THE JACKKNIFE AND THE BOOTSTRAP METHODOLOGIES IN THE ESTIMATION OF PARAMETERS OF RARE EVENTS

AS METODOLOGIAS JACKKNIFE E BOOTSTRAP NA ESTIMAÇÃO DE PARÂMETROS DE ACONTECIMENTOS RAROS

Autor: M. Ivette Gomes*
- Professora Catedrática no D.E.I.O., C.E.A.U.L., Faculdade de Ciências, Universidade de Lisboa

ABSTRACT:

- The main goal of this paper is to enhance the role of two well-known re-sampling methodologies, the *Bootstrap* of Efron (1979) and the *Jackknife* of Quenouile (1956) and Tukey (1958), in the *Statistical Theory of Extreme Values*. The Bootstrap will be used here to estimate *the optimal sample fraction* to be taken in semi-parametric estimation of *parameters of rare events*, and the Jackknife will be used to reduce the asymptotic *Bias* of estimators, without increasing *Mean Square Error*. The methodologies developed will be applied both to simulated and real data.

KEY-WORDS:

- *Statistical Theory of Extremes, Semi-parametric Estimation, Asymptotic Theory, Re-sampling Techniques, Jackknife and Bootstrap.*

RESUMO:

- O principal objectivo deste trabalho é enaltecer a importância de duas metodologias estatísticas bem conhecidas, o *Bootstrap* de Efron (1979) e o *Jackknife* de Quenouile (1956) e Tukey (1958), em *Teoria Estatística de Valores Extremos*. O Bootstrap será aqui usado para estimar *a fracção óptima* de estatísticas ordinais a considerar na estimação semi-paramétrica de *parâmetros de acontecimentos raros*, e o Jackknife será usado para reduzir o *Viés* assintótico dos estimadores, sem aumentar o *Erro Quadrático Médio*. As metodologias desenvolvidas serão aplicadas a dados simulados e a dados reais.

PALAVRAS-CHAVE:

- *Teoria Estatística de Valores Extremos, Estimação Semi-Paramétrica, Teoria Assintótica, Técnicas de Re-amostragem, Jackknife e Bootstrap.*

1. INTRODUCTION AND PRELIMINARIES

In *Statistical Extreme Value Theory* we are mainly interested in the estimation of *parameters of rare events*, like the *return period* of high levels, i.e. the mean time a stochastic process remains above the high level u, as $u \to \infty$, the *high quantiles* of the model $F(.)$ associated to an underlying stationary stochastic process, i.e., functionals $c_p(F) := F^{\leftarrow}(p)$, $p \to 1$, where $F^{\leftarrow}(y) := \inf\{x : F(x) \geq y\}$ is the generalized inverse function of $F(.)$, the *extremal index* $q(F)$, which in a certain sense measures the mean size of

* Research partially supported by FCT / PRAXIS XXI / FEDER.

the clusters of exceedances of high levels and, primarily, the *tail index* $\gamma = \gamma(F)$, the basic parameter of *rare events*, directly related to the *right tail weight* of the model *F*(.), the right tail 1–*F*(.) being heavier and heavier as γ increases. For $\gamma = 0$ we have a *Gumbel* model, with an exponential-type tail.

The *tail index* γ may be easily defined in terms of the asymptotic behaviour of the sequence of maximum values, $\{X_{n:n}\}_{n\geq 1} = \{\max(X_1, X_2, \ldots, X_n\}_{n\geq 1}$ of a random sample from a model *F*(.). Indeed, the sequence $X_{n:n}$ converges almost surely to the right endpoint $x_F := \sup\{x : F(x) < 1\}$ of the model *F*(.), but if there exists attraction coefficients $\{a_n > 0\}_{n\geq 1}$ and $\{b_n \in \Re\}_{n\geq 1}$ such that

$$P[X_{n:n} \leq a_n x + b_n] \underset{n\to\infty}{\to} G(x), \text{ for all } x \text{ in the set of continuity points of } G(.), \quad (1.1)$$

where *G*(*x*) is a non-degenerate distribution function (d.f.) then, up to location and scale, *G* has the functional form

$$G(x) \equiv G_\gamma(x) := \exp\{-(1+\gamma x)^{-1/\gamma}\}, \; 1+\gamma x > 0, \; \gamma \in \Re, \quad (1.2)$$

where $(1+\gamma x)^{-1/\gamma}$ is to be taken equal to $\exp\{-x\}$ for $\gamma = 0$ (Gnedenko, 1943).

Whenever (1.1) holds, with *G*(.) given in (1.2), we say that *F* is in the *domain of attraction* of G_γ, and write $F \in D(G_\gamma)$. A unified necessary and sufficient condition for $F \in D(G_\gamma)$, expressed in terms of the quantile function

$$U(t) := \begin{cases} 0 & t \leq 1 \\ F^{\leftarrow}(1-1/t) & t > 1 \end{cases}, \quad (1.3)$$

was given by de Haan (1984):

$$F \in D(G_\gamma) \text{ iff there is a function } a(t): \; \frac{U(tx)-U(t)}{a(t)} \to \frac{x^\gamma - 1}{\gamma}, \quad (1.4)$$

where, once again, $(x^\gamma - 1)/\gamma$ is to be taken equal to its limit, $\log x$, whenever $\gamma = 0$.

In *Statistical Theory of Extremes* the estimation of the tail index γ is usually done in a *semi-parametric* context, where we merely assume that $F \in D(G_\gamma)$ and base the estimation on a suitable number of top order statistics (o.s.), say $X_{n:n} \geq X_{n-1:n} \geq \ldots \geq X_{n-k:n}$, where $X_{i:n}$, $1 \leq i \leq n$, is the sample of the ascending o.s. associated to our original sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$, and $k = k_n$, must be suitably chosen, depending heavily on the estimator we intend to use and on the model underlying the data. Hence, an important question in *Statistical Extreme Value Theory* has been for a long time the choice of the "*nuisance*" parameter $k = k_n$, where *k* + 1 is the number of top o.s. considered.

