Subsampling techniques and the Jackknife methodology in the estimation of the extremal index

M. Ivette Gomes

DEIO and CEAUL, Universidade de Lisboa

Andreia Hall

CEA and DM, Universidade de Aveiro

and

M. Cristina Miranda

ISCA, Universidade de Aveiro and CEAUL, Universidade de Lisboa

December 16, 2005

Abstract. For a sequence of independent, identically distributed random variables any limiting point process for the time normalized exceedances of high levels is a *Poisson process*. However, for stationary dependent sequences, under general local and asymptotic dependence restrictions, any limiting point process for the time normalized exceedances of high levels is a *compound Poisson process*, i.e., there is a *clustering of high exceedances*, where the underlying Poisson points represent cluster positions, and the multiplicities correspond to the cluster sizes. For such classes of stationary sequences there exists and is well defined the *extremal index* θ, $0 \leq \theta \leq 1$, directly related to the clustering of exceedances of high values. We have $\theta = 1$ for independent, identically distributed sequences, i.e., high exceedances appear individually, and $\theta > 0$ for "almost all" cases of interest. Here, we are interested in the estimation of the extremal index through the use of the *Generalized Jackknife* methodology, possibly together with the use of subsampling techniques. A case-study in the field of finance will illustrate the performance of the new extremal index estimator comparatively to the classical one.

AMS 2000 subject classification. Primary 62G32; Secondary 65C05.
Keywords and phrases. *Statistics of Extremes*, *Generalized Jackknife*, *Subsampling techniques*, *Extremal index.*

1 Introduction and preliminaries

We shall assume throughout the paper that we are working with a strictly stationary sequence of random variables (r.v.'s), $\mathbf{X} = \{X_n\}_{n\geq 1}$, under general asymptotic and local dependence restrictions, to be specified in subsection 1.1. Whenever we need to perform a statistical analysis of extreme values, we are immediately confronted with a few parameters of interest, the so-called parameters of extreme or even rare events, which need to be suitably estimated on the basis of an available sample. Among those parameters we mention the *tail index* γ and, for dependent sequences, the *extremal index* θ, to be introduced later on, in subsection 1.2.

1.1 Long range and local dependence conditions

We shall assume that both the long-range dependence condition $\mathbf{D}$ (Leadbetter *et al.*, 1983) and the local dependence condition $\mathbf{D}$" (Leadbetter and Nandagopalan, 1989) hold for the stationary sequence $\mathbf{X}$. Let us denote $F_{i_1,i_2,\cdots,i_p}(u_1, u_2, \cdots, u_p) := \mathbb{P}\left(X_{i_1} \leq u_1, X_{i_2} \leq u_2, \cdots, X_{i_p} \leq u_p\right)$, the joint distribution function (d.f.) of $(X_{i_1}, X_{i_2}, \cdots, X_{i_p})$, for any arbitrary positive integers, $(i_1, i_2, \ldots, i_p)$.

Definition 1.1. *The* $\mathbf{D}(u_n)$ *condition holds for the stationary sequence* $\mathbf{X}$ *if for every integers* $p,\ q$ *and* $i_1 < i_2 < \cdots < i_p < j_1 < j_2 < \cdots j_q < n$*, for which* $j_1 - i_p \geq l = l_n$*, we have*

$$\left|F_{i_1,i_2,\cdots,i_p,j_1,j_2,\cdots,j_q}(u_n,\cdots,u_n) - F_{i_1,i_2,\cdots,i_p}(u_n,\cdots,u_n)\ F_{j_1,j_2,\cdots,j_q}(u_n,\cdots,u_n)\right| \le \alpha_{n,l},$$

with $\alpha_{n,l} \underset{n\to\infty}{\longrightarrow} 0$, *for some sequence* $\{l_n = o(n)\}$.

Definition 1.2. *The local dependence condition* $\mathbf{D}$"(u_n) *holds for a stationary* $\mathbf{D}(u_n)$*-sequence* $\mathbf{X}$ *if there exists a sequence of integers* $\{s_n\}$, *with* $s_n \to \infty$, $s_n\alpha_{n,l_n} \to 0$, $s_n\ l_n/n \to 0$, $s_n(1-F(u_n)) \to 0$, *as* $n \to \infty$, *such that*

$$\lim_{n\to\infty}\ n\sum_{j=3}^{r_n} \mathbb{P}\left(X_1 > u_n,\ X_{j-1} \le u_n < X_j\right) = 0,$$

with $r_n = [n/s_n]$.

Remark 1.1. *Note that conditions* $\mathbf{D}$ *and* $\mathbf{D}$*" hold trivially for independent, identically distributed (i.i.d.) data.*

1.2 The tail index and the extremal index

Let us denote $\mathbf{Y} = \{Y_n\}_{n\ge1}$, a sequence of i.i.d. r.v.'s from an underlying parent d.f. F. Let $Y_{i:n}$, $1 \le i \le n$, be the set of associated ascending order statistics (o.s.). The tail index γ may be roughly defined by the approximation

$$\mathbb{P}\left[Y_{n:n} := \max(Y_1, Y_2, \ldots, Y_n) \le x\right] = F^n(x) \approx EV_\gamma\left(\frac{x-b_n}{a_n}\right),$$

which holds for large values of n, and appropriate sequences $a_n > 0$, $b_n \in \mathbb{R}$, and where

$$EV_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\},\ 1+\gamma x > 0 & \text{if} \quad \gamma \ne 0 \\ \exp(-\exp(-x)),\ x \in \mathbb{R} & \text{if} \quad \gamma = 0 \end{cases} \tag{1.1}$$

is the *Extreme Value* (EV) d.f. The *tail index* γ is thus directly related to the weight of the right tail of the underlying model F: as γ increases the right tail becomes heavier and heavier. More formally, we may say that if the sequence of maximum values $Y_{n:n}$, linearly normalized, converges towards a non degenerate r.v., such a r.v. has an EV_γ d.f., defined by (1.1). We then say that F is in the domain of attraction for maxima of EV_γ, and denote such a fact by $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$. The same EV_γ model appears as the limiting d.f. of the normalized maximum for a large class of stationary sequences $\mathbf{X}$, like, for instance, the ones for which the mixing condition $\mathbf{D}$ holds (Leadbetter *et al.*, 1983).

Let us assume that the stationary sequence $\mathbf{X} = \{X_n\}_{n\geq 1}$ comes from an underlying d.f. F, being $\mathbf{Y} = \{Y_n\}_{n\geq 1}$ the associated i.i.d. sequence (i.e., an i.i.d. sequence from the same model F). Under adequate local dependence conditions, like $\mathbf{D}$", the limiting d.f. of the maximum $X_{n:n} := \max(X_1, X_2, \ldots, X_n)$ may be directly related to the limiting d.f. of the maximum, $Y_{n:n}$, of the i.i.d. associated sequence, through a new parameter, the so-called *extremal index*.

Definition 1.3. *The stationary sequence $\{X_n\}_{n\geq 1}$ is said to have an extremal index θ $(0 < \theta \leq 1)$ if, for all $\tau > 0$, we may find a sequence of levels $u_n = u_n(\tau)$ such that*

$$\mathbb{P}\left(Y_{n:n} \leq u_n\right) = F^n(u_n) \underset{n\to\infty}{\longrightarrow} e^{-\tau} \quad \textit{and} \quad \mathbb{P}\left(X_{n:n} \leq u_n\right) \underset{\to\infty}{\longrightarrow} e^{-\theta\ \tau}. \quad (1.2)$$

From Definition 1.3 it follows that the *extremal index* θ may be informally defined by the approximations:

$$P[\max(X_1, X_2, \ldots, X_n) \leq x] \approx F^{n\theta}(x) \approx EV_\gamma^\theta\left(\frac{x - b_n}{a_n}\right)$$
$$= EV_\gamma\left(\frac{x - b_n'}{a_n'}\right), \quad \begin{cases} a_n' = a_n\ \theta^\gamma \\ b_n' = b_n + a_n\ \left(\frac{\theta^\gamma - 1}{\gamma}\right) \end{cases}.$$

This means that there is a *shrinkage* of maximum values, but since the EV_γ d.f. is stable for maxima, the limiting d.f. of $X_{n:n}$, linearly normalized, is still an EV_γ.

