in-chief, Professor Ivette Gomes, (together with CEAUL, Center of Statistics and Applications at University of Lisbon, `http://www.ceaul.fc.ul.pt`), who had a prominent role on the organization of the IPMs and STCPMs under the auspices of ISI2007.

Finally our sincere thanks go to all those involved in making this project successful.

Jan Beirlant
Isabel Fraga Alves
Ross Leadbetter

INDEX

REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008, 1–20

MINIMUM-VARIANCE REDUCED-BIAS TAIL INDEX AND HIGH QUANTILE ESTIMATION

Authors: Frederico Caeiro
– F.C.T. (D.M.) and C.M.A.,
Universidade Nova de Lisboa, Portugal
fac@fct.unl.pt

M. Ivette Gomes
– F.C.U.L. (D.E.I.O.) and C.E.A.U.L.,
Universidade de Lisboa, Portugal
ivette.gomes@fc.ul.pt

Abstract:

- Heavy tailed-models are quite useful in many fields, like *insurance*, *finance*, *telecommunications*, *internet traffic*, among others, and it is often necessary to estimate a *high quantile*, i.e., a value that is exceeded with a probability p, small. The semi-parametric estimation of this parameter relies essentially on the estimation of the *tail index*, the primary parameter in *statistics of extremes*. Classical semi-parametric estimators of extreme parameters show usually a severe bias and are known to be very sensitive to the number k of top order statistics used in the estimation. For k small they have a high variance, and for large k a high bias. Recently, new second-order “shape” and “scale” estimators allowed the development of second-order reduced-bias estimators, which are much less sensitive to the choice of k. Here we shall study, under a third order framework, minimum-variance reduced-bias (MVRB) tail index estimators, recently introduced in the literature, and dependent on an adequate estimation of second order parameters. The improvement comes from the asymptotic variance, which is kept equal to the asymptotic variance of the classical Hill estimator, provided that we estimate the second order parameters at a level of a larger order than the level used for the estimation of the first order parameter. The use of those MVRB tail index estimators enables us to introduce new classes of reduced-bias high quantile estimators. These new classes are compared among themselves and with previous ones through the use of a small-scale Monte Carlo simulation.

Key-Words:

- *statistics of extremes; tail index; high quantiles; second-order reduced-bias semi-parametric estimation; third order framework.*

AMS Subject Classification:

- 62G32, 62E20, 65C05.

1. INTRODUCTION

Let $X_1, X_2, ..., X_n$ be a set of n independent and identically distributed (i.i.d.) random variables (r.v.'s), from a population with distribution function (d.f.) F, in the max domain of attraction of G_γ, $\gamma \in \mathbb{R}$, with

$$G_\gamma(x) = \begin{cases} \exp\Bigl[-(1+\gamma x)^{-\frac{1}{\gamma}}\Bigr], & 1+\gamma x > 0 \quad \text{if} \quad \gamma \neq 0\,, \\ \exp(-e^{-x})\,, & x \in \mathbb{R} \qquad\quad \text{if} \quad \gamma = 0\;. \end{cases}$$

The parameter γ is the *extreme value index* and we then use the notation $F \in D(G_\gamma)$. In this paper we shall work only with heavy-tailed models, i.e., models $F \in D(G_\gamma)$ with $\gamma > 0$. Then γ is often called *tail index*.

Let us define $U(t) := F^{\leftarrow}(1-1/t)$, $t > 1$, with $F^{\leftarrow}(x) := \inf\bigl\{y\colon F(y) \geq x\bigr\}$ denoting the generalized inverse function of F. We have

$$\text{(1.1)} \qquad F \in D(G_\gamma),\;\; \gamma > 0 \quad \Longleftrightarrow \quad 1-F \in RV_{-1/\gamma} \quad \Longleftrightarrow \quad U \in RV_\gamma$$

(Gnedenko, 1943; de Haan, 1970), where, for any real a, RV_a stands for the class of regularly varying functions at infinity with index of regular variation a, i.e. positive measurable functions g such that $\lim_{t\to\infty} g(tx)/g(t) = x^a$, for all $x > 0$.

We are interested in the estimation of a high quantile, χ_{1-p}, a typical parameter in the most diversified areas of application. Such a quantile is a value exceeded with a small probability p, i.e., such that $F(\chi_{1-p}) = 1-p$. More specifically, we want to extrapolate beyond the sample, and to estimate

$$\text{(1.2)} \qquad \chi_{1-p} = U(1/p)\,, \qquad p = p_n \to 0, \quad np_n \to K \quad \text{as } n \to \infty, \quad K \in [0,1]\;.$$

Denoting by $X_{1:n} < ... < X_{n:n}$ the order statistics (o.s.'s) from the original sample, Weissman (1978) proposed, for heavy-tailed models, the following semi-parametric estimator of χ_{1-p},

$$\text{(1.3)} \qquad Q^{(p)}_{\hat\gamma}(k) := X_{n-k:n}\, c_n^{\hat\gamma}\;, \qquad c_n := \tfrac{k}{np} \to \infty, \quad \text{as} \quad n \to \infty\;,$$

where $\hat\gamma$ is any consistent estimator of γ. For $\gamma \in \mathbb{R}$, we can find semi-parametric high quantile estimators in de Haan and Rootzén (1983), Ferreira *et al.* (2003) and Matthys and Beirlant (2003). As usual in semi-parametric estimation of parameters from extreme value models, we shall assume that $k = k_n$ is an *intermediate* sequence, i.e., a sequence of integer values in $[1, n]$, such that

$$\text{(1.4)} \qquad k_n \to \infty\;, \qquad k_n = o(n), \quad \text{as} \quad n \to \infty\;.$$

For heavy tails, the classical tail index estimator, usually the one which is plugged in (1.3) for a semi-parametric quantile estimation, is the Hill estimator $\hat{\gamma} = \hat{\gamma}(k) =: H(k)$ (Hill, 1975),

$$H(k) := \tfrac{1}{k}\sum_{i=1}^{k} V_{ik} = \tfrac{1}{k}\sum_{i=1}^{k} U_i \ , \tag{1.5}$$

the average of the log-excesses $V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}$, $1 \le i \le k < n$, as well as the average of the scaled log-spacings

$$U_i := i\big(\ln X_{n-i+1:n} - \ln X_{n-i:n}\big) \ , \qquad 1 \le i \le k < n \ . \tag{1.6}$$

We thus get the so-called classical quantile estimator, $Q_H^{(p)}(k)$, based on the Hill tail index estimator H. It is known that for intermediate k and if the first order condition (1.1) holds, $H(k)$ and $Q_H^{(p)}(k)$ are consistent for the estimation of γ and χ_{1-p}, respectively. The main problem with these semi-parametric estimators is a high variance for small k, i.e., high thresholds, and a high bias for large k.

To obtain information on the distributional behaviour of these estimators, we shall also assume a second order condition, that measures the rate of convergence of $\ln U(tx) - \ln U(t)$ to $\gamma \ln x$,

$$\begin{aligned} &\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho} \iff \\ &\qquad \iff \lim_{t\to\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma\, \frac{x^\rho - 1}{\rho} \ , \end{aligned} \tag{1.7}$$

for all $x > 0$, where $\rho \le 0$ is the shape second order parameter and the function $|A|$ must be of regular variation with index ρ (Geluk and de Haan, 1987). To be able to reduce the bias of these estimators, it is quite useful to assume that we are working in Hall's class of heavy-tailed models (Hall, 1982; Hall and Welsh, 1985) where, with $\gamma > 0$, $\rho < 0$, $C > 0$ and $D_1 \neq 0$,

$$U(t) = C t^\gamma \Big(1 + D_1 t^\rho + o(t^\rho)\Big) \ , \qquad t \to \infty \ . \tag{1.8}$$

Then, the second order condition (1.7) holds with $A(t) = \rho D_1 t^\rho := \gamma\, \beta\, t^\rho$.

Proposition 1.1 (de Haan and Peng, 1998). *Under the second order framework in* (1.7)*, and for intermediate k, i.e., whenever* (1.4) *holds, we may guarantee the asymptotic normality of $H(k)$ in* (1.5)*. Indeed, we may write,*

$$H(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k + \frac{A(n/k)}{1-\rho}\, \big(1 + o_p(1)\big) \ , \tag{1.9}$$

with $Z_k = \sqrt{k}\,\big(\sum_{i=1}^{k} E_i/k - 1\big)$, and $\{E_i\}$ i.i.d. standard exponential r.v.'s.

Consequently, if we choose k such that $\sqrt{k}\,A(n/k) \to \lambda \neq 0$, finite, as $n\to\infty$, $\sqrt{k}\,\big(H(k)-\gamma\big)$ is asymptotically normal, with variance equal to γ^2 and a non-null mean value given by $\lambda/(1-\rho)$.

The result in (1.9) has recently led researchers to consider the possibility of dealing with the asymptotic bias dominant term in an appropriate way, building second-order reduced-bias estimators, discussed by Peng (1998), Beirlant *et al.* (1999), Feuerverger and Hall (1999), Gomes *et al.* (2000), among others. In the above mentioned papers, authors have been able to remove the dominant component of the asymptotic bias, but with an increase of the asymptotic variance. More recently, Gomes *et al.* (2004b), Caeiro *et al.* (2005) and Gomes *et al.* (2007a) proposed minimum-variance reduced-bias (MVRB) estimators, based on an external estimation of second order parameters, built in such way that they were able to reduce the bias without increasing the asymptotic variance, which is kept equal to γ^2, the asymptotic variance of the Hill estimator.

If we look at (1.9), we see that the dominant component of the bias of Hill's estimator is $A(n/k)/(1-\rho) = \gamma\beta(n/k)^\rho/\rho$, for models in (1.8). This component can be easily estimated and removed from Hill's estimator, leading to any of the asymptotically equivalent estimators (Caeiro *et al.*, 2005),

$$\begin{aligned}
\overline{H}_{\hat\beta,\hat\rho}(k) &:= H(k)\left(1-\frac{\hat\beta}{1-\hat\rho}\left(\frac{n}{k}\right)^{\hat\rho}\right),\\
\overline{\overline{H}}_{\hat\beta,\hat\rho}(k) &:= H(k)\,\exp\!\left(-\frac{\hat\beta}{1-\hat\rho}\left(\frac{n}{k}\right)^{\hat\rho}\right),
\end{aligned} \tag{1.10}$$

where $\hat\rho$ and $\hat\beta$ need to be adequate consistent estimators of the second order parameters ρ and β, if we want to keep the asymptotic variance at γ^2. This requires an external estimation of the second order parameters using a number of top o.s.'s k_1, larger than the number of top o.s.'s, k, used for the tail index estimation, and an estimator $\hat\rho$ of ρ such that $\hat\rho-\rho = o_p(1/\ln n)$.

On the basis of the different papers dealing with high quantile semi-parametric estimation for heavy tails, among which we mention Gomes and Figueiredo (2006) and Caeiro and Gomes (2007), we can state the following result.

Proposition 1.2. *Under the conditions of Proposition 1.1, the validity of* (1.2)*, a known tail index γ and c_n defined in* (1.3)*,*

$$Q_\gamma^{(p)}(k) \overset{d}{=} \chi_{1-p}\left(1+\frac{\gamma}{\sqrt{k}}\,B_k + \frac{1-c_n^\rho}{\rho}\,A(n/k)\,\big(1+o_p(1)\big)\right), \tag{1.11}$$

with B_k an asymptotically standard normal r.v. Consequently, if $\sqrt{k}\,A(n/k)\to\lambda$, finite, $\sqrt{k}\,\big(Q_\gamma^{(p)}(k)/\chi_{1-p}-1\big)$ is asymptotically normal, with variance γ^2 and mean *value λ/ρ. If γ is unknown and is estimated by any consistent estimator $\hat\gamma$,*

$$(1.12)\qquad Q^{(p)}_{\hat\gamma}(k) \;\overset{d}{=}\; \chi_{1-p}\Bigg(1+(\hat\gamma-\gamma)\,\ln c_n+\frac{\gamma}{\sqrt{k}}\,B_k+\frac{1-c_n^{\rho}}{\rho}\,A(n/k)\,\big(1+o_p(1)\big)\Bigg)\,.$$

Consequently, if $\sqrt{k}\,A(n/k)\to\lambda$, finite, and $\ln c_n/\sqrt{k}\to 0$, as $n\to\infty$, then $\frac{\sqrt{k}}{\ln c_n}\big(Q^{(p)}_{\hat\gamma}(k)/\chi_{1-p}-1\big)$ has asymptotically the same distribution as $\sqrt{k}\,(\hat\gamma-\gamma)$.

From (1.12) it is obvious that the behaviour of $\hat\gamma$ rules strongly the behaviour of $Q^{(p)}_{\hat\gamma}$. The summand $(1-c_n^{\rho})\,A(n/k)/\rho$, asymptotically equivalent to $A(n/k)/\rho$ and the dominant component of the bias of $Q^{(p)}_{\gamma}$ in (1.11), does not influence the limiting distribution of $Q^{(p)}_{\hat\gamma}$. But, as already noticed in Matthys *et al.* (2004), the removal of this term for finite samples, typically leads to an improvement in the overall stability of the quantile estimates as a function of k. Since $\chi_{1-p}/X_{n-k:n} \overset{p}{\sim} c_n^{\gamma}\big(1+(c_n^{\rho}-1)\,A(n/k)/\rho\big)$, we shall consider the new estimators,

$$(1.13)\qquad \overline{Q}^{(p)}_{\hat\gamma}(k) \;=\; \overline{Q}^{(p)}_{\hat\gamma}(k;\hat\beta,\hat\rho) \;:=\; X_{n-k:n}\;c_n^{\hat\gamma}\left(1+\hat\gamma\,\hat\beta\left(\frac{n}{k}\right)^{\hat\rho}\frac{c_n^{\hat\rho}-1}{\hat\rho}\right),$$

asymptotically equivalent, up to the second order, to the estimators already proposed before by Matthys *et al.* (2004), Beirlant *et al.* (2006) and Gomes and Pestana (2007b),

$$(1.14)\qquad \overline{\overline{Q}}^{(p)}_{\hat\gamma}(k) \;=\; \overline{\overline{Q}}^{(p)}_{\hat\gamma}(k;\hat\beta,\hat\rho) \;:=\; X_{n-k:n}\;c_n^{\hat\gamma}\,\exp\!\left(\hat\gamma\,\hat\beta\left(\frac{n}{k}\right)^{\hat\rho}\frac{c_n^{\hat\rho}-1}{\hat\rho}\right).$$

We shall replace $\hat\gamma$ by any of the MVRB estimators $\overline{H}(k)=\overline{H}_{\hat\beta,\hat\rho}(k)$ and $\overline{\overline{H}}(k)=\overline{\overline{H}}_{\hat\beta,\hat\rho}(k)$, generally denoted by $\widetilde{H}(k)$, with $\overline{H}_{\hat\beta,\hat\rho}(k)$ and $\overline{\overline{H}}_{\hat\beta,\hat\rho}(k)$ given in (1.10).

Remark 1.1. Since $c_n^{\rho}\,\ln c_n=o(1)$, the asymptotic behavior of (1.13) and (1.14) does not change if we replace $c_n^{\hat\rho}$ by 0. In the simulation study, we did not notice any change in the performance of the estimators with this replacement. Anyway, we shall keep working with the quantile estimators defined in (1.13).

Is section 2, and assuming a third order framework in order to get full information on the leading terms of asymptotic bias, we study the tail index estimators $\widetilde{H}(k)$ in (1.10), as well as $\overline{Q}^{(p)}_{\widetilde{H}}$, with $\overline{Q}^{(p)}_{\hat\gamma}$ given in (1.13). In Section 3, a small-scale simulation study helps us to identify the behaviour of the quantile estimators in (1.13) for finite samples. Finally, in Section 4, we draw a short final conclusion.

2. ASYMPTOTIC PROPERTIES

2.1. Third order framework

In order to derive the asymptotic bias of the MVRB estimators under study, we shall work with a sub-class of Hall's class such that

$$(2.1)\qquad U(t) \;=\; C\,t^{\gamma}\Big(1 + D_1 t^{\rho} + D_2 t^{\rho+\rho^*} + o\big(t^{\rho+\rho^*}\big)\Big)\,, \qquad t \to \infty\;,$$

$C>0,\ D_1 \neq 0,\ \rho<0,\ \rho^*<0$. Note that, compared to Hall's class in (1.8) we merely specify the summand $o(t^{\rho})$. Note also that, with $h_\theta(x) := (x^\theta - 1)/\theta$, $\theta<0$, $A(t) = \rho\, D_1 t^{\rho} = \gamma\beta\, t^{\rho}$, $\rho' = \max(\rho,\rho^*) \geq \rho$ and

$$B(t) \;=\; \beta' t^{\rho'} = \begin{cases} \big((1+\rho^*/\rho)\,D_2/D_1\big)\,t^{\rho^*}, & \rho<\rho^*\,,\\ \big(2\,D_2/D_1 - D_1\big)\,t^{\rho}, & \rho=\rho^*\,,\\ -D_1\,t^{\rho}, & \rho>\rho^* \ \text{ or } \ D_2 = 0\;,\end{cases}$$

we may write for any $x>0$,

$$(2.2)\qquad \ln\frac{U(tx)}{U(t)} - \gamma\ln x \;=\; A(t)\,h_\rho(x) + A(t)\,B(t)\,h_{\rho+\rho'}(x)\,\big(1+o(1)\big)\;,$$

which is, for arbitrary ρ and ρ', the third order condition used in the paper by Gomes *et al.* (2004a), equivalent to the ones assumed in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003). As mentioned before, we shall essentially consider the validity of (2.1), which is equivalent to consider that (2.2) holds with $\rho \leq \rho'$ and $A(t) = \alpha\, t^{\rho}$ for some real α.

Remark 2.1. The class in (2.1) contains most of the heavy-tailed models used in applications, like the *Fréchet*, with $U(t) = \big(\ln(t/(t-1))\big)^{-\gamma}$, the *Burr*, with $U(t) = (t^{-\rho}-1)^{-\gamma/\rho}$, $t>1$, the *Generalized Pareto* (*GP*), with $U(t) = (t^{\gamma}-1)/\gamma$, $t>1$, and the *Student*'s-t_ν, $\nu>0$, with d.f.

$$F(x) \;=\; F(x\,|\,\nu) \;=\; \frac{\Gamma\big((\nu+1)/2\big)}{\Gamma(\nu/2)\,\sqrt{\pi\nu}} \int_{-\infty}^{x} (1+z^2/\nu)^{-(\nu+1)/2}\,dz\,, \qquad x\in\mathbb{R},\ \ \nu>0\;.$$

Although $\rho^* = \rho' = \rho$ for all these classical models, we have decided to work with a slight more general condition, the one in (2.1). Indeed, it is not so hard to find examples where $\rho' \neq \rho$. Gomes and Oliveira (2003) noticed that shifting the data can change the asymptotic behavior of the tail and the value of the second order parameters, i.e., if X is our original parent, and $Y = X+a$, then $U_Y(t) = U_X(t) + a$, and consequentially,

$$U_Y(t) \;=\; C\,t^{\gamma}\Big(1 + D_1 t^{\rho} + a\,t^{-\gamma}/C + D_2 t^{\rho+\rho^*} + o\big(t^{\rho+\rho^*}\big)\Big)\,, \qquad t\to\infty\;.$$

In Table 1 we present, for the above mentioned models, the values of the first, second and third order parameters in (2.1) and the values of β and β' in $A(t) = \gamma\beta\, t^\rho$ and $B(t) = \beta' t^\rho$. In this table, $c_\nu = (\nu\mathcal{B}(\nu/2, 1/2))^{1/\nu}$ ($c_1 = \pi$ leading to the usually called Cauchy d.f.), where $\mathcal{B}$ is the complete Beta function.

Table 1: Study of some distributions in Hall's class.

Distribution	C	D_1	D_2	γ	ρ	ρ^*	β	β'
Fréchet	1	$-\frac{\gamma}{2}$	$-\frac{\gamma}{12}$	γ	-1	-1	$\frac{1}{2}$	$\frac{5}{6}$
Burr	1	$\frac{\gamma}{\rho}$	$\frac{\gamma(\rho+\gamma)}{2\rho^2}$	γ	ρ	ρ	1	1
GP	$\frac{1}{\gamma}$	-1	0	γ	$-\gamma$	$-\gamma$	1	1
Student's t_ν	$\sqrt{v}\, c_\nu^{-1}$	$-\frac{(\nu+1)\, c_\nu^2}{2(\nu+2)}$	$-\frac{\nu\,(\nu+1)\,(\nu+3)\, c_\nu^4}{8\,(\nu+2)^2\,(\nu+4)}$	$\frac{1}{\nu}$	$-\frac{2}{\nu}$	$-\frac{2}{\nu}$	$\frac{(\nu+1)\, c_\nu^2}{\nu+2}$	$\frac{(\nu^2+4\nu+2)\, c_\nu^2}{(\nu+2)\,(\nu+4)}$

2.2. Estimation of second order parameters

The reduced-bias tail index and quantile estimators require the estimation of the second order parameters ρ and β, which will be now briefly discussed.

2.2.1. Estimation of the shape second order parameter ρ

We shall consider here particular members of the class of estimators of the second order parameter ρ proposed by Fraga Alves *et al.* (2003), but parameterized by a tuning real parameter τ (see Caeiro and Gomes, 2006). Denoting $M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^k V_{ik}^j$ the j-moment of the log-excesses, $j = 1, 2, 3$, these ρ-estimators depend on the statistics

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\Big(M_n^{(1)}(k)\Big)^\tau - \Big(M_n^{(2)}(k)/2\Big)^{\tau/2}}{\Big(M_n^{(2)}(k)/2\Big)^{\tau/2} - \Big(M_n^{(3)}(k)/6\Big)^{\tau/3}}\,, & \text{if } \ \tau \neq 0\,, \\[2ex] \dfrac{\ln\Big(M_n^{(1)}(k)\Big) - \frac{1}{2}\ln\Big(M_n^{(2)}(k)/2\Big)}{\frac{1}{2}\ln\Big(M_n^{(2)}(k)/2\Big) - \frac{1}{3}\ln\Big(M_n^{(3)}(k)/6\Big)}\,, & \text{if } \ \tau = 0\,, \end{cases}$$

which converge towards $3(1-\rho)/(3-\rho)$ for any real τ, whenever the second order condition (1.7) holds, k is such that (1.4) holds and $\sqrt{k}\,A(n/k) \to \infty$, as $n \to \infty$. The ρ-estimators considered have the functional expression,

$$\hat\rho_\tau(k) \;=\; \hat\rho(k;\tau) \;:=\; -\min\Bigl(0,\; 3\bigl(T_n^{(\tau)}(k)\bigr)-1\Bigr)\,/\,\bigl(T_n^{(\tau)}(k)-3\bigr)\;. \tag{2.3}$$

Proposition 2.1 (Fraga Alves *et al.*, 2003)**.** *If the second order condition* (1.7) *holds, with* $\rho<0$, (1.4) *holds and* $\sqrt{k}\,A(n/k)\to\infty$, *then* $\hat\rho(k;\tau)$ *in* (2.3) *converge in probability to* ρ, *as* $n\to\infty$. *Under the third order framework in* (2.2),

$$\hat\rho(k;\tau) \;\overset{d}{=}\; \rho + \left(\frac{\gamma\,\sigma_\rho W_k^\rho}{\sqrt{k}\,A(n/k)} + \upsilon_1 A(n/k) + \upsilon_2 B(n/k)\right)\bigl(1+o_p(1)\bigr)\;, \tag{2.4}$$

where W_k^ρ *is an asymptotically standard normal r.v.,* $\sigma_\rho = \frac{(1-\rho)^3}{\rho}\sqrt{(2\rho^2-2\rho+1)}$,

$$\upsilon_1 \;\equiv\; \upsilon_1(\gamma,\rho,\tau) \;=\; \frac{\rho\Bigl[\tau\,(1-2\rho)^2\,(3-\rho)\,(3-2\rho) \,+\, 6\rho\,\bigl(4\,(2-\rho)\,(1-\rho)^2-1\bigr)\Bigr]}{12\,\gamma\,(1-\rho)^2\,(1-2\rho)^2}\;,$$

$$\upsilon_2 \;=\; \frac{\rho'(\rho+\rho')(1-\rho)^3}{\rho(1-\rho-\rho')^3}\;.$$

Consequently, if $\sqrt{k}\,A^2(n/k)\to\lambda_{\scriptscriptstyle A}$ *and* $\sqrt{k}\,A(n/k)\,B(n/k)\to\lambda_{\scriptscriptstyle B}$, *finite, then* $\sqrt{k}\,A(n/k)\,(\hat\rho(k;\tau)-\rho) \;\xrightarrow{d}\; N(\lambda_{\scriptscriptstyle A}\upsilon_1+\lambda_{\scriptscriptstyle B}\upsilon_2,\,\gamma^2\sigma_\rho^2)$.

Corollary 2.1. *Under the third order framework in* (2.1), *if* (1.4) *holds,* $\sqrt{k}\,A(n/k)\to\infty$ *and* $\sqrt{k}\,A(n/k)\,B(n/k)\to\lambda_{\scriptscriptstyle B}$, *finite, then* $\hat\rho_n(k;\tau)-\rho = O_p\bigl(1/(\sqrt{k}\,A(n/k))\bigr)$. *But, if we chose* k *such that* $\sqrt{k}\,A(n/k)\,B(n/k)\to\infty$, *then* $1/(\sqrt{k}\,A(n/k)) = o(B(n/k))$ *and* $\hat\rho_n(k;\tau)-\rho = O_p(B(n/k))$.

A comment on the choice of the tuning parameter τ**.** From Proposition 2.1, we can conclude that the tuning parameter τ only affects $\hat\rho(k;\tau)$ asymptotic bias. If $\rho'=\rho$, and consequentially $B(n/k)=O(A(n/k))$, we can always choose $\tau=\tau_0$ so that the asymptotic bias $\upsilon_1 A(n/k)+\upsilon_2 B(n/k)$ in (2.4) is null, even when $\sqrt{k}\,A^2(n/k)\to\lambda_A>0$ and $\sqrt{k}\,A(n/k)\,B(n/k)\to\lambda_B\neq 0$. It is enough to choose the value τ_0 which is the solution of $\upsilon_1\gamma\,\beta+\upsilon_2\beta'=0$. Such a value is independent of γ and, with $\xi=\beta'/\beta$, is given by

$$\tau_0 \;\equiv\; \tau_0(\rho,\xi) \;=\; \frac{-6\,\Bigl[4\xi\,(1-\rho^5) + \rho\,(1-2\rho)\,\bigl(4\,(2-\rho)\,(1-\rho)^2-1\bigr)\Bigr]}{(1-2\rho)^3\,(3-\rho)\,(3-2\rho)}\;. \tag{2.5}$$

Although τ_0, as a function of ρ, is not always monotone, it converges to $3(1-\xi/2)$, as $\rho\to-\infty$ and to $-8\xi/3$, as $\rho\to 0$.

Using the available values ρ, β and β', from Table 1, we have for the *Fréchet* model, $\rho = -1$, $\xi = 5/3$ and $\tau_0 = -217/270 \simeq -0.8$. For models like the *Burr* and the *GP*, where $\beta' = \beta$ and consequently $\xi = 1$, we present in Figure 1 (left) $\tau_0(\rho,1)$ as function of ρ. For *Student*'s t_ν distribution, ρ, β and β' are functions of ν, and the value τ_0 in (2.5) can also be written as a function of ν:

$$\tau_0(\nu) \;=\; \frac{12\left(384 + 1216\,\nu + 1440\,\nu^2 + 720\,\nu^3 + 72\,\nu^4 - 61\,\nu^5 - 21\,\nu^6 - 2\,\nu^7\right)}{(1+\nu)\,(4+\nu)^4\,(2+3\nu)\,(4+3\nu)}\;.$$

This function $\tau_0(\nu)$ is shown in Figure 1 (right), as a function of ν.

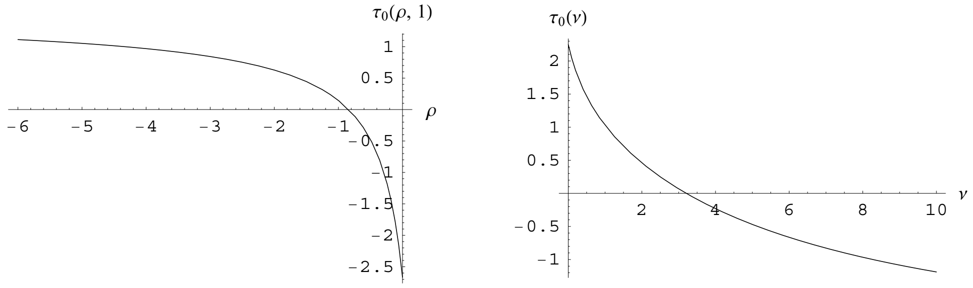


Figure 1: **Left:** $\tau_0(\rho, 1)$ as function of ρ. **Right:** $\tau_0(\nu)$ for Student's t_ν.

As an example, for the $GP(\gamma = 0.5)$, we have $\tau_0(-0.5, 1) = -213/448 \simeq -0.48$. In Figure 2 and to illustrate the comment above, we picture a sample path of $\hat\rho(k;\tau)$ with $\tau = \tau_0$ and $\tau = 0$, the value of τ most commonly suggested for models with $|\rho| < 1$. We conclude that $\hat\rho_{\tau_0}(k) = \hat\rho(k;\tau_0)$ is indeed more stable than $\hat\rho_0(k) = \hat\rho(k;0)$ around the true value $\rho = -0.5$.

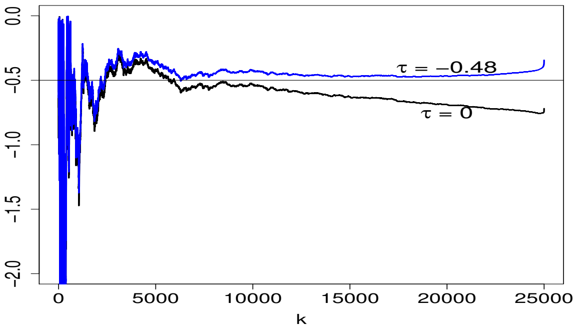


Figure 2: Sample path of the estimator $\widehat\rho(k;\tau)$, $\tau = 0, -0.48$, for one sample of size $n = 25000$ from the GP distribution with $\gamma = 0.5$.

Remark 2.2. Indeed, for an appropriate tuning parameter τ the ρ-estimators in (2.3) show highly stable sample paths as functions of k, the number of top o.s. used, for a wide range of large k-values. The theoretical and simulated results in Fraga Alves *et al.* (2003), together with the use of these estimators in different reduced-bias statistics, has led to advise in practice the estimation of ρ through the estimator in (2.3), computed at the value

$$k_1 := \left[n^{0.995}\right] , \tag{2.6}$$

not chosen in any optimal way, and the choice of the tuning parameter $\tau = 0$ for $\rho \in [-1, 0)$ and $\tau = 1$ for $\rho \in (-\infty, -1)$. As usual, $[x]$ denotes the integer part of x. However, practitioners should not choose blindly the value of τ in (2.3), and as pointed out in Caeiro and Gomes (2006), even negative values of τ should be possible candidates. It is indeed sensible to draw a few sample paths of $\hat\rho_\tau(k) = \hat\rho(k;\tau)$, as functions of k, electing the value of τ which provides the highest stability for large k, by means of any stability criterion, like the one suggested in Gomes *et al.* (2005) or Gomes and Pestana (2007a). For not too small n, we are frequently led to the above mentioned choice: $\hat\rho_0$ if $\rho \geq -1$ and $\hat\rho_1$ if $\rho < -1$, when we consider only the tuning parameters $\tau = 0$ and $\tau = 1$ as the possible alternatives. In practice, the adequate choice of τ is much more crucial than the choice of k_1, discussed in the following.

A few comments on the choice of the level k_1 for the ρ-estimation. On the basis of the results in Proposition 1.1 and Proposition 2.1, it seems sensible to estimate the second order ρ using a number k_1 of o.s.'s of a larger order than k, the number of o.s.'s used for the estimation of the tail index γ. We now make the following comments on the choice of the value k_1 that should be used for the estimation of the second order parameter ρ.

(1) The ideal situation would perhaps be the choice of an "optimal" k_1 for the estimation of ρ, in the sense of a value that enables the asymptotic normality of the ρ-estimator with a non-null asymptotic bias. For models in (2.1), k_1 is then such that $\sqrt{k_1}\, A(n/k_1)\, B(n/k_1) \to \lambda_{B_1}$, finite and non-null. We then get $k_1 = O\big(n^{-2(\rho+\rho')/(1-2(\rho+\rho'))}\big)$. Denoting $\hat\rho = \hat\rho(k_1;\tau)$ for any $\hat\rho(k;\tau)$ in (2.3), $\hat\rho - \rho$ is of the order of $1/\big(\sqrt{k_1}\, A(n/k_1)\big) = O\big(n^{\rho'/(1-2(\rho+\rho'))}\big) = o(1/\ln n)$, i.e.,

$$\hat\rho - \rho = o_p(1/\ln n), \qquad \text{as} \quad n \to \infty \ , \tag{2.7}$$

a condition needed later on. In practice, such a k_1 has only a "limited" interest at the current state-of-the-art. It is however of theoretical interest.