We shall assume here, for sake of simplicity, that we are dealing with heavy tails ($\gamma > 0$). We then have

$$F \in D(G_\gamma) \; (\gamma > 0) \quad \text{iff} \quad 1 - F \in RV_{-1/\gamma} \quad \text{iff} \quad U \in RV_\gamma, \quad (1.5)$$

where RV_a stands for the class of *regularly varying* functions at infinity with index of regular variation equal to a, i.e., functions $g(.)$ with infinite right endpoint, and such that $g(tx)/g(t) \underset{t\to\infty}{\to} x^a$, for all $x > 0$. The conditions in (1.5) characterize completely the first order behaviour of $F(.)$ [Gnedenko (1943), de Haan (1970)]. The second order theory has been worked out in full generality by de Haan and Stadtmüller (1996). Indeed, for a large class of models there exists a function $A(t)$ of constant sign for large values of t, such that

$$\frac{U(tx)/U(t) - x^g}{A(t)} \xrightarrow[t\to\infty]{} x^g \, \frac{x^r - 1}{r}, \tag{1.6}$$

for every $x > 0$, where r (≤ 0) is a second order parameter, which eventually also needs to be properly estimated from the original sample. The limit function in (1.6) must be of the stated form, and $|A(t)| \in RV_r$ (Geluk and de Haan, 1987).

From the above mentioned second order behaviour of $F(.)$ it follows that

$$\ln U(tx) - \ln U(t) = g \ln x + A(t)\frac{x^r - 1}{r}(1 + o(1)), \text{ as } t \to \infty. \tag{1.7}$$

The consideration of the empirical counterpart of the quantile function U defined in (1.3), leads then to Hill estimator (Hill, 1975)

$$g_n(k) := \frac{1}{k}\sum_{i=1}^{k} \left[\ln X_{n-i+1:n} - \ln X_{n-k:n}\right] = \frac{1}{k}\sum_{i=1}^{k} i\left[\ln X_{n-i+1:n} - \ln X_{n-i:n}\right]. \tag{1.8}$$

For a more complete study of Hill estimator, see Martins *et al* (1997).

In order to have consistency of Hill estimator, given in (1.8), we need to work with an intermediate number of o.s., i.e. we need to have $k = k_n \to \infty$, $k_n/n \to 0$. Indeed, since we have that for every r.v. X with d.f. $F(.)$, $X = F^{\leftarrow}\left(1 - \frac{1}{Y}\right) = U(Y)$, where Y is a Pareto(1) r.v with d.f $G_Y(y) = 1 - 1/y, y \geq 1$, $U(.)$ an increasing function, we thus have $X_{i:n} = U(Y_{i:n})$, $1 \leq i \leq n$. This property, together with (1.7), enables us to write the following distributional representation for Hill estimator (de Haan and Peng, 1995),

$$g_n(k) \overset{d}{=} g + \frac{g}{\sqrt{k}} P_n + \frac{1}{1-r} A(n/k) + o_p(A(n/k)), \tag{1.9}$$

where P_n is asymptotically a standard Normal r.v., i.e., $P_n \overset{d}{=} \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} W_i - 1\right)$, with $\{W_i\}$ a sequence of unit exponential r.v.'s. It thus follows that if $k = k_n \to \infty$, $k_n/n \to 0$, in such a way that $\sqrt{k}A(n/k) \to l$, as $n \to \infty$, then

$$\sqrt{k}\left[g_n(k) - g\right] \xrightarrow[n\to\infty]{d} N\left(\frac{l}{1-r},\ g^2\right),$$

i.e., we may have an *Asymptotically Normal* estimator of g with a *non-null* asymptotic bias, $l/(1-r)$.

On the other side, if $\sqrt{k}\left|A(n/k)\right| \to +\infty$, then $\dfrac{g_n(k) - g}{A(n/k)} \xrightarrow{p} \dfrac{1}{1-r}$.

Under the validity of (1.9), we have an *Asymptotic Mean Square Error* (*AMSE*) given by $AMSE[g_n(k)] = \dfrac{g^2}{k} + \dfrac{1}{(1-r)^2} A^2(n/k)$, and then

$$k_0(n) := \arg\min_k MSE[g_n(k)] = \frac{n}{s^{\leftarrow}\left(\dfrac{g^2(1-r)^2}{n}\right)}(1+o(1)), \qquad (1.10)$$

where $A^2(t) = \int_t^{+\infty} s(u)du(1+o(1))$. For the existence of this function *s*(.) see lemma 2.9 of Dekkers and de Haan (1993).

This semi-parametric estimator of the tail index and almost all usual semi-parametric estimators of parameters of rare events (and this happens more generally in Statistics, like for instance in the semi-parametric or non-parametric methods of density estimation (Devroye, 1985) have the same type of behaviour: high variance for small values of *k*, high bias for large values of *k*, and consistency only for intermediate ranks, i.e., we need to have $k = k_n \to \infty$, $k_n/n \to 0$, as $n \to \infty$.

And what does this mean from a practical point of view?

Consequently there are immediately two main questions put forward:

1. How to estimate the *optimal sample fraction*, i.e., given generally a semi-parametric estimator $x_n(k)$ of the functional of rare events $x(F)$, based on the k top o.s. of $\underline{X}_n$, how is it possible to estimate $k_0^{x_n}(n) := \arg\min_k MSE\{x_n(k)\}$, in order to estimate $x(F)$ by means of $x_n(k_0^{x_n}(n))$? This question has been addressed in several papers , among which we refer Dekkers and de Haan (1993), Berlaint *et al* (1996a, 1996b), Peng (1998), Drees and Kaufmann (1998), Gomes(1998).
2. Is it possible to reduce the bias of these semi-parametric estimators, and find other semi-parametric estimators of the parameter of rare events under consideration, with smaller *BIAS* and also smaller *MSE*, being conscious that then we eventually have to go further in the tail, and pay a price for the need to collect more observations? Under this context we refer here the papers of Peng (1998) and Gomes *et al*(1998).