In order to better understand the intuitive meaning of the *extremal index* let us think on the point process of exceedances over high thresholds for the sequence $\mathbf{X}$. Under independence (or even adequate weak dependence) this point process converges to a homogeneous Poisson process, as $n \to \infty$, but when there is a slightly stronger local dependence, clusters of exceedances may occur and the limiting process may be a compound Poisson process. Under $\mathbf{D}''$ an upcrossing is generally followed by a cluster of exceedances and therefore the clusters may be roughly identified by the occurrence of upcrossings. Indeed, Leadbetter and Nandagopalan(1989) proved that the *extremal index* may then also be defined as the reciprocal of the "mean time of duration of extreme events", being directly related to the exceedances of high levels. We have:

$$\theta = \frac{1}{\text{limiting mean size of clusters}}$$

$$= \lim_{n\to\infty} P(X_2 \le u_n | X_1 > u_n) \quad \text{(downcrossings)} \tag{1.3}$$

$$= \lim_{n\to\infty} P(X_1 \le u_n | X_2 > u_n) \quad \text{(upcrossings)}, \tag{1.4}$$

where u_n, a sequence of values such that

$$F(u_n) = 1 - \tau/n + o(1/n), \text{ as } n \to \infty, \tag{1.5}$$

is a so-called *normalized level*, i.e., a level such that $n(1 - F(u_n)) \to \tau > 0$, or equivalently, a level such that $F^n(u_n) \to \exp(-\tau)$, as $n \to \infty$, i.e., (1.2) holds.

The extremal index estimation is important not only by itself but also because of its influence in the estimation of other parameters of rare events, like the primary one, the tail index γ, both with great influence on the estimation of a *high quantile* χ_{1-p}, a value such that $F(\chi_{1-p}) = 1 - p$, $p < 1/n$, with n the sample size, or the *return period* of a high level, i.e., the mean waiting time between consecutive exceedances of a specified high level.

1.3 The extremal index estimation

Given a sample $(X_1, X_2, \ldots, X_n)$, the validity of (1.3) and (1.4) provides an obvious non-parametric estimator of θ (Leadbetter and Nandagopalan, 1989; Nandagopalan, 1990), given by the ratio between the number of downcrossings (or upcrossings) and the number of exceedances over a high threshold. Chosen a suitable threshold u, Nandagopalan's estimator is given by

$$\theta_n^N = \theta_n^N(u) := \frac{\sum_{j=1}^{n-1} I_{[X_j > u, X_{j+1} \le u]}}{\sum_{j=1}^{n} I_{[X_j > u]}} = \frac{\sum_{j=1}^{n-1} I_{[X_j \le u < X_{j+1}]}}{\sum_{j=1}^{n} I_{[X_j > u]}}, \tag{1.6}$$

where I_A denotes, as usual, the indicator function of the event A. In order to have consistency of this estimator the high level $u = u_n$ must be such that $n(1 - F(u_n)) = c_n \tau = \tau_n$, $\tau_n \to \infty$ and $\tau_n/n \to 0$ (Nandagopalan, 1990). We shall here consider a deterministic level $u \in [X_{n-k:n}, X_{n-k+1:n})$, the interval between the $(k+1)$ and the k-th top o.s. The extremal index estimator is then a function of k, the number of top o.s.'s higher than the chosen threshold. We shall thus work with the estimator

$$\widehat{\theta}_n^N(k) \equiv \theta_n^N(u) := \frac{1}{k} \sum_{j=1}^{n-1} I_{[X_j \le u < X_{j+1}]}, \quad u \in [X_{n-k:n}, X_{n-k+1:n}).$$

Equivalently, we may write

$$\widehat{\theta}_n^N(k) = \frac{1}{k} \sum_{j=1}^{n-1} I_{[X_j \le X_{n-k:n} < X_{j+1}]}. \tag{1.7}$$

We are thus placing ourselves in a situation similar to the one we find in the semi-parametric estimation of the tail index, being consistency attained only if k is intermediate, i.e.,

$$k = k_n \to \infty, \quad k = o(n), \text{ as } n \to \infty. \tag{1.8}$$

Indeed, k is replacing $\tau_n = c_n \tau$, $c_n \to \infty$, as $n \to \infty$.

1.4 Simulated models: the $ARCH$ processes

We shall here consider in the simulations a very rich class of stationary processes, the so-called $ARMAX$ processes, all from an underlying Fréchet parent, with d.f. $F(x) \equiv \Phi_\gamma(x) = \exp(-x^{-1/\gamma})$, $x > 0$, $\gamma > 0$. We shall also consider an i.i.d. sequence from the same Fréchet parent, $\Phi_\gamma(x)$.

An autoregressive for maxima ($ARMAX$) sequence is based on an i.i.d. sequence of innovations $\{Z_i\}_{i\geq 1}$, with d.f. H, and is defined through the relation,

$$X_i = \beta \max(X_{i-1}, Z_i), \; i \geq 1, \; 0 < \beta < 1. \tag{1.9}$$

From this definition, it is obvious that if the $ARMAX$ sequence has a stationary distribution F, it depends on the d.f. H of the i.i.d. sequence $\{Z_i\}_{i\geq 1}$ through the relation $F(\beta x)/F(x) = H(x)$. Alpuim (1989) has proved that such a markovian sequence possesses a stationary distribution $F(x)$ if and only if there exists $x > 0$, where $H(x/\beta) > 0$ and $0 < \sum_{j\geq 1}(1-H(x/\beta^j)) < \infty$. Consequently, the class $\mathbf{S}$ of stationary d.f.'s for the sequence $\mathbf{X}$ is a very huge class: any d.f. F such that

$$F(x) = F(x/\beta) \; H(x/\beta), \tag{1.10}$$

the stationarity equation, with a support in $\mathbb{R}^+$, and such that $\log F(\exp(x))$ is concave whenever $0 < F(\exp(x)) < 1$, belongs to $\mathbf{S}$. Conditions $\mathbf{D}$ and $\mathbf{D}$" hold for these sequences (Hall, 1996), and consequently they possess an extremal index $\theta < 1$, in a great variety of situations. Indeed, a stationary $ARMAX$ sequence possesses an extremal index if, for some $\gamma \in \mathbb{R}$, $F(x)$, the stationary distribution, belongs to $\mathcal{D}_{\mathcal{M}}(EV_\gamma)$, the domain of attraction for maximum values of the *Extreme Value* d.f. EV_γ in (1.1). Such sequences may have any extremal index in (0,1], and we have $\theta = 1$ if $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$, $\gamma \le 0$, and $0 < \theta < 1$ if $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$, $\gamma > 0$.

If we consider Fréchet innovations, such that $H(x) = \Phi_\gamma^{\beta^{-1/\gamma}-1}(x)$, we then get $F(x) = \Phi_\gamma(x)$, and

$$\theta = \lim_{x\to\infty} \frac{P(X_i > x, X_{i+1} \le x)}{P(X_i > x)} = 1 - \lim_{x\to\infty} \frac{1 - F(x/\beta)}{1 - F(x)} = 1 - \beta^{1/\gamma}.$$

For the particular case $\gamma = 1$, considered later on for illustration, we thus get $\theta = 1 - \beta$.

1.5 Scope of the paper

In section 2 of this paper, and essentially on the basis of the dominant component of the bias term of Nandagopalan's estimator for the stationary sequences specified in subsection 1.4, we make a realistic assumption on the bias term of the extremal index estimator in (1.7). That bias term increases drastically as k increases and similarly to the tail index estimation, there is an important tradeoff between bias and variance, linked to the choice of the threshold. At this point any contribution towards bias reduction and stability over a considerable set of thresholds (without affecting the variance too much) seems to be important and constitutes the major goal of this paper. Since such a bias term reveals two main components of different orders, we shall discuss, in section 3, the *Generalized Jackknife* methodology (Gray and Shucany, 1972), suggesting a way to use it in the estimation of the extremal index. In section 4 we shall obtain, via Monte Carlo methods, the main distributional properties of the proposed Generalized Jackknife estimator of the extremal index, not only for $ARMAX$ structures, but also for i.i.d. random samples. Section 5 is devoted to the effect of sub-sampling in the estimation of the extremal index and finally, in section 6, we provide an application in the field of finance.

2 Assumption on the data structures

Whenever we consider a level u in $[X_{n-k:n},\ X_{n-k+1:n})$, the extremal index estimator in (1.7) is a function of k, the number of o.s. higher than the chosen threshold. For a reasonably large number of dependent structures, like the ones in subsection 1.4, that altogether exhibit different features, peculiar to most of the data in practice, the bias function of $\widehat{\theta}_n(k)$ has two dominant components of orders k/n and $1/k$. We shall assume here that, as $n \to \infty$, and for intermediate k, i.e., whenever (1.8) holds,

$$Bias\left[\widehat{\theta}_n(k)\right] = \varphi_1(\theta)\left(\frac{k}{n}\right) + \varphi_2(\theta)\left(\frac{1}{k}\right) + o\left(\frac{1}{k}\right) + o\left(\frac{k}{n}\right), \qquad (2.1)$$

and we shall provide in subsection 2.1 illustrative examples that justify the assumption (2.1).

Since the bias term in (2.1) reveals two main components of different orders, we shall discuss in section 3 the *Generalized Jackknife* methodology (Gray and Schucany, 1972), suggesting a way to use it in the estimation of the extremal index. Notice that the Generalized Jackknife methodology has already been used with success in the estimation of a positive tail index γ (Gomes *et al.*, 2000, 2002, 2005).