(2) Assume next the validity of the following condition:

Condition U: There exist a *tuning parameter* τ^* and a level k_1, with $\sqrt{k_1}\, A(n/k_1)\, B(n/k_1) \to \infty$, such that, with $\hat\rho(k;\tau)$ defined in (2.3), $\hat\rho^* - \rho = \hat\rho(k_1;\tau^*) - \rho = O_p\big(1/(\sqrt{k_1}\, A(n/k_1))\big)$.

This is obviously a strong assumption, practically equivalent to saying that for any specific model there is a τ^* and a k_1 such that $\hat{\rho}^* = \hat{\rho}(k_1;\tau^*)$ is an unbiased estimator for ρ, so that the bias has no influence in the rate of convergence, which is kept at $1/\big(\sqrt{k_1}\,A(n/k_1)\big)$. Indeed, such a claim is made on the basis of the high stability of sample paths of the ρ-estimates in (2.3) for a specific $\tau=\tau^*$ and large values of k (see Figure 2 and the comment made above on the choice of τ). Then, the use of a value k_1 larger than the so-called "optimal" level in item 1., but intermediate, like for instance, the one suggested in Gomes and Martins (2002),

$$k_1 := \min\big(n-1,\, 2n/\ln\ln n\big)\ , \tag{2.8}$$

enables us to guarantee that $\hat{\rho}^* - \rho = o_p(1/\ln n)$. Indeed, if we assume the validity of *Condition U* for k_1 in (2.8), we get $\hat{\rho}^*-\rho = O_p\big(1/(\sqrt{k_1}\,A(n/k_1))\big) = O_p\big((\ln\ln n)^{(1-2\rho)/2}/\sqrt{n}\big)$, which is obviously of smaller order than $\{1/\ln n\}$, i.e., (2.7) holds. This will be the unique situation under which we may work with the k_1 suggested in Gomes and Martins (2002), i.e, the one in (2.8), and still guarantee the above mentioned property on the ρ-estimator, and a possible generalization of the third-order results derived for $\widetilde{H}_{\beta,\rho}$ to $\widetilde{H}_{\hat{\beta}^*,\hat{\rho}^*}$, with $\hat{\beta}^*$ an adequate β-estimator, to be specified later on, in Section 2.2.2.

(3) If we consider a level k_1 of the order of $n^{1-\epsilon}$, for some small $\epsilon > 0$, we may also guarantee that (2.7) holds for a large class of models, without the need to assume a condition as strong as *Condition U*. This is the reason why, such as done in Caeiro *et al.* (2004b), Gomes and Pestana (2007a,b) and Gomes *et al.* (2004b, 2007a), we advise in practice, as a compromise between theoretical and practical considerations, the use of an intermediate level like the one in (2.6) or any other level $k_1 = [n^{1-\epsilon}]$ for some $\epsilon > 0$, small.

2.2.2. Estimation of the scale second order parameter β

Let us introduce the notation $N_n^{(\alpha)}(k) := \frac{1}{k}\sum_{i=1}^{k}\big(\frac{i}{k}\big)^{\alpha-1}U_i$, with U_i defined in (1.6). For the estimation of β we shall here consider the estimator in Gomes and Martins (2002), with the functional expression,

$$\hat{\beta}_{\hat{\rho}}(k) \;=\; \hat{\beta}(k;\hat{\rho}) \;:=\; \Big(\frac{k}{n}\Big)^{\hat{\rho}}\, \frac{\bigg(\frac{1}{k}\sum\limits_{i=1}^{k}\big(\frac{i}{k}\big)^{-\hat{\rho}}\bigg) N_n^{(1)}(k) \;-\; N_n^{(1-\hat{\rho})}(k)}{\bigg(\frac{1}{k}\sum\limits_{i=1}^{k}\big(\frac{i}{k}\big)^{-\hat{\rho}}\bigg) N_n^{(1-\hat{\rho})}(k) \;-\; N_n^{(1-2\hat{\rho})}(k)}\ . \tag{2.9}$$

Theorem 2.1 (Gomes *et al.*, 2004b)**.** *If the second order condition* (1.7) *holds, with* $A(t) = \gamma\,\beta\,t^\rho$, $\rho < 0$, *if* (1.4) *holds, and if* $\sqrt{k}\,A(n/k) \to \infty$, *then, with* $\hat\rho_n(k;\tau)$ *and* $\hat\beta_{\hat\rho}(k)$ *given in* (2.3) *and* (2.9), *respectively, and* $\hat\rho = \hat\rho_n(k;\tau)$ *such that* (2.7) *holds, i.e.,* $\hat\rho - \rho = o_p(1/\ln n)$, *as* $n \to \infty$, $\hat\beta_{\hat\rho}(k)$ *is consistent for the estimation of* β*. Moreover,*

$$\hat\beta_{\hat\rho}(k) - \beta \;\overset{p}{\sim}\; -\beta\,\ln(n/k)\,(\hat\rho-\rho) \;=\; o_p(1)\;. \tag{2.10}$$

2.3. Asymptotic properties of the tail index estimators, under a third order framework

We shall study now the asymptotic behaviour, under a third order framework, of the MVRB estimators $\overline{H}$ and $\overline{\overline{H}}$, generally denoted $\widetilde{H}$. We assume first that we know the two second order parameters β and ρ. Next we estimate both second-order parameters externally at a level k_1 of a larger order than the level k at which we compute the tail index.

Theorem 2.2.

(**a**) *Under the second order framework in* (1.8), *and for intermediate* k, *i.e., whenever* (1.4) *holds, we may write,*

$$\widetilde{H}_{\beta,\rho}(k) \;\overset{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\,Z_k + o_p\big(A(n/k)\big)\;, \tag{2.11}$$

where Z_k *is the asymptotically standard normal r.v. in* (1.9)*. Also, if we choose* k *such that* $\sqrt{k}\,A(n/k) \to \lambda$, *finite, as* $n \to \infty$, $\sqrt{k}\,\big(\widetilde{H}_{\beta,\rho}(k) - \gamma\big)$ *are asymptotically normal, with variance* γ^2 *and a null mean value, even if* $\lambda \neq 0$.

(**b**) *If we further assume* (2.1), *more information can be given for the term* $o_p(A(n/k))$, *and we get the asymptotic distributional representations:*

$$\overline{H}_{\beta,\rho}(k) \;\overset{d}{=}\; \gamma+\frac{\gamma}{\sqrt{k}}Z^*_k+\frac{A(n/k)\,B(n/k)}{1-\rho-\rho'}\left(1-\frac{(1-\rho-\rho')\,A(n/k)}{\gamma\,(1-\rho)^2\,B(n/k)}\right)\big(1+o_p(1)\big)\,, \tag{2.12}$$

and

$$\overline{\overline{H}}_{\beta,\rho}(k) \;\overset{d}{=}\; \gamma+\frac{\gamma}{\sqrt{k}}Z^*_k+\frac{A(n/k)\,B(n/k)}{1-\rho-\rho'}\left(1-\frac{(1-\rho-\rho')\,A(n/k)}{2\gamma\,(1-\rho)^2\,B(n/k)}\right)\big(1+o_p(1)\big)\,, \tag{2.13}$$

with Z^*_k *asymptotically standard normal. If* $\sqrt{k}\,A(n/k)\,B(n/k) \to \lambda_{_B}$, *finite (and then,* $\sqrt{k}\,A^2(n/k) \to \lambda_{_A}$, *also finite),* $\sqrt{k}\big(\overline{H}_{\beta,\rho}(k)-\gamma\big)$ *and* $\sqrt{k}\big(\overline{\overline{H}}_{\beta,\rho}(k)-\gamma\big)$ *are asymptotically normal with the same variance, equal to* γ^2, *and asymptotic bias* $b_{\overline{H}} = \lambda_{_B}/(1-\rho-\rho')-\lambda_{_A}/(\gamma(1-\rho)^2)$ *and* $b_{\overline{\overline{H}}} = \lambda_{_B}/(1-\rho-\rho')-\lambda_{_A}/(2\gamma\,(1-\rho)^2)$, *respectively.*

Proof: The first part of the theorem has been proved in Caeiro *et al.* (2005). Regarding the second part: from the third order set-up in (2.2), we get

$$H(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k + \frac{A(n/k)}{1-\rho} + O_p\!\left(\frac{A(n/k)}{\sqrt{k}}\right) + \frac{A(n/k)\,B(n/k)}{1-\rho-\rho'}\,\big(1+o_p(1)\big)\;.$$

Consequently, as $\overline{H}_{\beta,\rho}(k) = H(k) \times \big(1 - A(n/k)/(\gamma\,(1-\rho))\big)$ for models in (2.1),

$$\overline{H}_{\beta,\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k + \left(\frac{A(n/k)\,B(n/k)}{1-\rho-\rho'} - \frac{A^2(n/k)}{\gamma\,(1-\rho)^2} + O_p\!\left(\frac{A(n/k)}{\sqrt{k}}\right)\right)\big(1+o_p(1)\big)\;,$$

$\overline{\overline{H}}_{\beta,\rho}(k) - \overline{H}_{\beta,\rho}(k) \overset{p}{\sim} A^2(n/k)/(2\gamma\,(1-\rho)^2)$, and the results in the theorem follow. Note that since $\sqrt{k}\, O_p\big(A(n/k)/\sqrt{k}\big) \to 0$, for the intermediate levels k considered, the term $O_p\big(A(n/k)/\sqrt{k}\big)$ is irrelevant for the asymptotic bias. □

Remark 2.3. Notice that $\overline{H}$ and $\overline{\overline{H}}$ have the same asymptotic variance and $b_{\overline{\overline{H}}} = b_{\overline{H}} + \lambda_A/(2\gamma\,(1-\rho)^2)$, with $\lambda_A \geq 0$. So if both bias are positive, $\overline{H}$ should have, asymptotically, a better performance than $\overline{\overline{H}}$.

Theorem 2.3.

(**a**) *Under the initial conditions of Theorem 2.2, let us consider the tail index estimators* $\widetilde{H}_{\hat\beta,\hat\rho}$ *with* $\hat\beta$ *and* $\hat\rho$ *consistent for the estimation of* β *and* ρ*, respectively, both computed at the level* k_1 *of a larger order than the level* k *at which we compute the tail index, and such that* (2.7) *holds. Then* $\sqrt{k}\,\big(\widetilde{H}_{\hat\beta,\hat\rho}(k) - \gamma\big)$ *are asymptotically normal, with variance equal to* γ^2 *and a null mean value, even if* $\sqrt{k}\, A(n/k) \to \lambda \neq 0$*, as* $n \to \infty$.

(**b**) *If we work under the third order framework in* (2.1)*, consider* $\hat\beta_{\hat\rho}(k)$ *in* (2.9)*,* $\hat\beta = \hat\beta_{\hat\rho}(k_1)$*, and choose* k *such that* $\sqrt{k}\, A(n/k) \to \infty$*, but* $\sqrt{k}\, A(n/k)\, B(n/k) \to \lambda_{\scriptscriptstyle B}$*, finite, then* $\sqrt{k}\,\big(\overline{H}_{\hat\beta,\hat\rho}(k) - \gamma\big)$ *and* $\sqrt{k}\,\big(\overline{\overline{H}}_{\hat\beta,\hat\rho}(k) - \gamma\big)$ *are asymptotically normal with variance* γ^2 *and asymptotic bias* $b_{\overline{H}}$ *and* $b_{\overline{\overline{H}}}$*, respectively, given in Theorem 2.2, provided that we can guarantee that* $(\hat\rho - \rho)\ln n = o_p\big(1/\sqrt{k}\, A(n/k)\big)$*. This last condition on* $\hat\rho$ *holds if we further assume the validity of Condition U for* k_1 *in* (2.8).

Proof: If we estimate consistently β and ρ through $\hat\beta$ and $\hat\rho$ under the conditions of the theorem, we may use Taylor's expansion series, and as $\partial\widetilde{H}_{\beta,\rho}/\partial\beta \overset{p}{\sim} A(n/k)/(\beta(1-\rho))$, $\partial\widetilde{H}_{\beta,\rho}/\partial\rho \overset{p}{\sim} -A(n/k)\,\big(\ln(n/k) + 1/(1-\rho)\big)/(1-\rho)$, we get

$$\widetilde{H}_{\hat\beta,\hat\rho}(k) - \widetilde{H}_{\beta,\rho}(k) \overset{p}{\sim} -\frac{A(n/k)}{1-\rho}\left\{\frac{\hat\beta-\beta}{\beta} + (\hat\rho-\rho)\left[\ln(n/k) + \frac{1}{1-\rho}\right]\right\}. \tag{2.14}$$

The first part of the theorem, related to levels k such that $\sqrt{k}\, A(n/k) \to \lambda$, finite, follows thus straightforwardly from (2.14).

Next, from (2.10), $(\hat\beta-\beta)/\beta \overset{p}{\sim} -\ln(n/k_1)\,(\hat\rho-\rho) = o_p\big(1/(\sqrt{k}\,A(n/k))\big)$, $\sqrt{k}\,\big(\widetilde{H}_{\hat\beta,\hat\rho}(k)-\widetilde{H}_{\beta,\rho}(k)\big) = o_p\big(1/\sqrt{k}\big)$ and the stated asymptotic normality of $\widetilde{H}_{\hat\beta,\hat\rho}$ follows as well. We may further write

$$\widetilde{H}_{\hat\beta,\hat\rho}(k)-\widetilde{H}_{\beta,\rho}(k) \overset{p}{\sim} -\frac{A(n/k)}{1-\rho}\,(\hat\rho-\rho)\left(\ln(k/k_1)+\frac{1}{1-\rho}\right). \tag{2.15}$$

If we assume the validity of *Condition U* for the level k_1 in (2.8) and consider $\widetilde{H}_{\hat\beta^*,\hat\rho^*}$, we straighforwardly guarantee that $\sqrt{k}\,(\hat\rho^*-\rho)\,A(n/k)\ln(k/k_1) = o_p(1)$. Consequently, the use of (2.15), with $(\hat\beta,\hat\rho)$ replaced by $(\hat\beta^*,\hat\rho^*)$, enables us to get the results in the theorem. □

2.4. Asymptotic properties of the reduced-bias quantile estimators, under a third order framework

We shall provide in theorems 2.4 and 2.5 the distributional behaviour of the quantile estimators under study, for models in (2.1).

Theorem 2.4. *Under the third order framework in* (2.1)*, for intermediate* k*, i.e., whenever* (1.4) *holds, and whenever* $\ln(np)=o(\sqrt{k})$*, we can write,*

$$\begin{aligned} Q^{(p)}_{H(k)}(k)/\chi_{1-p} &\overset{d}{=} 1+\big(H(k)-\gamma\big)\ln c_n+\frac{\gamma}{\sqrt{k}}\,B_k-h_\rho(c_n)\,A(n/k)+O_p\!\left(\frac{A(n/k)}{\sqrt{k}}\right)\\ &\quad -\left(h_{\rho+\rho'}(c_n)\,A(n/k)\,B(n/k)+\frac{1}{2}\,h^2_\rho(c_n)\,A^2(n/k)\right)\big(1+o_p(1)\big)\,, \end{aligned} \tag{2.16}$$

where B_k *is an asymptotically standard normal r.v.,* $h_\theta(x)=(x^\theta-1)/\theta$*,* $\theta<0$*. Consequently, if* $\sqrt{k}\,A(n/k)\to\lambda$*, finite, and* $\ln c_n/\sqrt{k}\to 0$*, as* $n\to\infty$*, then* $\frac{\sqrt{k}}{\ln c_n}\big(Q^{(p)}_{H(k)}(k)/\chi_{1-p}-1\big)$ *has asymptotically the same distribution as* $\sqrt{k}\,(H(k)-\gamma)$*, i.e., it is asymptotically normal, with variance* γ^2 *and mean value* $\lambda/(1-\rho)$*.*

Proof: From (2.2), and as $t\to\infty$, we get,

$$\frac{U(tx)}{U(t)} = x^\gamma\left\{1+h_\rho(x)\,A(t)+\Big(h_{\rho+\rho'}(x)\,A(t)\,B(t)+\frac{1}{2}\,h^2_\rho(x)\,A^2(t)\Big)\big(1+o(1)\big)\right\}. \tag{2.17}$$

Denoting by $\hat\gamma$ any consistent tail index estimator and since $X_{n-k:n}\overset{d}{=}U(Y_{n-k:n})$, where Y is a standard Pareto r.v., we can write

$$Q^{(p)}_{\hat\gamma(k)}(k)/\chi_{1-p} = \left(\frac{X_{n-k:n}}{U(1/p)}\right)c_n^{\hat\gamma(k)} = \left(\frac{X_{n-k:n}}{U(n/k)}\right)\left(\frac{U(n/k)}{U(nc_n/k)}\right)c_n^{\hat\gamma(k)}\,.$$

Using the delta method, together with the fact that $\ln c_n/\sqrt{k} \to 0$, as $n \to \infty$, $c_n^{\hat\gamma(k)} \overset{p}{\sim} c_n^\gamma\{1+(\hat\gamma(k)-\gamma)\ln c_n\}$. From (2.17), we obtain

$$\begin{aligned} Q^{(p)}_{H(k)}(k)/\chi_{1-p} &\overset{d}{=} \left(1+\frac{\gamma}{\sqrt{k}}B_k + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right)\right) \\ &\times \left\{1-h_\rho(c_n)\,A(n/k) - \left(h_{\rho+\rho'}(c_n)\,A(n/k)\,B(n/k) + h^2_\rho(c_n)\,A^2(n/k)/2\right)\left(1+o_p(1)\right)\right\} \\ &\times \left(1+\left(\hat\gamma(k)-\gamma\right)\ln c_n\right)\left(1+o_p(1)\right)\,, \end{aligned}$$

and, with $\hat\gamma$ replaced by H, (2.16) as well as the asymptotic normality follow. □

Theorem 2.5.

(**a**) *Under the conditions of Theorem 2.4, let us consider the tail index estimator $\widetilde H = \widetilde H_{\hat\beta,\hat\rho}$ with $(\hat\beta,\hat\rho)$ consistent estimators of (β,ρ), both computed at k_1, with $k=o(k_1)$ and such that $(\hat\rho-\rho)\ln n = o_p(1)$. Then, if $\sqrt{k}\,A(n/k)\to\lambda$, $\frac{\sqrt{k}}{\ln c_n}\big(\overline{Q}^{(p)}_{\widetilde H(k)}(k)/\chi_{1-p}-1\big)$ has asymptotically the same distribution as $\sqrt{k}\,\big(\widetilde H(k)-\gamma\big)$, i.e., they are both asymptotically normal, with variance equal to γ^2 and a null mean value (even if $\lambda\neq 0$).*

(**b**) *If we choose k such that $\sqrt{k}\,A(n/k)\to\infty$, but $\sqrt{k}\,A(n/k)\,B(n/k)\to\lambda_{{}_B}$, finite, $\frac{\sqrt{k}}{\ln c_n}\big(\overline{Q}^{(p)}_{\widetilde H(k)}(k)/\chi_{1-p}-1\big)$ and $\sqrt{k}\,\big(\widetilde H(k)-\gamma\big)$ also have asymptotically the same distributions, i.e., they are asymptotically normal, with variance equal to γ^2 and asymptotic bias given in Theorem 2.2, provided that we can guarantee that $(\hat\rho-\rho)\ln n = o_p\big(1/\sqrt{k}\,A(n/k)\big)$.*

Proof: Let as first assume to know β and ρ. Then, since $\overline{Q}^{(p)}_{\widetilde H_{\beta,\rho}}(k;\beta,\rho) = Q^{(p)}_{\widetilde H_{\beta,\rho}}(k)\left(1+\widetilde H_{\beta,\rho}(k)\,\beta\big(\tfrac{n}{k}\big)^\rho h_\rho(c_n)\right)$ for models in (2.1), we can use (2.16) and get

$$\begin{aligned} \overline{Q}^{(p)}_{\widetilde H_{\beta,\rho}}(k;\beta,\rho)/\chi_{1-p} \overset{d}{=}\; & 1+\left(H_{\beta,\rho}(k)-\gamma\right)\ln c_n + \frac{\gamma}{\sqrt{k}}B_k + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right) \\ & - \left(h_{\rho+\rho'}(c_n)\,A(n/k)\,B(n/k) + \frac{1}{2}\,h^2_\rho(c_n)\,A^2(n/k)\right)\left(1+o_p(1)\right)\,, \end{aligned} \tag{2.18}$$

Then $\frac{\sqrt{k}}{\ln c_n}\big(\overline{Q}^{(p)}_{\widetilde H_{\beta,\rho}(k;\beta,\rho)}(k)/\chi_{1-p}-1\big)$ has asymptotically the same distributions as $\sqrt{k}\,\big(\widetilde H_{\beta,\rho}(k)-\gamma\big)$. Since, $\widetilde H_{\hat\beta,\hat\rho}(k) = \gamma\,(1+o_p(1))$, $c_n^\rho\to 0$, $c_n^\rho\ln c_n\to 0$, for any intermediate k, we may use Cramer's delta-method, and write

$$\widetilde H_{\hat\beta,\hat\rho}(k)\,\hat\beta\left(\frac{n}{k}\right)^{\hat\rho} h_{\hat\rho}(c_n) \;\overset{p}{\sim}\; h_\rho(c_n)\,A(n/k)\left\{1+\frac{\hat\beta-\beta}{\beta}+(\hat\rho-\rho)\,\ln(n/k)\right\}.$$

Consequently,

$$\left(\overline{Q}^{(p)}_{\widetilde H_{\hat\beta,\hat\rho}}(k;\beta,\rho) - \overline{Q}^{(p)}_{\widetilde H_{\beta,\rho}}(k;\beta,\rho)\right)/\chi_{1-p} \;\overset{p}{\sim}\; \left(\widetilde H_{\hat\beta,\hat\rho}(k) - \widetilde H_{\beta,\rho}(k)\right)\ln c_n$$

and

$$\left(\overline{Q}^{(p)}_{\widetilde{H}_{\hat{\beta},\hat{\rho}}}(k;\hat{\beta},\hat{\rho}) - \overline{Q}^{(p)}_{\widetilde{H}_{\beta,\rho}}(k;\beta,\rho)\right)\Big/\chi_{1-p} \;\overset{p}{\sim}$$

$$\overset{p}{\sim}\; \left(\widetilde{H}_{\hat{\beta},\hat{\rho}}(k) - \widetilde{H}_{\beta,\rho}(k)\right)\ln c_n + h_\rho(c_n)\,A(n/k)\left\{\frac{\hat{\beta}-\beta}{\beta} + (\hat{\rho}-\rho)\,\ln(n/k)\right\}.$$

The remaining of the proof is analogous to the proof of Theorem 2.3. □

3. A SMALL-SCALE SIMULATION STUDY

We have implemented, for *Fréchet* underlying parents, a Monte Carlo simulation of size 5000 for $R_H \equiv Q^{(p)}_H/\chi_{1-p}$, $\overline{R}_{\overline{H}} \equiv \overline{Q}^{(p)}_{\overline{H}}/\chi_{1-p}$ and $\overline{R}_{\overline{\overline{H}}} \equiv \overline{Q}^{(p)}_{\overline{\overline{H}}}/\chi_{1-p}$. Results for $\overline{\overline{Q}}$, not presented, have also been simulated and almost overlap the ones for $\overline{Q}$. For every estimator $R = R(k)$, we have simulated for $p = 1/n$ and $p = 1/(n\ln n)$, the mean value, the root mean squared error (RMSE) and the optimal sample fraction, $OSF_R = k_0/n = \arg\min_k\{RMSE(R(k))\}/n$. The second order parameters were estimated through $\hat{\rho}_0 = \hat{\rho}(k_1;0)$ and $\hat{\beta}_0 = \hat{\beta}_{\hat{\rho}_0}(k_1)$, with $\hat{\rho}(k;\tau)$ and $\hat{\beta}_{\hat{\rho}}(k)$ defined in (2.3) and (2.9), respectively, and k_1 given in (2.6).

Table 2: Simulated mean values /RMSE at optimal levels.

n	100	500	1000	5000
Fréchet parent with $\gamma=0.25$ and $p=1/n$				
R_H	1.056 / 0.191	1.053 / 0.136	1.053 / 0.118	1.037 / 0.080
$\overline{R}_{\overline{H}}$	0.969 / 0.164	0.984 / 0.116	0.988 / 0.099	0.992 / 0.061
$\overline{R}_{\overline{\overline{H}}}$	1.007 / 0.154	1.006 / 0.108	1.004 / 0.092	1.004 / 0.057
Fréchet parent with $\gamma=0.25$ and $p=1/(n\ln n)$				
R_H	1.106 / 0.298	1.089 / 0.259	1.085 / 0.172	1.057 / 0.112
$\overline{R}_{\overline{H}}$	0.960 / 0.236	0.984 / 0.162	0.988 / 0.135	0.991 / 0.080
$\overline{R}_{\overline{\overline{H}}}$	1.009 / 0.224	1.013 / 0.152	1.009 / 0.127	1.009 / 0.076

A few remarks for Fréchet parents:

- For *Fréchet* parents, the RMSE of $\overline{R}_{\overline{H}(k)}(k)$ and $\overline{R}_{\overline{\overline{H}}(k)}(k)$ is always smaller (or equal) than the RMSE of the classical quantil estimator, $R_{H(k)}(k)$.
- Also, the normalized quantile estimator $\overline{R}_{\overline{\overline{H}}(k)}(k)$ has always the smallest mean squared error.

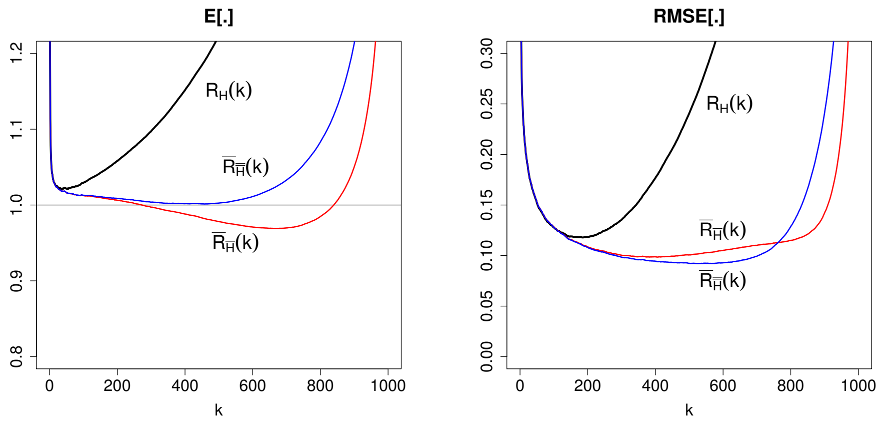


Figure 3: Underlying Fréchet parent with $\gamma = 0.25$, $p = 1/n$, and $n = 1000$.

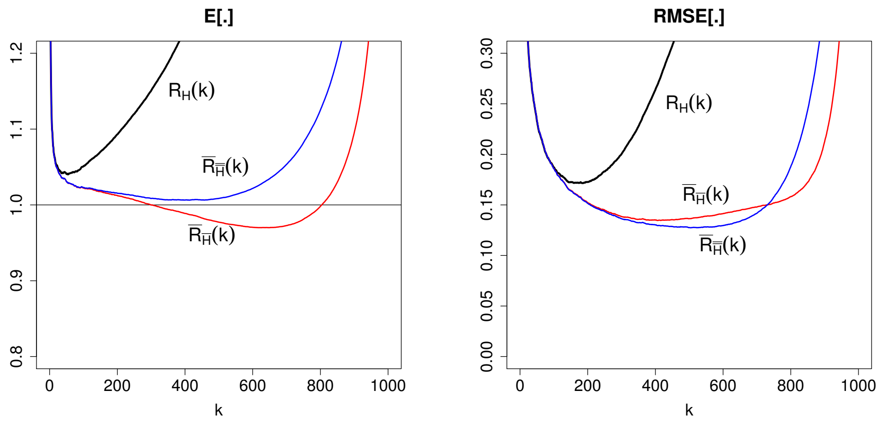


Figure 4: Underlying Fréchet parent with $\gamma = 0.25$, $p = 1/(n \ln n)$ and $n = 1000$.

4. CONCLUSION

The MVRB estimators proposed in this paper are bias-corrected Hill estimators which perform better than the classical Hill estimator for all k, the number of top o.s.'s used in the estimation of the tail index γ. Despite of this, it is sensible to understand their comparative behaviour at optimal levels, not only for finite sample size, but also asymptotically, as recently done in Gomes and Neves (2007) for some of the classical estimators, like the well-known Hill, moment, maximum likelihood and the recently introduced mixed moment estimator (Fraga Alves *et al.*, 2007). It is thus crucial to have information on the order of the dominant component(s) of their asymptotic bias, the main contribution in this paper, for the MVRB tail index estimators in (1.10) and the associated quantile estimators in (1.13). The adaptive choice of the threshold is now becoming feasible for a wide class of models, but it is outside of the scope of this paper.

ACKNOWLEDGMENTS

Research partially supported by FCT/POCTI and POCI/FEDER.

REFERENCES

[1] Beirlant, J.; Dierckx, G.; Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**(2), 177–200.

[2] Beirlant, J.; Figueiredo, F.; Gomes, M.I. and Vandewalle, B. (2006). Improved reduced-bias tail index and quantile estimators, *J. Statist. Plann. and Inference*, DOI:10.1016/j.jspi.2007.07.015, 2007.

[3] Caeiro, C. and Gomes, M.I. (2006). A new class of estimators of a scale second order parameter, *Extremes*, **9**, 193–211.

[4] Caeiro, C. and Gomes, M.I. (2007). *Semi-parametric second-order reduced-bias high quantile estimation*, Pré-Publicações CMA-UNL 8/2007 (submitted).

[5] Caeiro, C.; Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator, *Revstat*, **3**(2), 113–136.

[6] Ferreira, A.; Haan, L. de and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution, *Statistics*, **37**(5), 401–434.

[7] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *Ann. Statist.*, **27**, 760–781.

[8] Fraga Alves, M.I.; Gomes, M.I. and Haan, L. de (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**(2), 193–213.

[9] Fraga Alves, M.I.; Gomes, M.I.; Haan, L. de and Neves, C. (2007). *Mixed moment estimator and location invariant alternatives*, Notas e Comunicações CEAUL 14/2007 (submitted).

[10] Geluk, J. and Haan, L. de (1980). *Regular Variation, Extensions and Tauberian Theorems*, CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

[11] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.*, **44**(6), 423–453.

[12] Gomes, M.I.; Caeiro, F. and Figueiredo, F. (2004a). Bias reduction of a extreme value index estimator trough an external estimation of the second order parameter, *Statistics*, **38**(6), 497–510.

[13] Gomes, M.I. and Figueiredo, F. (2006). Bias reduction in risk modelling: semi-parametric quantile estimation, *Test*, **15**(2), 375–396.

[14] Gomes, M.I.; Figueiredo, F. and Mendonça, S. (2005). Asymptotically best linear unbiased tail estimators under a second order regular variation, *J. Statist. Planning and Inference*, **134**(2), 409–433.

[15] Gomes, M.I.; de Haan, L. and Henriques Rodrigues, L. (2004b). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses, *J. Royal Statistical Society B*, DOI: 10.1111/j.1467-9869.2007.00620.x, 2007.

[16] Gomes, M.I.; Haan, L. de and Peng, L. (2002). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour, *Extremes*, **5**(4), 387–414.

[17] 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.

[18] Gomes, M.I.; Martins, M.J. and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology, *Extremes*, **3**(3), 207–229.

[19] Gomes, M.I.; Martins, M.J. and Neves, M. (2007a). Improving second order reduced bias extreme value index estimation, *Revstat*, **5**(2), 177–207.

[20] Gomes, M.I. and Neves, C. (2007). Asymptotic comparison of the mixed moment and classical extreme value index estimators, *Statistics and Probability Letters*, DOI: 10.1016/j,spl.2007.07.026, 2007.

[21] Gomes, M.I. and Oliveira, O. (2003). How can non-invariant statistics work in our benefit in the semi-parametric estimation of parameters of rare events, *Comm. in Statist. — Simulation and Computation*, **32**(4), 1005–1028.

[22] Gomes, M.I. and Pestana, D. (2007a). A simple second order reduced bias' tail index estimator, *J. Statist. Comput. and Simulation*, **77**(6), 487–504.

[23] Gomes, M.I. and Pestana, D. (2007b). A sturdy reduced bias extreme quantile (VaR) estimator, *J. Amer. Statist. Assoc.*, **102**, 477, 280–292.