In section 2 of this paper we shall address the first question, and for data from a *Cauchy model* we shall illustrate the use of bootstrap methodology for the estimation of the optimal sample fraction by means of a bootstrap estimator of $k_0(n)$, of the type of the ones studied by de Haan *et al* (1997) and by Danielson *et al* (1997), but with the use of an *auxiliary statistic* of the type introduced in Gomes (1998), which is merely the difference of two estimators with the same functional form of the estimator under study, computed at two intermediate levels. De Haan's methodology has over Hall's bootstrap methodology (Hall(1990), Gomes (1994, 1998)) the advantage of overpassing the need of an initial consistent estimator of the tail index g by the consideration of an *auxiliary statistic*, with null mean value, which consequently has a *MSE* equal to its variance, and whose asymptotic properties are intimately close to the ones of the estimator under study. More than that: the estimated value of $k_0(n)$ may be used for the initial consistent estimation of the tail index g, needed in Hall's methodology.

In section 3 we shall consider the reduction of bias by means of the Generalized Jackknife theory (Gray and Schucany, 1972), and we study the behaviour of a Generalized Jackknife estimator of the type of the ones introduced by Gomes *et al* (1998), but where *Bias* is going to be estimated by means of the Bootstrap methodology developed in section 2.

Finally, in section 4, we shall consider an application to real data in the field of finance.

2. THE BOOTSTRAP METHODOLOGY AND THE ESTIMATION OF THE OPTIMAL SAMPLE FRACTION

The bootstrap methodology enables us to estimate the optimal sample fraction $k_0(n)/n$, $k_0(n) := \arg\min_k MSE[g_n(k)]$, in the following way [de Haan *et al* (1997), Danielson *et al* (1997), Gomes (1998)]: given the sample $\underline{X}_n = (X_1, \ldots, X_n)$ from an unknown model *F*, and the functional $g_n(k)$, $1 \le k \le n$, a consistent estimator of g, consider the bootstrap sample $\underline{X}^*_{n_1} = (X^*_1, \ldots, X^*_{n_1})$, $n_1 \le n$, from the model $F^*_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i \le x]}$,

the empirical d.f. associated to the original sample $\underline{X}_n$, to which we may associate the corresponding estimator $g^*_{n_1}(k_1)$, $1 \le k_1 \le n_1 - 1$.

Consider then an auxiliary statistic with null mean value, which consequently has a *MSE* equal to its variance, and whose asymptotic properties are intimately close to the ones of the estimator under study. We shall here consider

$$T_n(k) \coloneqq g_n(k/2) - g_n(k). \qquad (2.1)$$

From the joint behaviour of Hill estimator at two intermediate levels (Gomes (1998), Gomes *et al* (1998)) we get the distributional representation

$$T_n(k) \overset{d}{=} \frac{g}{\sqrt{k}} Z_n - \frac{1-2^r}{1-r} A(n/k) + o_p(A(n/k)),$$

where Z_n is asymptotically a Normal(0,1) r.v., i.e., asymptotically, the r.v. $T_n(k)$ has a variance g^2/k, and a bias $(2^r - 1)A(n/k)/(1-r)$.

Then, the fact that asymptotically,

$$AMSE(T_n(k)) = \frac{g^2}{k} + A^2(n/k)\frac{(1-2^r)^2}{(1-r)^2},$$

enables us to derive

$$k_{0T}(n) \coloneqq \arg\min_k \; MSE[T_n(k)] = \frac{n}{s^{\leftarrow}\left(\frac{g^2(1-r)^2}{n}\frac{1}{(1-2^r)^2}\right)}(1+o(1)). \qquad (2.2)$$

From (1.10) and (2.2), and from the fact that $s^{\leftarrow} \in RV_{-1/(1-2r)}$, it follows that

$$k_0(n) = \left(1-2^r\right)^{\frac{2}{1-2r}} k_{0T}(n)(1+o(1)), \text{ as } n \to \infty . \tag{2.3}$$

If we bootstrap $T_n(k)$, getting $T_n^*(k) \mid \underline{X}_n$, $k_{0T}^*(n) := \arg\min_k E\left(\left[T_n^*(k)\right]^2 \mid \underline{X}_n\right)$, it is immediate to ask if it is possible to replace $k_{0T}(n)$ by $k_{0T}^*(n)$ in (2.3). Theoretically, that is not possible. We must deal with samples of size $n_1 = O(n^{1-e})$, $0 < e < 1$, and with $k_1 \to \infty$, $k_1 / n_1 \to 0$, in order to have, as $n_1 \to \infty$,

$$k_{0T}^*(n_1) := \arg\min_{k_1} MSE\left[T_{n_1}^*(k_1) \mid \underline{X}_n\right] = \frac{n_1}{s^{\leftarrow}\left(\frac{g^2(1-r)^2}{n_1}\frac{1}{(1-2^r)^2}\right)}(1+o(1)) . \tag{2.4}$$

(see Peng (1998) for a proof).