Note that, as $n \to \infty$, the bias converges towards zero, as $n \to \infty$, if and only if $k \to \infty$ and $k/n \to 0$. Hence the need to assume (1.8). Moreover, and for a great variety of dependent sequences (Nandagopalan, 19900), there exists a function $\sigma(\theta)$ that enables us to write $Var\left[\widehat{\theta}_n(k)\right] \sim \sigma^2(\theta)/k$, which goes to zero if and only if $k \to \infty$.

2.1 The dominant component of the bias term of Nandagopalan's estimator

We may write the estimator $\theta_n^N(u)$ in (1.6) as $\theta_n^N(u) = U_n/V_n$, where

$$U_n = \sum_{j=1}^{n-1} I_{[X_j \le u < X_{j+1}]} =: \sum_{j=1}^{n-1} \widetilde{N}_j \quad \text{and} \quad V_n = \sum_{j=1}^{n} I_{[X_j > u]} =: \sum_{j=1}^{n} N_j.$$

For the sake of simplicity, let us assume that F is continuous. We may then choose the normalized level in (1.5) as the value u such that $1-F(u)=\tau/n$. Then, since $\mathbb{E}[N_j]=\mathbb{P}(X_j>u)=\tau/n$, $1\le j\le n$, we have, for any underlying stationary process, either dependent or independent, $\mathbb{E}[V_n]=\tau$. The use of the delta-method (Casela and Berger, 2002, pages 240-245) enables us to obtain the asymptotic approximations,

$$E\left[\theta_n^N(u)\right] \approx \frac{E[U_n]}{E[V_n]} + \frac{Var[V_n]\ E[U_n]}{E^3[V_n]} - \frac{Cov[U_n,V_n]}{2\ E^2[V_n]} \tag{2.2}$$

and

$$Var\left[\theta_n^N(u)\right] \approx \frac{Var[U_n]}{E^2[V_n]} + \frac{E^2[U_n]\ Var[V_n]}{E^4[V_n]} - \frac{2\ E[U_n]\ Cov[U_n,V_n]}{E^3[V_n]}. \tag{2.3}$$

2.1.1 Independent, identically distributed data

Under an i.i.d. set-up, straightforward computations lead us to

$$Var[V_n]=\tau\left(1-\frac{\tau}{n}\right) \quad \text{and} \quad E[U_n]=\tau\left(1-\frac{\tau}{n}\right)\left(1-\frac{1}{n}\right).$$

Next, since the summands of the r.v. U_n are 2-dependent, we get

$$\begin{aligned}
Var[U_n] &= \sum_{j=1}^{n-1} Var\left[\widetilde{N}_j\right] + 2\sum_{j=2}^{n-1} Cov\left[\widetilde{N}_{j-1},\widetilde{N}_j\right] \\
&= (n-1)\left(\frac{\tau}{n}\right)\left(1-\frac{\tau}{n}\right)\left(1-\frac{\tau}{n}\left(1-\frac{\tau}{n}\right)\right) \\
&\qquad +2(n-2)\left\{-\left(\frac{\tau}{n}\left(1-\frac{\tau}{n}\right)\right)^2\right\} \\
&= \tau\left(1-\frac{4\tau}{n}\right)(1+o(\tau/n)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
Cov[U_n,V_n] &= \sum_{j=1}^{n-1} Cov\left[\widetilde{N}_j,\ N_j\right] + \sum_{j=1}^{n-1} Cov\left[\widetilde{N}_j, N_{j+1}\right] \\
&= (n-1)\left(\frac{\tau}{n}\right)\left(1-\frac{\tau}{n}\right)^2 - (n-1)\left(\frac{\tau}{n}\right)^2\left(1-\frac{\tau}{n}\right) \\
&= \tau\left(1-\frac{3\tau}{n}\right)(1+o(\tau/n)).
\end{aligned}$$

The use of (2.2) leads us to the following bias term for $\theta_n^N(u)$ in (1.6), whenever the underlying data is i.i.d. ($\theta = 1$):

$$Bias[\theta_n^N(u)] = E\left[\theta_n^N(u) - 1\right] = \left(-\frac{\tau}{n} + \frac{1}{2\tau}\right)(1 + o(1)).$$

Consequently (2.1) holds, with $\varphi_1(\theta) = -1$ and $\varphi_2(\theta) = 1/2$. From (2.3), we may additionally get information on the variance of our estimator. For i.i.d. sequences the term of the order of $1/\tau$ disappears, and we get only a dominant term of the order of $1/n$, i.e., $Var[\theta_n^N(u)] = 1/n + o\,(1/n)$. However, for the dependent sequences under study, and for much more general dependent sequences, as specified in Nadagopalan (1990), $Var\left[\theta_n^N(u)\right]$ is indeed of the order of $1/\tau$.

2.1.2 The $ARMAX$ processes

Let us consider the $ARMAX$ model in (1.9). As said before, the use of the stationarity equation in (1.10), enables us to write,

$$F(x) = \exp\left(-x^{-1/\gamma}\right) \iff H(x) = \frac{F(\beta\; x)}{F(x)} = \exp\left(-x^{-1/\gamma}\left(\beta^{-1/\gamma} - 1\right)\right).$$

Then

$$F(u) = 1 - \tau/n \quad \iff \quad u = \left[-\ln\left(1 - \tau/n\right)\right]^{-\gamma}.$$

Consequently, $F(u/\beta) = (1 - \tau/n)^{\beta^{1/\gamma}}$, $H(u) = (1 - \tau/n)^{\beta^{-1/\gamma} - 1}$ and

$$H(u/\beta) = (1 - \tau/n)^{1 - \beta^{1/\gamma}} =: (1 - \tau/n)^{\theta}, \quad \theta = 1 - \beta^{1/\gamma}.$$

Next,

$$\begin{aligned} P(X_{i-1} \le u < X_i) &= F(u)(1 - H(u/\beta)) = \left(1 - \frac{\tau}{n}\right)\left[1 - \left(1 - \frac{\tau}{n}\right)^{\theta}\right] \\ &= \left(1 - \frac{\tau}{n}\right)\left(\frac{\theta\;\tau}{n} - \frac{\theta(\theta - 1)}{2}\left(\frac{\tau}{n}\right)^2 + o\left(\frac{\tau}{n}\right)^2\right) \\ &= \frac{\theta\;\tau}{n}\left(1 - \frac{(\theta + 1)\;\tau}{2\;n} + o\left(\frac{\tau}{n}\right)\right), \end{aligned}$$

i.e.,

$$E[U_n] = \theta\;\tau\left(1 - \frac{(\theta + 1)\;\tau}{2\;n}\right)\left(1 + o\left(\frac{\tau}{n}\right)\right).$$

To compute $Var\,[V_n]$, we use the fact that (Hsing and Leadbetter, 1988) $\mathbb{E}\,[V_n|U_n] = U_n(1+d_n)/\theta$ and $Var\,[V_n|U_n] = U_n(1-\theta)(1+o(1))/\theta^2$, with $d_n = o(1)$. But we may get more information on d_n. Indeed,

$$
\begin{aligned}
\tau &= \mathbb{E}\,[V_n] = \mathbb{E}\,[\mathbb{E}\,[V_n|U_n]] = \mathbb{E}\left[\frac{U_n(1+d_n)}{\theta}\right] \\
&= \left(\frac{1+d_n}{\theta}\right)\theta\tau\left(1-\frac{(\theta+1)\tau}{2n}+o\left(\frac{\tau}{n}\right)\right) \\
\Longleftrightarrow\ & d_n = \frac{\frac{(\theta+1)\tau}{2n}+o(\tau/n)}{1-\frac{(\theta+1)\tau}{2n}+o\left(\frac{\tau}{n}\right)} = \frac{(\theta+1)\tau}{2n}+o\left(\frac{\tau}{n}\right).
\end{aligned}
$$

Consequently,

$$
\mathbb{E}\,[V_n|U_n] = \frac{U_n}{\theta}\left(1+\frac{(\theta+1)\tau}{2n}+o\left(\frac{\tau}{n}\right)\right).
$$

Also, as $n\to\infty$,

$$
Var\,[U_n] = \theta\ \tau(1+o(1))
$$

and

$$
\begin{aligned}
Var\,[V_n] &= \mathbb{E}\,[Var\,(V_n|U_n)] + Var\,[\mathbb{E}\,(V_n|U_n)] \\
&= \mathbb{E}\left[\frac{U_n(1-\theta)}{\theta^2}(1+o(1))\right] + Var\left[\frac{U_n}{\theta}\left(1+\frac{(\theta+1)\tau}{2n}+o\left(\frac{\tau}{n}\right)\right)\right] \\
&= \frac{(1-\theta)\tau}{\theta}(1+o(1)) + \frac{\tau}{\theta}(1+o(1)) = \frac{\tau(2-\theta)}{\theta}(1+o(1)).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
Cov\,[U_n,V_n] &= Cov\,[\mathbb{E}\,(U_n|V_n), V_n] = Cov\left[\frac{V_n}{\theta}\left(1+\frac{(\theta+1)\tau}{2\ n}+o\left(\frac{\tau}{n}\right)\right), V_n\right] \\
&= \frac{1}{\theta}\left(1+\frac{(\theta+1)\tau}{2\ n}+o\left(\frac{\tau}{n}\right)\right) Var\,(V_n) \\
&= \tau\left(1+\frac{(\theta+1)\tau}{2\ n}\right)\ (1+o(1)) = \tau(1+o(1)),
\end{aligned}
$$

and the use of (2.2) and (2.3), enables us to write the approximations,

$$
\mathbb{E}\left[\theta_n^N(u)\right] = \theta - \frac{\theta(\theta+1)}{2}\left(\frac{\tau}{n}\right) + \frac{3-2\ \theta}{2\ \tau} + o\left(\frac{1}{\tau}\right) + o\left(\frac{\tau}{n}\right)
$$

and

$$
Var\left[\theta_n^N(u)\right] = \frac{\theta(1-\theta)}{\tau}(1+o(1)).
$$

Once again (2.1) holds, with $\varphi_1(\theta) = -\theta(\theta+1)/2$ and $\varphi_2(\theta) = (3-2\theta)/2$.