[24] Haan, L. de (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Amsterdam.

[25] Haan, L. de and Peng, L. (1998). Comparison of tail index estimators, *Statistica Neerlandica*, **52**, 60–70.

[26] Haan, L. de and Rootzén, H. (1983). On the estimation of high quantiles, *J. Statist. Planning and Inference*, **35**, 1–13.

[27] Hall, P. (1982). On some simple estimates of an exponent of regular variation, *J. R. Statist. Soc.*, **44**(1), 37–42.

[28] Hall, P. and Welsh, A.H. (1985). Adaptative estimates of parameters of regular variation, *Ann. Statist.*, **13**, 331–341.

[29] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**(5), 1163–1174.

[30] Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models, *Statistica Sinica*, **13**, 853–880.

[31] Matthys, G.; Delafosse, M.; Guillou, A. and Beirlant, J. (2004). Estimating catastrophic quantile levels for heavy-tailed distributions, *Insurance: Mathematics and Economics*, **34**, 517–537.

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

[33] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.

REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008, 21–32

A GEOSTATISTICAL EXPLORATORY ANALYSIS OF PRECIPITATION EXTREMES IN SOUTHERN PORTUGAL

Authors: Ana Cristina Costa
– ISEGI, Universidade Nova de Lisboa, Portugal
ccosta@isegi.unl.pt

Rita Durão
– CERENA, Instituto Superior Técnico, Portugal
rmdurao@ist.utl.pt

Amílcar Soares
– CERENA, Instituto Superior Técnico, Portugal
ncmrp@alfa.ist.utl.pt

Maria João Pereira
– CERENA, Instituto Superior Técnico, Portugal
maria.pereira@ist.utl.pt

Abstract:

- In Mediterranean climate regions, prolonged periods of unusually dry conditions reduce the availability of water resources and affect vegetation cover; while other areas can be affected by an increase in the number of heavy precipitation events, with an increase in the flood risk. Issues such as drought and erosive rainfall have been raising concern about the risks of land degradation and desertification. The main objective of this paper is to provide an insight of the geographic distribution of extreme precipitation events in the Southern region of continental Portugal, as a basis for a future study of the relationships between extreme rainfall patterns, both spatial and temporal, and desertification processes. The data used in this study are a set of 105 station records with daily precipitation observations for the period 1940–1999. This 60-year period was chosen to optimize data availability across the region, taking into consideration the quality control analysis performed. Among the numerous indices of extreme precipitation described in the literature, we selected three of them for an exploratory analysis: one index representing dry conditions, another one representing extremely heavy precipitation events and another index representing flood events. For each of these three indices, yearly trends and decadal space-time patterns are investigated. The results show no significant trends in the regional extreme indices. The geostatistical study concluded that the spatial patterns are more continuous in the last decade than the other ones before. The preliminary results of this study agree with other similar studies of the same region reported in the literature.

Key-Words:

- *indices of precipitation extremes; geostatistical approach; southern Portugal.*

AMS Subject Classification:

- 86A32, 62P12.

1. INTRODUCTION

In Mediterranean climate regions, prolonged periods of unusually dry conditions reduce the availability of water resources and affect vegetation cover; while other areas can be affected by an increase in the number of heavy precipitation events, with an increase in the flood risk ([9]). Issues such as drought and erosive rainfall have been raising concern about the risks of land degradation and desertification ([13]). The main objective of this paper is to provide an insight of the geographic distribution of extreme precipitation events in the southern region of continental Portugal, as a basis for a future study of the relationships between extreme rainfall patterns, both spatial and temporal, and desertification processes.

The data used in this study are a set of 105 station records with daily precipitation observations for the period 1940–1999. This 60-year period was chosen to optimize data availability across the region, taking into consideration the quality control analysis performed. Among the numerous indices of extreme precipitation described in the literature, we selected three of them for an exploratory analysis: one index representing dry conditions, another one representing extremely heavy precipitation events and another index representing flood events.

Most of the studies analysing extreme precipitation indices focus on temporal linear trends, rather than space-time patterns, because many of them aim to assess climate changes, whereas for others a spatial analysis is not feasible due to the sparse number of monitoring stations over large study regions (e.g. [11]). However, that kind of analysis is extremely important for impact studies related with the desertification phenomenon and therefore, for each of the three extreme precipitation indices calculated, yearly trends and decadal space-time patterns are investigated.

2. STUDY AREA AND DATA

The study area is located in the South of continental Portugal, and an original set of 106 monitoring stations with daily precipitation data was selected. Most of them were extracted from the National System of Water Resources Information (SNIRH Sistema Nacional de Informao de Recursos Hdricos) database (`http://snirh.inag.pt`), and three of them were compiled from the European Climate Assessment (ECA) dataset (`http://eca.knmi.nl`). Each station series data was quality controlled by several procedures: gross error check (e.g. check negative precipitation and non-existent dates); records flagging using several criteria (data outlying pre-fixed thresholds and graphical analysis); cross control among highly correlated series (consider as possible errors the data that markedly disagree with the rainfall in other stations highly correlated); set to missing the observations considered as erroneous or extremely suspicious; "flat line" check procedures, which identify data of the same value for at least three consecutive days (not applied to zero precipitation data).

Furthermore, for each station, annual precipitation series were computed and studied for homogeneity through the application of six statistical tests, by means of the hybrid approach proposed by Wijngaard *et al.* ([15]). In addition, 62% of the long-term series were also checked through relative approaches (testing procedures that use records from reference stations), comprising the application of five homogeneity tests which are capable of locating the year where a break is likely ([2]; [3]). One station's data was then rejected because multiple breakpoints were identified and the homogeneous periods were too short and unreliable. Hence, the data used in this study are a set of 105 station records with daily precipitation observations within the period 1940–1999 (Figure 1). Only the longest homogeneous period was used to build the extreme precipitation indices whenever at least one of the relative tests identified a break year.

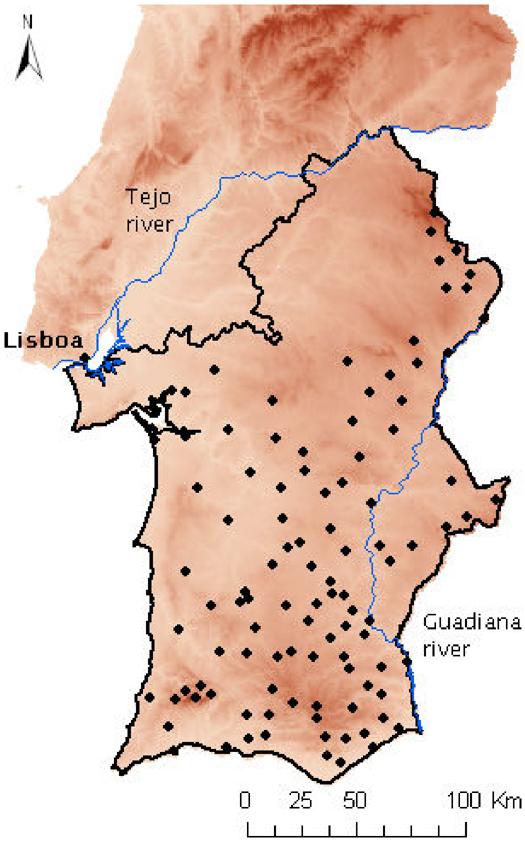


Figure 1: Study domain and stations' locations with selected daily precipitation series.

The extreme precipitation indices are sensitive to the number of missing days, thus the selected stations satisfy the following criterion. The daily records are as complete as possible, with less than 16% of data missing in each year. Hence, for each station, the indices for a specific year were set to missing if there were more than 16% of the days missing for that year ([8]).

3. EXTREME PRECIPITATION INDICES

Numerous extreme precipitation indices are described and analyzed in the literature. Some indices involve arbitrary fixed thresholds, such as the number of days per year with daily precipitation exceeding 10mm or 20mm (e.g. [11]; [12]). Other indices are based on statistical quantities such as percentiles, which are more appropriate for regions that contain a broad range of climates ([7]; [11]). Indices based on the count of days crossing certain fixed thresholds are beneficial for impact studies as they can be related with extreme events that affect human society and the natural environment ([11]). Since numerous extreme rainfall indices described in the literature are largely inter-correlated, we selected one index representing dry conditions (RL10) and two for wet conditions (R30 and R5D).

The index RL10 is defined as the number of days per year with precipitation amount below 10mm, thus measures the frequency of dry events. The index R30 measures the frequency of extremely heavy precipitation events and is defined as the number of days per year with precipitation amount above or equal to 30mm. The index R5D is defined as the highest consecutive 5-day precipitation total in each year and provides a measure (in mm) of the magnitude of strong precipitation events. In the present study only annually specified indices are considered.

4. METHODOLOGY FOR ASSESSING TRENDS AND SPACE-TIME PATTERNS IN PRECIPITATION EXTREMES

The extreme precipitation indices were calculated for each station and were then averaged over all the stations to obtain the regional-average of each extreme index per year. The average number of stations series used by year to build the regional-average series is equal to 47. The slopes of the trends in the indices of precipitation extremes were calculated by least squares linear fitting and trends' significance determined using Student's t-tests. The regression coefficient b (slope) multiplied by 10 gives the change per decade ([11]).

The spatial interpolation of precipitation has been the focus of much research (e.g. [14]; [6]; [4]). However, the number of studies analysing space-time patterns of extreme precipitation indices is very limited, as the large majority of the studies focus on the temporal linear trends in the indices. An exception is the work of Hundecha and Bárdossy ([10]) in which the station values of the daily precipitation were interpolated on a 5 km × 5 km grid through external drift kriging, using a digital elevation model as secondary information and, afterwards, several extreme precipitation indices were calculated on grids of 5, 10, 25 and 50 km.

Geostatistical estimators, known as kriging (family of generalized least-squares regression algorithms), provide statistically unbiased estimates of surface values from a set of observations at recorded locations, using the estimated spatial and temporal covariance model of the observed data. Stationarity assumptions on kriging are traditionally accounted for by using local search neighbourhoods so that the dependence on stationarity becomes local ([5]). The most commonly applied forms of kriging use a variogram — inverse function of the spatial (and temporal) covariance. This is a key function of geostatistics and represents the variability of the spatial and temporal patterns of physical phenomena. Usually, a mathematical variogram model is fitted to the empirical semi-variogram values (experimental semi-variogram) calculated for given angular and distance classes. The most common models are the linear, spherical, exponential and Gaussian models ([5]). These models are known as *transitive* variograms, because the spatial correlation structure varies with the distance h.

The parameters of the variogram model (sill, range and nugget) are then used to assign optimal weights for spatial prediction using kriging. The nugget is determined when h approaches 0. The nugget effect results from high variability at short distances that can be caused by lack of samples, or sampling inaccuracy. The range is one of the most important parameters as it is related with the spatial (or space-time) extent of continuity of the phenomenon. For the case of a spherical model, the range of the variogram is the distance h beyond which the variance no longer shows spatial dependence. At h, the sill value is reached. Observations separated by a distance larger than the range are spatially independent observations.

In this study, the extreme precipitation indices were calculated for each station using data within the baseline period 1940–1999. Afterwards, the space-time patterns of the indices were assessed through ordinary kriging on a 800 m × 800 m grid, for each year, using a different space-time variogram model for each decade. The way in which the variogram models are chosen and their parameters are estimated is controversial ([5]). In this study we chose exponential models that capture the major spatial features of each attribute under study by subjectively fitting the models to the experimental semi-variogram values taking into account physical knowledge of the area and phenomenon.

It is recognized that topography and other geographical factors are responsible for considerable spatial heterogeneity of the precipitation distribution at the sub-regional scale. Precipitation generally increases with elevation because of the orographic effect of mountainous terrain. Ordinary kringing was preferred over other methods incorporating secondary information (e.g., altitude or distance to the coastline) such as cokriging, because of the poor linear relationships (assessed through Pearson's correlation coefficients) between the three indices and variables such as the elevation and the geographical coordinates of the stations' locations.

5. RESULTS AND DISCUSSION

5.1. Time trend analysis

The frequency of dry events, measured by the regional average of RL10 within the period 1940–99, is increasing (the change per decade is equal to 0.582 days), but the trend is not statistically significant (the *p*-value of the *t*-test is equal to 0.214). The regional average of R30 shows a decrease in the frequency of extremely heavy precipitation events (the change per decade is equal to -0.009 days) but it is not statistically significant (the p-value of the t-test is equal to 0.937). The magnitude of strong precipitation events, measured by the regional average of R5D, is also decreasing (the change per decade is equal to -0.132 mm), but the trend is not statistically significant (the *t*-test *p*-value is equal to 0.939).

The variability of the regional-average series of RL10, R30 and R5D does not show significant trends either (Figure 2, left graphics).

5.2. Space-time continuity analysis

The variogram is an inverse measure of the correlation for a given vector distance h. Variograms, which are inverse functions of covariances, describe how the spatial continuity changes as a function of the distance and direction (where anisotropy is considered) between any pair of points in space and time. Variogram's values increase with increasing distance of separation until they reach a maximum, named sill, at a distance known as the range. Here, the variogram models are a function of these two parameters only, the range (denoted by a) and the sill (denoted by c).

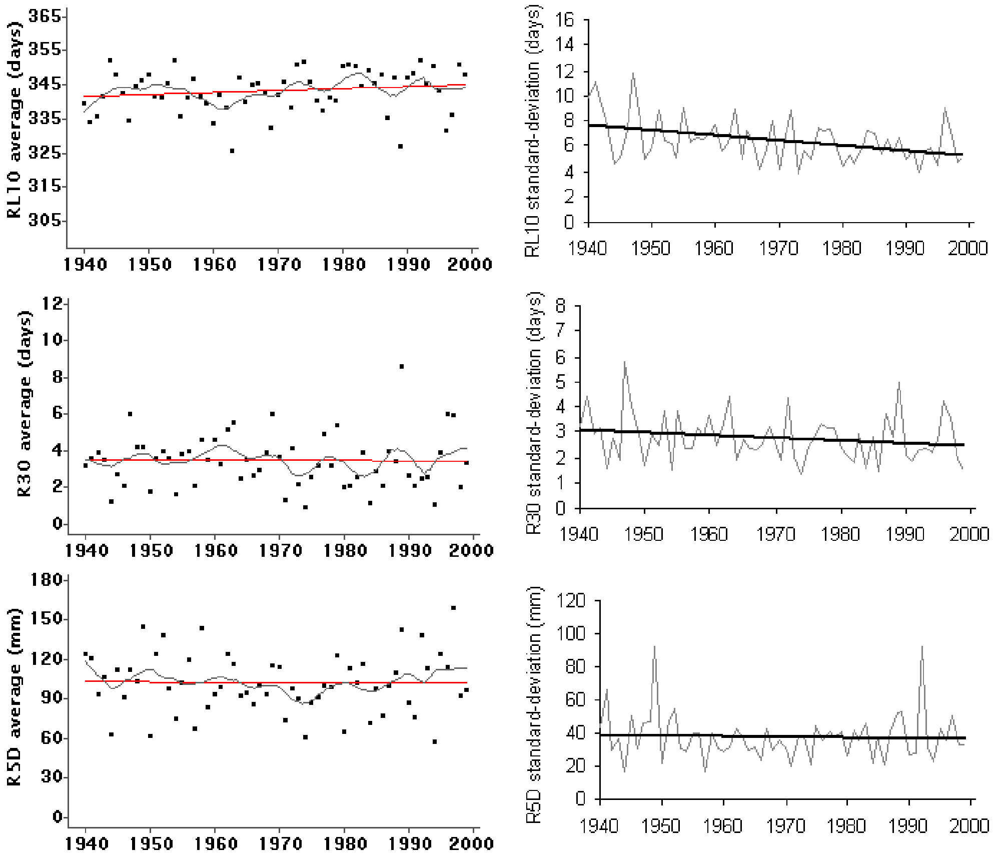


Figure 2: Left graphs: least squares linear fitting (red line) and weighted local polynomial fitting (LOWESS smoothing introduced by [1]) with a time span of 10 years (grey line) for each regional-average series. Right graphs: least squares linear fitting of the regional standard-deviation of RL10, R30 and R5D.

The space-time analysis was done on periods of 10 years: 1940–49, 1945–54, 1950–59, 1954–65, 1960–69, 1965–74, 1970–79, 1975–84, 1980–89, 1985–94, 1990–99. Experimental space-time semi-variograms were calculated for the eleven decades for each extreme precipitation index and exponential models fitted (the spatial component was modelled as isotropic). For sake of simplicity only the variograms of RL10 for 3 decades are shown in Figure 3. The parameters fitted for each variogram are summarized in Table 1, showing that there are no relevant tendencies in what concerns the temporal component of the semi-variograms, which is consistent with what was previously discussed about the series time trends. However, the ranges of the exponential models fitted to the experimental semi-variograms, which express the extent of spatial continuity of the phenomena, are generally increasing along the decades for all indices. This means that extreme events tend to be more spatially homogeneous along time in this region.

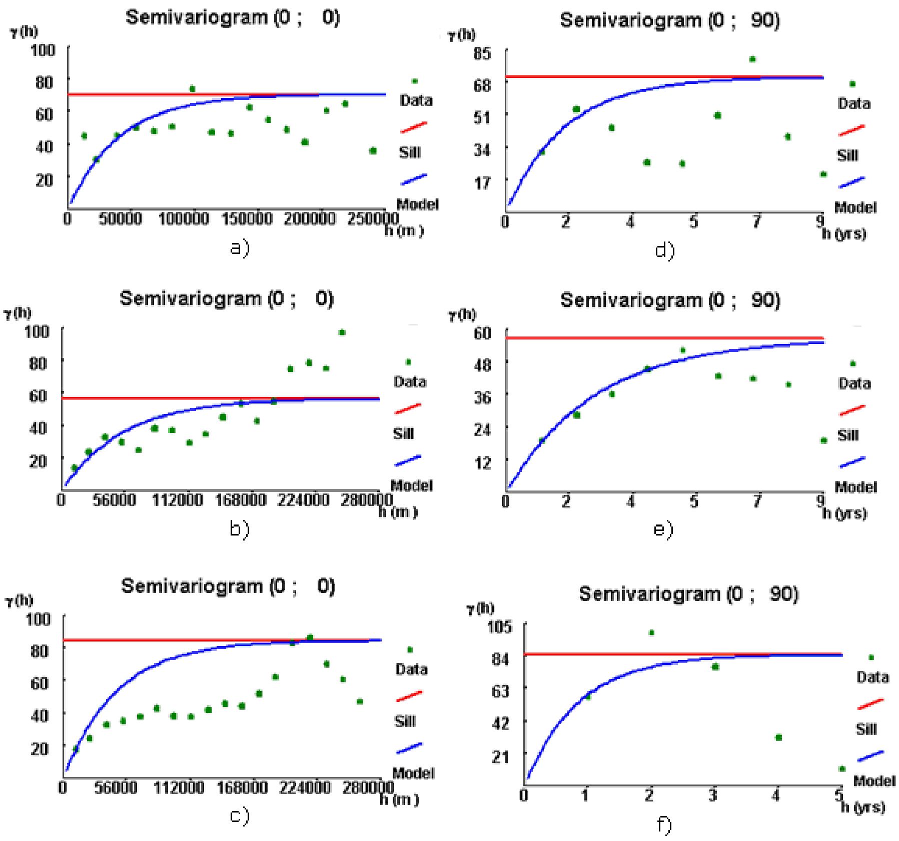


Figure 3: Space-time variograms of RL10 and exponential models fitted. Graphs a), b), c) show the spatial component, and graphs d), e), f) show the temporal component for the decades 1945–54, 1975–85 and 1990–99, respectively.

Table 1: Parameters of the space-time variograms for each precipitation index by decade.

Index	**Decades**	**Spatial range** a (m)	**Temporal range** a (years)	**Sill** c
RL10	1945–54 1975–85 1990–99	120 000 160 000 275 000	5 7.5 4	70.406 56.461 71.366
R30	1945–54 1975–85 1990–99	60 000 100 000 150 000	4 4.5 4.5	11.598 7.675 8.984
R5D	1945–54 1975–85 1990–99	120 000 120 000 160 000	2.5 3.5 1.4	2885.47 1704.018 2803.646

5.3. Space-time inference

Ordinary kriging was used to estimate in space and time the extreme precipitation indices, producing one map for each index per year. One map from the decade 1945–54 and another one from the decade 1990–99 of each extreme precipitation index are shown in the following figures. The estimated maps of RL10 represent the driest years of those decades, namely 1954 and 1992 (Figure 4). While the estimated maps of R30 represent the wetter years of those decades, namely 1947 and 1996 (Figure 5). The estimated maps of R5D refer to the years with more accumulated precipitation in five consecutive days, namely 1949 and 1997 (Figure 6).

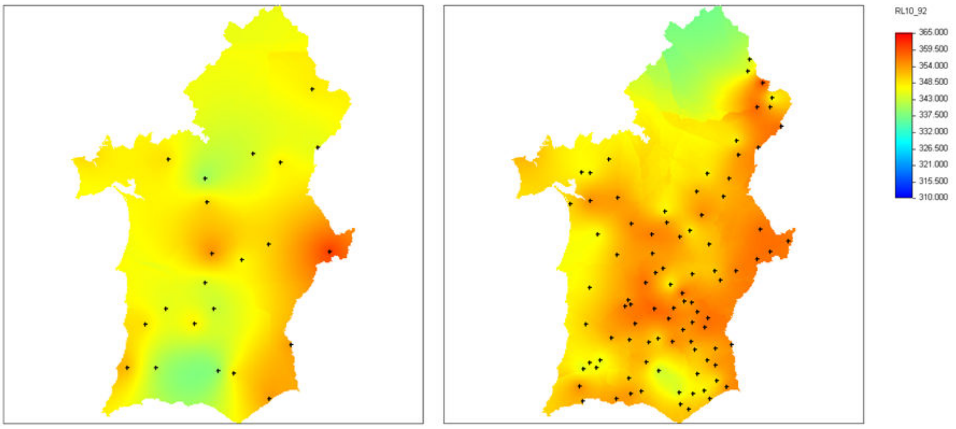


Figure 4: Spatial distribution of the RL10 index for 1954 (left figure) and 1992 (right figure).

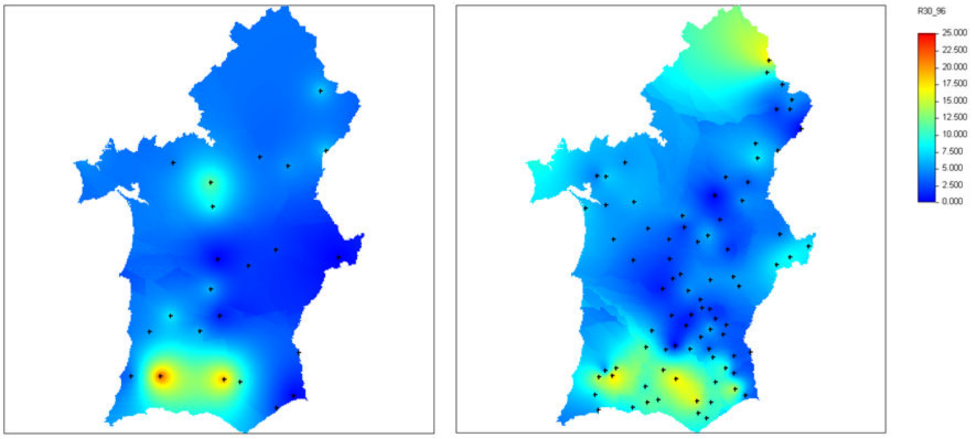


Figure 5: Spatial distribution of the R30 index for 1947 (left figure) and 1996 (right figure).

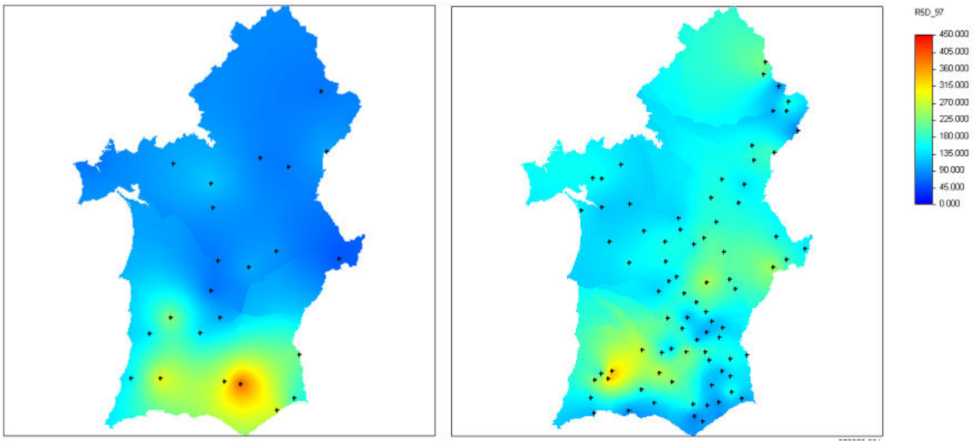


Figure 6: Spatial distribution of the R5D index for 1949 (left figure) and 1997 (right figure).

According with the experimental space-time variograms in the previous section, the spatial distributions of the three precipitation indexes show an increase of the spatial continuity along time as expected.

6. FINAL REMARKS

The results show no statistically significant trends in the regional-average series within the period 1940–99, although the signs of the slopes are as expected ([11]; ECA project, `http://eca.knmi.nl`, retrieved April 2007): for the index representing dry conditions the slope is positive and both indices representing wet conditions show negative slopes. However, there is no significant change in the temporal variability of the three regional extreme indices.

All extreme precipitation indices analysed show increased spatial continuity along time.

The results of this application open perspectives for new approaches of the analysis of extreme climate events, particularly in the context of impact studies related with the desertification phenomenon.

REFERENCES

[1] Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, **74**, 829–836.

[2] Costa, A.C. and Soares, A. (2006). *Identification of inhomogeneities in precipitation time series using SUR models and the Ellipse test.* In "Proceedings of Accuracy 2006 — 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences" (M. Caetano and M. Painho, Eds.), Instituto Geogrfico Português, 419–428.

[3] Costa, A.C.; Negreiros, J. and Soares, A. (2008). *Identification of inhomogeneities in precipitation time series using stochastic simulation.* In "geoENV VI – Geostatistics for Environmental Applications" (A. Soares, M.J. Pereira and R. Dimitrakopoulos, Eds.), Springer, 271–278.

[4] Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets, *International Journal of Climatology*, **26**(6), 707–721.

[5] Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press.

[6] Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *Journal of Hydrology*, **228**, 113–129.

[7] Haylock, M. and Nicholls, N. (2000). Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998, *International Journal of Climatology*, **20**(13), 1533–1541.

[8] Haylock, M. and Goodess, C.M. (2004). Interannual variability of European extreme winter rainfall and links with mean large-scale circulation, *International Journal of Climatology*, **24**(6), 759–776.

[9] Hidalgo, J.C.G.; De Luís, M.; Raventós, J. and Sánchez, J.R. (2003). Daily rainfall trend in the Valencia region of Spain, *Theoretical and Applied Climatology*, **75**, 117–130.

[10] Hundecha, Y. and Bárdossy, A. (2005). Trends in daily precipitation and temperature extremes across western Germany in the second half of the 20th century, *International Journal of Climatology*, **25**(9), 1189–1202.

[11] Klein Tank, A.M.G. and Können, G.P. (2003). Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99, *Journal of Climate*, **16**(22), 3665–3680.

[12] Kostopoulou, E. and Jones, P.D. (2005). Assessment of climate extremes in the Eastern Mediterranean, *Meteorology and Atmospheric Physics*, **89**, 69–85.

[13] Lázaro, R.; Rodrigo, F.S.; Gutiérrez, L.; Domingo, F. and Puigdefábregas, J. (2001). Analysis of a 30-year rainfall record (1967–1997) in semiarid SE Spain for implications on vegetation, *Journal of Arid Environments*, **48**(3), 373–395.

[14] Prudhomme, C. (1999). Mapping a statistic of extreme rainfall in a mountainous region, *Physics and Chemistry of the Earth*, **24B**, 79–84.

[15] Wijngaard, J.B.; Klein Tank, A.M.G. and Können, G.P. (2003). Homogeneity of 20th century European daily temperature and precipitation series, *International Journal of Climatology*, **23**(6), 679–692.

REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008, 33–50

IMPROVING PROBABILITY-WEIGHTED MOMENT METHODS FOR THE GENERALIZED EXTREME VALUE DISTRIBUTION

Authors: Jean Diebolt
– CNRS, Université de Marne-la-Vallée,
5, boulevard Descartes, 77454 Marne-la-Vallée Cedex 2, France

Armelle Guillou
– Université de Strasbourg, IRMA,
7, rue René Descartes, 67084 Strasbourg Cedex, France
guillou@math.u-strasbg.fr

Philippe Naveau
– Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS,
Orme des Merisiers / Bat. 709, 91191 Gif-sur-Yvette, France
philippe.naveau@lsce.ipsl.fr

Pierre Ribereau
– Université de Montpellier, Équipe de Probabilités et Statistique,
CC 051, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
pribere@math.univ-montp2.fr

Abstract:

- In 1985 Hosking *et al.* estimated with the so-called Probability-Weighted Moments (PWM) method the parameters of the Generalized Extreme Value (GEV) distribution, the latter being classically fitted to maxima of sequences of independent and identically distributed random variables. Their approach is still very popular in hydrology and climatology because of its conceptual simplicity, its easy implementation and its good performance for most distributions encountered in geosciences. Its main drawback resides in its limitations when applied to strong heavy-tailed densities. Whenever the GEV shape parameter is larger than 0.5, the asymptotic properties of the PWMs cannot be derived and consequently, asymptotic confidence intervals cannot be obtained. To broaden the validity domain of the PWM approach, we take advantage of a recent extension of PWM to a larger class of moments, called Generalized PWM (GPWM). This allows us to derive the asymptotic properties of our estimators for larger values of the shape parameter. The performance of our approach is illustrated by studying simulations of small, medium and large GEV samples. Comparisons with other GEV estimation techniques used in hydrology and climatology are performed.

Key-Words:

- *empirical processes; maximum likelihood estimators.*

AMS Subject Classification:

- 62G32, 60G70, 62G20.

1. INTRODUCTION

In climatology and hydrology, maxima of temperatures, precipitation and river discharges have been recorded for many decades. The block maxima size (hourly, daily, weekly, monthly or yearly) varies according to instrumental constraints, seasonalities and the application at hand. Extreme Value Theory (EVT) provides a theoretical framework to model the distribution of such block maxima (e.g. Embrechts *et al.*, 1997; Beirlant *et al.*, 2004; de Haan and Ferreira, 2006). Since the work of Fisher and Tippett in 1928, it is known that the only possible limiting form of a normalized maximum of a random sample (when a non-degenerate limit exists) is captured by the Generalized Extreme Value distribution (GEV)

$$G(x;\sigma,\gamma,\mu) = \begin{cases} \exp\left(-\left\{1+\gamma\,\dfrac{x-\mu}{\sigma}\right\}^{-1/\gamma}\right), & \text{if } 1+\gamma\,\dfrac{x-\mu}{\sigma}>0,\ \gamma\neq 0\ , \\ \exp\left(-\exp\left\{-\,\dfrac{x-\mu}{\sigma}\right\}\right), & \text{if } x\in\mathbb{R},\ \gamma=0\ , \end{cases}$$

with $\mu \in \mathbb{R}$, $\sigma > 0$ and γ are called the location, scale and shape parameters, respectively.

Whenever all observations from a given sample are available, it is statistically more efficient to disregard the block maxima modeling approach and instead to analyze exceedances above a high fixed threshold. The exceedances amplitudes can be asymptotically modeled by the Generalized Pareto Distribution (GPD) (e.g., Pickands, 1975; Davison, 1984). In the last four decades, a wide range of methods have been proposed to estimate the GPD scale and shape parameters (e.g. Embrechts *et al.*, 1997; Beirlant *et al.*, 2004; de Haan and Ferreira, 2006). But, for some specific cases, a GEV based approach may still be preferred to a GPD one for at least three reasons. Firstly, block maxima may be the only measurements available to the practitioner (this is specially true for long historical records). Secondly, climatologists frequently face a computational problem. A very high number time series have to be analyzed. For example, General Circulation Models, complex computer codes simulating the atmospheric circulation through resolving the equations representing the Earths atmospheric dynamics provide synthetic temperature time series on a spherical grid. The number of points on such a grid can easily be greater than the hundreds. Consequently, it is computationally easier to only focus on block maxima. This strategy bypasses the difficult problem of choosing a high threshold for each grid point (Kharin, 2007). The latter task is already difficult for a single time series. The third reason to work with blocks of a given size centers on the interpretability of the estimated parameters. For example, a block size of one year makes sense for the Earth scientist because inter-annual physical processes are often very different than decadal ones. For these three reasons, modeling block maxima with a GEV distribution remains a very frequent procedure in hydrology and climatology.