Since $k_{0T}^*(n_1) \in RV_{-2r/(1-2r)}$,

$$\frac{k_{0T}^*(n_1)}{k_{0T}(n)} = \left(\frac{n_1}{n}\right)^{\frac{2r}{2r-1}}(1+o(1)), \text{ as } n \to \infty . \tag{2.5}$$

Thus, for another sample size $n_2 = \frac{n_1^2}{n}$ (chosen in this way in order to have independence of r), we have

$$k_{0T}(n) = \frac{\left[k_{0T}^*(n_1)\right]^2}{k_{0T}^*(n_2)}(1+o(1)), \text{ as } n \to \infty . \tag{2.6}$$

Several estimators of r have been proposed in the literature. We use here the bootstrap estimator of Danielson *et al* (1997), also used in Gomes (1998): since $k_{0T}^* \in RV_{-2r/(1-2r)}$, it follows that $\frac{\ln k_{0T}^*}{\ln n_1} \to \frac{2r}{2r-1}$, as $n \to \infty$. The bootstrap estimator is

$$r^* := \frac{\ln \bar{k}_{0T}^*(n_1)}{2\ln(\bar{k}_{0T}^*(n_1)/n_1)}, \tag{2.7}$$

where $\bar{k}_{0T}^*(n_1)$ denotes, the sample counterpart of $k_{0T}^*(n_1)$, i.e. for *B* generated bootstrap samples we take $\bar{k}_{0T}^*(n_1) = \arg\min_k \sum_{i=1}^{B} \left[T_{n_1,i}^*(k)\right]^2$.

We then have

$$\hat{k}_0(n \mid n_1, r^*) := \frac{\left[\bar{k}^*_{0T}(n_1)\right]^2}{\bar{k}^*_{0T}(n_1^2/n)}\left[1-2^{r^*}\right]^{\frac{2}{1-2r^*}}, \quad \text{and} \quad g_n^{*(1)}(n_1; r^*) := g_n(\hat{k}_0(n \mid n_1, r^*)). \quad (2.8)$$

In Gomes (1998) the robustness of the estimator in (2.8), regarding the choice of the sub-sample size n_1, was exhibited by simulation in a Fréchet model. Here, we generate a sample of size 1000 from a Cauchy model with null mean value, and we present in Figures 1 and 2 the sample path of $\hat{k}_0(n \mid n_1, r)$ and of $g_n^{*(1)}(n_1; r^*)$, respectively — both for r estimated through (2.7) and for r assumed to be known and equal to –2, just as happens in a Cauchy model, and for values of n_1 = 50(5)1000. We have used a multi-sample Bootstrap procedure of 10 replicates of B = 100 runs each. Comparatively to a one-sample procedure, the multi-sample procedure provided a higher stability of the sample paths. The simulated mean value of $k_0(n)$, on the basis of 20 replicas of 5000 runs each, for n = 1000 and for a Cauchy model is 132.85, with a 95% confidence interval given by (130.64, 135.06). The simulated mean value of Hill estimator at the optimal level is given by 1.0411, being the 95% confidence interval, (1.0397, 1.0425).

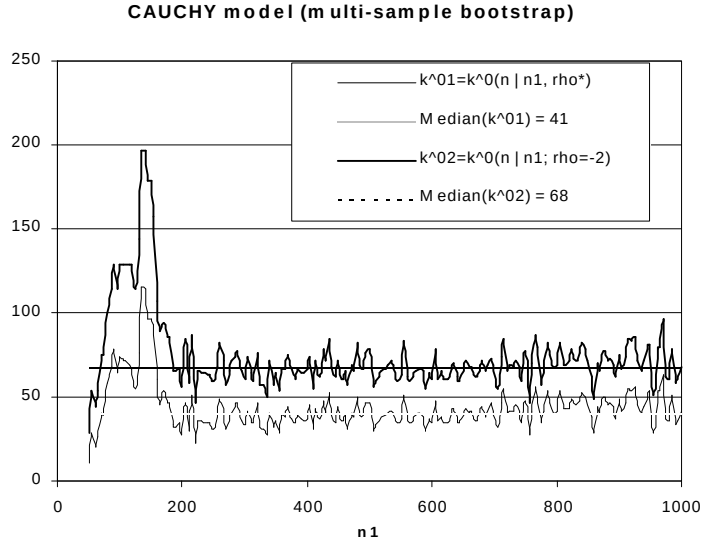

Figure 1. Sample path of $\hat{k}_0(n \mid n_1, r)$ for estimated r (through r^*) and for r assumed to be known ($r = -2$), and for values of n_1 = 50(5)1000.

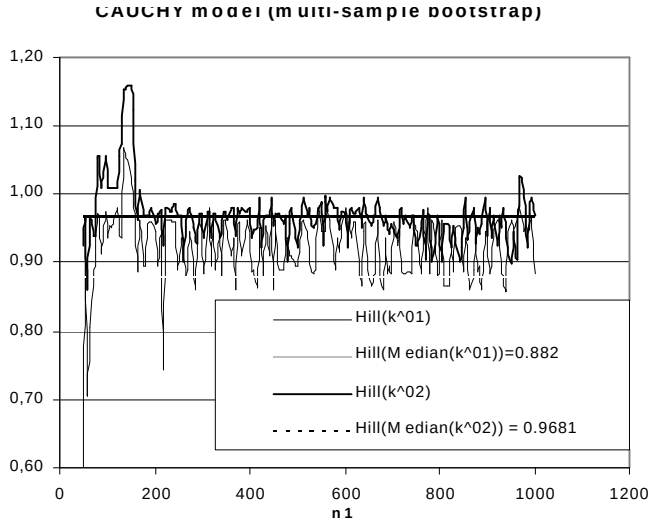

Figure 2. Sample path of $g_n^{*(1)}(n_1; r)$ for estimated r (through r^*) and for r assumed to be known ($r = -2$), and for values of n_1 = 50(5)1000.

As pointed out in Gomes (1998) the high stability obtained along the sub-sample size n_1, both for the optimal sample fraction and for Hill estimator at the optimal level, enhanced in Figures 1 and 2, together with a high stability of the MSE of $g_n^{*(1)}(n_1; r^*)$, quite close to the minimal *MSE* of Hill estimator (see Gomes (1998) for details) suggest the real practical importance of this estimation procedure of the optimal sample fraction to be taken in a semi-parametric estimation of a parameter of rare events.