3 The Generalized Jackknife methodology

The main objectives of the *Jackknife methodology* are:

1. Bias and variance estimation, together with the derivation of the sampling distribution of a certain statistic, only through manipulation of the observed data $\underline{x}$.
2. The building of estimators with bias and mean squared error smaller than those of an initial set of estimators.

The Jackknife is thus a resampling methodology, which gives usually a positive answer to the question of whether the combination of information may improve the quality of estimators of a certain parameter or functional.

Let $\underline{X} = (X_1, \ldots, X_n)$ be a sample from an underlying model F, and let $T_n = T_n(\underline{X}, F)$ be an estimator of a functional $\theta(F)$, or of a parameter θ, in the case F is known up to unknown parameters. Looking at the bias in (2.1) we notice that we have two main terms we would like to reduce (one of the order of k/n and another of the order of $1/k$). According to the *Generalized Jackknife* methodology we thus need to have access to three estimators of θ, $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$, with the same type of bias. Then,

Definition 3.1. *Given three biased estimators of θ, $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$, such that*

$$E\left[T_n^{(i)} - \theta\right] = d_1(\theta)\ \varphi_1^{(i)}(n) + d_2(\theta)\ \varphi_2^{(i)}(n),\ i = 1, 2, 3, \tag{3.1}$$

the Generalized Jackknife statistic (of order 2) is given by

$$T_n^{GJ} := \frac{\begin{Vmatrix} T_n^{(1)} & T_n^{(2)} & T_n^{(3)} \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{Vmatrix}}{\begin{Vmatrix} 1 & 1 & 1 \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{Vmatrix}}, \tag{3.2}$$

with $||A||$ denoting, as usual, the determinant of the matrix A.

Straightforwardly, one may prove that:

Proposition 3.1. *Under the validity of* (3.1), *the statistic T_n^{GJ} in* (3.2) *is unbiased for the estimation of θ.*

Moreover, although the variance of T_n^{GJ} is always larger than the variance of the original estimators, the mean squared error (MSE) of T_n^{GJ} is often (unfortunately, not in general!) smaller than that of any of the statistics $T_n^{(i)}$, $i = 1, 2, 3$.

3.1 An extremal index Generalized Jackknife estimator

The information given in (2.1), on the bias of the extremal index estimator $\widehat{\theta}_n^N(k)$ in (1.7), led us to consider first the Generalized Jackknife estimator of order 2, based on the estimator $\widehat{\theta}_n^N(k)$ computed at the three levels, k, $[k/2]+1$ and $[k/4]+1$, not at all chosen with any kind of optimality criterion, and where $[x]$ denotes, as usual, the integer part of x (Gomes and Miranda, 2003). We then got the new estimator

$$\begin{aligned}
\widehat{\theta}_n^{GJ_0}(k) \quad &:= \quad \frac{\begin{Vmatrix} \widehat{\theta}_n^N([k/4]+1) & \widehat{\theta}_n^N([k/2]+1) & \widehat{\theta}_n^N(k) \\ 4 & 2 & 1 \\ 1/4 & 1/2 & 1 \end{Vmatrix}}{\begin{Vmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 1/4 & 1/2 & 1 \end{Vmatrix}} \\
&= \quad 5\ \widehat{\theta}_n^N\left([k/2]+1\right) - 2\left(\widehat{\theta}_n^N\left([k/4]+1\right) + \widehat{\theta}_n^N(k)\right). \qquad (3.3)
\end{aligned}$$

This estimator $\widehat{\theta}_n^{GJ_0}(k)$ is "asymptotically unbiased" for the estimation of θ, in the sense that with its use we are able to remove the two dominant components of bias, the ones of orders k/n and $1/k$. But this estimator has very stable sample paths around the target value θ at expenses of a very high variance, which does not enable it to overpass the original estimator in (1.7), regarding mean squared errors at optimal levels. Computationally, we have then analyzed the performance of different classes of Generalized Jackknife extremal index estimators. All these classes have revealed to be highly efficient in the reduction of bias, but again, most of them did not enable us to beat the original estimator, regarding mean squared errors at optimal levels. We have finally decided to consider the levels k, $[\delta k]+1$ and $[\delta^2 k]+1$, dependent of a *tuning* parameter δ, $0<\delta<1$. We then get the class of estimators,

$$\widehat{\theta}_n^{GJ(\delta)}(k) \quad = \quad \frac{\left\|\begin{array}{ccc} \widehat{\theta}_n^N([\delta^2 k]+1) & \widehat{\theta}_n^N([\delta\ k]+1) & \widehat{\theta}_n^N(k) \\ \delta^2 & \delta & 1 \\ 1/\delta^2 & 1/\delta & 1 \end{array}\right\|}{\left\|\begin{array}{ccc} 1 & 1 & 1 \\ \delta^2 & \delta & 1 \\ 1/\delta^2 & 1/\delta & 1 \end{array}\right\|},$$

i.e.,

$$\widehat{\theta}_n^{GJ(\delta)}(k) := \frac{(\delta^2+1)\ \widehat{\theta}_n^N\ ([\delta k]+1) - \delta\left(\widehat{\theta}_n^N\ \left([\delta^2 k]+1\right) + \widehat{\theta}_n^N(k)\right)}{(1-\delta)^2}. \tag{3.4}$$

Note that $\widehat{\theta}_n^{GJ(1/2)}(k) = \widehat{\theta}_n^{GJ_0}(k)$, with $\widehat{\theta}_n^{GJ_0}$ given in (3.3), i.e., the class in (3.4) generalizes the class in (3.3). Among the members of the class in (3.4), we have been led to the heuristic choice $\delta = 1/4$. The distributional properties of $\widehat{\theta}_n^{GJ} \equiv \widehat{\theta}_n^{GJ(1/4)}$ will be presented in the following section and, so far, due to technical difficulties, have been obtained only through simulation techniques.

4 Simulated distributional behaviour of the Generalized Jackknife extremal index estimator

We shall here obtain the distributional properties of the *Generalized Jackknife* extremal index estimator $\widehat{\theta}_n^{GJ}(k) \equiv \widehat{\theta}_n^{GJ(1/4)}(k)$, the estimator in (3.4) with $\delta = 1/4$, for the above mentioned types of stationary structures, all from an underlying Fréchet parent, with d.f. $F(x) \equiv \Phi_\gamma(x) = \exp(-x^{-1/\gamma}),\ x > 0$, and $\gamma = 1$.

For samples of size $n = 100, 200, 500, 1000, 2000, 5000$ and 10000, from the different simulated structures, we have performed a multi-sample simulation with 5000 runs and 10 replicates. For details on multi-sample simulation see, for instance, Gomes and Oliveira (2001). We have simulated, for both the original estimator $\widehat{\theta}_n^N(k)$ in (1.7) and the Generalized Jackknife estimator $\widehat{\theta}_n^{GJ}(k)$, the estimator in (3.4) with $\delta = 1/4$, the Mean Value ($E^\bullet$), the Mean Squared Error ($MSE^\bullet$), the Optimal Sample Fraction, $k_0^\bullet/n$, with $k_0^\bullet := \arg\min_k MSE[\widehat{\theta}_n^\bullet(k)]$, and two indicators at optimal levels, the Relative Efficiency ($REFF_0^{GJ}$), defined as

$$REFF_0 \equiv REFF_0^{GJ} = \sqrt{\frac{MSE_s\left[\widehat{\theta}_{n0}^N\right]}{MSE_s\left[\widehat{\theta}_{n0}^{GJ}\right]}}, \tag{4.1}$$

and a Bias Reduction Indicator (BRI_0^{GJ}), defined as

$$BRI_0 \equiv BRI_0^{GJ} = \left|\frac{BIAS_s\left[\widehat{\theta}_{n0}^N\right]}{BIAS_s\left[\widehat{\theta}_{n0}^{GJ}\right]}\right|. \tag{4.2}$$

with $\widehat{\theta}_{n0}^\bullet = \widehat{\theta}_n^\bullet(k_{0s}^\bullet(n))$ and where the subscript s denotes "simulated". Also, to illustrate the loss of sensitivity of the new estimators to the choice of the level k, we have simulated another indicator, related to the stability with k of the mean value, and denoted

$$STI^{GJ} = \frac{\sum_{k=1}^{n-1} I_{\left\{\left|E_s(\widehat{\theta}_n^{GJ}(k))-\theta\right|\le 0.01\right\}}}{\sum_{k=1}^{n-1} I_{\left\{\left|E_s(\widehat{\theta}_n^{N}(k))-\theta\right|\le 0.01\right\}}}. \tag{4.3}$$

Note that the higher the indicators, the better the GJ estimator is.