To estimate the GEV parameters in the independent and identically distributed (iid) setting, there exists a wide variety of approaches. In this paper we focus on the two most popular ones used in hydrology and climatology: method-of-moments types (e.g. Hosking *et al.*, 1985) and likelihood based procedures (e.g. Coles and Dixon, 1999; Katz *et al.*, 2002). For the former, hydrologists frequently analyze their maxima with the so-called Probability Weighted Moments (PWM) method introduced by Landwehr *et al.* (1979) and Greenwood *et al.* (1979). The main idea of this approach is to match the moments

$$\mathbb{E}\Big[X^p\big(F(X)\big)^r\big(1-F(X)\big)^s\Big], \qquad \text{with } p,\, r \text{ and } s \text{ real numbers },$$

with their empirical functionals, similarly to the classical method-of-moments. For the GEV distribution, it is easy to show (Hosking *et al.*, 1985) that $\mathbb{E}\big[X(F(X))^r\big]$ can be written as

$$\text{(1.1)} \qquad \beta_r = \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\gamma}\Big[1-(r+1)^\gamma\,\Gamma(1-\gamma)\Big]\right\}, \qquad \gamma<1 \quad\text{and}\quad \gamma\neq 0\ .$$

Consequently, the PWM estimators $(\widehat{\sigma},\widehat{\gamma},\widehat{\mu})$ of the GEV parameters (σ,γ,μ) are simply the solution of the following system of equations

$$\begin{cases} \beta_0 \;=\; \mu - \dfrac{\sigma}{\gamma}\Big(1-\Gamma(1-\gamma)\Big) \\[2mm] 2\beta_1-\beta_0 \;=\; \dfrac{\sigma}{\gamma}\,\Gamma(1-\gamma)\,(2^\gamma-1) \\[2mm] \dfrac{3\beta_2-\beta_0}{2\beta_1-\beta_0} \;=\; \dfrac{3^\gamma-1}{2^\gamma-1} \end{cases}$$

in which β_r has to be replaced by the unbiased estimator proposed by Landwehr *et al.* (1979)

$$\widehat{\beta}_r \;=\; \frac{1}{n}\sum_{j=1}^{n}\left(\prod_{\ell=1}^{r}\frac{j-\ell}{n-\ell}\right)X_{j,n}$$

where $(X_{1,n},...,X_{n,n})$ represents the ordered GEV distributed sample. The properties and performances of $(\widehat{\beta}_0,\widehat{\beta}_1,\widehat{\beta}_2)$ and $(\widehat{\sigma},\widehat{\gamma},\widehat{\mu})$ were studied in details by Hosking *et al.* (1985) who showed the asymptotic normality of these estimators for $\gamma<0.5$. Hoskings and his co-workers also asserted that PWMs estimators performed better than a classical maximum likelihood estimation (MLE) for small samples (see also Hosking and Wallis, 1987). Its conceptual simplicity, its practicability and its good properties for small samples can explain the success of the PWM approach in geosciences (e.g. Katz *et al.*, 2002). Furrer and Naveau (2007) derived some PWMs properties for small GPD distributed samples.

Despite its qualities, the PWM approach has been criticized by Coles and Dixon (1999). In particular, these authors first argued that the PWM estimator assumes *a priori* that the GEV shape parameter is smaller than one, equivalent to specifying that the studied distribution has finite mean. Then they deduced that, if this prior information is available, then a penalized likelihood approach with the constraint $\gamma < 1$ should be preferred. In this case, a simulation study indicated that the penalized MLE outperformed the PWM estimators. But one has to be careful with such a reasoning because PWM estimators are still computable even when $\gamma > 1$ (like the sample mean $\overline{X}$ can be calculated even when the mean is not finite). A penalized MLE with the constraint $\gamma < 1$ will never be able to provide a shape estimator greater than one. In addition, the classical and penalized MLE approaches impose a restriction on the lower values of γ. We need $\gamma > -0.5$ to have regularity of the MLE based estimators and the numerical solutions of the MLE equations are erratic for γ close to -0.5. Although it is rare to work with bounded upper tails, they can be encountered in geophysics. For example, atmospheric scientists can be interested in relative humidity maxima, a bounded random variable. In this context, we argue that it is always better to try removing restrictions on γ than adding ones because we never know in practice the true value of the shape parameter. Hence one of our goals is to extend the validity of method-of-moments based procedures. Still, we agree with Coles and Dixon (1999) on the inherent flexibility of the maximum likelihood and that the conditions on moments existence have to be carefully examined and discussed to understand the limits of the PWMs approach. Our main point is not to sell one estimator in favor of another, but rather to know how to improve a simple approach frequently used in geosciences. With this objective in mind, we recall that Diebolt *et al.* (2007) have recently proposed a wider class of PWMs (called Generalized PWMs) for the GPD. In this paper, our aims are threefold. Firstly, we propose GPWM estimators for the GEV parameters. Secondly, we establish the asymptotic properties of our new estimators under general conditions ensuring the validity of the method for a large range of values of γ. Thirdly, we compare their performances with MLE and classical PWMs.

2. ASYMPTOTIC PROPERTIES OF THE GENERALIZED PWM ESTIMATORS

The generalized probability-weighted moments (GPWM) recently introduced by Diebolt *et al.* (2007) can be described in the following way

$$\nu_\omega = \mathbb{E}\big(X\,\omega(G)\big) = \int_{-\infty}^{\infty} x\,\omega\big(G(x)\big)\,dG(x)~,$$

where ω is a suitable continuous function. By changing variables, this moment can be rewritten as

$$\nu_\omega = \int_0^1 G^{-1}(u)\,\omega(u)\,du~.$$

Let W be the primitive of ω, null at 0, i.e. $W(t) = \int_0^t \omega(u)\, du$. We propose to estimate ν_ω by

$$\widehat{\nu}_{\omega,n} = \int_0^1 \mathbb{F}_n^{-1}(u)\, \omega(u)\, du \tag{2.1}$$

where $\mathbb{F}_n$ denotes the classical empirical distribution function based on a sample $(X_1, ..., X_n)$. We are interested in the asymptotic properties of $\widehat{\nu}_{\omega,n}$ for the GEV distribution. To reach this goal, we select a function ω such that

$$\omega(t) = O\big((1-t)^b\big) \qquad \text{for } t \text{ close to } 1,\ \ b \geq 0 \tag{2.2}$$

and

$$\omega(t) = O(t^{a'}) \qquad \text{for } t \text{ close to } 0,\ \ a' > 0\ . \tag{2.3}$$

These assumptions tie down the functions $G^{-1}(t)$ and $\mathbb{F}_n^{-1}(t)$ at $t = 0$ and $t = 1$. An example of such a function is $\omega(t) = t^a(-\log t)^b$, $a > a'$. In this case, the GPWM for the GEV distribution can be rewritten (see Appendix) as

$$\nu_\omega = \frac{\sigma}{\gamma}\, \frac{1}{(a+1)^{b-\gamma+1}}\, \Gamma\big(b-\gamma+1\big) - \Big(\frac{\sigma}{\gamma} - \mu\Big)\, \frac{1}{(a+1)^{b+1}}\, \Gamma\big(b+1\big)\ . \tag{2.4}$$

Compared to Equality (1.1) derived by Hosking *et al.* (1985), we have a more general expression, Equation (1.1) can be obtained by taking $b = 0$ in (2.4). As for the PWMs method, a system of three equations for three different values of a and/or b has to be solved in order to obtain estimators for σ, γ, μ.

Under the conditions (2.2) and (2.3), the GPWM ν_ω exists as soon as $\gamma < b+1$. This means that the domain of validity for the asymptotic normality of the GPWM estimators has been extended from the set $(\gamma < 1/2)$ to the larger set $\gamma < \frac{1}{2} + b$. More precisely, the following theorem summarizes our findings.

Theorem 2.1. *Let $(X_1, ..., X_n)$ be a sample of maxima whose marginal follows a GEV distribution. Let ω_1, ω_2 and ω_3 be any three continuous functions satisfying (2.2) and (2.3). If $\gamma < \frac{1}{2} + \min(b_1, b_2, b_3)$ for some $b_i \geq 0$, then the rescaled trivariate GPWM estimator vector defined by (2.1) and denoted by*

$$\sqrt{n} \begin{pmatrix} \widehat{\nu}_{\omega_1,n} - \nu_{\omega_1} \\ \widehat{\nu}_{\omega_2,n} - \nu_{\omega_2} \\ \widehat{\nu}_{\omega_3,n} - \nu_{\omega_3} \end{pmatrix}$$

converges in distribution towards the trivariate vector

$$\begin{pmatrix} \sigma \displaystyle\int_0^1 \frac{B(t)}{t} \big(-\log t\big)^{-\gamma-1} \omega_1(t)\, dt \\ \sigma \displaystyle\int_0^1 \frac{B(t)}{t} \big(-\log t\big)^{-\gamma-1} \omega_2(t)\, dt \\ \sigma \displaystyle\int_0^1 \frac{B(t)}{t} \big(-\log t\big)^{-\gamma-1} \omega_3(t)\, dt \end{pmatrix} \tag{2.5}$$

where B denotes a Brownian bridge and $n \to \infty$. The elements of the variance-covariance matrix, Γ, of this limiting vector are given by

$$\begin{aligned}(2.6)\qquad & \int_0^1 \frac{1}{t}\,\big(-\log t\big)^{-\gamma-1}\omega_i(t) \int_0^t \big(-\log s\big)^{-\gamma-1}\omega_j(s)\; ds\,dt \\ &+ \int_0^1 \big(-\log t\big)^{-\gamma-1}\omega_i(t) \int_0^t \frac{1}{s}\,\big(-\log s\big)^{-\gamma-1}\omega_j(s)\; ds\,dt \\ &- \int_0^1 \big(-\log t\big)^{-\gamma-1}\omega_i(t)\; dt \int_0^1 \big(-\log t\big)^{-\gamma-1}\omega_j(t)\; dt\;,\end{aligned}$$

where $i = 1,2,3$ *and* $j = 1,2,3$.

The proof of this theorem is postponed to the appendix and is based on empirical process arguments. From this result, we can deduce estimators for the three parameters σ, γ, μ of the GEV distribution by applying the delta-method. In order to assess the performance of our approach, we analyze simulated and real data in the next section.

3. ANALYSIS OF SIMULATED AND REAL DATA

Theorem 2.1 is a general result. In practice, we have to select the three function ω_1, ω_2 and ω_3. In this section, we opt for $\omega(t) = t^a(-\log t)^b$ with the three pairs $(a,b) = (1,1), (1,2), (2,1)$. This choice is justified by the fact that an estimator of γ can be deduced for these functions by solving the following equation

$$\frac{\widehat{\gamma}}{1-\left(\frac{3}{2}\right)^{\widehat{\gamma}}} = \frac{2\left[\widehat{\omega}_{11} - \widehat{\omega}_{12}\right]}{\widehat{\omega}_{11} - \frac{9}{4}\,\widehat{\omega}_{21}}\;,$$

where

$$\widehat{\omega}_{ab} = \int_0^1 \mathbb{F}_n^{-1}(u)\, u^a(-\log u)^b\, du\;.$$

Two estimators of σ and μ can be obtained from the relations

$$\widehat{\sigma} = 2^{3-\widehat{\gamma}}\,\frac{\widehat{\omega}_{11} - \widehat{\omega}_{12}}{\Gamma(2-\widehat{\gamma})} \qquad \text{and} \qquad \widehat{\mu} = \frac{\widehat{\sigma}}{\widehat{\gamma}} - \frac{\widehat{\sigma}}{\widehat{\gamma}}\, 2^{\widehat{\gamma}}\,\Gamma(2-\widehat{\gamma}) + 4\,\widehat{\omega}_{11}\;.$$

From Theorem 2.1, the asymptotic normality of these three estimators of the GEV parameters can be derived. It is possible to show the existence of a C^1-diffeomorphism T which transforms the GPWMs $(\omega_{11}, \omega_{12}, \omega_{21})$ into (σ, γ, μ). Direct but lengthy computations lead to the following Jacobian matrix, M, associated to this diffeomorphism

$$\begin{pmatrix} \dfrac{2^{\gamma-2}}{\gamma}\Gamma(2-\gamma)-\dfrac{1}{4\gamma} & \dfrac{\sigma}{\gamma}2^{\gamma-2}\Big[\Big(\log 2-\dfrac{1}{\gamma}\Big)\Gamma(2-\gamma)-\Gamma'(2-\gamma)\Big]+\dfrac{\sigma}{4\gamma^2} & \dfrac{1}{4} \\ \Big(\dfrac{2}{\gamma}-1\Big)2^{\gamma-3}\Gamma(2-\gamma)-\dfrac{1}{4\gamma} & \dfrac{\sigma}{\gamma}2^{\gamma-3}\Big[\Big(-\dfrac{2}{\gamma}+\log 2\,(2-\gamma)\Big)\Gamma(2-\gamma)-(2-\gamma)\Gamma'(2-\gamma)\Big]+\dfrac{\sigma}{4\gamma^2} & \dfrac{1}{4} \\ \dfrac{3^{\gamma-2}}{\gamma}\Gamma(2-\gamma)-\dfrac{1}{9\gamma} & \dfrac{\sigma}{\gamma}3^{\gamma-2}\Big[\Big(\log 3-\dfrac{1}{\gamma}\Big)\Gamma(2-\gamma)-\Gamma'(2-\gamma)\Big]+\dfrac{\sigma}{9\gamma^2} & \dfrac{1}{9} \end{pmatrix}.$$

Under the same assumptions stated in Theorem 2.1, we can deduce that the limiting variance-covariance matrix of the trivariate vector $\sqrt{n}\begin{pmatrix}\widehat{\sigma}-\sigma\\ \widehat{\gamma}-\gamma\\ \widehat{\mu}-\mu\end{pmatrix}$ can be written as $M^{-1}\Gamma(M^{-1})'$. This matrix will be useful for computing asymptotic confidence intervals for our GPWM estimators for our application.

3.1. A simulation study

The aim of this simulation study is to show that our method performs adequately for a wide range of values of γ (we will test $\gamma=-0.2$, 0, 0.2 and 1.2) and for small and medium samples sizes ($n=15$, 25, 50 and 100). The quality of our estimators will be compared to the two most common approaches used in hydrology (MLE and PWM). These three estimation methods (MLE, PWM and GPWM) are invariant under linear transformations of the data, so without loss of generality the location and scale parameters are set to $\mu=0$ and $\sigma=1$ in all the simulations. For each combination of values of n and γ, 10 000 random samples are generated from the GEV distribution, and for each sample of parameters, μ, σ and γ were estimated by each of the three methods.

We also implemented the penalized likelihood procedure proposed by Coles and Dixon (1999) but it did not produced valuable results for $\gamma=-0.2$ and 1.2 for all sample sizes. Consequently, we will not show figures about the penalized likelihood procedure (they are available upon request).

Figure 1 shows the estimation results for four different shape parameters γ. Each vertical panel corresponds to the estimations obtained from $\gamma=-0.2$, 0, 0.2 and 1.2 (from bottom to top). The gray, yellow and white boxplots derived from 10 000 GEV samples represent the performance of the MLE, PWM and GPWM estimators, respectively. The x-axis corresponds to different sample sizes $n=15$, 25, 50 and 100. This graph indicates at least three things for the estimation of the shape parameter. For $\gamma=-0.2$, 0, 0.2, the GPWM and MLE behave fairly similarly for all sample sizes, while the PWM method tends to a smaller interquartile but a larger bias. For strong heavy tail ($\gamma=1.2$), PWM does not perform well. The MLE provides a very large interquartile (even for $n=100$),

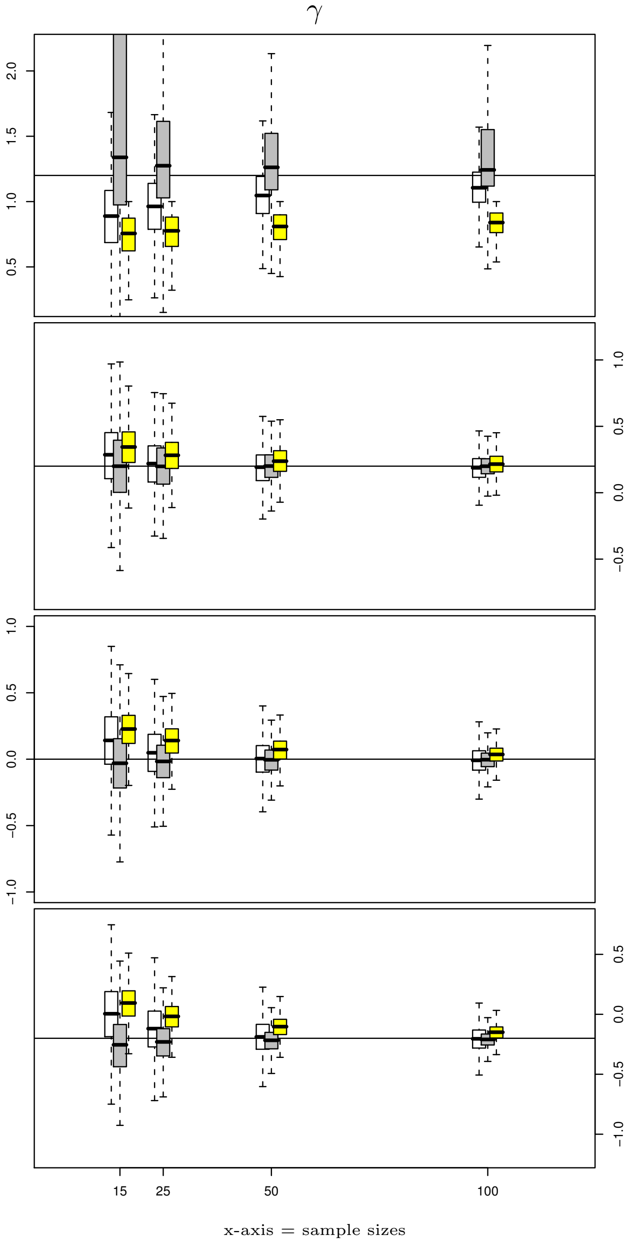


Figure 1: Estimation of γ: The gray, yellow and white boxplots from 10 000 GEV samples represent the performance of the MLE, PWM and GPWM estimators, respectively. The x-axis corresponds to different sample sizes $n = 15$, 25, 50 and 100. Each vertical panel represents the estimations obtained from a different value of $\gamma = -0.2$, 0, 0.2 and 1.2 (from bottom to top).

while the GPWM gives reasonable results for medium sample sizes $n = 50$ and $n = 100$. But this may not be the whole story because one also has to look at the two other GEV parameters. Figure 2 displays the estimation results for $\mu = 0$ (left panels) and $\sigma = 1$ (right panels). As in Figure 1, each vertical panel represents the estimations obtained from $\gamma = -0.2$, 0, 0.2 and 1.2 (from bottom to top).

Figure 2 confirms the remarks raised from Figure 1. The GPWM estimators seem to outperform the PWM ones in all cases. That means that our generalization of the PWM has widened the domain of validity without deteriorating the estimation of the parameters. The MLE approach works adequately but not for $\gamma = 1.2$. For this latter case, the estimation of σ even for $n = 100$ does not seem to provide reasonable estimates.

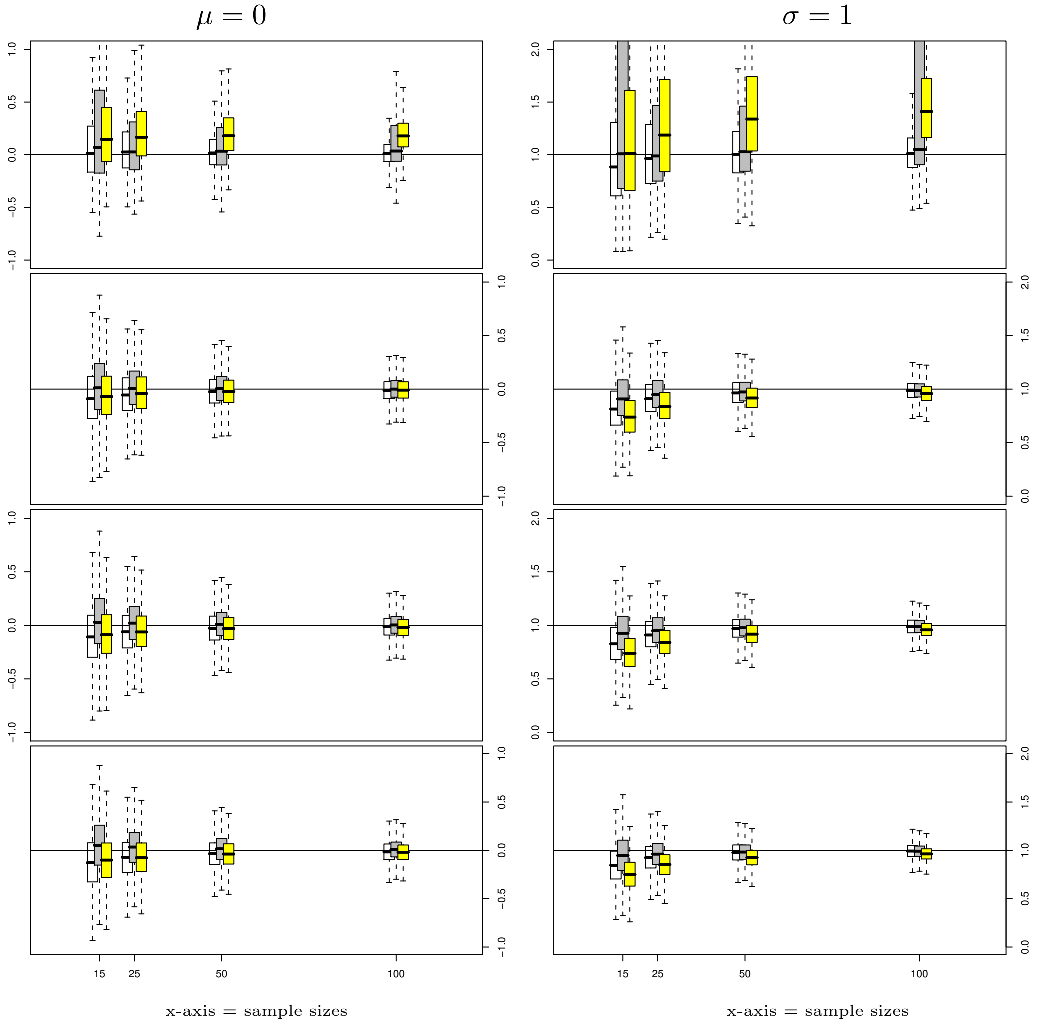


Figure 2: Estimation of $\mu = 0$ (left panels) and $\sigma = 1$ (right panels): The gray, yellow and white boxplots from 10 000 GEV samples represent the performance of the MLE, PWM and GPWM estimators, respectively. The x-axis corresponds to different sample sizes n = 15, 25, 50 and 100. Each vertical panel represents the estimations obtained from a different value of $\gamma = -0.2$, 0, 0.2 and 1.2 (from bottom to top).

3.2. A real data set

One weather station in the city of Fort-Collins (Colorado, USA) recorded annual daily precipitation maxima (in mm) from 1948 to 2001. Figure 3 displays these precipitation maxima. The year 1997 stands up because a storm caused extensive flood damage to this city on July 28th 1997. In order to fit a GEV distribution to this series of yearly maxima, we apply the three estimation methods (MLE, PWM and GPWM) to our data. For each method and for each parameter, 95% asymptotic confidence intervals were obtained. Table 1 summarizes our findings. The three estimation methods (PWM, MLE and GPWM) give similar value for the shape parameter γ, around 0.3. For this type of value and type of sample size (around 50), we know from our simulation study that the three methods should provide similar results in terms of estimation and confidence intervals. The results presented in Table 1 tends to confirm this fact.

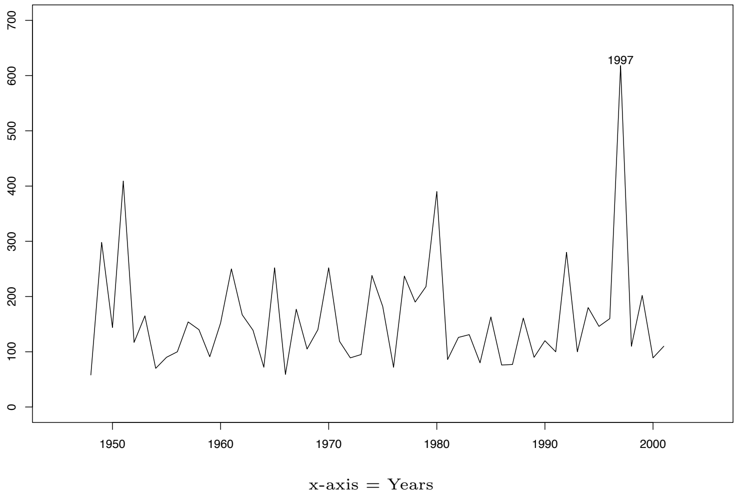


Figure 3: Annual daily precipitation maxima (in mm) recorded in Fort Collins (Colorado, USA) from 1948 to 2001.

Table 1: GEV parameters fitted to the annual daily precipitation maxima in Fort-Collins, Colorado, USA. For each estimation method and parameters, the 95% asymptotic confidence intervals are shown into brackets.

	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
PWM	112.47 $[96.61, 128.33]$	50.57 $[36.43, 64.71]$	0.27 $[-0.01, 0.55]$
MLE	111.31 $[96.45, 126.17]$	47.39 $[34.43, 60.35]$	0.35 $[0.07, 0.63]$
GPWM	112.01 $[106.40, 117.62]$	50.24 $[38.12, 62.36]$	0.32 $[-0.08, 0.73]$

4. CONCLUDING REMARKS

In this paper, we extend the PWM method of Hosking *et al.* (1985) for the GEV distribution. As observed in the simulation part, the validity domain is not only broadened but also the performance of our new method is improved over the classical PWM, especially for large values of the shape parameter. The latter situation is not favorable to the ML approach for small and medium sample sizes. Still, while it is clear that GPWM should be favored to classical PWM, it is difficult to disregard the MLE because it can bring a powerful flexibility in the presence of covariates and/or non-stationarity. In the iid case, the hydrologist and the climatologist may prefer to estimate their GEV parameters with GPWMs because the latter are based on the same method-of-moment approach as the PWM. PWM has been used in their communities for decades and is well understood. The GPWM conserves the PWM conceptual simplicity and its easy implementation. Consequently, it could be quickly integrated in the toolbox of the hydrologist. One remaining challenge for the statistician is to extend such method-of-moment procedures to non-stationary situations.

APPENDIX

Proof of equality (2.4)

$$\begin{aligned}
\nu_\omega &= \int_{-\infty}^{\infty} x\bigl(G(x)\bigr)^a \bigl(-\log G(x)\bigr)^b \, dG(x) \\
&= \int_0^1 G^{-1}(u)\, u^a \bigl(-\log u\bigr)^b \, du \\
&= \int_0^1 \left\{ \frac{\sigma}{\gamma} \Bigl[\bigl(-\log u\bigr)^{-\gamma} - 1\Bigr] + \mu \right\} u^a \bigl(-\log u\bigr)^b \, du \\
&= \int_0^\infty \left\{ \frac{\sigma}{\gamma} \Bigl[x^{-\gamma} - 1\Bigr] + \mu \right\} e^{-(a+1)x} x^b \, dx \\
&= \frac{\sigma}{\gamma} \int_0^\infty x^{b-\gamma} e^{-(a+1)x} \, dx \;-\; \left(\frac{\sigma}{\gamma} - \mu\right) \int_0^\infty x^b e^{-(a+1)x} \, dx \\
&= \frac{\sigma}{\gamma} \, \frac{1}{(a+1)^{b-\gamma+1}} \, \Gamma\bigl(b-\gamma+1\bigr) - \left(\frac{\sigma}{\gamma} - \mu\right) \frac{1}{(a+1)^{b+1}} \, \Gamma\bigl(b+1\bigr) \, .
\end{aligned}$$

Proof of our Theorem 2.1

We consider the difference

$$\begin{aligned}
\widehat{\nu}_{\omega,n} - \nu_\omega \;&\overset{d}{=}\; \int_0^1 \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&= \int_0^{\frac{a_n}{n}} \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&\quad + \int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&\quad + \int_{1-\frac{a_n}{n}}^{1} \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&=:\; T_{1,n} + T_{2,n} + T_{3,n}\;,
\end{aligned}$$

where $(a_n)_n$ is defined by $a_n = ([9\log\log n] + 1)^2$ and $\mathbb{G}_n^{-1}$ denotes the empirical quantile function of independent uniform random variables on $(0,1)$. We study the different terms separately. We can easily prove that, if $\gamma \neq 0$, we have

$$G^{-1}(t) \;=\; \frac{\sigma}{\gamma}\Big[\big(-\log t\big)^{-\gamma} - 1\Big] + \mu\;.$$

The case $\gamma = 0$ can be viewed as the limiting case, letting $\gamma \to 0$.

Term $T_{1,n}$

$$\begin{aligned}
T_{1,n} \;&=\; \int_0^{\frac{a_n}{n}} \frac{\sigma}{\gamma}\Big[\big(-\log \mathbb{G}_n^{-1}(t)\big)^{-\gamma} - \big(-\log t\big)^{-\gamma}\Big]\,\omega(t)\;dt \\
&=\; \frac{\sigma}{\gamma}\int_0^{\frac{a_n}{n}} \big(-\log \mathbb{G}_n^{-1}(t)\big)^{-\gamma}\omega(t)\;dt \;-\; \frac{\sigma}{\gamma}\int_0^{\frac{a_n}{n}} \big(-\log t\big)^{-\gamma}\omega(t)\;dt \\
&=:\; T_{1,n}^{(1)} + T_{1,n}^{(2)}\;.
\end{aligned}$$

By changing variables, it is clear that

$$T_{1,n}^{(2)} \;=\; -\frac{\sigma}{\gamma}\int_{\log\frac{n}{a_n}}^{\infty} x^{-\gamma}\,e^{-x}\,\omega(e^{-x})\;dx\;.$$

Consequently

$$\Big|T_{1,n}^{(2)}\Big| \;\le\; \frac{\sigma}{\gamma}\,\frac{a_n}{n}\int_{\log\frac{n}{a_n}}^{\infty} x^{-\gamma}\big|\omega(e^{-x})\big|\;dx\;.$$

Therefore, we have, under the assumption

$$\text{(A.1)} \qquad \int_0^{\infty} x^{-\gamma}\,\big|\omega(e^{-x})\big|\;dx \;<\; \infty\;,$$

that

$$\sqrt{n}\,\Big|T^{(2)}_{1,n}\Big| = O\bigg(\frac{a_n}{\sqrt{n}}\bigg) \;\longrightarrow\; 0 \;.$$

Of course, (A.1) is satisfied since we have (2.2) and (2.3). Now, concerning the term $T^{(1)}_{1,n}$, we use the following decomposition

$$\begin{aligned}
T^{(1)}_{1,n} &= \frac{\sigma}{\gamma}\left[\int_0^{1/n}\big(-\log\mathbb{G}_n^{-1}(t)\big)^{-\gamma}\omega(t)\,dt + \cdots + \int_{(a_n-1)/n}^{a_n/n}\big(-\log\mathbb{G}_n^{-1}(t)\big)^{-\gamma}\omega(t)\,dt\right] \\
&= \frac{\sigma}{\gamma}\sum_{i=1}^{a_n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\big(-\log\mathbb{G}_n^{-1}(t)\big)^{-\gamma}\omega(t)\,dt \\
&= \frac{\sigma}{\gamma}\sum_{i=1}^{a_n}\big(-\log U_{i,n}\big)^{-\gamma}\left[W\Big(\frac{i}{n}\Big) - W\Big(\frac{i-1}{n}\Big)\right] \\
&= \frac{\sigma}{\gamma}\,\frac{1}{n}\sum_{i=1}^{a_n}\big(-\log U_{i,n}\big)^{-\gamma}\omega\bigg(\frac{\xi_{i,n}}{n}\bigg)
\end{aligned}$$

with $i-1 \le \xi_{i,n} \le i$ and $U_{1,n} \le ... \le U_{n,n}$ the order statistics of a sample of n independent random variables from a uniform distribution on $(0,1)$. Since $|\omega(t)| \le C t^{a'}$ for t close to 0, we have

$$\Big|T^{(1)}_{1,n}\Big| \;\le\; \frac{\sigma}{\gamma}\,\frac{C}{n^{a'+1}}\left\{\big(-\log U_{1,n}\big)^{-\gamma} + \sum_{i=2}^{a_n}\big(-\log U_{i,n}\big)^{-\gamma}\, i^{a'}\right\}.$$

Using the following bounds (see Shorack and Wellner, 1986, p. 408 & 420), we have, for n large enough:

- for $U_{1,n}$:

$$\frac{1}{n(\log n)^{1+\varepsilon}} \;\le\; U_{1,n} \;\le\; (1+\varepsilon')\,\frac{\log\log n}{n} \qquad \text{a.s.}$$

- for $U_{i,n}$:

$$\max_{1\le i\le n}\frac{i}{n\,U_{i,n}} \;\le\; \big(\log n\big)^{1+\varepsilon} \qquad \text{a.s.}\,,$$

$$\max_{1\le i\le n}\frac{n\,U_{i+1,n}}{i} \;\le\; (1+\varepsilon')\log\log n \qquad \text{a.s.}\,.$$

Therefore, it is clear that

$$\sqrt{n}\,\Big|T^{(1)}_{1,n}\Big| = O\bigg(\frac{(\log n)^{-\gamma}}{n^{a'+1/2}}\bigg) + O\bigg(\frac{a_n^{a'+1}}{n^{a'+1/2}}\big(\log n\big)^{-\gamma}\bigg) \;\longrightarrow\; 0\;,$$

by definition of a_n.