We now suggest the following more intrincate procedure: after using an auxiliary statistic to estimate the optimal level, just as done before, use that optimal value as k_{aux} in Hall's Bootstrap methodology (Hall (1990), Gomes (1994, 1998)), and then estimate the *MSE* and the *BIAS* of the estimator under study. The value k_{aux} is the number of top o.s. needed initially to get a first rough consistent estimate of the tail index g .

We would like to point out the following: the bootstrap estimator $g_n^*(k)$ "smooths" the sample path of the original estimator $g_n(k)$, $1 \le k \le n-1$ — in a way similar to what does the moving average procedure of Resnick and Starica (1997); we may easily obtain bootstrap estimates of the *MSE* and of the *BIAS* of $g_n(k)$:

$$\left.\begin{array}{l} \hat{BIAS}\left[g_n(k) \mid k_{aux}\right] :\overset{B}{=} \hat{E}\left\{g_n^*(k) - g_n(k_{aux}) \mid \underline{X}_n\right\} \\ \hat{MSE}\left[g_n(k) \mid k_{aux}\right] :\overset{B}{=} \hat{E}\left\{\left[g_n^*(k) - g_n(k_{aux})\right]^2 \mid \underline{X}_n\right\} \end{array}\right\} \qquad (2.9)$$

We have here used two different values for k_{aux} :

$$k_{aux}^{(1)} = \underset{n_1=50(5)1000}{Median}\left(\hat{k}_0(n \mid n_1, r^*)\right), \text{ and } k_{aux}^{(2)} = \underset{n_1=50(5)1000}{Median}\left(\hat{k}_0(n \mid n_1, r=-2)\right)$$

which turn out to be appealing from a practical point of view, due to the stability of the sample path in Figure 1.

In case we had not such a stability, we might use the value suggested by Danielson *et al* (1998) in such situations, and given by

$$k_{aux} = \arg\min_{n_1} \; R(n_1),$$

where

$$R(n_1) = \frac{MSE^2\left[T^*_{n_1}(k^*_{0T}(n_1)) \mid \underline{X}_n\right]}{MSE\left[T^*_{n_2}(k^*_{0T}(n_2)) \mid \underline{X}_n\right]}, \quad n_2 = n_1^2 / n,$$

is an estimate of $MSE\left[T_n(k_{0T}(n))\right]$.

After removing bias, we obtain the new estimators

$$g_n^{*(2)}(k \mid k_{aux}) \;:=\; g_n(k) - \hat{BIAS}\left[g_n(k) \mid k_{aux}\right], \quad k_{aux} = k_{aux}^{(1)},\; k_{aux}^{(2)}. \qquad (2.10)$$

Always working with the same sample of size n = 1000 underlying Figure 1, from a Cauchy model, for which g =1, we present in Figure 3 (*A*) the sample path of Hill estimates $g_n(k)$, of Hill bootstrap estimates $g_n^*(k)$, and of $g_n^{*(2)}(k \mid k_{aux}^{(1)})$ and $g_n^{*(2)}(k \mid k_{aux}^{(2)})$, in (2.10), both for $r := r^*$ and for $r = -2$. Figure 3 (*B*) is a zoom of Figure 3 (*A*). In these Figures, we show also two other sample paths of Jackknife estimators $g_{n,B}^{G_1}(k)$ and $g_{n,B}^{G_2}(k)$, described in the next section.

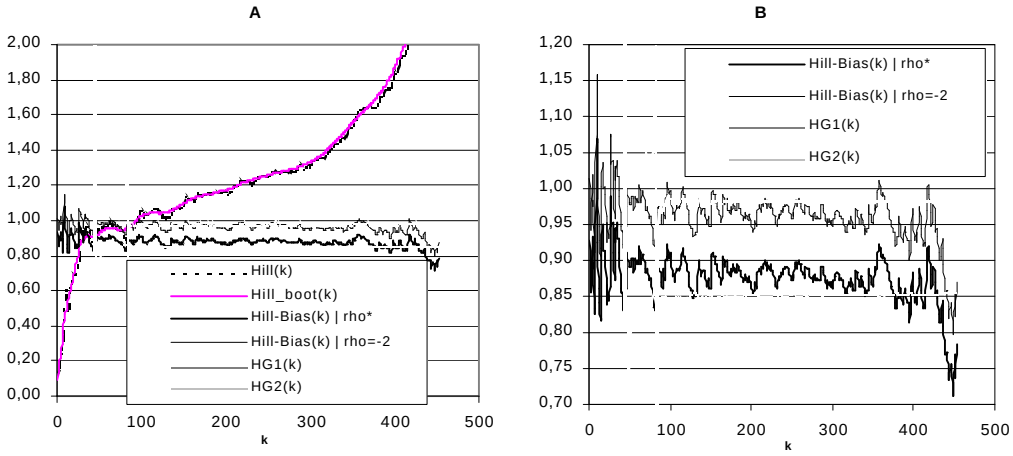

Figure 3. Sample path of Hill estimate $g_n(k)$, Bootstrap Hill estimate $g_n^*(k)$, $g_n^{*(2)}(k \mid k_{aux}^{(i)})$ and $g_{n,B}^{G_i}(k)$, i = 1, 2 for a *Cauchy* sample, n=1000 and k = 1(1)n-1.

In Figure 4 we show the Bootstrap estimates, given in (2.9), of $MSE[g_n(k)]$ and of $BIAS[g_n(k)]$, respectively. For comparison, we show also the simulated *MSE* and *BIAS* of Hill estimator at different levels.