4.1 The $ARMAX$ processes

To enhance the richness of the $ARMAX$ processes in (1.9), regarding the clustering of exceedances of high levels, we first present, in Figure 1, sample paths of stationary Fréchet(1) $ARMAX$ structures for $\beta = 0.2,\ 0.5$ and 0.8 ($\theta = 0.8,\ 0.5$ and $0.2,$ respectively).

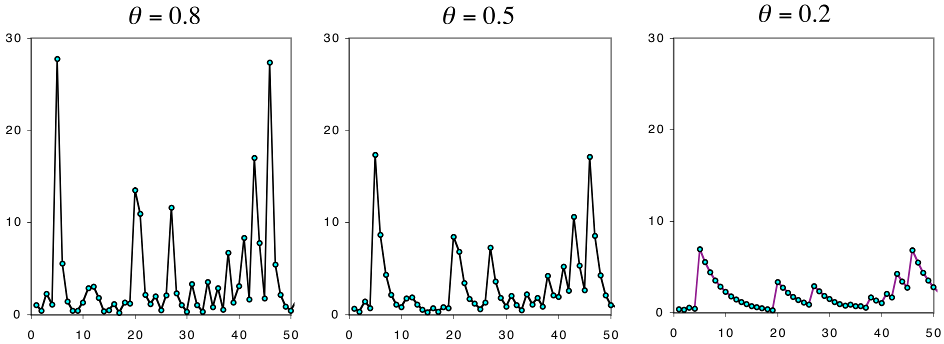

Figure 1: Sample paths of the stationary Fréchet($\gamma = 1$) $ARMAX$ processes in (1.9), for $\beta = 0.2$ (*left*), $\beta = 0.5$ (*center*) and $\beta = 0.8$ (*right*).

Notice the "shrinkage" of maximum values, together with the exhibition of larger and larger clusters of exceedances of high values, as θ decreases (note particularly the clusters of exceedances in the figure at the right, where the limiting mean cluster size should be 5).

We next present in Figure 2 a sample path of $\widehat{\theta}_n^N(k)$ in (1.7) and $\widehat{\theta}_n^{GJ}(k)$, the estimator in (3.4) with $\delta = 1/4$, for a stationary Fréchet(1) $ARMAX$ sample of size $n = 5000$, with $\theta = 0.5$. Comparatively to the behavior of Nandagopalan's estimator, note the reasonably high stability of the sample path of the Generalized Jackknife estimator, around the target value $\theta = 0.5$, for a wide range of k-values.

In Figure 3, to exhibit the influence of the tuning parameter δ in the Generalized Jackknife estimator in (3.4), we present the mean values and

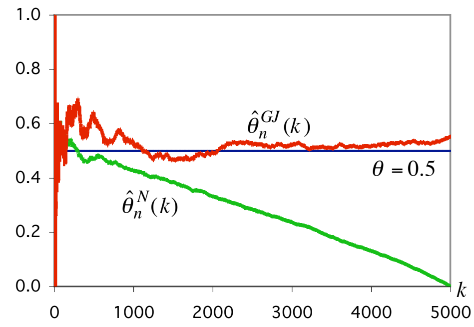

Figure 2: Sample path of the extremal index estimators under study for a sample of size$n = 5000$ from an $ARMAX$ *Fréchet*(1) sequence ($\theta = 0.5$).

MSE's of the Generalized Jackknife estimators in (3.4), associated with $\delta = 0.1,\ 0.2,\ 0.4$ and 0.5, for the same structure as before.

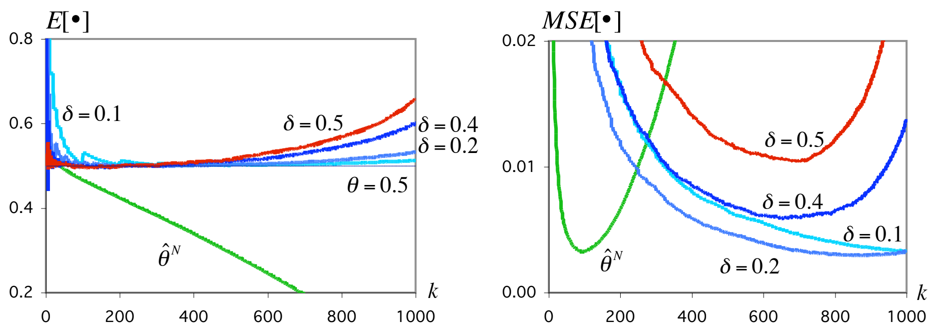

Figure 3: Mean values (*left*) and mean squared errors (*right*) of $\widehat{\theta}_n^{GJ(\delta)}$ in (3.4), $\delta = 0.1, 0.2, 0.4$ and 0.5 and samples of size $n = 1000$ from an $ARMAX$ *Fréchet* (1) sequence ($\theta = 0.5$).

Remark 4.1. *From Figure 3 (left), we see that for all values of* δ *in* $\widehat{\theta}_n^{GJ(\delta)}(k)$ *there is a mean value stability around the target value* $\theta = 0.5$*, for a wide range of* k*-levels. Such a stability contrasts strongly to what happens with* $\widehat{\theta}_n^N(k)$*. Simulations have shown that this is true for all* θ*.*

We present next, in Figures 4, 5 and 6, the simulated mean values and MSE's of $\widehat{\theta}_n^N(k)$ in (1.7) and $\widehat{\theta}_n^{GJ}(k)$, again the estimator in (3.4) with $\delta = 1/4$, for a sample size $n = 1000$ from the $ARMAX$ sequences mentioned before, with extremal indexes $\theta = 0.2$, $\theta = 0.5$ and $\theta = 0.8$, respectively.

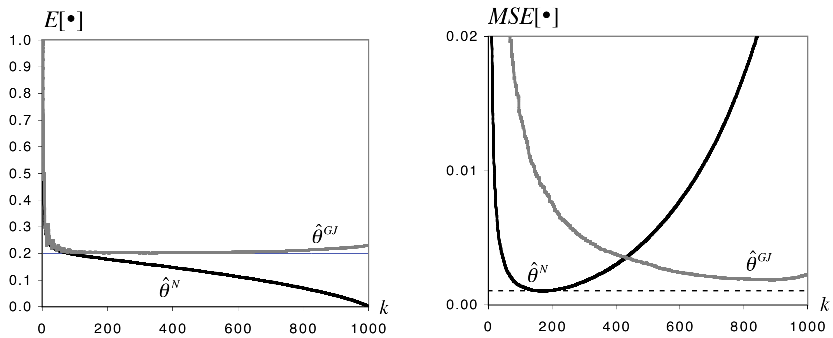

Figure 4: Simulated mean values (*left*) and mean squared errors (*right*), for sample sizes $n = 1000$, from an *ARMAX Fréchet* sequence ($\theta = 0.2$).

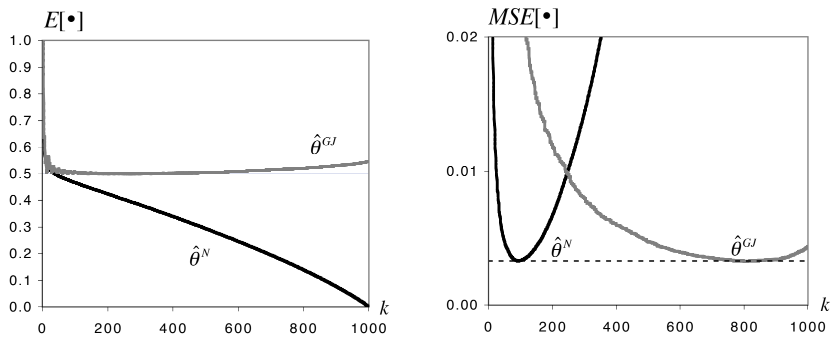

Figure 5: Simulated mean values (*left*) and mean squared errors (*right*), for sample sizes $n = 1000$, from an *ARMAX Fréchet* sequence ($\theta = 0.5$).