Term $T_{2,n}$

$$\begin{aligned}\sqrt{n}\,T_{2,n} &= \sqrt{n}\,\frac{\sigma}{\gamma}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\left[\Big(-\log\Big(t+\frac{\beta_n(t)}{\sqrt{n}}\Big)\Big)^{-\gamma}-\big(-\log t\big)^{-\gamma}\right]\omega(t)\;dt\\ &= \sigma\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{\beta_n(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\;dt\\ &\quad+\frac{\sigma}{2\sqrt{n}}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{\beta_n^2(t)}{\xi_{t,n}^2}\,\big(-\log\xi_{t,n}\big)^{-\gamma-2}\,\big(\log\xi_{t,n}+\gamma+1\big)\,\omega(t)\;dt\;,\end{aligned}$$

where β_n is the uniform empirical quantile process and $\xi_{t,n}\in\Big[\min\Big(t,t+\frac{\beta_n(t)}{\sqrt{n}}\Big),\max\Big(t,t+\frac{\beta_n(t)}{\sqrt{n}}\Big)\Big]$. Our aim now is to use a result due to Csörgő *et al.* (1983) (see e.g. Shorack and Wellner, 1986, p. 500). There exists a sequence of Brownian bridges B_n such that, for $\nu\in[0,\frac{1}{2}[$:

$$\begin{aligned}\sqrt{n}\,T_{2,n} &= \sigma\,n^{-\nu}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}n^{\nu}\,\frac{\beta_n(t)-B_n(t)}{\big[t(1-t)\big]^{\frac{1}{2}-\nu}}\,\frac{1}{t}\,\big(-\log t\big)^{-\gamma-1}\,\big[t(1-t)\big]^{\frac{1}{2}-\nu}\omega(t)\;dt\\ &\quad+\sigma\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{B_n(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\;dt\\ &\quad+\frac{\sigma}{2\sqrt{n}}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{\beta_n^2(t)}{\xi_{t,n}^2}\big(-\log\xi_{t,n}\big)^{-\gamma-2}\,\big(\log\xi_{t,n}+\gamma+1\big)\,\omega(t)\;dt\\ &=:\;T_{2,n}^{(1)}+T_{2,n}^{(2)}+T_{2,n}^{(3)}\;,\end{aligned}$$

with

$$\Big|T_{2,n}^{(1)}\Big|\;\le\;O_{\mathbb{P}}(n^{-\nu})\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{1}{t}\,\big(-\log t\big)^{-\gamma-1}\,\big[t\,(1-t)\big]^{\frac{1}{2}-\nu}\omega(t)\;dt\;.$$

Therefore, under the conditions (2.2) and (2.3), $T_{2,n}^{(1)}$ tends to 0 as soon as $\gamma<b+\frac{1}{2}$. Now, we consider $T_{2,n}^{(2)}$. We can use the fact that

$$\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{B_n(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\;dt\;\overset{d}{=}\;\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{B(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\;dt$$

with

$$|B(t)|\;\le\;C\,\sqrt{\big[t\,(1-t)\big]\log\log\frac{1}{\big[t\,(1-t)\big]}}\qquad\text{a.s., for } t \text{ close to 0 and 1}\;.$$

Here and in all the paper, C represents a generic constant.

Therefore, we have

$$\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}}\frac{B_n(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\;dt\;\xrightarrow{d}\;\int_0^1\frac{B(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\,dt\;,\qquad n\to\infty\,.$$

Now, we have to study $T_{2,n}^{(3)}$. According to Shorack and Wellner (1986, p. 616), we have

$$\big|\beta_n(t)\big| \le C\sqrt{t(1-t)}\,\sqrt{\log\log n} \quad \text{a.s., uniformly on } \left[9\,\frac{\log\log n}{n}, 1-9\,\frac{\log\log n}{n}\right].$$

Therefore

$$\left|T_{2,n}^{(3)}\right| \le C\,\frac{\log\log n}{\sqrt{n}} \int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{t\,(1-t)}{\xi_{t,n}^2}\,\big(-\log\xi_{t,n}\big)^{-\gamma-2}\big(\log\xi_{t,n}+\gamma+1\big)\,\omega(t)\;dt\;.$$

This integral can be divided into three parts: from $\frac{a_n}{n}$ to ε, from ε to $1-\varepsilon$ and from $1-\varepsilon$ to $1-\frac{a_n}{n}$, where ε is fixed. We denote these integrals by $T_{2,n}^{(3,1)}, T_{2,n}^{(3,2)}$ and $T_{2,n}^{(3,3)}$ respectively. We start with $T_{2,n}^{(3,1)}$. Note that for $t\in[\frac{a_n}{n},\varepsilon]$, we have

$$\frac{\big|\beta_n(t)\big|}{t\sqrt{n}} \le C\sqrt{\frac{1-t}{t}}\,\sqrt{\frac{\log\log n}{n}} \tag{A.2}$$

$$\le \frac{C}{\sqrt{\log\log n}}\;. \tag{A.3}$$

Therefore

$$\frac{\xi_{t,n}}{t} = 1+o(1) \qquad\text{and}\qquad \log\xi_{t,n} = \big(1+o(1)\big)\,\log t\;, \tag{A.4}$$

where the $o(1)$-terms are uniform in t. Consequently,

$$\begin{aligned}\left|T_{2,n}^{(3,1)}\right| \le\; & -\,C\,\frac{\log\log n}{\sqrt{n}}\int_{\frac{a_n}{n}}^{\varepsilon}\frac{1-t}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\;dt\;\big(1+o(1)\big)\\ & +\;C\,(1+\gamma)\,\frac{\log\log n}{\sqrt{n}}\int_{\frac{a_n}{n}}^{\varepsilon}\frac{1-t}{t}\,\big(-\log t\big)^{-\gamma-2}\omega(t)\;dt\;\big(1+o(1)\big)\;.\end{aligned}$$

Using (2.3), $T_{2,n}^{(3,1)}$ tends clearly to 0. Similarly, since (A.2) and (A.4) are true for $t\in\big[1-\varepsilon, 1-\frac{a_n}{n}\big]$, we have

$$\left|T_{2,n}^{(3,3)}\right| = O\Big(\frac{\log\log n}{\sqrt{n}}\Big)\left\{\int_{1-\varepsilon}^{1-\frac{a_n}{n}}(1-t)^{b-\gamma}\,dt \;+\; \int_{1-\varepsilon}^{1-\frac{a_n}{n}}(1-t)^{b-\gamma-1}\,dt\right\},$$

by (2.2). The right-hand side of the last equality tends to 0 as soon as $\gamma < b+\frac{1}{2}$. For the central part, $T_{2,n}^{(3,2)}$, similar arguments lead to its negligibility.

Term $T_{3,n}$

$$\begin{aligned}\sqrt{n}\,T_{3,n} \;=\; & \frac{\sigma}{\gamma}\sqrt{n}\int_{1-\frac{a_n}{n}}^{1}\big(-\log\mathbb{G}_n^{-1}(t)\big)^{-\gamma}\omega(t)\;dt \;-\; \frac{\sigma}{\gamma}\sqrt{n}\int_{1-\frac{a_n}{n}}^{1}\big(-\log t\big)^{-\gamma}\omega(t)\;dt\\ =:\; & T_{3,n}^{(1)}+T_{3,n}^{(2)}\;.\end{aligned}$$

The term $T^{(2)}_{3,n}$ is of order

$$\sqrt{n}\int_{1-\frac{a_n}{n}}^{1}(1-t)^{-\gamma+b}\,dt$$

which tends to 0 as soon as $\gamma < b+\frac{1}{2}$. For $T^{(1)}_{3,n}$, we decompose again the integral as follows, with $j-1 \le \xi_j \le j$:

$$\begin{aligned}
T^{(1)}_{3,n} &= \frac{\sigma}{\gamma}\sqrt{n}\,\frac{1}{n}\sum_{j=n-a_n+1}^{n}\big(-\log U_{j,n}\big)^{-\gamma}\,\omega\Big(\frac{\xi_j}{n}\Big)\\
&= \frac{\sigma}{\gamma}\sqrt{n}\,\frac{1}{n}\sum_{j=1}^{a_n}\big(1-U_{n-j+1,n}\big)^{-\gamma}\Big(1+O_{\mathbb{P}}\Big(\frac{j}{n}\Big)\Big)\,\omega\Big(\frac{\xi_{n-j+1}}{n}\Big)\\
&\overset{d}{=} \frac{\sigma}{\gamma}\sqrt{n}\,\frac{1}{n}\sum_{j=1}^{a_n}\big(U_{j,n}\big)^{-\gamma}\Big(1+O_{\mathbb{P}}\Big(\frac{j}{n}\Big)\Big)\,\omega\Big(\frac{\xi_{n-j+1}}{n}\Big)\\
&= O\Big(\big(\log n\big)^{|\gamma|\,(1+\varepsilon)}\,n^{\gamma-b-\frac{1}{2}}\,a_n^{b-\gamma+1}\Big) \;=\; o(1)\ ,
\end{aligned}$$

as soon as $\gamma < b+\frac{1}{2}$.

From all the above convergences, using Serfling (1980, page 18), we deduce (2.5), and therefore the expression of the generic term at position (i,j), $1 \le i,j \le 3$, of the limiting variance-covariance matrix given in (2.6).

Combining these results, Theorem 2.1 follows.

ACKNOWLEDGMENTS

This work was supported by the french ANR-AssimilEx project, the european E2-C2 grant, and by The Weather and Climate Impact Assessment Science Initiative at the National Center for Atmospheric Research (NCAR). The authors would also like to credit the contributors of the R project. Moreover, the authors are grateful to the referee for her/his constructive comments from which the paper has benefitted a lot.

REFERENCES

[1] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley Series in Probability and Statistics.

[2] Coles, S.G. and Dixon, M.J. (1999). Likelihood-based inference of extreme value models, *Extremes*, **2**, 5–23.

[3] Csörgő, M.; Csörgő, S.; Horváth, L. and Mason, D.M. (1983). An asymptotic theory for empirical reliability and concentration processes, *Unpublished manuscript.*

[4] Davison, A.C. (1984). *Modelling excesses over high thresholds, with an application.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), Vol. 131, pp. 461–482, Dordrecht, D. Reidel Publishing Co..

[5] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer Series in Operations Research.

[6] Diebolt, J.; Guillou, A. and Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *J. Statist Plann. Inference*, **137**, 841–857.

[7] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Vol. 33 of "Applications of Mathematics", Springer-Verlag, Berlin.

[8] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution in the largest particle size and smallest member of a sample, *Proc. Camb. Phil. Soc.*, **24**, 180–190.

[9] Furrer, R. and Naveau, P. (2007). Probability weighted moments properties for small samples, *Stat. Probab. Letters*, **70**, 190–195.

[10] Greenwood, J.A.; Landwehr, J.M.; Matalas, N.C. and Wallis, J.R. (1979). Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, **15**, 1049–1054.

[11] Hosking, J.R.M.; Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, **27**, 251–261.

[12] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.

[13] Katz, R.; Parlange, M. and Naveau, P. (2002). Extremes in hydrology, *Advances in Water Resour.*, **25**, 1287–1304.

[14] Kharin, V.V.; Zwiers, F.W.; Zhang, X. and Hegerl, G.C. (2007). Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *Journal of Climate*, **20**, 1419–1444.

[15] Landwehr, J.; Matalas, N. and Wallis, J. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resour. Res.*, **15**, 1055–1064.

[16] Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

[17] Serfling, R.J. 1980. *Approximation Theorems of Mathematical Statistics*, Wiley & Son.

[18] Shorack, G.R. and Wellner, J.A. 1986. *Empirical Processes with Applications to Statistics*, Wiley, New York.

[19] Zhang J. (2007). Likelihood Moment Estimation for the Generalized Pareto Distribution, *Aust. N. Z. J. Stat.*, **49**, 69–77.

REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008, 51–69

LINKING PARETO-TAIL KERNEL GOODNESS-OF-FIT STATISTICS WITH TAIL INDEX AT OPTIMAL THRESHOLD AND SECOND ORDER ESTIMATION

Authors: Yuri Goegebeur
– Department of Statistics,
University of Southern Denmark,
J.B. Winsløws Vej 9B, 5000 Odense C, Denmark
yuri.goegebeur@stat.sdu.dk

Jan Beirlant
– Department of Mathematics and Leuven Statistics Research Center,
Catholic University of Leuven,
Celestijnenlaan 200B, 3001 Heverlee, Belgium
jan.beirlant@wis.kuleuven.be

Tertius de Wet
– Department of Statistics and Actuarial Science,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa
tdewet@sun.ac.za

Abstract:

- In this paper the relation between goodness-of-fit testing and the optimal selection of the sample fraction for tail estimation, for instance using Hill's estimator, is examined. We consider this problem under a general kernel goodness-of-fit test statistic for assessing whether a sample is consistent with the Pareto-type model. The derivation of the class of kernel goodness-of-fit statistics is based on the close link between the strict Pareto and the exponential distribution, and puts some of the available goodness-of-fit procedures for the latter in a broader perspective. Two important special cases of the kernel statistic, the Jackson and the Lewis statistic, will be discussed in greater depth. The relationship between the limiting distribution of the Lewis statistic and the bias-component of the asymptotic mean squared error of the Hill estimator is exploited to construct a new tail sample fraction selection criterion for the latter. The methodology is illustrated on a case study.

Key-Words:

- *extreme value statistics; Pareto-type distribution; goodness-of-fit; threshold selection.*

AMS Subject Classification:

- 62G32, 62G30, 62E20.

1. INTRODUCTION

Extreme value theory focuses on characteristics related to the tail of a distribution function such as indices describing tail decay, extreme quantiles and small tail probabilities. In the process of making inferences about the far tail of a distribution function, it is necessary to extend the empirical distribution function beyond the available data. This is typically done by only considering the upper k order statistics, which then entails the issue of how to select a good, or, if possible, an optimal, k-value. Many proposals to tackle this issue have been made in the literature, see for instance Drees and Kaufmann (1998), Danielsson *et al.* (2001), Guillou and Hall (2001), and Beirlant *et al.* (2002). In this paper we use recently introduced kernel goodness-of-fit statistics for Pareto-type behavior as a basis for proposing a new procedure for selecting k.

Consider random variables $X_1, ..., X_n$ independent and identically distributed (i.i.d.) according to some distribution function F and let $X_{1,n} \le ... \le X_{n,n}$ denote the corresponding ascending order statistics. If for sequences of constants $(a_n > 0)_n$ and $(b_n)_n$

$$(1.1) \qquad \lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \le x\right) = \lim_{n\to\infty} F^n(b_n + a_n x) = G(x)$$

at all continuity points of G, for G some non-degenerate distribution function, then G has to be of the generalized extreme value (GEV) type:

$$(1.2) \qquad G_\gamma(x) = \begin{cases} \exp\bigl(-(1+\gamma x)^{-1/\gamma}\bigr), & 1+\gamma x > 0, \quad \gamma \ne 0 \ , \\ \exp\bigl(-\exp(-x)\bigr), & x \in \mathbb{R}, \quad \gamma = 0 \ . \end{cases}$$

Note that the behavior of this distribution function is governed by the single parameter γ, called the extreme value index. If F satisfies (1.1)–(1.2), then it is said to belong to the max-domain of attraction of G_γ, denoted $F \in \mathcal{D}(G_\gamma)$. An important subclass of the max-domain of attraction of the GEV distribution is the class of the Pareto-type models. These are characterized by heavy tailed distribution functions with infinite right endpoints, having $\gamma > 0$.

For Pareto-type distributions the *first order condition* (1.1) can be expressed in an equivalent way in terms of the survival function $1-F$:

$$(1.3) \qquad 1 - F(x) = x^{-1/\gamma}\, \ell_F(x) \ , \qquad x > 0 \ ,$$

where ℓ_F denotes a slowly varying function at infinity, i.e.

$$(1.4) \qquad \frac{\ell_F(\lambda x)}{\ell_F(x)} \to 1 \quad \text{as } x \to \infty \qquad \text{for all } \lambda > 0 \ .$$

In terms of the tail quantile function U, defined as $U(x) = \inf\{y\colon F(y) \geq 1-1/x\}$, $x > 1$, we then have that

$$U(x) = x^{\gamma}\,\ell_U(x) \;, \tag{1.5}$$

where ℓ_U again denotes a slowly varying function at infinity (Gnedenko, 1943). Pareto-type tails are systematically used in certain branches of non-life insurance, as well as in finance (stock returns), telecommunication (file sizes, waiting times), geology (diamond values, earthquake magnitudes), and many others. In the analysis of heavy tailed distributions the estimation of γ, and the subsequent estimation of extreme quantiles, assume a central position. Several estimators for γ have been proposed in the literature, and their asymptotic distributions established, usually under a *second order condition* on the tail behavior (see e.g. Beirlant *et al.*, 2004, and de Haan and Ferreira, 2006). This condition specifies the rate of convergence of ratios of the form $\ell(\lambda x)/\ell(x)$, with ℓ a slowly varying function, to their limit (see Bingham *et al.*, 1987).

Second order condition $(\mathcal{R}_\ell)$**.** *A slowly varying function ℓ satisfies a second order condition if there exists a real constant $\rho \leq 0$ and a rate function b satisfying $b(x) \to 0$ as $x \to \infty$, such that for all $\lambda \geq 1$, as $x \to \infty$,*

$$\frac{\ell(\lambda x)}{\ell(x)} - 1 \;\sim\; b(x)\,\frac{\lambda^\rho - 1}{\rho} \;.$$

In the context of estimation of γ, it is then typically assumed that the slowly varying function ℓ_U in (1.5) satisfies a second order condition. Of interest for the subsequent development of a procedure for selecting a threshold, is testing of the hypothesis that the underlying distribution is of Pareto-type together with a second order condition holding. Formally, this hypothesis can be stated as

$$H_0\colon\; F \text{ is of Pareto-type with } \ell_U \text{ satisfying } \mathcal{R}_\ell \;. \tag{1.6}$$

It is well known that the log-transform of a (strict) Pareto random variable has an exponential distribution. Our approach to testing H_0 is to exploit this fact by considering goodness-of-fit tests for exponentiality as possible test statistics. The literature on goodness-of-fit tests for the exponential distribution is quite elaborate, see e.g. Henze and Meintanis (2005) for a recent overview of this literature. Such tests often take the form of the ratio of two estimators for the exponential scale parameter. In a similar way, one can construct test statistics as ratios of two estimators for the extreme value index γ.

Of course it is intuitively clear that goodness-of-fit procedures should enable one to choose an appropriate threshold $X_{n-k,n}$ for tail index estimation. Hill (1975) already recognized this idea, see also Beirlant *et al.* (1996). Typically, however, goodness-of-fit based procedures are too conservative with respect to the null hypothesis, leading to too high values of k (or equivalently too low thresholds) with respect to the asymptotic mean squared error (AMSE) criterion. Based on the limiting distribution of our kernel goodness-of-fit statistic, we propose an estimator for the bias component of the AMSE of the Hill estimator, yielding an alternative method to select the threshold $X_{n-k,n}$.

The remainder of this paper is organized as follows. In Section 2 we introduce a general kernel goodness-of-fit statistic for assessing whether a sample is consistent with the Pareto-type model, and state its main properties. Section 3 deals with the link between goodness-of-fit testing and the selection of the optimal tail sample fraction, for instance when using the Hill estimator. In Section 4 we illustrate the methodology with a practical example.

2. A KERNEL GOODNESS-OF-FIT STATISTIC FOR PARETO-TYPE BEHAVIOR

Consider $X_1, ..., X_n$ i.i.d. $Pa(1/\gamma)$ random variables, where $Pa(1/\gamma)$ denotes the strict Pareto distribution with Pareto index $1/\gamma$, i.e. $F(x) = 1 - x^{-1/\gamma}$, $x > 1$, and the corresponding ascending order statistics $X_{1,n} \leq ... \leq X_{n,n}$. Then the ratios $Y_{j,k} = X_{n-k+j,n}/X_{n-k,n}$, $j = 1, ..., k$, are jointly distributed as the order statistics of a random sample of size k from the $Pa(1/\gamma)$ distribution. Consequently, $Y^*_{j,k} = \log Y_{j,k}$ behave as $Exp(1/\gamma)$ order statistics, where $Exp(1/\gamma)$ denotes the exponential distribution with mean γ. In case the data originate from a Pareto-type distribution these properties hold approximately above a sufficiently high threshold. This close link between the Pareto-type and the exponential model will be exploited in the derivation of goodness-of-fit tests for the former. The literature on testing whether a sample is consistent with an exponential distribution is quite extensive, see for instance Stephens (1986) and Henze and Meintanis (2005), and the references therein. These exponential goodness-of-fit test statistics are quite often a ratio of two estimators for the exponential scale parameter (e.g. Lewis, 1965, Jackson, 1967, de Wet and Venter, 1973). Inspired by this and based on the above properties of $Pa(1/\gamma)$ order statistics, we apply a similar ratio to the k largest order statistics, leading to the following test statistic

$$\frac{\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) Z_j}{H_{k,n}}\,, \tag{2.1}$$

with K denoting a kernel function satisfying $\int_0^1 K(u)\,du = 0$, $Z_j = j(\log X_{n-j+1,n} - \log X_{n-j,n})$, and $H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} Z_j$, the Hill estimator for γ (Hill, 1975).

In Goegebeur *et al.* (2007), generalizing Beirlant *et al.* (2006), the statistic in (2.1) was proposed and its limiting distribution derived under the hypothesis stated in (1.6), some mild regularity conditions on K, and an intermediate

k sequence, i.e. $k = k_n \to \infty$, $k_n = o(n)$ as $n \to \infty$. We use $\log^+ u$ to denote $\max\{\log u, 1\}$.

Theorem 2.1. *Consider $X_1, ..., X_n$ i.i.d. random variables according to distribution function F, where $F \in \mathcal{D}(G_\gamma)$ for some $\gamma > 0$. Assume ℓ_U satisfies $\mathcal{R}_\ell$ and let $K(t) = \frac{1}{t}\int_0^t u(v)\,dv$ for some function u satisfying $\left|k\int_{(j-1)/k}^{j/k} u(t)\,dt\right| \le f\left(\frac{j}{k+1}\right)$ for some positive continuous function f defined on $(0,1)$ such that $\int_0^1 \log^+(1/w)\,f(w)\,dw < \infty$ in case $\rho < 0$ and $\int_0^1 w^{-\xi} f(w)\,dw < \infty$ for some small $\xi > 0$ in case $\rho = 0$, $\int_0^1 |K(w)|^{2+\delta}\,dw < \infty$ for some $\delta > 0$ and $\frac{1}{\sqrt{k}}\sum_{j=1}^k K\left(\frac{j}{k+1}\right) \to 0$ as $k \to \infty$. Then as $k, n \to \infty$, $k/n \to 0$ and $\sqrt{k}\,b(n/k) \to c$,*

$$\frac{\sqrt{k}}{H_{k,n}}\,\frac{1}{k}\sum_{j=1}^k K\left(\frac{j}{k+1}\right) Z_j \;\xrightarrow{\mathcal{L}}\; N\left(\frac{c}{\gamma}\int_0^1 K(u)\,u^{-\rho}\,du,\; \int_0^1 K^2(u)\,du\right).$$

Using this theorem, the decision rule for testing the hypothesis (1.6) at the significance level α is to reject H_0 if

$$\sqrt{k}\left|\frac{1}{k\,H_{k,n}}\sum_{j=1}^k K\left(\frac{j}{k+1}\right) Z_j - \frac{b(n/k)}{\gamma}\int_0^1 K(u)\,u^{-\rho}\,du\right| > \\ > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\int_0^1 K^2(u)\,du}\;,$$

where Φ^{-1} denotes the standard normal quantile function. However, for practical application this rule is not very helpful as it depends on the unknown function b as well as on the parameters γ and ρ. A way out of this is to choose k relatively small, i.e. small enough to guarantee that $\sqrt{k}\,b(n/k) \approx 0$, which then leads to the rule to reject H_0 if

$$\frac{\sqrt{k}}{H_{k,n}}\left|\frac{1}{k}\sum_{j=1}^k K\left(\frac{j}{k+1}\right) Z_j\right| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\int_0^1 K^2(u)\,du}\;.$$

For a detailed description of the fundamental properties of the goodness-of-fit statistic and for an evaluation of its small sample performance through a simulation study, we refer to Goegebeur *et al.* (2007). We will now describe two important special cases of this kernel-type goodness-of-fit statistic, the Jackson (Jackson, 1967) and the Lewis (Lewis, 1965) statistics, in more detail.

2.1. Jackson kernel function

We modify the Jackson statistic (Jackson, 1967), originally proposed as a goodness-of-fit statistic for testing exponentiality, in such a way that it measures the linearity of the k largest observations on the Pareto quantile plot. Consider $X_1,, ..., X_n$ i.i.d. $Exp(\delta)$ random variables. The Jackson statistic is given by

$$T_{\mathrm{J}} \;=\; \frac{\sum_{j=1}^{n} t_{j,n}\, X_{j,n}}{\sum_{j=1}^{n} X_j} \tag{2.2}$$

where $t_{j,n} = \delta E(X_{j,n}) = \sum_{i=1}^{j}(n-i+1)^{-1}$. The numerator is clearly a sum of cross products of order statistics and their expected values. The denominator is introduced to eliminate the dependence on the nuisance parameter δ. The Jackson statistic can hence be considered as a 'correlation like' statistic based on the exponential quantile plot. The limiting distribution of the appropriately normalized Jackson statistic was derived by Jackson (1976), in particular $\sqrt{n}(T_{\mathrm{J}}-2) \xrightarrow{\mathcal{D}} N(0,1)$, as $n \to \infty$. For our purposes it is more convenient to express (2.2) in terms of the standardized spacings $V_j = (n-j+1)(X_{j,n} - X_{j-1,n})$, $j = 1, ..., n$. From the Rényi representation these are known to be i.i.d. $Exp(\delta)$ random variables. Rearranging terms of (2.2), it can be shown that

$$T_{\mathrm{J}} \;=\; \frac{\sum_{j=1}^{n} C_{j,n}\, V_j}{\sum_{j=1}^{n} V_j}$$

where $C_{1,n} = 1$ and $C_{j,n} = 1 + t_{j-1,n}$, $j = 2, ..., n$.

We will now adjust the Jackson statistic in such a way that it measures the linearity of the k upper order statistics on the Pareto quantile plot. Consider a random sample $X_1, ..., X_n$ of Pareto-type distributed random variables. Application of the Jackson statistic to $Y^*_{j,k}$, $j = 1, ..., k$, yields, after suitable normalization and rearranging terms,

$$T^{\mathrm{J}}_{k,n} \;=\; \sqrt{k}\;\frac{\frac{1}{k}\sum_{j=1}^{k} K_{\mathrm{J}}\Big(\frac{j}{k+1}\Big)\, Z_j}{H_{k,n}}$$

where $K_{\mathrm{J}}(u) = -1 - \log u$, see also Beirlant *et al.* (2006). The kernel function K_{J} satisfies the conditions of Theorem 2.1 with $u(s) = -2 - \log s$, and hence we can state the following proposition.

Proposition 2.1. *Assume $X_1, ..., X_n$ i.i.d. random variables according to distribution function F, where $F \in \mathcal{D}(G_\gamma)$ for some $\gamma > 0$ and ℓ_U satisfying $\mathcal{R}_\ell$. Then as $k, n \to \infty$, $k/n \to 0$ and $\sqrt{k}\, b(n/k) \to c$,*

$$\frac{\sqrt{k}}{H_{k,n}}\,\frac{1}{k}\sum_{j=1}^{k} K_{\mathrm{J}}\bigg(\frac{j}{k+1}\bigg) Z_j \;\xrightarrow{\mathcal{L}}\; N\bigg(\frac{c\,\rho}{\gamma(1-\rho)^2}\,,\,1\bigg)\;.$$

Note that the normal limit is not necessarily centered at zero, i.e. the statistic may exhibit some bias. The centering depends, besides γ, on the slowly varying function ℓ_U through the parameters ρ and c.

2.2. Lewis kernel function

As a second example we study the Lewis goodness-of-fit statistic. Consider a sample $X_1, ..., X_n$ of i.i.d. $Exp(\delta)$ random variables. The Lewis statistic is given by

$$T_{\mathrm{L}} = \frac{\sum_{j=1}^{n} \frac{j}{n+1} V_{n-j+1}}{\sum_{j=1}^{n} X_j} ,$$

and $\sqrt{n}(T_{\mathrm{L}} - 1/2) \xrightarrow{\mathcal{L}} N(0, 1/12)$, as $n \to \infty$ (Lewis, 1965). In case of a random sample $X_1, ..., X_n$ of Pareto-type random variables, we can apply the Lewis statistic to $Y^*_{j,k}$, $j = 1, ..., k$, yielding, after appropriate normalization and rearranging terms,

$$T^{\mathrm{L}}_{k,n} = \sqrt{k}\, \frac{\frac{1}{k} \sum_{j=1}^{k} K_{\mathrm{L}}\Big(\frac{j}{k+1}\Big) Z_j}{H_{k,n}} ,$$

with $K_{\mathrm{L}}(u) = u - 0.5$. The function K_{L} satisfies the conditions of Theorem 2.1 with $u(s) = 2s - 0.5$, leading to the following proposition:

Proposition 2.2. *Assume $X_1, ..., X_n$ i.i.d. random variables according to distribution function F, where $F \in \mathcal{D}(G_\gamma)$ for some $\gamma > 0$ and ℓ_U satisfying $\mathcal{R}_\ell$. Then as $k, n \to \infty$, $k/n \to 0$ and $\sqrt{k}\, b(n/k) \to c$,*

$$\frac{\sqrt{k}}{H_{k,n}} \frac{1}{k} \sum_{j=1}^{k} K_{\mathrm{L}}\Big(\frac{j}{k+1}\Big) Z_j \xrightarrow{\mathcal{L}} N\Big(-\frac{c\,\rho}{2\,\gamma\,(1-\rho)(2-\rho)}, \frac{1}{12}\Big) .$$

Note that for the same value of c, the absolute value of the asymptotic bias of the Lewis statistic is smaller than the absolute bias of the Jackson statistic.

2.3. Bias-correction

As mentioned above, the bias of the kernel statistics may make it difficult to evaluate the nature of the tail behavior. It is, however, possible to derive, for a given kernel function K, a bias-corrected kernel function, denoted $K_{\mathrm{BC}}(\cdot; \rho)$, i.e. a kernel satisfying $\int_0^1 K_{\mathrm{BC}}(u; \rho)\, u^{-\rho}\, du = 0$. To obtain such a bias-corrected kernel, note that both the numerator and the denominator of the general kernel statistic (2.1) are weighted averages of the Z_j, $j = 1, ..., k$. Within the framework of Pareto-type tails and assuming condition $\mathcal{R}_\ell$ on ℓ_U holds, with $\rho < 0$, Beirlant *et al.* (1999) derived the following approximate representation for log-spacings of successive order statistics

$$Z_j \sim \gamma + b_{n,k} \Big(\frac{j}{k+1}\Big)^{-\rho} + \varepsilon_j , \qquad j = 1, ..., k , \tag{2.3}$$

where $b_{n,k} = b(n/k)$ and ε_j, $j = 1, ..., k$, are zero centered error terms, or, equivalently

$$Z_j - b_{n,k}\left(\frac{j}{k+1}\right)^{-\rho} \sim \gamma + \varepsilon_j \ , \qquad j = 1, ..., k \ .$$

This then motivates the following bias-corrected statistic

$$\text{(2.4)} \qquad \sqrt{k}\ \frac{\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(Z_j - \hat{b}_{\mathrm{LS},k}(\rho)\left(\frac{j}{k+1}\right)^{-\rho}\right)}{\hat{\gamma}_{\mathrm{LS},k}(\rho)} \ ,$$

with $\hat{\gamma}_{\mathrm{LS},k}(\rho)$ and $\hat{b}_{\mathrm{LS},k}(\rho)$ the least squares estimators for respectively γ and $b_{n,k}$ obtained from (2.3), taking ρ as fixed:

$$\text{(2.5)} \qquad \hat{\gamma}_{\mathrm{LS},k}(\rho) \ = \ \frac{1}{k}\sum_{j=1}^{k} Z_j - \frac{\hat{b}_{\mathrm{LS},k}(\rho)}{1-\rho} \ ,$$

$$\text{(2.6)} \qquad \hat{b}_{\mathrm{LS},k}(\rho) \ = \ \frac{(1-\rho)^2\,(1-2\rho)}{\rho^2}\,\frac{1}{k}\sum_{j=1}^{k}\left(\left(\frac{j}{k+1}\right)^{-\rho} - \frac{1}{1-\rho}\right) Z_j \ .$$

After some additional straightforward manipulations on (2.4), we obtain the bias-corrected kernel function:

$$\text{(2.7)} \qquad K_{\mathrm{BC}}(u;\rho) \ = \ K(u) - \frac{(1-\rho)^2\,(1-2\rho)}{\rho^2}\left(u^{-\rho} - \frac{1}{1-\rho}\right)\int_0^1 K(v)\,v^{-\rho}\,dv \ .$$

It is easy to verify that for kernel functions K satisfying the conditions of Theorem 2.1, K_{BC} will also satisfy these conditions with $\int_0^1 K_{\mathrm{BC}}(u;\rho)\,u^{-\rho}\,du = 0$, hence leading to an asymptotic normal distribution with null mean value, stated in the next theorem (Goegebeur *et al.*, 2007).