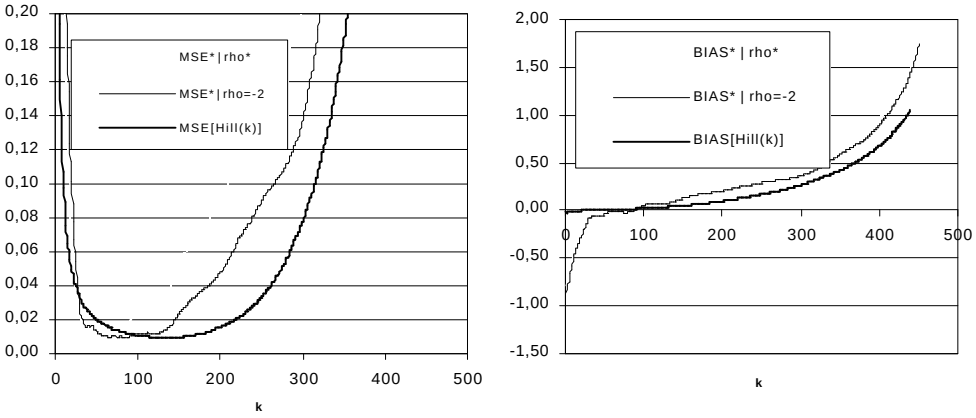

Figure 4. Bootstrap estimates of $MSE[g_n(k)]$ and of $BIAS[g_n(k)]$ for estimated r and r = -2 (one-sample), and simulated $MSE[g_n(k)]$ and $BIAS[g_n(k)]$ (5000 runs), for a *Cauchy* sample, n=1000 and k = 1(1) n-1.

This procedure is easy to implement and the new estimator has, from a practical point of view, a much more appealing sample path than that of Hill estimate, which is often referred to as "the Horror show of Hill estimates". It is also appealing to have bootstrap estimates of MSE and BIAS on the basis of the available sample, as we shall partially see in the next section.

3. THE JACKKNIFE METHODOLOGY AND THE REDUCTION OF THE BIAS TERM

The Jackknife methodology (Tukey, 1958) is a non-parametric re-sampling technique, essentially in the field of exploratory data analysis, largely used to reduce the bias of an estimator, by means of the construction of an auxiliary estimator based on Quenouille's re-sampling technique (Quenouille, 1956), and the consideration of a suitable combination of the two estimators. The *Generalized Jackknife* statistic of Gray and Schucany (1972) is, more generally, based on two different estimators of the same functional, with similar bias properties. More precisely, and as a particular case of the Jackknife theory, if we have two different biased consistent estimators $g_n^{(1)}$ and $g_n^{(2)}$ of the functional $g(F)$, and if $E[g_n^{(1)}] = g + j(g)d_1(n)$, $E[g_n^{(2)}] = g + j(g)d_2(n)$, then, denoting by $q_n := \frac{BIAS[g_n^{(1)}]}{BIAS[g_n^{(2)}]} = \frac{d_1(n)}{d_2(n)}$, the *Generalized Jackknife* statistic associated to $\left(g_n^{(1)}, g_n^{(2)}\right)$ is $g_n^G(g_n^{(1)}, g_n^{(2)}) = \frac{g_n^{(1)} - q_n g_n^{(2)}}{1 - q_n}$, which is an unbiased consistent estimator of $g(F)$, provided $q_n \neq 1$, for every n. (For a detailed application of the Jackknife methodology to the estimation of parameters of rare events, see Gomes *et al* (1998)).

In *Statistical Theory of Extremes*, whenever we are dealing with semi-parametric estimators of the tail index, or even other parameters of rare events, we have usually information about the asymptotic bias of those estimators, just as we have seen in (1.9) for Hill estimator. We may thus choose estimators with similar asymptotic properties, and construct the associated *Generalized Jackknife* estimator.

Let us now think on Hill estimator at two different levels, $k_1 < k_2$, and let

$$Q_n(k_1,k_2) \;:=\; \frac{BIAS_\infty\left[g_n(k_1)\right]}{BIAS_\infty\left[g_n(k_2)\right]} = \frac{A(n/k_1)}{A(n/k_2)} = \left(\frac{k_1}{k_2}\right)^{-r}(1+o(1)), \qquad (3.1)$$

be the quotient of the asymptotic bias of Hill estimators at those intermediate levels. Consider then

$$g_n^{\,G}(k_1,k_2) := \frac{g_n(k_1) - Q_n(k_1,k_2)\; g_n(k_2)}{1-Q_n(k_1,k_2)}, \qquad (3.2)$$

As was shown in Gomes *et al* (1998) this estimator is asymptotically normal with a null asymptotic bias whenever we choose k_1 in such a way that $\sqrt{k_1}\,A(n/k_1) \underset{n\to\infty}{\to} l_1$, finite, but we have degeneracy at g, whenever $\sqrt{k_1}\left|A(n/k_1)\right| \underset{n\to\infty}{\to} +\infty$, i.e. we have

$$\left[g_n^{\,G}(k_1,k_2) - g\right] / A(n/k_1) \;\overset{d}{=}\; o_p(1) \;\text{ whenever }\; \sqrt{k_1}\left|A(n/k_1)\right| \underset{n\to\infty}{\to} +\infty. \qquad (3.3)$$

The key to get a better estimator seems then to be: choose $k_1 > k_0(n)$, where $k_0(n) := \arg\min_k MSE\left[g_n(k)\right]$ may be estimated by means of the bootstrap techniques of the previous section, and study the properties of $g_n^{\,G}(k_1,k_2)$, $k_2 > k_1$. We are here going to work with the estimator

$$g_{n,B}^{\,G}(k) \;:=\; \frac{g_n(\hat{k}_0(n)) - \hat{Q}_n(\hat{k}_0(n),k)\; g_n(k)}{1-\hat{Q}_n(\hat{k}_0(n),k)}, \;\; k > \hat{k}_0(n), \qquad (3.4)$$

where $\hat{Q}_n(\hat{k}_0(n),k)$, a suitable estimator of the quotient of biases in (3.1), is going here to be estimated also by means of the bootstrap techniques of section 2.