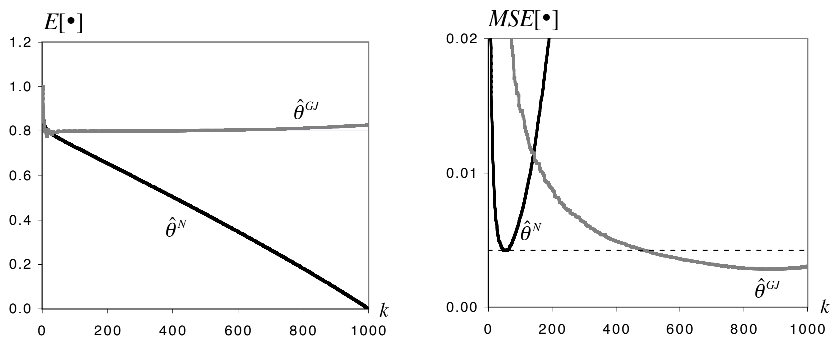

Figure 6: Simulated mean values (*left*) and mean squared errors (*right*), for sample sizes $n = 1000$, from an *ARMAX Fréchet* sequence ($\theta = 0.8$).

Remark 4.2. *From Figure 4 (right), we see that the Generalized Jackknife estimator* $\widehat{\theta}_n^{GJ}$*, i.e., the estimator in* (3.4) *with* $\delta = 1/4$*, may not overpass, for* $n = 1000$*, the original estimator* $\widehat{\theta}_n^{N}$ *in* (1.7)*, regarding mean squared error at optimal levels. Indeed, for small values of* θ*, i.e.,* $\theta \leq 0.2$*, this happens for all the simulated values of* n *(even for* $n = 10000$*), as may be easily seen from the entry* $REFF_0$ *in Table 1 (always smaller than one). Some extra investment is thus needed on the "optimal" choice of the three levels to be used in the building of a Generalized Jackknife extremal index estimator. Anyway, the result related to mean value stability around the target value* θ*, for a wide range of* k*-levels, is true for all values of* θ*, and quite relevant in practice.*

In Table 1 we present the main distributional properties of the estimators under study, for this same set of dependent structures.

Remark 4.3. *For these ARMAX sequences we have*

$$\widehat{\theta}_n(k) - \theta \ \stackrel{d}{=}\ \left(\frac{\sigma(\theta)\ P_k}{\sqrt{k}} + \frac{\varphi_1(\theta)\ k}{n} + \frac{\varphi_2(\theta)}{k}\right)(1 + o_p(1)),$$

where P_k *is asymptotically standard normal, and*

$$\sigma^2(\theta) = \theta(1-\theta)/2;\ \varphi_1(\theta) = -\theta(\theta+1)/2;\ \varphi_2(\theta) = (3 - 2\ \theta)/2.$$

If, for this type of sequences, we try to minimize the usual approximation for the asymptotic mean squared error of the estimator in (1.7)*,*

$$MSE_\infty(k) = \frac{\theta(1-\theta)}{k} + \left(\frac{3 - 2\ \theta}{2k} - \frac{\theta(\theta+1)\ k}{2\ n}\right)^2,$$

we get

$$k_{0n}^N \sim \left\{\frac{2(1-\theta)\ n^2}{\theta(1+\theta)^2}\right\}^{1/3} =: k_{ass}^N,$$

and this result is not a long way from the ones presented in Table 1, as may be seen in Table 2, where we present, for the same values of n *as before,* k_{ass}^N/n *and the simulated value of* k_0^N/n *for the ARMAX Fréchet*(1) *sequences with* $\theta = 0.2,\ 0.5$ *and* 0.8.

Table 1: Optimal sample fractions, mean values and mean squared errors of the estimators under study at their optimal levels for $ARMAX$ Fréchet(1) sequences, with $\theta = 0.2,\ 0.5$ and 0.8, together with the three indicators $REFF_0$, BRI_0 and STI in (4.1), (4.2) and (4.3), respectively.

n	100	200	500	1000	2000	5000	10000
$\theta = \mathbf{0.2}$							
k_0^N/n	0.3510	0.2820	0.2132	0.1686	0.1354	0.0997	0.0788
k_0^{GJ}/n	0.9600	0.9760	0.9224	0.8668	0.8052	0.7162	0.6499
E_0^N	0.1721	0.1755	0.1795	0.1831	0.1859	0.1892	0.1913
E_0^{GJ}	0.2246	0.2247	0.2198	0.2158	0.2124	0.2090	0.2069
MSE_0^N	0.0036	0.00253	0.0015	0.00104	0.0007	0.0004	0.0003
MSE_0^{GJ}	0.0145E	0.0076	0.0034	0.0019	0.0010	0.0005	0.0003
$REFF_0$	0.4984	0.5775	0.6743	0.7499	0.8179	0.9065	0.9744
BRI_0	1.1404	1.0031	1.0488	1.0756	1.1394	1.2079	1.2671
STI	2.4857	6.5038	8.3058	8.8092	9.0403	9.0755	9.1332
$\theta = \mathbf{0.5}$							
k_0^N/n	0.1990	0.1590	0.1178	0.0943	0.0743	0.0552	0.0443
k_0^{GJ}/n	0.9600	0.9660	0.8840	0.8248	0.7316	0.6265	0.5668
E_0^N	0.4462	0.4537	0.4629	0.4689	0.4750	0.48081	0.48432
E_0^{GJ}	0.5360	0.5373	0.5287	0.5234	0.5171	0.51101	0.5079
MSE_0^N	0.0125	0.0085	0.0049	0.0033	0.0021	0.0012	0.0008
MSE_0^{GJ}	0.0240	0.0130	0.0061	0.0034	0.0019	0.0009	0.0005
$REFF_0$	0.7225	0.8093	0.9029	0.9824	1.0509	1.1622	1.2582
BRI_0	1.5061	1.2449	1.3041	1.3408	1.4862	1.7742	1.9921
STI	12.1833	17.7367	21.3432	22.0213	22.3163	22.5525	22.9920
$\theta = \mathbf{0.8}$							
k_0^N/n	0.1140	0.0910	0.0680	0.0527	0.0427	0.0312	0.0245
k_0^{GJ}/n	0.9900	0.9780	0.9680	0.8596	0.7928	0.7360	0.6424
E_0^N	0.7383	0.7470	0.7577	0.7661	0.7719	0.7790	0.7832
E_0^{GJ}	0.8229	0.8211	0.8243	0.8166	0.8126	0.8106	0.8082
MSE_0^N	0.0183	0.0118	0.0065	0.0042	0.0027	0.0015	0.0009
MSE_0^{GJ}	0.0186	0.0111	0.0053	0.0030	0.0016	0.0007	0.0004
$REFF_0$	0.9905	1.0330	1.1080	1.1951	1.3008	1.4341	1.5286
BRI_0	2.6933	2.5280	1.7491	2.1133	2.2482	1.9793	2.0786
STI	25.7500	40.3000	46.0357	50.1709	50.0586	49.9929	50.2329

Table 2: Asymptotic and simulated optimal sample fractions for Nandagopalan's estimator.

n	$k_{ass}^N\|\theta = 0.2$	$k_0^N\|\theta = 0.2$	$k_{ass}^N\|\theta = 0.5$	$k_0^N\|\theta = 0.5$	$k_{ass}^N\|\theta = 0.8$	$k_0^N\|\theta = 0.8$
100	0.3816	0.3510	0.2071	0.1990	0.1156	0.1140
200	0.3029	0.2820	0.1644	0.1590	0.0917	0.0910
500	0.2231	0.2132	0.1211	0.1178	0.0676	0.0680
1000	0.1771	0.1686	0.0961	0.0943	0.0536	0.0527
2000	0.1406	0.1357	0.0763	0.0743	0.0426	0.0427
5000	0.1036	0.0988	0.0562	0.0552	0.0314	0.0312
10000	0.0822	0.0787	0.0446	0.0443	0.0249	0.0245

4.2 The i.i.d. framework

Figure 7 is equivalent to Figures 4, 5 and 6, but now for a *Fréchet*(1) i.i.d. parent.

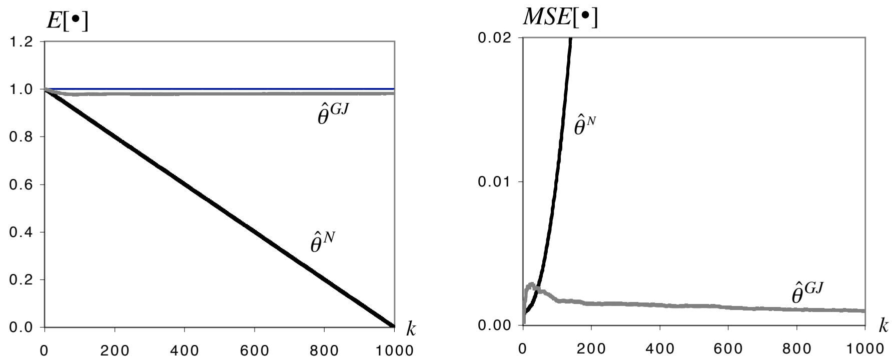

Figure 7: Simulated mean values and *MSE's* of the extremal index estimators under study, for an i.i.d. sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 1$ $(\theta = 1)$.