Theorem 2.2. *Consider $X_1, ..., X_n$ i.i.d. random variables according to distribution function F, where $F \in \mathcal{D}(G_\gamma)$ for some $\gamma > 0$, and with ℓ_U satisfying $\mathcal{R}_\ell$, fixed $\rho < 0$. If K satisfies the conditions of Theorem 2.1, then if $k, n \to \infty$, $k/n \to 0$ and $\sqrt{k}\,b(n/k) \to c$,*

$$\frac{\sqrt{k}}{\hat{\gamma}_{\mathrm{LS},k}(\rho)}\,\frac{1}{k}\sum_{j=1}^{k} K_{\mathrm{BC}}\left(\frac{j}{k+1};\rho\right) Z_j \ \xrightarrow{\mathcal{L}} \ N\left(0, \int_0^1 K^2_{\mathrm{BC}}(u;\rho)\,du\right) .$$

The bias-correcting effect of the above described operation can be readily seen from the limiting distribution: whatever c the normal limit is centered at zero. In case of the bias-corrected Lewis kernel function, denoted K_{BCL}, obtained by plugging K_{L} into (2.7), $\int_0^1 K^2_{\mathrm{BCL}}(u;\rho)\,du = 0$ if $\rho = -1$, leading to a degenerate distribution at zero. When dealing with this kernel function we exclude the value $\rho = -1$.

3. SELECTION OF THE NUMBER OF UPPER ORDER STATISTICS FOR TAIL INDEX ESTIMATION

In this section we discuss the use of the kernel goodness-of-fit statistic for selecting the optimal threshold in tail index estimation. The discussion will be focused on the Hill estimator, but the idea can of course be equally well applied to other estimators for $\gamma > 0$. The basic idea is to exploit the relationship between the bias component of the asymptotic mean squared error of the Hill estimator, denoted $AMSE(H_{k,n})$, and the kernel goodness-of-fit statistics introduced above. It is well known that for the Hill estimator

$$\begin{aligned} AMSE(H_{k,n}) &= \frac{\gamma^2}{k} + \left(\frac{b_{n,k}}{1-\rho}\right)^2 \\ &= \gamma^2 \left[\frac{1}{k} + \left(\frac{b_{n,k}}{\gamma(1-\rho)}\right)^2\right] . \end{aligned}$$

From Theorem 2.1, we have for the general kernel goodness-of-fit statistic, for k, n large and k/n small,

$$\frac{\frac{1}{k}\sum_{j=1}^{k} K\Big(\frac{j}{k+1}\Big) Z_j}{H_{k,n}} \sim \frac{b_{n,k}}{\gamma} \int_0^1 K(u)\, u^{-\rho}\, du \ ,$$

and hence, provided $\int_0^1 K(u)\, u^{-\rho}\, du \neq 0$,

$$\frac{b_{n,k}}{\gamma(1-\rho)} \sim \frac{\frac{1}{k}\sum_{j=1}^{k} K\Big(\frac{j}{k+1}\Big) Z_j}{(1-\rho)\, H_{k,n} \int_0^1 K(u)\, u^{-\rho}\, du} \ , \tag{3.1}$$

leading to the following approximation to $AMSE(H_{k,n})$

$$\widehat{AMSE}(H_{k,n}) = \gamma^2 \left\{\frac{1}{k} + \left[\frac{\frac{1}{k}\sum_{j=1}^{k} K\Big(\frac{j}{k+1}\Big) Z_j}{(1-\rho)\, H_{k,n} \int_0^1 K(u)\, u^{-\rho}\, du}\right]^2\right\} .$$

The optimal choice of k is then approximated by

$$\hat{k}_{\text{opt}} = \arg\min \left\{\frac{1}{k} + \left[\frac{\frac{1}{k}\sum_{j=1}^{k} K\Big(\frac{j}{k+1}\Big) Z_j}{(1-\rho)\, H_{k,n} \int_0^1 K(u)\, u^{-\rho}\, du}\right]^2\right\} . \tag{3.2}$$

Note that the squared goodness-of-fit statistic is to be complemented by a penalty $1/k$ in order to prevent choosing too small values of k. Also the role of ρ is important: typically, the smaller $|\rho|$ the heavier the ρ-factor with the test statistic leading to small values of k.

In the remainder of this section we will concentrate on the Lewis goodness-of-fit statistic, but of course similar results can be easily obtained for other kernel functions. For the Lewis statistic, $K_{\mathrm{L}}(u) = u - 0.5$, and hence, $\int_0^1 K(u)\, u^{-\rho}\, du =$

$|\rho|/\big[2\,(1-\rho)\,(2-\rho)\big]$, which leads to minimizing

$$\frac{1}{k} + \left[\frac{2\,(2-\rho)}{|\rho|\,\sqrt{k}}\;T^{\mathrm{L}}_{k,n}\right]^2 \tag{3.3}$$

with respect to k.

Practical implementations based on (3.2) or (3.3) require of course an estimate for the unknown parameter ρ. Gomes *et al.* (2002) proposed ratios involving different powers of statistics $M^{(r)}_{k,n}$, with

$$M^{(r)}_{k,n} \;=\; \frac{1}{k}\sum_{j=1}^{k}\Big(\log X_{n-j+1,n} - \log X_{n-k,n}\Big)^r$$

to derive estimators for ρ. In a similar fashion, we propose an estimator for ρ using a ratio of two kernel goodness-of-fit statistics. Define

$$T_i \;=\; \frac{1}{k}\sum_{j=1}^{k} K_i\left(\frac{j}{k+1}\right) Z_j \;, \qquad i=1,2\;,$$

where the indices 1 and 2 refer to the Jackson and Lewis goodness-of-fit statistics, for instance. From Proposition 2.1 and Proposition 2.2, we have, in probability,

$$\begin{aligned} T_1 \;&\sim\; b_{n,k}\,\frac{\rho}{(1-\rho)^2}\;,\\ T_2 \;&\sim\; -b_{n,k}\,\frac{\rho}{2\,(1-\rho)\,(2-\rho)}\;, \end{aligned}$$

and hence

$$\frac{T_1}{T_2} \;\sim\; -\frac{2\,(2-\rho)}{1-\rho}\;,$$

which can be solved for ρ, yielding

$$\hat\rho_k \;=\; \frac{4T_2+T_1}{2T_2+T_1}\;. \tag{3.4}$$

The asymptotic properties of this estimator will be discussed elsewhere.

As an alternative goodness-of-fit based procedure, the optimal k could be derived from comparing observed and fitted values on the Pareto quantile plot, for instance minimizing a weighted Cramér–von Mises statistic

$$\frac{1}{H^2_{k,n}}\,\frac{1}{k}\,\sum_{j=1}^{k}\frac{j}{k-j+1}\left(\log\frac{X_{n-j+1,n}}{X_{n-k,n}} + H_{k,n}\log\frac{j}{k+1}\right)^2\;. \tag{3.5}$$

Criteria of this type were considered in, for instance, Beirlant *et al.* (1996), and Dupuis and Victoria-Feser (2003). Unlike the goodness-of-fit based threshold selection procedure described above, this prediction error criterion does not require the estimation of the nuisance parameter ρ, but even asymptotically it will not minimize the AMSE.

The Lewis based AMSE criterion and the prediction error criterion will now be compared on the basis of a small sample simulation study. For the Lewis based AMSE criterion we consider three cases: ρ fixed at -1, correct specification of ρ and the case where ρ is replaced by (3.4). We simulated 500 samples of size $n = 500$ from the Burr(η, τ, λ) distribution, with distribution function given by

$$F(x) \;=\; 1 - \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda, \qquad x > 0, \quad \eta, \tau, \lambda > 0 \;,$$

for which $\gamma = 1/(\lambda\tau)$ and $\rho = -1/\lambda$. In Table 1, we summarize the results of the simulation study by the empirical mean squared errors (MSE) of $H_{\hat{k}_{\mathrm{opt}},n}$. For both procedures considered, the γ estimates deteriorate with increasing values of ρ. Clearly, the Lewis based AMSE approximation outperforms the prediction error criterion. Moreover, the gains in MSE tend to increase in ρ. Note that, although the Lewis based approximation of the AMSE requires an estimate for ρ, the results are quite insensitive with respect to the specification of ρ.

Table 1: Empirical MSE of $H_{\hat{k}_{\mathrm{opt}},n}$.

Distribution	γ	ρ	Lewis based AMSE criterion			Prediction error criterion
			$\rho = -1$	correct ρ	$\hat{\rho}$	
Burr$(1, 2, 0.5)$	1	-2	0.0103	0.0100	0.0109	0.0109
Burr$(1, 1, 1)$	1	-1	0.0275	0.0275	0.0288	0.0359
Burr$(1, 0.5, 2)$	1	-0.5	0.1178	0.1018	0.1199	0.1996
Burr$(1, 0.25, 4)$	1	-0.25	0.6869	0.5156	0.7195	1.1239
Burr$(1, 4, 0.5)$	0.5	-2	0.0029	0.0025	0.0030	0.0027
Burr$(1, 2, 1)$	0.5	-1	0.0069	0.0069	0.0072	0.0089
Burr$(1, 1, 2)$	0.5	-0.5	0.0299	0.0271	0.0308	0.0464
Burr$(1, 0.5, 4)$	0.5	-0.25	0.1771	0.1316	0.1741	0.2756

Besides this prediction error criterion we will also compare our goodness-of-fit based approach with some other criteria recently proposed. The computational complexity of some of these is such that they are not easy to implement for comparison purposes. Beirlant *et al.* (2002) performed an extensive simulation study and we will refer to some of their results, along with those from Beirlant *et al.* (1996) and Matthys and Beirlant (2003).

To summarize, the procedures that will be compared are:

- Method 1: the Lewis based criterion given by (3.3),
- Method 2: the prediction error criterion given by (3.5),
- Method 3: Beirlant *et al.* (2002),

- Method 4: Danielsson *et al.* (2001),
- Method 5: Drees and Kaufmann (1998),
- Method 6: Guillou and Hall (2001).

The performance of these procedures is evaluated on the basis of a small sample simulation study. In this simulation we use, next to the Burr(η, τ, λ) distribution introduced above, the following distributions:

1. The Fréchet(α) distribution,

$$F(x) = \exp(-x^{-\alpha}) , \qquad x > 0, \quad \alpha > 0 ,$$

with $\gamma = 1/\alpha$ and $\rho = -1$. We set $\alpha = 2$.

2. The $|T_\nu|$ distribution,

$$F(x) = \int_0^x \frac{2\,\Gamma\big(\frac{\nu+1}{2}\big)}{\sqrt{\nu\pi}\;\Gamma\big(\frac{\nu}{2}\big)} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}} dy , \qquad x > 0, \quad \nu > 0 ,$$

with $\gamma = 1/\nu$ and $\rho = -2/\nu$. We took $\nu = 6$.

3. The loggamma(λ, α) distribution,

$$F(x) = \int_1^x \frac{\lambda^\alpha}{\Gamma(\alpha)} (\log y)^{\alpha-1}\, y^{-\lambda-1}\, dy , \qquad x > 1, \quad \lambda, \alpha > 0 ,$$

with $\gamma = 1/\lambda$ and $\rho = 0$. We set $\lambda = 1$ and $\alpha = 2$.

For each of the above models, 500 datasets of size $n = 500$ are simulated. The results of the simulation are summarized in Table 2 where we show the empirical mean squared error of $H_{\hat{k}_{\rm opt},n}$ for the different methods and distributions considered. As is clear from Table 2 no single criterion performs uniformly best. The Lewis based approximation is clearly competitive and maintains itself in the first half of the methods considered.

Table 2: Empirical MSE of $H_{\hat{k}_{\rm opt},n}$.

Method	Fréchet(2)	Burr$(1, 0.5, 2)$	$\lvert T_6\rvert$	loggamma$(1, 2)$
1	0.0047	0.1178	0.0148	0.0873
2	0.0054	0.1996	0.0242	0.1105
3	0.0052	0.0930	0.0110	0.0602
4	0.0109	0.1459	0.0176	0.0904
5	0.0041	0.1239	0.0129	0.0784
6	0.0049	0.1452	0.0190	0.0689

4. Case study: diamond data

Our case study can be situated in a geostatistical context and concerns the valuation of diamonds. The profitability of a mining exploration largely depends on the occurrence of precious stones, and consequently, accurate modeling of the tail of the diamond value distribution is of crucial importance. The data set considered here contains the value (in USD) of a sample of 1914 diamonds obtained from a kimberlite deposit. These data are publicly available at `http://ucs.kuleuven.be/Wiley/Data/diamond.txt`. Figure 1 (a) shows the exponential quantile plot for the variable value; Figure 1 (b) is the corresponding mean excess plot. The convex shape of the exponential quantile plot and the means excess function that is decreasing when considered as a function of k indicate sub-exponential tail behavior. To assess the hypothesis of Pareto-type behavior we also construct the Pareto quantile plot, see Figure 1 (c). The Pareto quantile plot is clearly approximately linear in the largest observations indicating a good fit of the value distribution by a Pareto-type model. The mean excess function of the log-transformed data, which is in fact the Hill estimator, given in Figure 1 (d), confirms this in the sense that it clearly shows a constant slope at the smaller $\log k$ values.

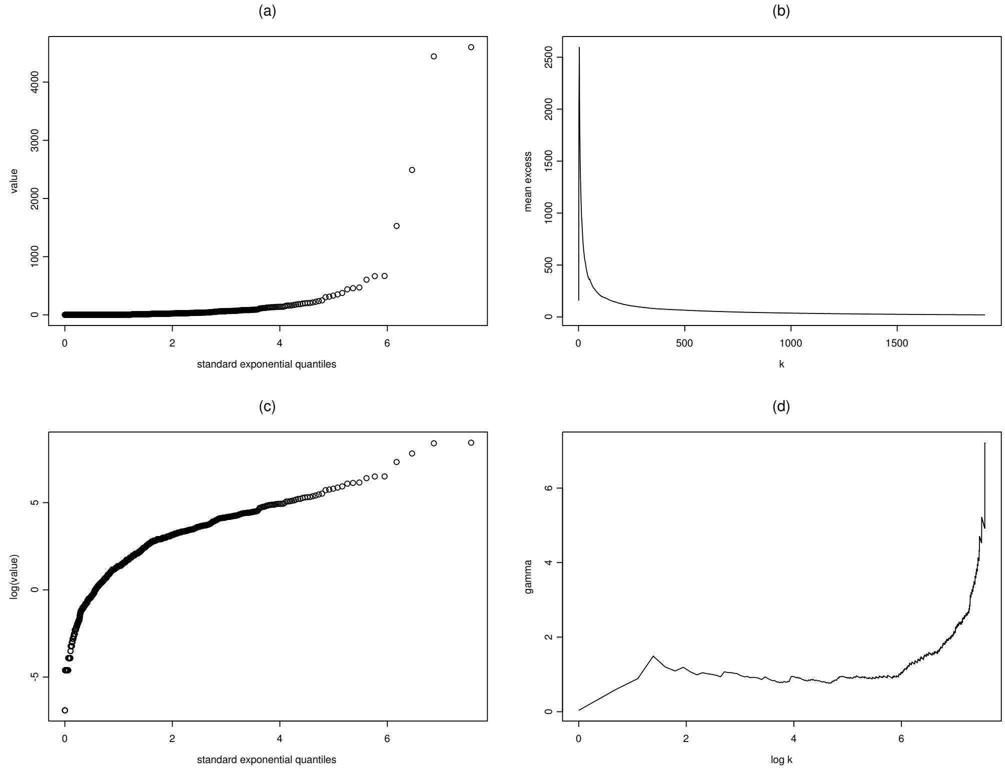


Figure 1: Diamond data: **(a)** exponential quantile plot, **(b)** mean excess plot, **(c)** Pareto quantile plot and **(d)** $H_{k,n}$ as a function of $\log k$.

In Figure 2 we show the four goodness-of-fit statistics together with the critical values of pointwise 5% hypothesis tests, as derived from the limiting distributions of the test statistics. For the ease of comparison we show the statistics in standardized format, i.e. we show $T^{\mathrm{J}}_{k,n}$, $\sqrt{12}\;T^{\mathrm{L}}_{k,n}$, $(1-\hat\rho)/|\hat\rho|\,T^{\mathrm{BCJ}}_{k,n}$ and $(2-\hat\rho)/|1+\hat\rho|\,\sqrt{12}\;T^{\mathrm{BCL}}_{k,n}$, where the scaling factors follow from the asymptotic variance expression $\int_0^1 K^2(u)\,du$, and where $T^{\mathrm{BCJ}}_{k,n}$ and $T^{\mathrm{BCL}}_{k,n}$ denote the bias-corrected Jackson and Lewis statistic, respectively, obtained by plugging K_{J} and K_{L} in (2.7). Globally, up to approximately $k = 380$, all statistics fail to reject H_0 of Pareto-type behavior with $\mathcal{R}_\ell$ on ℓ_U. The bias-corrected Lewis statistic shows two exceptions to this overall pattern, namely at the positions $k = 53$ and $k = 128$. These positions are indicated on the Pareto quantile plot given in Figure 3.

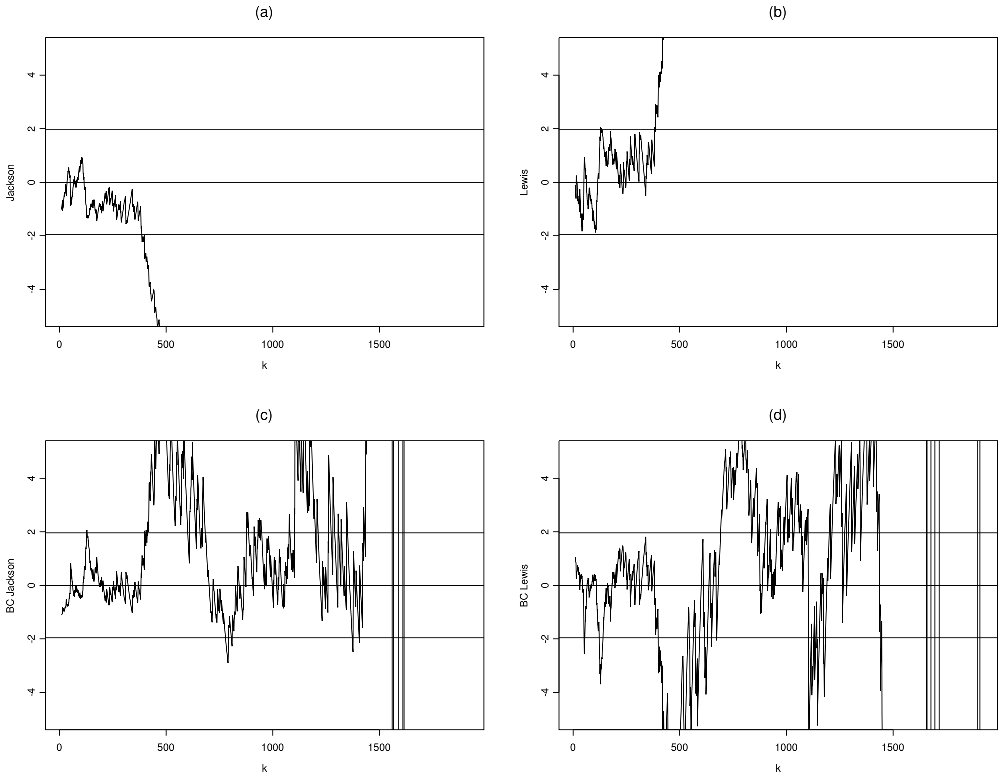


Figure 2: Diamond data: **(a)** $T^{\mathrm{J}}_{k,n}$, **(b)** $\sqrt{12}\,T^{\mathrm{L}}_{k,n}$, **(c)** $(1-\hat\rho)/|\hat\rho|\;T^{\mathrm{BCJ}}_{k,n}$, **(d)** $(2-\hat\rho)/|1+\hat\rho|\;\sqrt{12}\,T^{\mathrm{BCL}}_{k,n}$ as a function of k.

Clearly, at these positions the Pareto quantile plot makes vertical jumps. Beyond $k = 380$ the uncorrected statistics diverge and move outside the acceptance region, while the bias-corrected statistics fluctuate heavily and show portions of reasonable length both inside and outside the acceptance region, and hence give a more nuanced picture of the distributional behavior. A plausible explanation for this pattern can be found in the Pareto quantile plot. The Pareto quantile plot shows more or less linear segments on both the left- and right-hand side of the observation $k = 380$, although with different slopes. The uncorrected statistics can only handle the ultimate linear part of this plot and hence beyond this point they diverge. The bias-corrected statistics, through the inclusion of the second order tail condition, are also able to deal with the curved part. However, the portions inside and outside the acceptance region indicate special features of these data. Looking back at the Pareto quantile plot we find that also deeper in the data, i.e. at larger k-values, other linear portions with different slopes can be distinguished. This may indicate that the diamond value distribution is a mixture of several Pareto-type models with different Pareto indices. In fact, the tail of the diamond value distribution is known to be influenced by factors such as, among others, size and color (Beirlant and Goegebeur, 2003). In this analysis we ignored this information. It is a nice feature of the bias-corrected statistics that they indicate this change in distributional regime and give, compared to the uncorrected statistics, a more subtle view on the tail behavior.

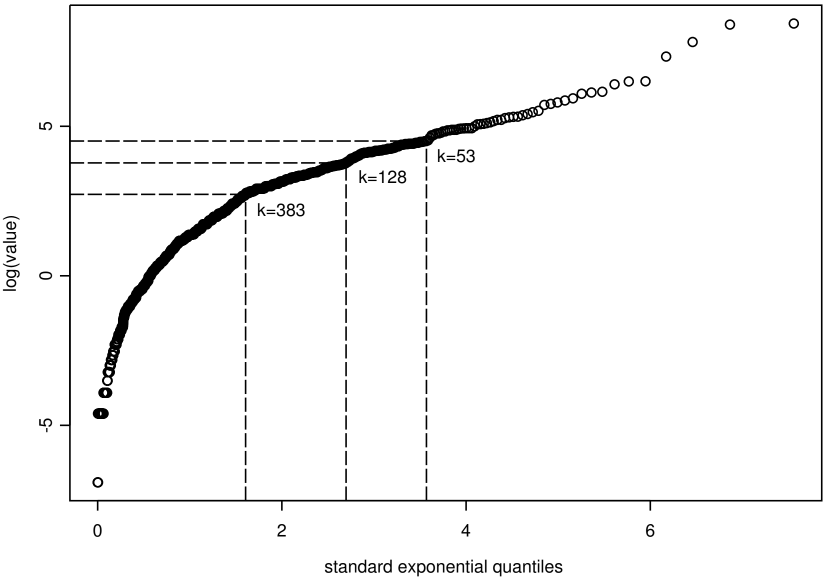


Figure 3: Diamond data: Pareto quantile plot with the positions where H_0 of Pareto-type behavior is rejected.

Given that the Pareto-type model provides a plausible explanation of these data, the analysis can be carried one step further, focusing on the estimation of the tail index γ. In this respect, Figure 4 (a) shows the Lewis based approximation to the asymptotic mean squared error of the Hill estimator, $\widehat{AMSE}(H_{k,n})$, obtained with $\hat{\rho} = -2.724$, as a function of k. This ρ-value is obtained from (3.4) together with the rule of thumb proposed by Gomes *et al.* (2002) that the k for the estimation of ρ can be taken as $k = \lfloor n^{0.995} \rfloor$, see also Figure 4 (e). The minimum value of $\widehat{AMSE}(H_{k,n})$ is reached at $\hat{k}_{\text{opt}} = 343$ and $H_{343,1914} = 0.917$. Note that

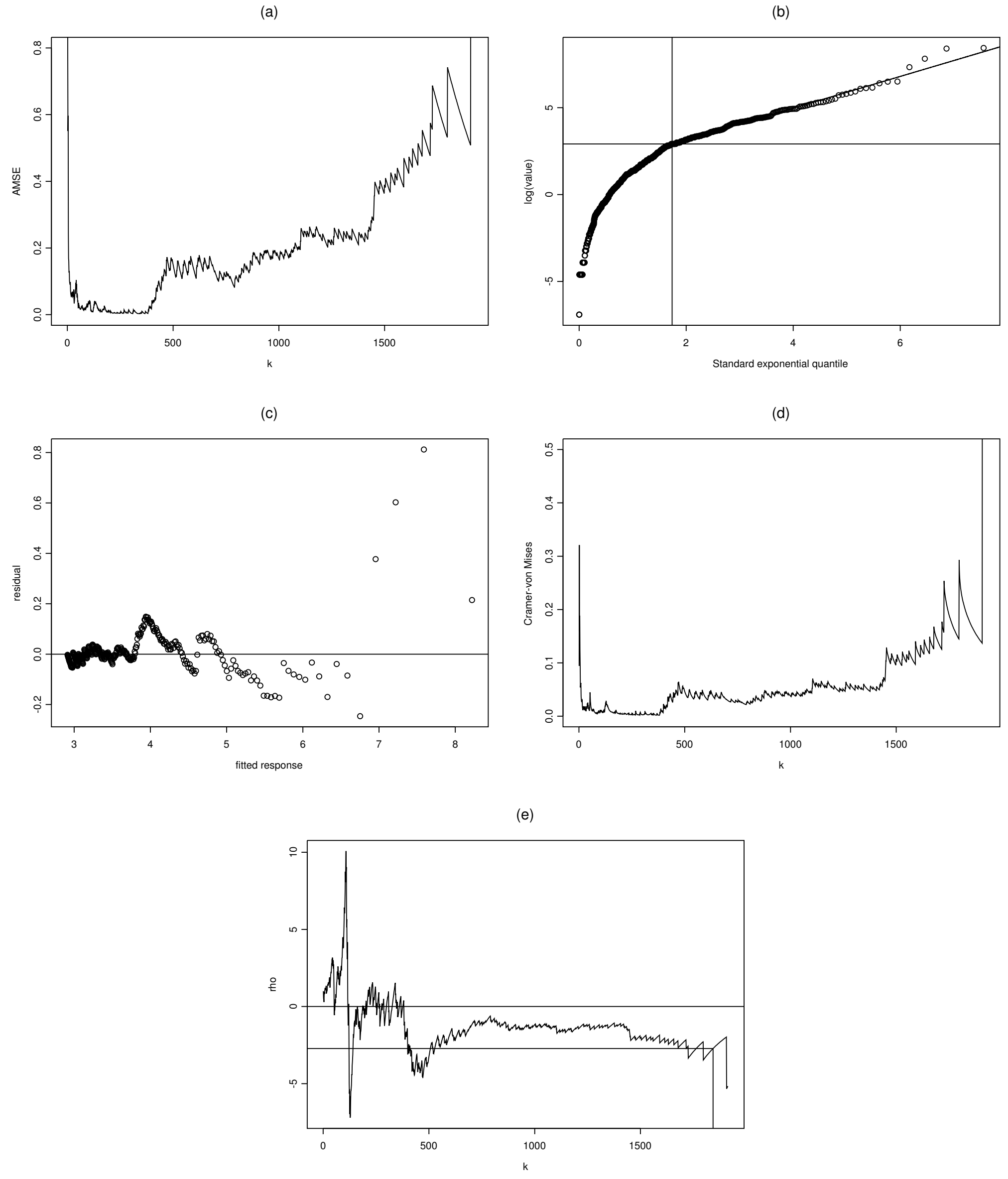


Figure 4: Diamond data: **(a)** $\widehat{AMSE}(H_{k,n})$ as a function of k, **(b)** Pareto quantile plot, **(c)** residuals versus fitted responses, **(d)** prediction error criterion, **(e)** $\hat{\rho}_k$ as a function of k.

the k-value minimizing the asymptotic mean squared error is smaller than the k-value beyond which goodness-of-fit tests consistently reject the null hypothesis given in (1.6). Alternatively, based on Figure 4 (e), we could also have taken $\hat{\rho} = -1$, which would result in $\hat{k}_{\text{opt}} = 336$ and $H_{336,1914} = 0.912$, results that are in line with those obtained with the former ρ-value. In Figure 4 (b) we indicate the 343$^{\text{th}}$ largest observation on the Pareto quantile plot of the variable value together with a straight line through this point and with slope $H_{\hat{k}_{\text{opt}},n}$. Clearly, the straight line summarizes the upper right portion of the Pareto quantile plot quite well, see also the Figure 4 (c) showing the residuals resulting from this line fit. Finally, in Figure 4 (d), we show the prediction error criterion (3.5) as a function of k. The prediction error reaches its minimum deeper in the data, at $k = 379$ and $H_{379,1914} = 0.940$, results that are comparable with the minimization of the asymptotic mean squared error.

5. CONCLUSION

In this paper we examined the relationship between Pareto-type goodness-of-fit testing and the selection of the upper sample fraction when estimating the tail index, for instance using Hill's estimator. To this end we considered the class of kernel statistics introduced in Goegebeur *et al.* (2007). Typically, goodness-of-fit tests are too conservative with respect to the null hypothesis, entailing too high k-values (or too small thresholds) relative to the minimum AMSE criterion, which led us to follow another route, exploiting the relationship between the kernel statistic and the bias component of the AMSE of the Hill estimator. The procedure was evaluated on a small sample simulation study and showed to be competitive with some of the better performing currently available algorithms. As a nice side result, we obtained a new estimator for the second order parameter ρ, of which the in-depth investigation is a topic of current research.

ACKNOWLEDGMENTS

Yuri Goegebeur's research was supported by a grant from the Danish Natural Science Research Council. Tertius de Wet's research was supported by National Research Foundation grant 2053332 and a research grant from the University of Stellenbosch.

REFERENCES

[1] Beirlant, J.; de Wet, T. and Goegebeur, Y. (2006). A goodness-of-fit statistic for Pareto-type behaviour, *Journal of Computational and Applied Mathematics*, **186**, 99–116.

[2] Beirlant, J.; Dierckx, G.; Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.

[3] Beirlant, J.; Dierckx, G.; Guillou, A. and Stărică, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**, 157–180.

[4] Beirlant, J. and Goegebeur, Y. (2003). Regression with response distributions of Pareto-type, *Computational Statistics and Data Analysis*, **42**, 595–619.

[5] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes – Theory and Applications*, Wiley Series in Probability and Statistics.

[6] Beirlant, J.; Vynckier, P. and Teugels, J.L. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics, *Journal of the American Statistical Association*, **91**, 1659–1667.

[7] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

[8] Danielsson, J.; de Haan, L.; Peng, L. and de Vries, C. (2001). Using a bootstrap method to choose sample fraction in tail index estimation, *Journal of Multivariate analysis*, **76**, 226–248.

[9] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer.

[10] de Wet, T. and Venter, J.H. (1973). A goodness-of-fit test for a scale parameter family of distributions, *South African Statistical Journal*, **7**, 35–46.

[11] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation, *Stochastic Processes and their Applications*, **75**, 149–172.

[12] Dupuis, D. and Victoria-Feser, M.-P. (2003). *A prediction error criterion for choosing the lower quantile in Pareto-index estimation*, Cahiers de Recherche HEC no. 2003.19, University of Geneva.

[13] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, **44**, 423–453.

[14] Goegebeur, Y.; Beirlant, J. and de Wet, T. (2007). *Kernel goodness-of-fit statistics for Pareto-type behavior*, submitted.

[15] Gomes, M.I.; de Haan, L. and Peng, L. (2002). Semi-parametric estimators of the second order parameter in statistics of extremes, *Extremes*, **5**, 387–414.