We have considered the bootstrap estimators of Bias obtained in the previous section, and we have obtained the sample paths shown in Figure 3, where the index *i* = 1, 2 refers to the use of an estimate of r or the assumption of a known r, respectively.

The sample path of these estimators is even smoother than the sample paths of the estimators in (2.10), and although a justification for this is beyond the scope of this paper, we would like to refer here that the bootstrap estimator of Bias is not sufficiently accurate to increase, in terms of *MSE*, the performance of the estimator in (3.4) relatively to Hill estimator at the optimal level. Anyway, the sample path is indeed quite appealing from a practical point of view.

4. AN APPLICATION TO EXCHANGE RATES

For a long time, and essentially due to the mathematical tractability of *the Gumbel model* $(g = 0)$, statisticians have developed a strong bias towards this particular limiting model in the max-scheme. This is similar to what happens towards the *Normal model* as a limiting model in an additive scheme, although such indiscriminate use has been sistematically questioned by economists and financial analysts. This is particularly so in what concerns the distribution of log stock price changes (Fama, 1963). From Fama's analysis of

monthly stock price changes — $\log(p_t / p_{t-1})$ — it is evident that an appropriate model must be highly peaked and heavy tailed when compared either to *Normal* or to *Gumbel* models, i.e., there is empirical evidence of *"heavy tails"* ($g > 0$).

In Fraga Alves and Gomes (1996) the analysis of the observed values of Gumbel statistical choice test statistic (Gomes, 1987), when applied to data published by *Banco de Portugal* (1984-1993), namely on changes of the monthly exchange rate of the U.S. Dollar and the Dutch Guilder, enhanced that although there is not any strong evidence against $g = 0$, there is a slight indication of an heavy tail, particularly for the Dutch Guilder.

We here apply the re-sampling techniques of sections 2 and 3. In Figures 5 and 6 it is shown the stability of the sample paths of the estimate of the optimal sample fraction and of Hill's estimate at that optimal level for the exchange rate returns of the US Dollar and of the Dutch Guilder, respectively. In Figure 7 we show the sample paths of the different estimates of the tail index, for both sets of data.

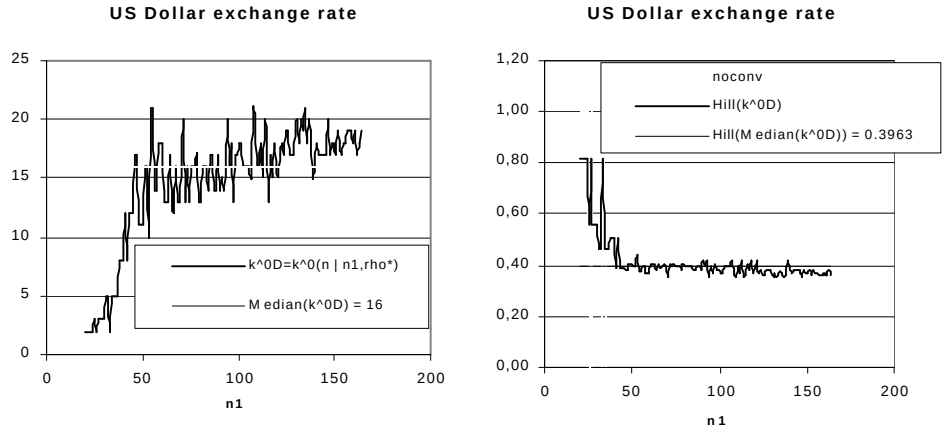

Figure 5 Sample path of estimates of optimal sample fraction and Hill estimates at optimal levels, for the US Dollar exchange rate returns and for sub-sample sizes n_1 = 25(1)163.

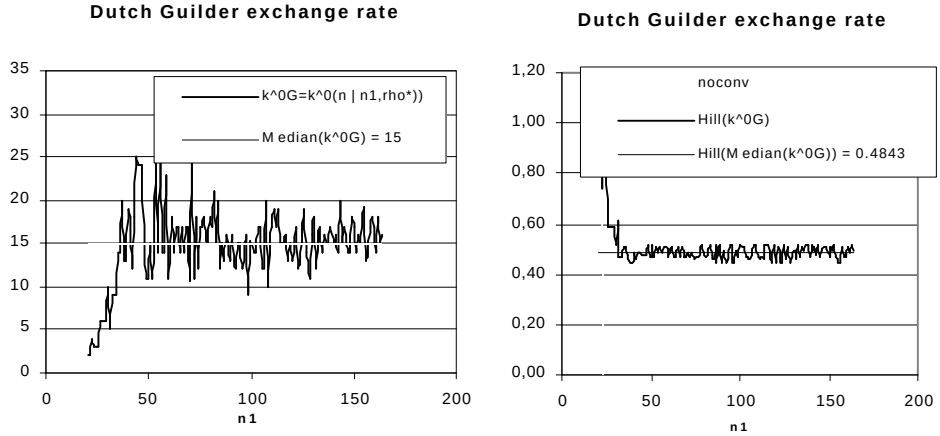

Figure 6. Sample path of estimates of optimal sample fraction and Hill estimates at those optimal levels, for the Dutch Guilder exchange rate changes and for sub-sample sizes n_1 = 25(1)163.