Remark 4.4. *Again, the mean value stability of the Generalized Jackknife estimator, in this i.i.d. set-up, is incredible and was not at all expected. And even if we choose a logarithmic scale for the k-axis, as suggested for the tail index estimator by Drees et al.* (2000)*, the Generalized Jackknife estimator turns out to be preferable to the classical estimator. Figure 8 is the same as Figure 7, but with a logarithmic k-scale. It is then possible to notice that the mean squared error of the Generalized Jackknife estimator* $\widehat{\theta}_n^{GJ}$ *is smaller than the mean squared error of* $\widehat{\theta}_n^{N}$ *for very small values of* k*, as expected. However, the* MSE *pattern looks a bit strange, in the sense that it is an increasing function of* k*, for small values of* k*.*

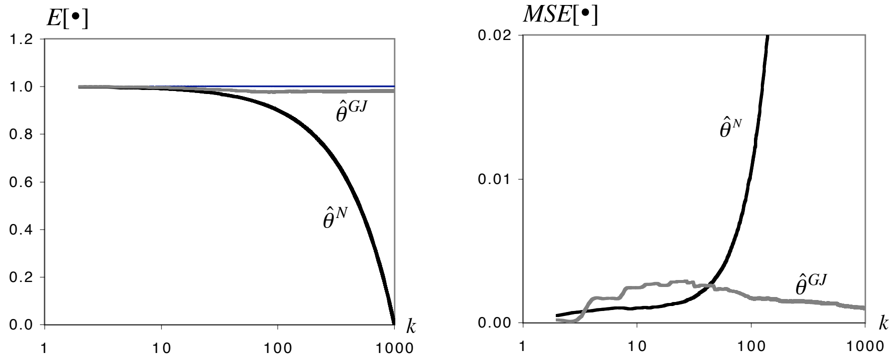

Figure 8: Simulated mean values and *MSE's* of the extremal index estimators under study in a logarithmic k-scale, for an i.i.d. sample size $n = 1000$, from a *Fréchet* parent with $\gamma = 1$ ($\theta = 1$).

5 Effect of the sampling frequency on the extremal index of an $ARMAX$ process

From the papers of Robison and Tawn (2000), Ferreira and Martins (2003), Scotto and Ferreira (2003), Scotto *et al.* (2003) and Martins and Ferreira (2004), we get the following result for stationary sequences under **D** and **D"** conditions: If we consider the sub-sample $\mathbf{V} = \left\{X_{(n-1)T}\right\}_{n\geq 1}$ we have, for levels u such that (1.5) holds,

$$\theta_V = \theta_X + \lim_{n\to\infty} \frac{\sum_{i=0}^{T-2} P(X_0 \leq u < X_1, X_{T-i} > u)}{\tau/n},$$

and consequently, for $ARMAX$ sequences,

$$\theta_{\mathbf{V}} = 1 - (1-\theta_{\mathbf{X}})^T \iff \theta_{\mathbf{X}} = 1 - (1-\theta_{\mathbf{V}})^{1/T}.$$

Subsampling may thus improve the performance of an estimator, through the consideration of averages, which often enable a decrease in variance, and consequently in mean squared error. After the implementation of different subsampling algorithms, we here propose the following algorithm:

- Fix T (possibly $T = 2$), and compute $r = [n/T]$;

- Consider the T subsamples of size r, $\mathbf{V}_i = \big(X_i, X_{T+i}, \cdots, X_{(r-1)T+i}\big)$, for $i = 1, 2, \cdots, T$, and compute the estimates $\widehat{\theta}^{GJ}_{\mathbf{V}_i}(j)$, $j = 1, 2, \ldots, r-1$, with $\widehat{\theta}^{GJ}_{\mathbf{V}}$, the estimator in (3.4) with $\delta = 1/4$, applied to the subsample under consideration;
- Compute $\widehat{\theta}^{GJ}_{sub|T}(k) = 1 - \frac{1}{T}\sum_{i=1}^{T}\left(1 - \widehat{\theta}^{GJ}_{\mathbf{V}_i}(j)\right)^{1/T}$, for thresholds $k = (j-1)T+1, \ldots, j\,T$, $j = 1, 2, \ldots, r-1$.

The use of the previous algorithm in the estimator $\widehat{\theta}^{GJ_0}$, in (3.3), enables the estimator $\widehat{\theta}^{GJ_0}_{sub|T}$ to achieve, at optimal levels, a mean squared error smaller than that of the original estimator $\widehat{\theta}^{GJ_0}$, though in general not smaller than that of the Generalized Jackknife estimator $\widehat{\theta}^{GJ(\delta)}_n$ in (3.4), with $\delta = 1/4$. However, the use of the same algorithm in the later estimator, $\widehat{\theta}^{GJ(1/4)}_n$, did not enable us to achieve that same result for values of $\theta > 0.2$ and not too large values of n, as may be seen from Figure 9. Note however that these are the situations under which the Generalized Jackknife estimator was already able to overpass the original estimator, regarding mean squared errors at optimal leves.

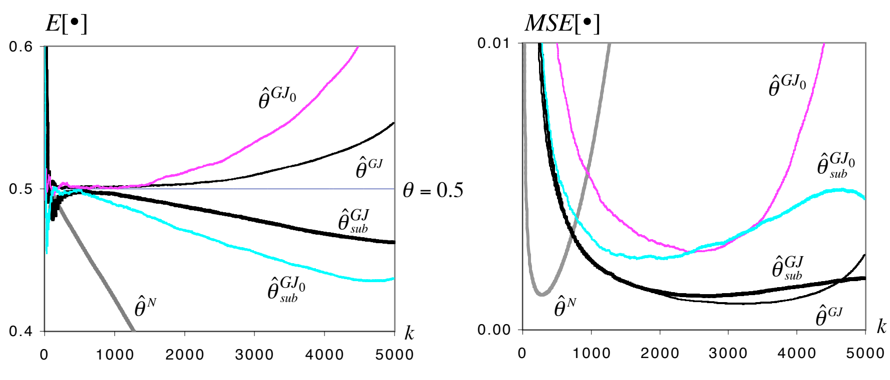

Figure 9: Effect of subsampling in the estimation of the extremal index, for an $ARMAX$ Fréchet(1) sequence, with $\theta = 0.5$, $n = 5000$.

If $\theta \le 0.2$ (here illustrated for $\theta = 0.2$, in Figure 10), we are able to overpass the original estimator at optimal levels, when we consider the Generalized Jackknife statistic in (3.4) with $\delta = 1/4$, together with the use of subsampling techniques with $T = 2, 3$ or 4.

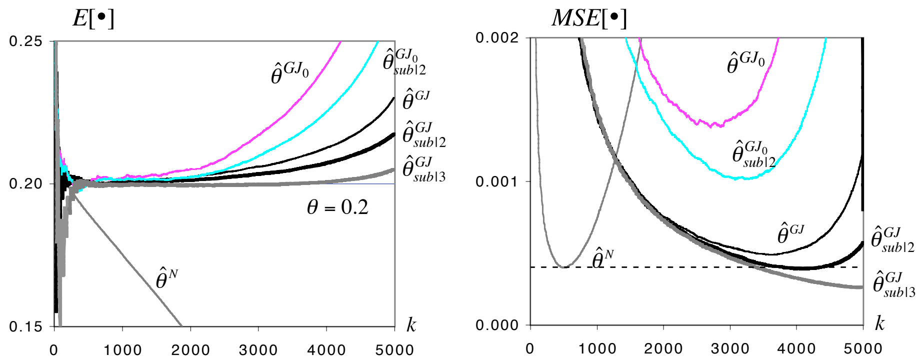

Figure 10: Effect of subsampling in the estimation of the extremal index for an $ARMAX$ Fréchet(1) sequence with $\theta = 0.2$.

In Table 3, for these same $ARMAX$ sequences (with $\theta = 0.2$), we present at each entry four values of the three indicators $REFF$, BRI and STI in (4.1), (4.2) and (4.3), respectively, when we consider in this order, the Generalized Jackknife estimator $\widehat{\theta}^{GJ}$, i.e., the estimator in (3.4) with $\delta = 1/4$, and its three subsampling variants $\widehat{\theta}^{GJ}_{sub|T}$, for $T = 2, 3$ and 4.

Table 3:

n	$REFF_0$	BRI_0	STI
1000	0.7499/0.8340/0.9308/0.9187	1.0756/1.2419/5.3496/1.8271	8.6981/10.4223/9.8730/4.7712
2000	0.8178/0.9106/1.0612/1.0592	1.1420/1.3488/3.3129/1.8682	8.9894/10.6499/11.1869/9.8626
5000	0.9066/1.0138/1.2377/1.2047	1.1845/1.3329/2.5174/1.4783	9.0553/10.6898/11.8725/11.3921
10000	0.9743/1.0110/1.3781/1.2786	1.2698/1.8357/2.6470/1.2543	9.1219/10.2783/12.1436/11.9106

For all values of n and for all indicators considered, we get first an increase, followed by a decrease in their values. The choice $T = 2$ or $T = 3$ seems to be a sensible one. The value of T should however be considered as a *tuning parameter*: the drawing of the sample paths for different values of T will then enable to choose the most stable sample path, according to any stability criterion, like the one in Gomes and Pestana (2005).