[16] Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis, *Journal of the Royal Statistical Society B*, **63**, 293–305.

[17] Henze, N. and Meintanis, S.G. (2005). Recent and classical tests for exponentiality: a partial review with comparisons, *Metrika*, **61**, 29–45.

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

[19] Jackson, O.A.Y. (1967). An analysis of departures from the exponential distribution, *Journal of the Royal Statistical Society B*, **29**, 540–549.

[20] Lewis, P.A.W. (1965). Some results on tests for Poisson processes, *Biometrika*, **52**, 67–77.

[21] Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models, *Statistica Sinica*, **13**, 853–880.

[22] Stephens, M.A. (1986). *Tests for the exponential distribution.* In "Goodness-Of-Fit Techniques" (R.B. D'Agostino and M.A. Stephens, Eds.), Marcel Dekker Inc., 421–459.

REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008, 71–81

ON EXTREME VALUE ANALYSIS OF A SPATIAL PROCESS

Authors: Laurens de Haan
– Erasmus University Rotterdam and University Lisbon,
The Netherlands and Portugal
ldehaan@few.eur.nl

Chen Zhou
– Tinbergen Institute, Erasmus University Rotterdam,
The Netherlands
zhou@few.eur.nl

Abstract:

- One common way to deal with extreme value analysis in spatial statistics is by using the max-stable process. By employing a representation of simple max-stable processes in de Haan and Ferreira ([3]), we propose a stationary max-stable process as a model of the dependence structure in two-dimensional spatial problems. We calculate its two-dimensional marginal distributions, which creates the opportunity to estimate the dependence parameter. The model is used in Buishand, de Haan and Zhou ([1]) for a spatial rainfall problem.

Key-Words:

- *spatial extremes; max-stable process.*

AMS Subject Classification:

- 62M30, 60G70.

1. INTRODUCTION

Problems of spatial statistics connected with high values of the spatial process need to be dealt with using extreme value theory (EVT), since the dependence between locations at high levels may differ from the dependence at moderate levels.

A case in point is the estimation of high quantiles of the total rainfall in a certain area. Engineers often need extreme rainfall statistics for the design of structures for flood protection. The observed rainfall data is only available on a few fixed monitoring stations. In order to study the high quantiles of the total rainfall, it is necessary to model the extreme rainfall process with dependence.

Considering the dependence structure, Cooley, Nychka and Naveau ([2]) used a Bayesian hierarchical model: locally the extreme rainfall is modeled by a one-dimensional EVT distribution and the parameters of this distribution follow some spatial dependence model.

A different way of introducing dependence is via a max-stable process. The mathematical setting of a spatial model for extreme rainfall is as follows. Consider independent replications of a stochastic process with continuous sample paths

$$\bigl\{X_n(t)\bigr\}_{t\in\mathbb{R}} \ ,$$

$n = 1, 2, \ldots$. Suppose that the process is in the domain of attraction of a max-stable process, that is, there are sequences of continuous functions $a_n > 0$ and b_n such that as $n \to \infty$

$$\left\{\frac{\max_{1\le i\le n} X_i(t) - b_n(t)}{a_n(t)}\right\}_{t\in\mathbb{R}} \xrightarrow{w} \bigl\{\tilde{\eta}(t)\bigr\}_{t\in\mathbb{R}} \tag{1.1}$$

in C-space. Necessary and sufficient conditions have been given by de Haan and Lin ([4]). The limit process $\{\tilde{\eta}(t)\}$ is a max-stable process. Without loss of generality we can assume that the marginal distribution of $\tilde{\eta}$ can be written as

$$\exp\Bigl\{-\bigl(1+\gamma(t)\,x\bigr)^{-1/\gamma(t)}\Bigr\}$$

for all x with $1+\gamma(t)\,x > 0$ where the function γ is continuous.

Buishand, de Haan and Zhou ([1]) simulated extreme rainfall from a max-stable process. Combining simulations of extreme rainfall with resampling from the non-extreme observations, an overview on the total rainfall can be generated. This is a novel solution for problems connected to both spatial statistics and extreme value analysis.

A major difficulty in the above methodology is to find a reasonable model for the max-stable process. With a suitable standardization, we can restrict ourselves to discussing the standardized process, called simple max-stable,

$$\big\{\eta(t)\big\} := \Big\{\big(1+\gamma(t)\,\tilde{\eta}(t)\big)_+^{1/\gamma(t)}\Big\}\,,$$

whose marginal distribution functions are all standard Fréchet: $\exp(-1/x)$, $x > 0$.

For application, it would be nice to have a stationary simple max-stable process. There are two different representations of stationary simple max-stable processes in literature. We consider one of them as follows, see Corollary 9.4.5, de Haan and Ferreira ([3]).

All simple max-stable process in $C^+(\mathbb{R})$ (the positive continuous functions on $\mathbb{R}$) can be generated in the following way. Consider a Poisson point process on $(0, +\infty]$ with mean measure dr/r^2. Let $\{Z_i\}_{i=1}^{\infty}$ be a realization of this point process. Further consider i.i.d. stochastic processes $V, V_1, V_2, \ldots$ in $C^+(\mathbb{R})$ with $EV(s) = 1$ for all $s \in \mathbb{R}$ and $E\,\sup_{s\in I} V(s) < \infty$ for all compact interval I. Let the point process and the sequence $V, V_1, V_2, \ldots$ be independent. Then

$$\big\{\eta(s)\big\}_{s\in\mathbb{R}} \overset{d}{=} \Big\{\max_{i\geq 1} Z_i\,V_i(s)\Big\}_{s\in\mathbb{R}} \tag{1.2}$$

is a simple max-stable process. Conversely each simple max-stable process has such a representation.

We use this result in a two-dimensional context and propose the following model

$$\eta(s_1, s_2) := \max_{i\geq 1} Z_i \exp\Big\{W_{1i}(\beta s_1) + W_{2i}(\beta s_2) - \beta\big(|s_1| + |s_2|\big)/2\Big\} \tag{1.3}$$

for $(s_1, s_2) \in \mathbb{R}^2$. The processes $W_{11}, W_{21}, W_{12}, W_{22}, W_{13}, W_{23}, \ldots$ are independent copies of double-sided Brownian motions W defined as follows. Take two independent Brownian motions B_1 and B_2. Then

$$W(s) := \begin{cases} B_1(s), & s \geq 0\,; \\ B_2(-s), & s < 0\,. \end{cases} \tag{1.4}$$

The positive constant β reflects the amount of spatial dependence at high levels of local observation: "β small" means strong dependence and "β large" means weak dependence. For this model, we shall prove that the dependence between extreme observations at two locations depends only on the distance between the locations.

The process η satisfies the requirements as follows:

$$E\,\exp\Big\{W_1(\beta s_1)+W_2(\beta s_2)-\beta\big(|s_1|+|s_2|\big)/2\Big\} = 1 \qquad \text{for } (s_1,s_2)\in\mathbb{R}^2\ ,$$

and

$$E \sup_{\substack{a_1\le s_1\le b_1\\ a_2\le s_2\le b_2}} \exp\Big\{W_1(\beta s_1)+W_2(\beta s_2)-\beta\big(|s_1|+|s_2|\big)/2\Big\} < \infty \qquad \text{for all } a_1<b_1,\ a_2<b_2 \text{ real}\,.$$

Meanwhile, the one-dimensional marginal distribution functions of (1.3) are all $e^{-1/x}$, $x>0$. Notice that only a one-dimensional Poisson point process is used in η. Thus, this process is easy to simulate.

Similar to de Haan and Pereira ([5]), in order to use this model in studying spatial extremes, we have to prove that the process η is shift stationary and we have to calculate the two-dimensional marginal distributions.

Since the two-dimensional process η is a combination of two one-dimensional processes, for the stationarity it is sufficient to prove the same for the one-dimensional version, i.e. that the process

$$\eta'(s) := \max_{i\ge1} Z_i \exp\Big\{W_{1i}(\beta s_1)-\beta|s_1|/2\Big\} \tag{1.5}$$

is stationary. This follows from the fact that the process η' can be obtained as the limit of the pointwise maximum of i.i.d. Ornstein–Uhlenbeck processes (cf. e.g. Example 9.8.2, de Haan and Ferreira ([3])). The stationarity follows from the stationarity of the Ornstein–Uhlenbeck process.

It remains to calculate the two-dimensional marginal distributions. This is done in Section 2.

2. THE TWO-DIMENSIONAL MARGINAL DISTRIBUTION OF η

The two-dimensional marginal distribution of η' in (1.5) is calculated in de Haan and Ferreira ([3]), section 9.8. We state it as the following proposition.

Proposition 2.1. *Suppose $\{\eta'(s)\}_{s\in\mathbb{R}}$ is defined as in (1.5). Then for $x,y\in\mathbb{R}$ and $s_1,s_2\in\mathbb{R}$,*

$$-\log P\Big(\eta'(s_1)\le e^x,\ \eta'(s_2)\le e^y\Big) = \\ = e^{-x}\,\Phi\left(\frac{\sqrt{|s_1-s_2|}}{2}+\frac{-x+y}{\sqrt{|s_1-s_2|}}\right)+e^{-y}\,\Phi\left(\frac{\sqrt{|s_1-s_2|}}{2}+\frac{x-y}{\sqrt{|s_1-s_2|}}\right).$$

This is useful in similar calculation for the two-dimensional process η. Besides Proposition 2.1, we need the following Lemma.

Lemma 2.1. *Suppose N is normally distributed with mean 0, variance u, then with non-random constants $a > 0$ and b,*

$$E\, e^{N-u/2}\, \Phi(aN+b) \;=\; \Phi\left(\frac{a\,u+b}{\sqrt{a^2 u+1}}\right) . \tag{2.1}$$

Proof: Suppose N_1 is standard normally distributed, and independent of N, then we have

$$E\, e^{N-u/2}\, 1_{N_1 \le aN+b} \;=\; E_N\, E\Big(e^{N-u/2}\, 1_{N_1\le aN+b}\,|\, N\Big) \;=\; E\, e^{N-u/2}\, \Phi(aN+b) \;,$$

which is the left side of (2.1). By Fubini's Theorem, it can be recalculated in the following way

$$\begin{aligned} E\, e^{N-u/2}\, 1_{N_1\le aN+b} \;&=\; E_{N_1} E\Big(e^{N-u/2}\, 1_{N_1\le aN+b}\,|\, N_1\Big) \\ &=\; E_{N_1} \int_{\frac{N_1-b}{a}}^{\infty} e^{t-u/2}\, \frac{1}{\sqrt{2\pi u}}\, e^{-\frac{t^2}{2u}}\; dt \\ &=\; E_{N_1} \int_{\frac{N_1-b}{a}}^{\infty} \frac{1}{\sqrt{2\pi u}}\, e^{-\frac{(t-u)^2}{2u}}\; dt \\ &=\; E_{N_1}\left(1-\Phi\left(\frac{N_1-b}{a\sqrt{u}}-\sqrt{u}\right)\right) . \end{aligned}$$

By a similar trick — introducing a standard normal variable N_2 independent of N_1, the calculation can be finished to prove the lemma.

$$\begin{aligned} E_{N_1}\left(1-\Phi\left(\frac{N_1-b}{a\sqrt{u}}-\sqrt{u}\right)\right) \;&=\; E_{N_1} E\Big(1_{N_2\ge \frac{N_1-b}{a\sqrt{u}}-\sqrt{u}}\,|\, N_1\Big) \\ &=\; E_{N_1,N_2}\; 1_{N_2\ge \frac{N_1-b}{a\sqrt{u}}-\sqrt{u}} \\ &=\; P\left(N_2 \ge \frac{N_1-b}{a\sqrt{u}}-\sqrt{u}\right) \\ &=\; \Phi\left(\frac{a\,u+b}{\sqrt{a^2 u+1}}\right) . \end{aligned}$$

We remark that the last calculation is similar to that of Lemma 2.1 in Gupta, González-Farías and Domínguez-Molina ([6]). □

The lemma can be used to derive the two-dimensional marginal distributions as follows. As in the proof of Proposition 2.1 (cf. de Haan and Ferreira ([3]), Section 9.8), we have

$$
(2.2) \qquad -\log P\Big(\eta(u_1,u_2)\le e^x,\ \eta(v_1,v_2)\le e^y\Big) =
$$

$$
\begin{aligned}
&= E\ \max\Big(e^{W_1(\beta u_1)+W_2(\beta u_2)-(|\beta u_1|+|\beta u_2|)/2-x},\ e^{W_1(\beta v_1)+W_2(\beta v_2)-(|\beta v_1|+|\beta v_2|)/2-y}\Big)\\
&= E_{W_1}\, E\Bigg(\max\Big(e^{W_1(\beta u_1)+W_2(\beta u_2)-(\beta|u_1|+\beta|u_2|)/2-x},\\
&\qquad\qquad e^{W_1(\beta v_1)+W_2(\beta v_2)-(\beta|v_1|+\beta|v_2|)/2-y}\Big)\ \Big|\ W_1\Bigg)\\
&= E\ e^{-x+W_1(\beta u_1)-\beta|u_1|/2}\\
&\quad\cdot\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)\\
&\quad+E\ e^{-y+W_1(\beta v_1)-\beta|v_1|/2}\\
&\quad\cdot\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{x-y+W_1(\beta v_1)-W_1(\beta u_1)-\beta|v_1|/2+\beta|u_1|/2}{\sqrt{\beta|u_2-v_2|}}\right).
\end{aligned}
$$

Now we can calculate the two parts in (2.2) separately. Without loosing generality, we only focus on the first part.

Case 1: $0\le u_1\le v_1$.
In this case $e^{-x+W_1(\beta u_1)-\beta|u_1|/2}$ is independent of the other part. Hence,

$$
\begin{aligned}
&E\ e^{-x+W_1(\beta u_1)-\beta|u_1|/2}\\
&\cdot\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)=\\
&= e^{-x}E\ \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x-\Big(W_1(\beta v_1)-W_1(\beta u_1)-\beta(v_1-u_1)/2\Big)}{\sqrt{\beta|u_2-v_2|}}\right)\\
&= e^{-x}P\left(N\le\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x-\Big(W_1(\beta v_1)-W_1(\beta u_1)-\beta(v_1-u_1)/2\Big)}{\sqrt{\beta|u_2-v_2|}}\right)\\
&= e^{-x}\ \Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta(v_1-u_1)}}{2}+\frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta(v_1-u_1)}}\right).
\end{aligned}
$$

Case 2: $0\le v_1<u_1$.
Note that $E\,e^{W_1(\beta v_1)-\beta v_1/2}=1$ and $W_1(\beta v_1)$ is independent of $W_1(\beta u_1)-W_1(\beta v_1)$,

we have

$$E\,e^{-x+W_1(\beta u_1)-\beta|u_1|/2} \cdot \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right) =$$

$$= e^{-x}\,E\,e^{W_1(\beta u_1)-W_1(\beta v_1)-\beta(u_1-v_1)/2} \cdot \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right).$$

Since $W_1(\beta u_1)-W_1(\beta v_1)$ is normally distributed with mean 0, variance $\beta(u_1-v_1)$, we can apply Lemma 2.1 with the constants $a=1/\sqrt{\beta|u_2-v_2|}$, $u=\beta(u_1-v_1)$ and

$$b = \frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x-\beta u_1/2+\beta v_1/2}{\sqrt{\beta|u_2-v_2|}}.$$

The final result is

$$E\,e^{-x+W_1(\beta u_1)-\beta|u_1|/2} \cdot \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right) =$$

$$= e^{-x}\,\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta(u_1-v_1)}}{2}+\frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta(u_1-v_1)}}\right).$$

Case 3: $v_1<u_1<0$ and $u_1\le v_1<0$.

These two cases are similar to Case 1 and 2 respectively. The final results are all the same as follows.

$$E\,e^{-x+W_1(\beta u_1)-\beta|u_1|/2} \cdot \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right) =$$

$$= e^{-x}\,\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta|u_1-v_1|}}{2}+\frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta|u_1-v_1|}}\right).$$

Case 4: u_1 and v_1 have different signs.

In this case $W_1(\beta u_1)$ and $W_1(\beta v_1)$ are independent, we can calculate the expectation with respect to $W_1(\beta v_1)$ first, then with respect to $W_1(\beta u_1)$.

$$E\,e^{-x+W_1(\beta u_1)-\beta|u_1|/2} \cdot \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right) =$$

$$= e^{-x}\,E\,e^{W_1(\beta u_1)-\beta|u_1|/2}\,\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}{2}+\frac{y-x+W_1(\beta u_1)-\beta|u_1|/2}{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}\right).$$

Now we can again apply Lemma 2.1 with the constants $a = 1/\sqrt{\beta|u_2-v_2|+\beta|v_1|}$, $u = \beta|u_1|$ and

$$b = \frac{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}{2} + \frac{y-x-\beta|u_1|/2}{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}$$

to get that

$$\begin{aligned} & E\, e^{-x+W_1(\beta u_1)-\beta|u_1|/2} \\ & \cdot \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2} + \frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right) = \\ & \quad = e^{-x}\,\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta\big(|u_1|+|v_1|\big)}}{2} + \frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta\big(|u_1|+|v_1|\big)}}\right). \end{aligned}$$

Notice that due to the different signs of u_1 and v_1, $|u_1-v_1| = |u_1|+|v_1|$.

By defining $h = |u_1-v_1|+|u_2-v_2|$, all these cases can be combined together as

$$\begin{aligned} & E\, e^{-x+W_1(\beta u_1)-\beta|u_1|/2} \\ & \cdot \Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2} + \frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right) = \\ & \quad = e^{-x}\,\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{y-x}{\sqrt{\beta h}}\right). \end{aligned}$$

Symmetrically, the second part of (2.2) can be simplified as

$$e^{-y}\,\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{x-y}{\sqrt{\beta h}}\right).$$

Combining these two parts, we get the following theorem about the two-dimensional marginal distribution of η.

Theorem 2.1. *Suppose the simple max-stable process η is defined in (1.3). Given any two coordinates (u_1,u_2) and (v_1,v_2) on $\mathbb{R}^2$, denote the distance between them as $h := |u_1-v_1|+|u_2-v_2|$. Then the two-dimensional distribution function of $\big(\eta(u_1,u_2), \eta(v_1,v_2)\big)$ is*

$$\begin{aligned} (2.3) \quad & P\Big(\eta(u_1,u_2) \le e^x,\ \eta(v_1,v_2) \le e^y\Big) = \\ & \quad = \exp\left\{-\left(e^{-x}\,\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{y-x}{\sqrt{\beta h}}\right) + e^{-y}\,\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{x-y}{\sqrt{\beta h}}\right)\right)\right\}, \end{aligned}$$

where Φ is the standard normal distribution function and $x, y \in \mathbb{R}$.

Note that the two-dimensional marginal distribution depends on only h. It agrees with the shift stationarity discussed in Section 1.

Similar to de Haan and Pereira ([5]), Theorem 2.1 is useful in estimating β. By taking $x = y = 0$, we get that

$$P\Big(\eta(u_1,u_2) \le 1,\, \eta(v_1,v_2) \le 1\Big) \;=\; \exp\left\{-2\,\Phi\left(\frac{\sqrt{\beta h}}{2}\right)\right\} .$$

Consequently, we have that

$$\beta \;=\; \frac{4}{h}\left(\Phi^{\leftarrow}\left(-\frac{1}{2}\,\log P\Big(\eta(u_1,u_2) \le 1,\, \eta(v_1,v_2) \le 1\Big)\right)\right)^2 .$$

Hence we can estimate β if we know how to estimate

$$L_{(u_1,u_2),(v_1,v_2)}(1,1) \;:=\; -\log P\Big(\eta(u_1,u_2) \le 1,\, \eta(v_1,v_2) \le 1\Big) .$$

In fact, this problem has been solved by Huang and Mason (cf. Huang ([8]), Drees and Huang ([7])). Suppose we have i.i.d. observations of η as $\eta_1, \eta_2, \ldots$. Write $\big\{\eta_{i,n}(s_1,s_2)\big\}_{i=1}^n$ for the order statistics at location (s_1,s_2). Then the estimator

$$\hat{L}^{(k)}_{(u_1,u_2),(v_1,v_2)}(1,1) \;:=\; \frac{1}{k}\sum_{j=1}^{n} 1_{\big\{\eta_j(u_1,u_2)\ge\eta_{n-k+1,n}(u_1,u_2) \;\text{ or }\; \eta_j(v_1,v_2)\ge\eta_{n-k+1,n}(v_1,v_2)\big\}}$$

is consistent provided $k = k(n) \to \infty$, $k(n)/n \to 0$, $n \to \infty$. It is asymptotically normal under certain mild extra conditions.

Hence, from the two-dimensional marginal distribution, we can estimate β when we have the observation at two specific locations. An application of this method is in Buishand, de Haan and Zhou ([1]), Section 5.

ACKNOWLEDGMENTS

Research partially supported by FCT/POCI 2010 and EXES–Project PTDC/MAT/64924/2006.

REFERENCES

[1] Buishand, T.A.; de Haan, L. and Zhou, C. (2008). On spatial extremes: with application to a rainfall problem, *Ann. Appl. Statist.*, To appear.

[2] Cooley, D.; Nychka, D. and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels, *J. Amer. Statist. Association*, **102**, 479, 824–840.

[3] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer.

[4] de Haan, L. and Lin, T. (2001). On convergence towards an extreme value distribution in C[0,1], *Ann. Prob.*, **29**, 467–483.

[5] de Haan, L. and Pereira, T.T. (2006). Spatial Extremes: Models for the stationary case, *Ann. Statist.*, **34**, 146–168.

[6] Gupta, A.; González-Farías, G. and Domínguez-Molina, J. (2004). A multivariate skew normal distribution, *J. Multivariate Anal.*, **89**, 181–190.

[7] Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function, *J. Multivariate Anal.*, **64**, 25–47.

[8] Huang, X. (1992). *Statistics of Bivariate Extreme Values*, PhD Thesis, Tinbergen Institute.

REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008, 83–100

TESTING EXTREME VALUE CONDITIONS — AN OVERVIEW AND RECENT APPROACHES

Authors: Cláudia Neves
– Department of Mathematics and UIMA,
University of Aveiro, Portugal
claudia.neves@ua.pt

M. Isabel Fraga Alves
– DEIO and CEAUL, Faculty of Sciences,
University of Lisbon, Portugal
isabel.alves@fc.ul.pt

Abstract:

- The aim of this paper is to give a brief overview about several tests published in the context of statistical choice of extreme value domains and for assessing extreme value conditions. Some of the most recent testing procedures encompassed in this framework will be illustrated using a teletraffic data set.

Key-Words:

- *extreme values; POT and PORT methodologies; location/scale invariance; statistical testing.*

AMS Subject Classification:

- Primary: 62G32; Secondary: 62F03, 62G10.

1. INTRODUCTION

Statistical inference about rare and damaging events can fairly be designed upon those observations which are considered extreme in some sense. There are different ways of mapping such observations yielding alternative approaches to statistical inference on extreme values: the classical Gumbel parametric method of block of *Annual Maxima*, *Peaks-Over-Threshold* (POT) parametric methods and the recently denominated *Peaks-Over-Random-Threshold* (PORT) semi-parametric methods, which is nothing more than a fairly small variant of POT for statistical inference conditionally on an intermediate random threshold.

However, regardless of the specific approach we intend to follow, statistical inference is clearly improved if one makes *a priori* assumptions about the most appropriate type of decay of the underlying tail distribution function $1-F$, i.e., about whether it decays exponentially fast, is polynomially decreasing or exhibits a light tail with finite right endpoint. This is supported by Extreme Value Theory, stemming from the fundamental Theorem of Fisher and Tippett (1928), which ascertains that all possible non-degenerate weak limit distributions of partial maxima of independent and identically distributed random variables $X_1, X_2, ...$ are (Generalized) Extreme Value distributions.

The Generalized Extreme Value distribution (GEVd) comprises Fréchet, Weibull and Gumbel distributions. A distribution function (d.f.) F that belongs to the Fréchet domain of attraction is called a heavy-tailed distribution, the Weibull domain encloses light-tailed distributions with finite right endpoint and the particularly interesting case of the Gumbel domain embraces a great variety of tail distribution functions ranging from light to moderately heavy, whether detaining finite right endpoint or not.

Hence, separating statistical inference procedures according to the most suitable domain of attraction for the underlying distribution has become a usual practice in the literature either by following a parametric or a semi-parametric approach. Following a semi-parametric approach, the only assumption made is that the underlying d.f. is in the domain of attraction of the GEVd. In this setup, any inference concerning the tail of the underlying distribution is based exclusively on those observations lying above an intermediate random threshold, giving rise to the PORT method. The latter compares with the alternative setup of restricting attention to a random number of observations exceeding a given high increasing deterministic level u, an approach engraved in the POT method.

Our aim here is to give a brief overview of several well-known testing procedures in the context of statistical choice of extreme value conditions, along with some recent proposals using location/scale invariant statistics that have been built on the k excesses above a random threshold. This random threshold is consensually an intermediate order statistic. The development of statistical procedures and techniques with the specific intention of dealing with extreme data in a more systematic and reliable way renders, to our best knowledge, a challenge that many applied fields such as environmetrics, climatology, telecommunications or finance hold in common.

The paper proceeds as follows. Section 2 contains some notation and sets general ground rules in the context of extreme value analysis. For analyzing extreme values there are different approaches, according to the underlying assumptions on F and the specific observations of the random sample available for statistical inference purposes. In this sequence, Sections 3 and 4 provide references and brief descriptions of several contributions in both parametric and semi-parametric setup. Finally, Section 5 brings the PORT-method into focus by means of an application to real data.

2. PRELIMINARIES AND SOME NOTATION

When we are interested in modeling large observations, we are usually confronted with two extreme value models:

- Generalized Extreme Value distribution (GEVd) with d.f.

$$G_\gamma(x) := \begin{cases} \exp\bigl(-(1+\gamma x)^{-1/\gamma}\bigr)\,, & 1+\gamma x>0 \quad \text{if} \quad \gamma\neq 0\,,\\ \exp\bigl(-\exp(-x)\bigr)\,, & x\in\mathbb{R} \quad \text{if} \quad \gamma=0\;. \end{cases} \tag{2.1}$$

- Generalized Pareto distribution (GPd) with d.f.

$$H_\gamma(x) := \begin{cases} 1-(1+\gamma x)^{-1/\gamma}\,, & 1+\gamma x>0 \text{ and } x\in\mathbb{R}^+ \quad \text{if} \quad \gamma\neq 0\,,\\ 1-\exp(-x)\,, & x\in\mathbb{R}^+ \quad \text{if} \quad \gamma=0\;. \end{cases} \tag{2.2}$$

The introduction of scale $\delta>0$ and location $\lambda\in\mathbb{R}$, results in the full GEV and GP families of distributions given by $G_\gamma(x;\lambda,\delta)=G_\gamma\bigl((x-\lambda)/\delta\bigr)$ and $H_\gamma(x;\lambda,\delta)=H_\gamma\bigl((x-\lambda)/\delta\bigr)$, respectively, which play a central role in statistical inference of extreme values.

GEVd and MAX-Domain: The Fisher–Tippett theorem of extreme values (Fisher and Tippett, 1928) states that all possible non-degenerate weak limit distributions of partial maxima of independent and identically distributed (i.i.d.) random variables $X_1, X_2, ...$ are (Generalized) Extreme Value distributions. That is, assume there exist normalizing constants $a_n>0$ and $b_n\in\mathbb{R}$ such that, for all x

$$\lim_{n\to\infty} P\Big\{a_n^{-1}\big(\max(X_1, ..., X_n) - b_n\big) \le x\Big\} = G(x) \ , \tag{2.3}$$

where G is some non-degenerate distribution function, we can redefine the constants in such a way that the limit G is one of the GEV family of distributions given by (2.1) in the von Mises–Jenkinson form (von Mises, 1936; Jenkinson, 1955). We then say that $G = G_\gamma$ and the underlying d.f. F is in the domain of attraction of G_γ (notation: $F \in \mathcal{D}(G_\gamma)$). In case of $\gamma < 0$, $\gamma = 0$ or $\gamma > 0$, the G_γ reduces to Weibull, Gumbel or Fréchet distribution function, respectively.

GPd and POT-Domain: The use of GPd is suggested by the result of Balkema and de Haan (1974) and Pickands (1975), who proved that $F \in \mathcal{D}(G_\gamma)$ if and only if the upper tail of F is, in a certain sense, close to the upper tail of H_γ. While restricting attention to a top portion of the original sample, the GPd comes into play since it appears as the limiting distribution for the excesses $Y_i = X_i - u\,|X_i > u$, $i = 1, ..., k_u$ over a sufficiently high threshold u (POT method). For $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$, the H_γ d.f. in (2.2) reduces to Beta, Exponential and Pareto distribution functions, respectively. In both classes, the extreme value index γ is closely related to the tail heaviness of the distribution. In that sense, the value $\gamma = 0$ (*exponential tail*) can be regarded as a change point: $\gamma < 0$ refers to short tails with finite right endpoint $x^F := \sup\big\{x \colon F(x) < 1\big\}$, whereas for $\gamma > 0$ d.f.'s are heavy tailed. In many applied sciences where extremes come into play, it is assumed that the extreme value index γ of the underlying d.f. equals 0, and statistical inference procedures concerning rare events on the tail of F, such as the estimation of small exceedance probabilities or return periods, bear on this assumption. Moreover, Gumbel and exponential models are also preferred because of the greater simplicity of inference associated with Gumbel or exponential populations.

Here and throughout this paper, let us denote by $X_{1:n} \le ... \le X_{n:n}$ the order statistics pertaining to the i.i.d. random variable $X_1, X_2, ..., X_n$, after arranging these by nondecreasing order.

3. TESTING EXTREMES UNDER A PARAMETRIC APPROACH

In a parametric set-up, the main assumption regards the existence of a suitable class of models for describing the random variable attached to the process that is generating the data under study. These only three possible classes are motivated by Extreme Value Theory, and depend mainly on the shape parameter γ, and eventually on location and scale parameters.

Annual Maxima (AM): Suppose that the maximum of a random sample can be obtained in each of k equally spaced observation periods. The class of GEVd functions, G_γ, may be prescribed in order to model maxima of k subsamples taken from a given set of data of size $k \cdot n$, that is,

$$Z_i := X_{n:n}^{(i)} = \max\Big\{X_1^{(i)}, ..., X_n^{(i)}\Big\}\ , \qquad i = 1, ..., k\ . \tag{3.1}$$

A typical course of action lying in this classical Gumbel method is to take annual maxima. In this AM setup, the following testing problem has been treated extensively in the literature, with main emphasis on testing the Gumbel hypothesis for the d.f. of the $\{Z_i\}_{i=1}^k$ defined in (3.1):

$$H_0\colon\ \gamma = 0 \qquad \text{vs.} \qquad H_1\colon\ \gamma \neq 0\ . \tag{3.2}$$

The testing problem (3.2) has received much attention in the literature; in fact, the hydrologists have long made use of extreme value distributions for estimating probabilities of flood events and the correct choice of the GEVd under approach is of crucial importance, since the three types differ considerably in their right tails. Among the papers concerned with the special testing problem G_0 against $\{G_\gamma\colon\ \gamma \neq 0\}$, or against one-sided alternatives $\{G_\gamma\colon\ \gamma > 0\}$, $\{G_\gamma\colon\ \gamma < 0\}$, we refer to Van Montfort (1970), Bardsley (1977), Otten and Van Montfort (1978), Tiago de Oliveira (1981), Gomes (1982), Tiago de Oliveira (1984), Tiago de Oliveira and Gomes (1984), Hosking (1984), Marohn (1994), Wang *et al.* (1996) and Marohn (1998a). Somewhat connected with the problem (3.2), there is the problem of goodness-of-fit tests for the Gumbel model, which has received the attention of Stephens (1976), Stephens (1977), Stephens (1986) and Kinnison (1989). The tests therein considered are mostly based on the well known goodness-of-fit statistics: Kolmogorov, Cramér–von Mises and Anderson–Darling statistics.