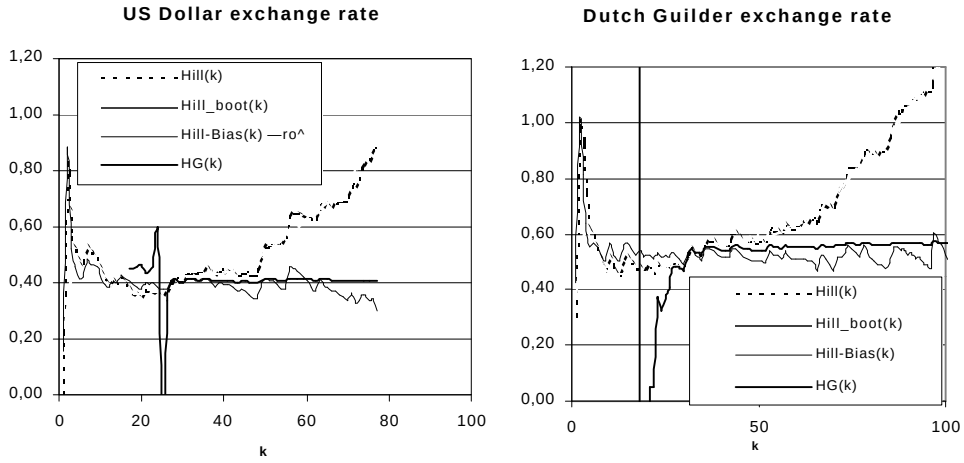

Figure 7. Sample path of Hill estimate, Bootstrap Hill estimate, Unbiased Bootstrap Hill estimate and Jackknife Hill estimates of the tail index for US Dollar and Dutch Guilder exchange rate changes, and for $k = 1(1)n\text{-}1$.

The second order parameter is around $r = -1$ for both sets of data, and we are in the presence of heavy tails. The estimate of the tail index g is equal to 0.4 for the US Dollar and equal to 0.58 for the Dutch Guilder. Despite the shortness of the size of the samples under study, $n = 164$, the stability of the sample path of the new estimates is quite striking!

REFERENCES

BEIRLANT, J., P. VYNCKIER and J. L. TEUGELS (1996a). "Excess function and estimation of the extreme-*Bernoulli* **2**, 293-318.

BEIRLANT, J., P. VYNCKIER and J. L. TEUGELS (1996b). "Tail index estimation, Pareto quantile plots, and regression diagnostics". *J. Amer. Statist. Assoc.* **91**, 1659-1667.

DANIELSON, J., L. de HAAN, L. PENG and C.G. de VRIES (1997). *Using a bootstrap method to choose the sample fraction in the tail index estimation*. TI97-016/4. Erasmus University Rotterdam.

DEKKERS, A.L.M. and L. de HAAN (1993). "Optimal choice of sample fraction in extreme-value estimation". *J. Multivariate Analysis* **47**, 173-195.

DEVROYE, L.P. and L. GYORFI (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.

DREES, H. and E. KAUFFMAN (1998). "Selecting the optimal sample fraction in univariate extreme value estimation". *Stoch. Proc. Appl* **75**, 149-172.

EFRON, B. (1979)."Bootstrap methods – another look at *Ann. Statist.* **7**, 1-26.

FAMA, E. (1963). "Mandelbrot and the stable Paretian hypothesis". *J. Business* **36**, 420-429.

FAMA, E. (1965). "The behaviour of stock market prices". *J. Business* **38**, 34-105.

FRAGA ALVES, M.I. and M.I. GOMES (1996). "Statistical choice of extreme value domains of attraction — a comparative analysis". *Comm. Statist. — Theory Meth. 25, 789-811.*

GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd edition). Krieger.

GELUK, J. and L. de HAAN (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.

GNEDENKO, B.V. (1943). "Sur la distribution limite du terme maximum d'une série aléatoire". *Ann. Math.* **44**, 423-453.

GOMES, M.I. (1987). "Extreme Value Theory — Statistical choice". In P. Revesz *et al* (eds.*), Goodness-of-Fit*. 195-209, North-Holland.

GOMES, M.I. (1994). "Metodologias Jackknife e Bootstrap em Estatística de Extremos". *Actas do II Congresso Anual da Sociedade Portuguesa de Estatística*, 31-46.

GOMES, M.I. (1998). *The bootstrap methodology in Statistical Extremes — the choice of the optimal sample fraction*. Notas e Comunicações CEAUL 15/98.

GOMES, M.I., M.J. MARTINS and M. NEVES (1998). *Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology*. Notas e Comunicações CEAUL 18/98

GRAY, H.L. and W.R. SCHUCANY (1972). *The Generalized Jackknife Statistic*. Marcel Dekker.

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

HAAN, L. de (1984). "Slow variation and characterization of domains of attraction". In J. Tiago de Oliveira (ed.) *Statistical Extremes and Applications*, 31-38. D. Reidel.

HAAN, L. de and L. PENG (1995). "Comparison of tail index estimators". *To appear* in *Statistica Neerlandica*.

HAAN, L. de, L. PENG and T.T. PEREIRA (1997) "A bootstrap-based method to achieve optimality in estimating *In* Lian Peng, *Second Order Condition and Extreme Value Theory*, Ph.D. Thesis, Erasmus Universiteit.

HAAN, L. de and U. STADTMÜLLER (1996). "Generalized regular variation of second order". *J. Austral. Math. Soc*. (A) **61**, 381-395.

HALL, P. (1990). "Using the bootstrap to estimate mean squared error and selecting parameter in nonparametric problems". *J. Multivariate Analysis* **32**, 177-203.

HILL, B.M. (1975). "A simple general approach to inference about the tail of a distribution". *Ann. Statist.* **3**, 1163-1174.

MARTINS, M. J., GOMES, M. I. and M. NEVES (1997). "Some results on the behaviour of Hill's estimator". To appear in *J. Statist. Comput. and Simulation*.

PENG, L. (1998). *Second Order Condition and Extreme value theory.* Ph.D. Thesis, Erasmus Universiteit Rotterdam.

QUENOUILLE, B. (1956). "Notes on bias in estimation". *Biometrika* **43**, 353-360.

RESNICK, S. and C. STARICA (1997). "Smoothing the Hill estimator". *Adv. Appl. Probab*. **29**, 271-293.

TUKEY, J. (1958). "Bias and confidence in not quite large samples". *Ann. Math. Statist.* **29**, 614.