6 Some overall conclusions

1. The most attractive feature of the proposed Generalized Jackknife estimator of the extremal index is its stable sample path (for a wide region of k values), close to the target value θ (look at figures shown in sections 4 and 5, where the mean values provide a hint of what happens to the sample paths).

2. The second most attractive feature of the GJ estimators is the wide "bath-tube" patterns of their $MSE(k)$ functions, which make less relevant the choice of the optimal sample fraction k_0/n. Overall and in practice, the "bath-tube" pattern of any of the GJ estimators may make them preferable to the original estimator, even when the $REFF_0^{GJ}$ indicator is smaller than one.

3. Indeed, regarding the mean squared error, the initial Generalized Jackknife estimator in (3.3) does not overpass the original estimator, when both estimators are considered at their optimal levels. To obtain $REFF$ indicators higher than 1, we had to proceed to a different choice of the three levels under play, arriving at the class of estimators in (3.4). Even with such a choice, and whenever θ is small, the $REFF$ indicator becomes greater than one only with the extra use of a subsampling algorithm. An alternative theoretical choice of these three levels is an interesting topic of research, but totally overpasses the aim of this paper.

4. For the first choice in (3.3), the BRI indicator is also sometimes smaller than 1, but the STI indicator is always greater than 1. For the estimator in (3.4) with $\delta = 1/4$ or their subsampling variants with a moderate value of T, both the BRI and the STI indicators are greater than 1 for all simulated models and sample sizes. This insensitivity of the mean value (and sample path) to changes in k is indeed one of the nice features of this new extremal index estimators.

7 A case-study in the field of finance

We shall here consider the performance of the above mentioned estimators in the analysis of Euro-UK Pound daily exchange rates from January 4, 1999, until December 14, 2004. This data has been collected by the European System of Central Banks, and was obtained from *http://www.bportugal.pt/*. In Figure 11 we picture the Daily Exchange Rates x_t over the above mentioned period and the Log-Returns, $r_t = 100 \times (\ln x_t - \ln x_{t-1})$, the data to be analyzed.

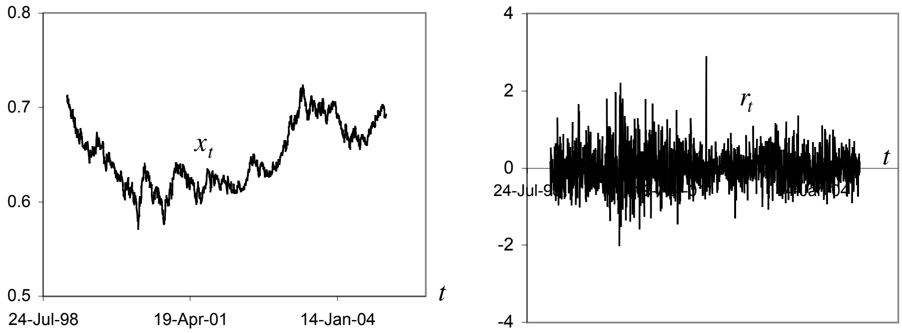

Figure 11: Daily Exchange Rates (*left*) and Daily Log-Returns (*right*) on Euro-UK Pound Exchange Rate.

In Figure 12, working with the $n_0 = 725$ positive log-returns, we now picture the sample paths of $\widehat{\theta}^N(k)$ in (1.7) and $\widehat{\theta}^{GJ}(k)$ in (3.4), as functions of k (*left*), together with those same sample paths as functions of the ascending o.s.'s associated to the log-returns (*right*), which is practically equivalent to produce the left graph in a log-scale, as suggested in Drees *et al.* (2000).

The graphs in Figure 12 enable us to draw the following conclusions:

1. First, the higher stability of the *Generalized Jackknife* estimates, around the value $\theta = 1$, is clear. Indeed, that stability appears both for small and large values of k, whereas for the classical estimates in (1.7), the sample path exhibts a very small region of stability around the value $\theta = 1$.

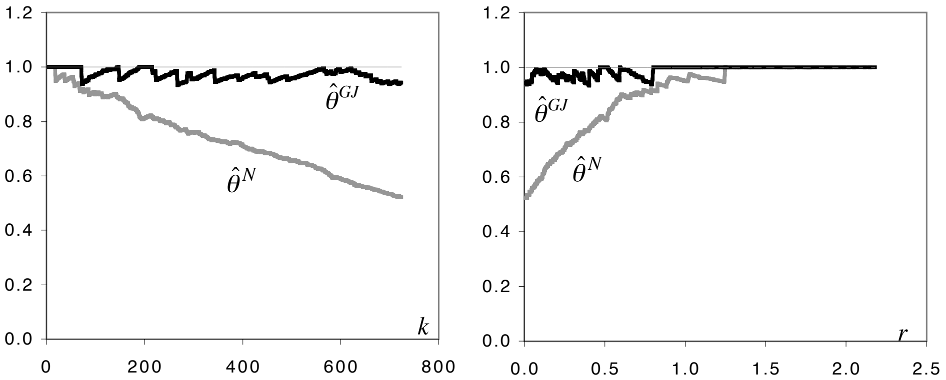

Figure 12: Estimates of the extremal index estimators under study as functions of k (*left*) and of the positive ascending o.s. of the sample of log-returns (*right*) for the Daily Log-Returns on the Euro-UK Pound.

2. The conclusion that the extremal index is equal to 1 enables us to deal with the sample of log-returns as approximately i.i.d., i.e., the inferential procedures related to i.i.d. samples may be applied.

References

[1] Alpuim, M.T. (1989). An extremal markovian sequence. *J. Appl. Probab.* **26**, 219-232.

[2] Casela, G. and R. L. Berger (2002). *Statistical Inference*, 2nd edition. Duxbury Press.

[3] Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.* **28**, 254-274.

[4] Ferreira, H. and Martins, A. P. (2003). The extremal index of sub-sampled periodic sequences. *RevStat* **1**, 15-24.

[5] 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.

[6] Gomes, M. I., Martins, M. J. and Neves, M. (2002). Generalized Jackknife semi-parametric estimators of the tail index. *Portugaliae Mathematica* **59**:4,393-408.

[7] Gomes, M. I. and Miranda, M. C. (2003). A metodologia jackknife na estimação do índice extremal. In Brito, P. et al. (eds.), *Literacia e Estatística*, Edições S.P.E., 299-310 (in portuguese).

[8] Gomes, M. I., Miranda, C. and Pereira, H. (2005). Revisiting the role of the Jackknife methodology in the estimation of a positive tail index. *Comm. in Statistics – Theory and Methods* **34**, 1-20.

[9] Gomes, M. I. and Oliveira, O. (2001) The bootstrap methodology in Statistics of Extremes — choice of the optimal sample fraction. *Extremes* **4**:4, 331-358.

[10] Gomes, M. I. and Pestana, D. (2004). A simple second order reduced bias estimator. To appear at *J. Statist. Comp. and Simul.*

[11] Gray, H. L. and Schucany, W. R. (1972). *The Generalized Jackknife Statistic.* Marcel Dekker.

[12] Hall, A. (1996). Maximum term of a particular autoregressive sequence with discrete margins. *Comm. in Statistics – Theory and Methods* **25**: 4, 721-736.

[13] Hsing, T. and Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probabb. Th. and Related Fields* **78**, 97-112.

[14] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Series.* Springer-Verlag, New York.

[15] Leadbetter, M.R. and Nandagopalan, S. (1989). On exceedance point processes for stationary sequences under mild oscillation restrictions. In Hüsler, J. and R.-D. Reiss (eds.), *Extreme Value Theory*, Springer-Verlag, 69-80.

[16] Martins, A. P. and Ferreira, H. (2004). The extremal index of sub-sampled processes. *J. Statist. Planning and Inference* **124**, 145-152.

[17] Nandagopalan, S. (1990). *Multivariate Extremes and Estimation of the Extremal Index.* Ph.D. Thesis, Univ. North Carolina at Chapel Hill.

[18] Robinson, M. E. and Tawn, J. A. (2000). Extremal analysis of processes sampled at different frequencies. *J. Royal Statist. Soc.* **B 62**, 117-135.

[19] Scotto, M, and Ferreira, H. (2003). Extremes of deterministic sub-sampled moving averages with heavy-tailed innovations. *Appl. Stochastic Models Bus. Ind.* **19**, 303-313.

[20] Scotto, M. Turkman, K. F. and Anderson, C. W. (2003). Extremes of some sub-sampled time series. *J. Time Series Analysis* **24**, 5, 505-512.