Largest Observations (LO): It may be that, when considering yearly data, some years contain several values that are larger then the maxima of other years. Although the requirement of only a simplified data summary carries reduction of possible dependencies in the sampled data, the loss of information provided by the largest observations in the sample can, by itself, motivate this alternative approach. Hence, suppose we take the k largest observations in the sample. If the underlying d.f. $F \in \mathcal{D}(G_\gamma)$, the non-degenerate joint limiting behavior of the k largest random variables determines the probability density function (p.d.f.)

$$f_\gamma(z_1, z_2, ..., z_k) \;=\; g_\gamma(z_k) \prod_{i=1}^{k-1} \frac{g_\gamma(z_i)}{G_\gamma(z_i)}\ , \qquad z_1 > z_2 > ... > z_k\ , \tag{3.3}$$

where $g_\gamma(z) = \partial G_\gamma(z)/\partial z$, in the sense that, after appropriately normalized with constants $a_n > 0$ and b_n,

$$\left(\frac{X_{n:n} - b_n}{a_n}, \frac{X_{n-1:n} - b_n}{a_n}, ..., \frac{X_{n-k+1:n} - b_n}{a_n}\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \Big(Z_1, Z_2, ..., Z_k\Big)\ .$$

This multivariate model, introduced in Weissman (1978), has received the general designation of *extremal process*. In light of this result, statistical procedures to discern between Gumbel and Fréchet or Weibull distributions have been considered, for instance, in Gomes and Alpuim (1986), Hasofer and Wang (1992), Wang (1995) and Wang *et al.* (1996). The key insight for the testing problem G_0 against $\{G_\gamma \colon \gamma \neq 0\}$, or against the one-sided alternatives $\{G_\gamma \colon \gamma > 0\}$, $\{G_\gamma \colon \gamma < 0\}$ is thus to assume that, with k fixed, the joint stochastic behavior of the largest k random variables tends to be properly described by the p.d.f (3.3) pertaining to $\gamma = 0$, i.e.,

$$f_0(z_1, z_2, ..., z_k) \;=\; \exp\left(-\exp(-z_k) - \sum_{i=1}^{k} z_i\right) , \qquad z_1 > z_2 > ... > z_k \;,$$

which enables replacement of the normalized top order statistics with $(Z_1, Z_2, ..., Z_k)$. Under this assumption, Hasofer and Wang (1992) prove that the following test statistic

$$W(k) \;:=\; \frac{1}{k-1}\, \frac{\left(\frac{1}{k}\sum\limits_{i=1}^{k} Z_i - Z_k\right)^2}{\frac{1}{k}\sum\limits_{i=1}^{k}\left(Z_i - Z_k\right)^2 - \left(\frac{1}{k}\sum\limits_{i=1}^{k} Z_i - Z_k\right)^2} \;,$$

akin to the Shapiro–Wilk goodness-of-fit statistic (see Shapiro and Wilk, 1965), can be considered as approximately normal with mean $(k-1)^{-1}$ and variance $2^2\,(k-2)\,(k-1)^{-2}\,\big((k+1)\,(k+2)\big)^{-1}$.

Despite the above results concern a fixed number k of top observations, we can find in Hasofer and Wang (1992) an attempt to make k to increase with n, but at a much slower rate, through the specification of $k = c_1\, n^{c_2}$ in the simulation study. Wang (1995) also mention the case where $k \to \infty$ and $k = o(n)$, as $n \to \infty$. Furthermore, Wang (1995) relies on the Hasofer and Wang test to select the number k of largest order statistics for suitable statistical inference in the Gumbel domain. In general, if G is a goodness-of-fit statistic, then at a certain nominal level of the test α, say, choose $k+1 = \min\big\{i \colon\, g(i) \in \text{critical region of } G(i)\big\}$, provided the adopted statistic G is scale and location invariant and, of course, sensitive to small deviations from the null hypothesis.

Combination of AM and LO: In Gomes (1989), for instance, the testing problem specifying the Gumbel d.f. G_0 in the (simple) null hypothesis is handled with a combination of blocking split of the sample data and the k largest observations in each of the m blocks through what is called the *multidimensional — GEV$_\gamma$ model*, as follows: a set of independent, identically distributed k-dimensional random vectors $\{\mathbf{X}_i \colon\, i = 1, ..., m\}$, and after suitable normalization, with common p.d.f of the vectors $\mathbf{Z}_i = (\mathbf{X}_i - \lambda)/\delta$ is given by $f_\gamma(\mathbf{z})$ defined in (3.3). Note that both AM and LO approaches can be particular cases of this multidimensional model, taking $k=1$ and $m=1$, respectively. In Gomes (1987) a truncated sample of the largest values of a sample, whose size increases to infinity and whose limiting distribution is in the class of GEVd is considered for goodness-of-fit purposes. Using the reduction to the exponentials of Gumbel distributions the author develops two-sided tests of exponentiality for the transformed variables, the tests being the Kolmogorov–Smirnov, Cramér–von Mises and Stephens goodness-of-fit test.

Peaks Over Threshold (POT): Suppose we pick up those observations exceeding a fixed high threshold u. As described in Section 2, given a random sample $(X_1, X_2, ..., X_n)$ from the d.f. F, the GPd is regarded as a good approximation for the distribution of the excesses $W_i := X_i - u$ over a sufficiently high threshold u if and only if $F \in \mathcal{D}(G_\gamma)$. A clear difference between the designated AM and POT setups is that the k yearly maxima do not necessarily carry over as to yield the k largest observations from the original sample.

In this POT setup the following testing problem has been frequently considered, rendering priority to testing the Exponential hypothesis for the d.f. of the excesses $\{W_i\}_{i=1}^{k_u}$, i.e., $H_0\colon \gamma=0$ *versus* $H_1\colon \gamma\neq 0$. The maximum likelihood method may then be applied under the assumption that those k_u observations over the threshold u follow exactly a GPd, provided a scale normalization σ_u, i.e.

$$H_{\gamma,\sigma_u}(w) \;=\; 1-\bigl(1+\gamma w/\sigma_u\bigr)^{-1/\gamma} \ ,$$

for all positive w such that $1+\gamma/\sigma_u\, w>0$. The parametrization $\tau=-\gamma/\sigma$ (Davison and Smith, 1990; Grimshaw, 1993) can be used for reducing dimensionality and therefore construct a likelihood ratio test based on the log-profile likelihood. In view of applications, the problem of detecting the presence of exponential distribution, under the POT approach, has received particular attention from hydrologists. Davison and Smith (1990) addresses this testing problem in the context of river-flow exceedances. Van Montfort and Witter (1986) illustrates the "lack"-of-fit statistic towards exponentiality $\hat\gamma/\sqrt{\hat{\mathrm{var}}(\hat\gamma)}$, where $\hat\gamma$ denotes the Maximum Likelihood (ML) estimator of γ, in the sequence of a thorough application of the POT method to rainfall data. Among the numerous works connected with the special problem of testing exponential against other GPd upon the tail we mention, for instance, Van Montfort and Witter (1985), Gomes and Van Montfort (1986) and Brilhante (2004). Chaouche and Bacro (2004) introduce the test statistic $S=\overline{W}/(\overline{W}-W_{n:n})$, where again W_i are independent random variables with the same d.f. H_{γ,σ_u}, and obtain its empirical distribution via simulation. Moreover, when using Probability Weighted Moments of different orders to adapt S, a method to purge the influence of σ_u off these new test statistics is provided. Giving heed to the Local Asymptotic Normality theory, Falk (1995) followed by Marohn (1998b) and Marohn (2000), aim at asymptotically optimal tests for discriminating between different values of the extreme value index γ.

Goodness-of-fit tests for the Generalized Pareto distribution

Fitting the GPd function to data, which we expect to be lying far away in the tail, has been worked out in Castillo and Hadi (1997). The problem of goodness-of-fit tests for the GP model has been studied by Choulakian and Stephens (2001), with the following proposals for Cramér–von Mises and Anderson–Darling statistics:

$$W^2 = \sum_{i=1}^{k} \left(H_{\hat\gamma,\hat\sigma}(X_{n-i+1:n}) - \frac{2(k-i)+1}{2k} \right)^2 + \frac{1}{12\,k} ,$$

$$A^2 = -k - \frac{1}{k} \sum_{i=1}^{k} \Big(2(k-i)+1 \Big) \Big(\log H_{\hat\gamma,\hat\sigma}(X_{n-i+1:n}) + \log\big(1 - H_{\hat\gamma,\hat\sigma}(X_{i:n})\big) \Big) ,$$

where $(\hat\gamma, \hat\sigma)$ are ML estimators. A table of critical points is provided with good accuracy for $k \geq 25$. Konstantinides and Meintanis (2004) assess the presence of a GPd by means of a transformation of the data to reduce to exponential, then search for traces of exponentiality in the empirical Laplace transform. They also adapt the critical points leading to what promises to be a more accurate level of the test, pursuing the path of Davison and Smith (1990) claim that tables for testing the presence of a exponential distribution (see Van Montfort and Witter (1986) lack of fit statistic mentioned upstairs) give in general critical values which are too high, thus resulting in a very conservative test. Comparison with Choulakian and Stephens (2001) are also present by means of a simulation study. Luceño (2006) assigns more weight to the tails than the usual practice relating Cramér–von Mises and Aderson–Darling statistics goodness-of-fit test statistics and considers a maximum goodness-of-fit estimation method, which enables us to deal successfully with the estimation of GPd parameters, overcoming the occasional lack of convergence in ML estimation.

4. TESTING EXTREMES UNDER A SEMI-PARAMETRIC APPROACH

Following a semi-parametric approach, the only assumption made is that the extreme value condition (2.3) is satisfied, i.e., the underlying d.f. $F \in \mathcal{D}(G_\gamma)$. In this framework, the extreme value index γ is the parameter of prominent interest since, in both GEV and GP classes of distributions, it determines the shape of the tail of the underlying distribution function F.

To this extent, $\gamma = 0$ can be regarded as a benchmark value, since a negative γ is inevitably associated with short tails with finite right endpoint, while a positive (tail index) γ is connected with the presence of a heavy-tailed distribution. In many applied sciences where extremes are relevant, the case of simplest inference $\gamma = 0$ is assumed and bearing on this assumption, extreme characteristics such as exceedance probabilities or return periods are easily estimated.

As a matter of fact, separating statistical inference procedures according to the most suitable domain of attraction for the sampled distribution has become a usual practice. Methodologies for testing the Gumbel domain against Fréchet or Weibull max-domains have been of great usefulness. This fit-of-attraction problem, crafted from a semi-parametric setup, can be rephrased as a test for

$$H_0\colon\ F \in \mathcal{D}(G_0) \quad \textit{versus} \quad H_1\colon\ F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0}\ . \tag{4.1}$$

or against one-sided alternatives $F \in \mathcal{D}(G_\gamma)_{\gamma<0}$ or $F \in \mathcal{D}(G_\gamma)_{\gamma>0}$.

Statistical tests that tackle the problem (4.1) can be traced back to the seminal papers by Galambos (1982) and Castillo *et al.* (1989). The latter presents a cunning procedure for fit of attraction diagnostics from the curvature of the graph of the sample distribution function hinged on the Gumbel probability paper. Predicated on this (so-called) curvature method, the authors introduce a test to assess whether the upper tail distribution function might be classified as convex, concave or a straight line.

Further testing procedures for (4.1) can be found in Fraga Alves and Gomes (1996), Fraga Alves (1999). Segers and Teugels (2000) have recently suggested a large sample test for the Gumbel domain with asymptotics deriving from the limiting distribution of Galton's ratio under the extreme value condition (2.3), which Rao's test statistic (see e.g. Serfling, 1980) for simple null hypothesis was applied to, with the ulterior aim of establishing a decision rule. In the process, the authors were confronted with the need of blocking the original sample of size n into m subsamples, each of size n_i, $i = 1, ..., m$ also under pledge of largeness.

Recently, Neves *et al.* (2006) and Neves and Fraga Alves (2007) have introduced two testing procedures that are based on the sample observations lying above a random threshold. More specifically, in the last two references, the designed statistics for testing (4.1) are based on the k excesses over the $(n-k)$-th ascending intermediate order statistic $X_{n-k:n}$, where $k = k_n$ is such that $k \to \infty$ and $k = o(n)$ as $n \to \infty$. Clearly, the latter only differs from the POT approach on the absence of a parametric model and on the fact that the *intermediate* random threshold is now playing the role of the deterministic *sufficiently high* threshold u which, only by itself, we find relevant enough to motive the Peaks Over Random Threshold (PORT) methodology. Now following a semi-parametric approach supported on concepts from the theory of regularly varying functions, Neves and Fraga Alves (2007), reformulate the asymptotic properties of the Hasofer and Wang test statistic (denoted below with $W_n(k)$) in case $k = k_n$ behaves as an intermediate sequence rather than remaining fixed while the sample size n increases (which was case covered by Hasofer and Wang, 1992). In the process, a new Greenwood-type test statistic $G_n(k)$ (cf. Greenwood, 1946) proves to be useful in assessing the presence of heavy-tailed distributions.

Furthermore, motivated by eventual differences in the relative contribution of the maximum to the sum of the k excesses over the random threshold at different tail heaviness, a complementary test statistic $R_n(k)$ was introduced by Neves *et al.* (2006) in order to discern between max-domains of attraction.

Under the null hypothesis of Gumbel domain of attraction plus extra mild second order conditions on the upper tail of F and on the growth of the intermediate sequence k_n, we have that

$$\text{(4.2)} \qquad \text{[Ratio-test]} \qquad R_n(k) := \frac{X_{n:n}-X_{n-k:n}}{\frac{1}{k}\sum\limits_{i=1}^{k}\big(X_{n-i+1:n}-X_{n-k:n}\big)} - \log k \xrightarrow[n\to\infty]{d} \Lambda ,$$

$$\text{(4.3)} \qquad \text{[Gt-test]} \qquad G_n(k) := \frac{\frac{1}{k}\sum\limits_{i=1}^{k}\big(X_{n-i+1:n}-X_{n-k:n}\big)^2}{\left(\frac{1}{k}\sum\limits_{i=1}^{k}X_{n-i+1:n}-X_{n-k:n}\right)^2} ,$$

$$\sqrt{k/4}\,\big(G_n(k)-2\big) \xrightarrow[n\to\infty]{d} N(0,1) ,$$

$$\text{(4.4)} \qquad \text{[HW-test]} \qquad W_n(k) := \frac{1}{k}\left[1-\frac{G_n(k)-2}{1+\big(G_n(k)-2\big)}\right],$$

$$\sqrt{k/4}\,\big(k\,W_n(k)-1\big) \xrightarrow[n\to\infty]{d} N(0,1) ,$$

where Λ stands for a Gumbel random variable. The critical regions for testing the two-sided alternative (4.1), at a nominal size α, are given by $V_n(k) < v_{\alpha/2}$ or $V_n(k) > v_{1-\alpha/2}$, where V has to be conveniently replaced by T, R, or W and v_ε denotes the ε-quantile of the corresponding limiting distribution. The limiting distribution of $G_n(k)$ [*resp.* $W_n(k)$] shifts towards the right [*resp.* left] for distributions in the Fréchet domain of attraction ($F \in \mathcal{D}(G_\gamma)_{\gamma>0}$) and towards the left [*resp.* right] for distributions lying in the Weibull domain ($F \in \mathcal{D}(G_\gamma)_{\gamma<0}$). Notice that the test statistic S in Chaouche and Bacro (2004) may be seen as the POT-counterpart of $\big(1-R_n(k)\big)^{-1}$. An extensive simulation study involving Ratio, Gt and HW tests, let us to perceived the following guidelines:

- The test based on the G_n^* is shown to good advantage when testing the presence of heavy-tailed distributions is in demand.
- While the Gt-test barely detects small negative values of γ, the HW is the most powerful test under study with respect to alternatives in the Weibull domain of attraction.
- The simulations have emphasized the admonition for controlling the actual size of the test to apply, keeping low within acceptable bounds the probability of incorrect rejection of the null hypothesis. Since the test based on the very simple Ratio statistic tends to be a conservative test and yet detains a reasonable power, it proves to be a valuable complement to the remainder procedures.

Testing Extreme Value conditions

From its grounds, any inferential methodology considered in the field of Extreme Values is inextricably bound to the validity of an extreme value condition. Inevitably, the methods of the previous sections do not escape such a requirement. Hence, assessing whether the hypothesis that "$F \in \mathcal{D}(G_\gamma)$" is strongly supported by the data at hand, becomes an impending problem. On this matter, Dietrich *et al.* (2002) introduce the test statistic

$$(4.5) \qquad E_n(k) := k \int_0^1 \left(\frac{\log X_{n-[kt]:n} - \log X_{n-k:n}}{\hat{\gamma}_+} - \frac{t^{-\hat{\gamma}_-} - 1}{\hat{\gamma}_-} \, (1-\hat{\gamma}_-) \right)^2 t^\eta \, dt \, ,$$

for some $\eta > 0$, where $\hat{\gamma}_+$ and $\hat{\gamma}_-$ are the same estimators of $\gamma_+ = \max(0,\gamma)$ and $\gamma_- = \min(\gamma, 0)$ as in Dekkers *et al.* (1989). Furthermore, in case we wish to test the null hypothesis that $F \in \mathcal{D}(G_\gamma)_{\gamma \geq 0}$, a simple version is available:

$$PE_n(k) := k \int_0^1 \left(\frac{\log X_{n-[kt]:n} - \log X_{n-k:n}}{\hat{\gamma}_+} + \log t \right)^2 t^\eta \, dt \, .$$

Under extra mild condition upon the growth of k, the limit distributions of $E_n(k)$ and $PE_n(k)$ are attainable, with their specific forms being established by using an asymptotic expansion for the tail empirical quantile function due to Drees (1998). A table of critical points several values of γ is provided, although some corrections have become available in Hüsler and Li (2006). Aside from the latter, Drees *et al.* (2006) deal with the testing of extreme value conditions pertaining to $\gamma > -1/2$, via the statistic

$$(4.6) \qquad T_n(k) := k \int_0^1 \left(\frac{n}{k} \, \overline{F}_n \left(\hat{a}\Big(\frac{n}{k}\Big) \, \frac{x^{-\hat{\gamma}} - 1}{\hat{\gamma}} + \hat{b}\Big(\frac{n}{k}\Big) \right) - x \right)^2 x^{\eta-2} \, dx \, ,$$

for some $\eta > 0$, with $\overline{F}_n = 1 - F_n$. The use ML estimators for γ and a as in Drees *et al.* (2004) is recommend, while $\hat{b}(n/k) := X_{n-k:n}$. Similarly as before, under mild restrictions upon the growth of k, the limit distribution of $T_n(k)$ is attainable and its specific form can be established using a tail approximation to the empirical distribution function. Again, tables of critical points at quite good accuracy are provided in Hüsler and Li (2006), where an exhaustive simulation study is carried out in order to draw general guidelines for the adequate specification of η in the most suitable test for the problem at hand.

Notwithstanding, if we strongly suspect we are dealing with heavy tailed phenomena, Beirlant *et al.* (2006) provide a goodness-of-fit procedure for testing the inherent Pareto-type behavior upon the tail of the underlying distribution function F.

5. AN ILLUSTRATIVE EXAMPLE

The potential of Extreme Value theory in assessing statistical models for tail-related values has gained widespread recognition in fields ranging from hydrology to insurance, finance and, more recently, in telecommunications and engineering.

As an illustrative example of methodologies embraced in the previous section, consider the 36 699 file lengths, in bytes, extracted from the Internet Traffic Archive (`http://ita.ee.lbl.gov/index.html`). In the light of extreme value analysis, the main concern here is not towards the accumulation of many file lengths, none of these being dominant (in which case the normal assumption would reasonably follow from the Central Limit Theorem), but the interest goes instead to the transmission of such huge batches of data that could possibly compromise the capacity of the system, thus making the normal distribution inadequate to describe the small set of data arising with such individual large and, therefore, dominant contributors. This same data set is analyzed in a paper by Tsourti and Panaretos (2004). Their exploratory analysis for independence seems to ascertain that an application of the testing procedures mentioned in this paper, to the available data set, will not be hindered by the pernicious effects of seasonality and clustering.

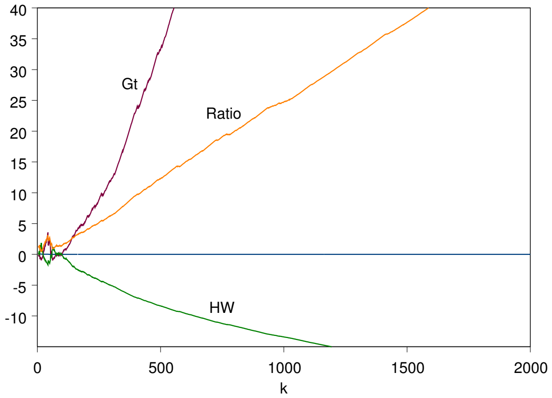
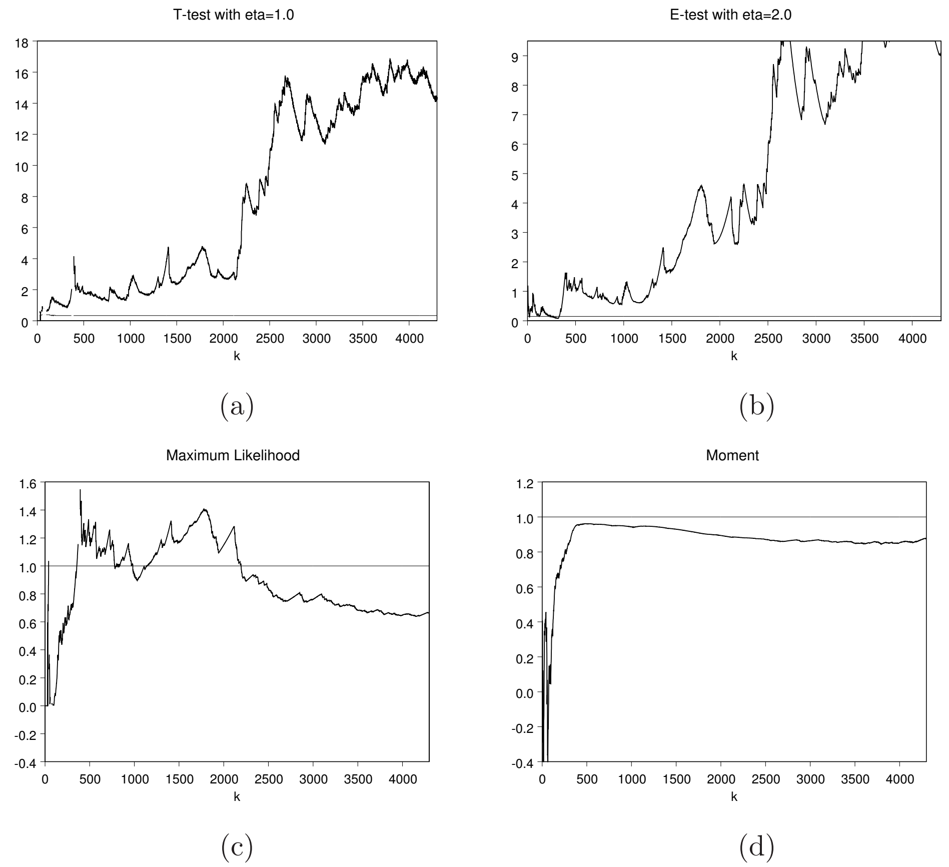


Figure 1: Plot of the sample paths returned by the three test statistics for the Gumbel domain.

Hence, we have found it reasonable to proceed with three tests in (4.2)–(4.4). The results are depicted in Figure 1. All the tests point towards a definite rejection of the null hypothesis that the underlying distribution function F belongs to the Gumbel domain. Nevertheless, the validity of condition (2.3) is still questionable. So far, we have only found evidences in the data of that F can be in any domain except for the Gumbel domain, but the question “does the underlying d.f. F belongs to any domain of attraction at all?” remains unanswered.



(a) (b)

(c) (d)

Figure 2: (a)–(b) Plot of the sample paths for the T-test and E-test statistics with accompanying critical points; (c) Plot of the ML estimates; (d) Plot of the Moment estimates.

Owing to these last remarks and following practical recommendations of Hüsler and Li (2006), we have furthermore considered application of the T-test and the E-test, given in (4.6) and (4.5), with $\eta = 1$ and $\eta = 2$, respectively. Figure 2 displays the results with respect to a significance level $\alpha = 0.05$. Although the moment estimator yields a stable plateau near $\gamma = 1$ for quite long, the conjunction of the two testing procedures seems to advise rejection of the null hypothesis on that the tail of F obeys the dictates of an extreme value law.

ACKNOWLEDGMENTS

Research partially supported by FCT/POCI 2010 and POCI — ERAS Project MAT/58876/2004.

Cláudia Neves is grateful to the organization of the Invited Paper Meeting Session on Extremes, Risk and the Environment, from the ISI 2007 — 56th Session of the International Statistical Institute, which took place in Lisbon, for the opportunity to amplify her contribution to the Session in this paper.

REFERENCES

[1] Balkema, A.A. and de Haan, L. (1974). Residual life time at great age, *Ann. Probab.*, **2**, 792–804.

[2] Bardsley, W.E. (1977). A test for distinguishing between extreme value distributions, *J. Hydrology*, **34**, 377–381.

[3] Beirlant, J.; de Wet, T. and Goegebeur, Y. (2006). A goodness-of-fit statistic for Pareto-type behaviour, *J. Comput. Appl. Math.*, **186**, 99–116.

[4] Brilhante, M.F. (2004). Exponentiality versus generalized Pareto — a resistant and robust test, *RevStat*, **2**, 1–13.

[5] Castillo, E.; Galambos, J. and Sarabia, J.M. (1989). *The selection of the domain of attraction of an extreme value distribution from a set of data.* In "Extreme value theory (Oberwolfach, 1987) — Lecture Notes in Statistics 51" (J. Hüsler and R.-D. Reiss, Eds.), Springer, Berlin–Heidelberg, pp. 181–190.

[6] Castillo, E. and Hadi, A.S. (1997). Fitting the Generalized Pareto distribution to data, *J. Amer. Statist. Assoc.*, **92**, 1609–1620.

[7] Chaouche, A. and Bacro, J.-N. (2004). Fitting the Generalized Pareto distribution to data, *Comput. Statist. Data Anal.*, **45**, 787–803.

[8] Choulakian, V. and Stephens, M.A. (2001). Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, **43**, 478–484.

[9] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *J. Roy. Statist. Soc. Ser. B*, **52**, 393–442.

[10] Dekkers, A.L.M.; Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Ann. Statist.*, **17**, 1833–1855.

[11] Dietrich, D.; de Haan, L. and Husler, J. (2002). Testing extreme value conditions, *Extremes*, **5**, 71–85.

[12] Drees, H. (1998). On smooth statistical tail functionals, *Scand. J. Statist.*, **25**, 187–210.

[13] Drees, H.; de Haan, L. and Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions, *J. Statist. Plann. Inference*, **136**, 3498–3538.

[14] Drees, H.; Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Ann. Probab.*, **14**, 1179–1201.

[15] Falk, M. (1995). On testing the extreme value index via the POT-method, *Ann. Statist.*, **23**, 2013–2035.

[16] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample, *Proc. Camb. Phil. Soc.*, **24**, 180–190.

[17] Fraga Alves, M.I. (1999). Asymptotic distribution of Gumbel statistic in a semi-parametric approach, *Port. Math.*, **56**, 282–298.

[18] Fraga Alves, M.I. and Gomes, M.I. (1996). Statistical choice of extreme value domains of attraction — a comparative analysis, *Commun. Statist. — Theory Meth.*, **25**, 789–811.

[19] Galambos, J. (1982). *A statistical test for extreme value distributions.* In "Nonparametric Statistical Inference" (B.W. Gnedenko *et al.*, Eds.), North Holland, Amsterdam, pp. 221–230.

[20] Gomes, M. (1982). *A note on statistical choice of extremal models.* In "Proc. IX Jornadas Mat. Hispano-Lusas", Salamanca, pp. 653–655.

[21] Gomes, M. and Alpuim, M. (1986). Inference in a multivariate generalized extreme value model — asymptotic properties of two test statistics, *Scand. J. Statist.*, **13**, 291–300.

[22] Gomes, M.I. (1987). *Extreme value theory — statistical choice.* In "Colloq. Math. Soc. János Bolyai 45", Debrecen, pp. 195–210.

[23] Gomes, M.I. (1989). *Comparison of extremal models through statistical choice in multidimensional backgrounds.* In "Extreme value theory (Oberwolfach, 1987) — Lecture Notes in Statistics 51" (J. Hüsler and R.-D. Reiss, Eds.), Springer, Berlin–Heidelberg, pp. 191–203.

[24] Gomes, M.I. and Van Montfort, M.A.J. (1986). *Exponentiality versus generalized* P*areto-quick tests.* In "Proc. III Internat. Conf. Statistical Climatology", pp. 185–195.

[25] Greenwood, M. (1946). The statistical study of infectious diseases, *J. Roy. Statist. Soc. Ser. A*, **109**, 85–109.

[26] Grimshaw, S.D. (1993). Computing the maximum likelihood estimates for the Generalized Pareto distribution, *Technometrics*, **35**, 185–191.

[27] Hasofer, A. and Wang, J.Z. (1992). A test for extreme value domain of attraction, *J. Amer. Statist. Assoc.*, **87**, 171–177.

[28] Hosking, J. (1984). Testing whether the shape parameter is zero in the generalized extreme value distribution, *Biometrika*, **71**, 367–374.

[29] Hüsler, J. and Li, D. (2006). On testing extreme value conditions, *Extremes*, **9**, 69–86.

[30] Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quart. J. Roy. Meteo. Soc.*, **81**, 158–171.

[31] Kinnison, R. (1989). Correlation coefficient goodness-of-fit test for the Extreme-Value distribution, *Amer. Statist.*, **43**, 98–100.

[32] Konstantinides, D. and Meintanis, S.G. (2004). *A test of fit for the generalized* P*areto distribution based on transforms.* In "Proc. 3$^{\text{rd}}$ Conf. in Actuarial Science and Finance in Samos".

[33] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators, *Comput. Statist. Data Anal.*, **51**, 904–917.

[34] Marohn, F. (1994). *On testing the exponential and* G*umbel distribution.* In "Extreme Value Theory" (J. Galambos, Ed.), Kluwer, Dordrecht, pp. 159–174.

[35] Marohn, F. (1998a). An adaptive efficient test for Gumbel domain of attraction, *Scand. J. Statist.*, **25**, 311–324.

[36] Marohn, F. (1998b). Testing the Gumbel hypothesis via the POT-method, *Extremes*, **1**(2), 191–213.

[37] Marohn, F. (2000). Testing extreme value models, *Extremes*, **3**(4), 363–384.

[38] Neves, C. and Fraga Alves, M.I. (2007). Semi-parametric approach to Hasofer-Wang and Greenwood statistics in extremes, *Test*, **16**, 297–313.

[39] Neves, C.; Picek, J. and Fraga Alves, M.I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction, *J. Statist. Plann. Inference*, **136**(4), 1281–1301.

[40] Otten, A. and Van Montfort, M. (1978). The power of two tests on the type of distribution of extremes, *J. Hydrology*, **37**, 195–199.

[41] Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

[42] Segers, J. and Teugels, J. (2000). Testing the Gumbel hypothesis by Galton's ratio, *Extremes*, **3**(3), 291–303.

[43] Serfling, R.J. (1980). *Approximations Theorems of Mathematical Statistics*, Wiley, New York.

[44] Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591–611.

[45] Stephens, M.A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters, *Ann. Statist.*, **4**, 357–369.

[46] Stephens, M.A. (1977). Goodness-of-fit for the extreme value distribution, *Biometrika*, **64**, 583–588.

[47] Stephens, M.A. (1986). *Tests for the exponential distribution*, In "Goodness-of-Fit Techniques" (R.B. D'Agostinho and M.A. Stephens, Eds.), Marcel Dekker, New York, pp. 421–459.

[48] Tiago de Oliveira, J. (1981). *The selection of the domain of attraction of an extreme value distribution from a set of data.* In "Statistical Distributions in Scientific Work", Vol. 6 (C. Taillie, Ed.), D. Reidel, Dordrecht, pp. 367–387.

[49] Tiago de Oliveira, J. (1984). *Univariate extremes: Statistical choice.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, Dordrecht, pp. 91–107.

[50] Tiago de Oliveira, J. and Gomes, M.I. (1984). *Two statistics for choice of univariate extreme value models.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), D. Reidel, Dordrecht, pp. 651–668.

[51] Tsourti, Z. and Panaretos, J. (2004). Extreme-value analysis of teletraffic data, *Comput. Statist. Data Anal.*, **45**, 85–103.

[52] Van Montfort, M.A.J. (1970). On testing that the distribution of extremes is of type i when type ii is the alternative, *J. Hydrology*, **11**, 421–427.

[53] Van Montfort, M.A.J. and Witter, J.V. (1985). Testing exponentiality against Generalized Pareto distribution, *J. Hydrology*, **78**, 305–315.

[54] Van Montfort, M.A.J. and Witter, J.V. (1986). The Generalized Pareto distribution applied to ranfall depths, *J. Hydrology*, **31**, 151–162.

[55] von Mises, R. (1936). *La distribution de la plus grande de n valeurs.* Reprinted in Selected Papers Volumen II, American Mathematical Society, Providence, R.I., 1954, pp. 271–294.

[56] Wang, J.Z. (1995). Selection of the k largest order statistics for the domain of attraction of the gumbel distribution, *J. Amer. Statist. Assoc.*, **90**, 1055–1061.

[57] Wang, J.Z.; Cooke, P. and Li, S. (1996). Determination of domains of attraction based on a sequence of maxima, *Austral. J. Statist.*, **38**, 173–181.

[58] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.