# High Quantile Estimation and the PORT Methodology*

Lígia Henriques Rodrigues

Instituto Politécnico de Tomar and CEAUL

E-mail: Ligia.Rodrigues@aim.estt.ipt.pt

M. Ivette Gomes

Universidade de Lisboa, DEIO and CEAUL

E-mail: ivette.gomes@fc.ul.pt

December 10, 2008

**Abstract:** In many areas of application, a typical requirement is to estimate a *high quantile* $\chi_{1-p}$ of probability $1-p$, a value, high enough, so that the chance of an exceedance of that value is equal to $p$, small. The semi-parametric estimation of high quantiles depends not only on the estimation of the tail index $\gamma$, the primary parameter of extreme events, but also on an adequate estimation of a scale first order parameter. The great majority of semi-parametric quantile estimators in the literature, do not enjoy the adequate behaviour, in the sense that they do not suffer the appropriate linear shift in the presence of linear transformations of the data. Recently, and for heavy tails ($\gamma > 0$), a new class of quantile estimators was introduced with such a behaviour. They were named PORT-quantile estimators, with PORT standing for *peaks over random threshold*. In this paper, also for heavy tails, we introduce a new class of PORT-quantile estimators with the above mentioned behaviour, using the PORT methodology and incorporating Hill and moment PORT-classes of *tail index* estimators in one of most recent classes of quantile estimators suggested in the literature. Under convenient restrictions on the underlying model, these classes of estimators are consistent and asymptotically normal for adequate $k$, the number of top order statistics used in the semi-parametric estimation of $\chi_{1-p}$.

*Research partially financed by FCT / POCTI, POCI and PPCDT / FEDER.

**AMS 2000 subject classification**. Primary 62G32, 62E30; Secondary 65C05.

**Keywords and phrases**. *Heavy tails*, *high quantiles*, *semi-parametric estimation*, *extremes*, *PORT methodology*, *asymptotic behaviour*.

## 1 Introduction

A model $F$ is said to have a heavy right-tail whenever the right *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation $\alpha = -1/\gamma$, i.e., for every $x > 0$, $\lim_{t\to\infty} \overline{F}(tx)/\overline{F}(t) = x^{-1/\gamma}$. Then we are in the domain of attraction for maxima of an *extreme value* ($EV$) distribution function (d.f.),

$$EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \qquad x > -1/\gamma, \quad \gamma > 0,$$

and we write $F \in \mathcal{D}(EV_{\gamma>0})$. The parameter $\gamma$ is the *tail index*, one of the primary parameters of rare events.

In a context of heavy tails, and with the notation $U(t) = F^{\leftarrow}(1-1/t),\ t \geq 1$, $F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$ the generalized inverse function of the underlying model $F$, the first order parameter (or tail index) $\gamma$ ($> 0$) appears, for every $x > 0$, as the limiting value, as $t \to \infty$, of the quotient $(\ln U(tx) - \ln U(t))/\ln x$ (de Haan, 1970). Indeed, with the usual notation $RV_\alpha$ for the class of regularly varying functions with an index of regularly variation $\alpha$, i.e., positive measurable functions $g$ such that $g(tx)/g(t) \to x^\alpha$, as $t \to \infty$ and for all $x > 0$, we can further say

$$F \in \mathcal{D}(EV_{\gamma>0}) \quad \text{iff} \quad U \in RV_\gamma \quad \text{iff} \quad 1 - F \in RV_{-1/\gamma} \text{ (Gnedenko, 1943).} \tag{1.1}$$

Heavy-tailed distributions have recently been accepted as realistic models for various phenomena in economics, ecology, bibliometrics and biometry, among others. See, for instance, the recent books on the topic by Markovich (2007) and Resnick (2007).

For small values of $p$, we want to extrapolate beyond the sample, estimating a typical parameter in many areas of application, a high quantile $\chi_{1-p}$, i.e., a value such that $F(\chi_{1-p}) = 1 - p$ or equivalently

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

We are going to base inference on the largest $k+1$ order statistics (o.s.), and as usual in semi-parametric estimation of parameters of extreme events, we shall assume that $k$ is an *intermediate* sequence of integers in $[1, n[$, i.e.,

$$k = k_n \to \infty, \quad k/n \to 0, \quad \text{as} \quad n \to \infty. \tag{1.3}$$

In order to derive the asymptotic non-degenerate behaviour of semi-parametric estimators of parameters of extreme events, we need more than the first-order condition, $U \in RV_\gamma$ in (1.1). A convenient condition is the following second-order condition, which guarantees that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \tag{1.4}$$

which we assume to hold for every $x > 0$, being $\rho \leq 0$ the "shape" or, more properly, the generalised shape second order parameter. The limit function in (1.4) is necessarily of this given form and $|A| \in RV_\rho$ (Geluk and de Haan, 1987). Sometimes, only for the sake of simplicity, we shall assume to be working in a sub-class of Hall-Welsh's class of models (Hall and Welsh, 1985), where there exist $\gamma > 0$, $\rho < 0$, $C > 0$ and $\beta \neq 0$, such that, as $t \to \infty$,

$$U(t) = C\, t^\gamma \Big(1 + \frac{A(t)}{\rho}(1 + o(1))\Big), \quad \text{with} \quad A(t) = \gamma\beta t^\rho. \tag{1.5}$$

Typical heavy-tailed models, such as the Fréchet, the Generalized Pareto and the Student-$t_\nu$ belong to such a class. Then, the second-order condition in equation (1.4) holds, with $A(t) = \gamma\beta t^\rho$, $\beta \neq 0$, $\rho < 0$. The parameters $\beta$ and $\rho$ are the so-called generalised scale and shape second-order parameters, respectively.

Condition (1.1) is equivalent to $\lim_{t\to\infty} U(tx)/U(t) = x^\gamma$. Then $U(t) \sim Ct^\gamma$, as $t \to \infty$, for a real $C > 0$, and from (1.2), we have

$$\chi_{1-p} = U(1/p) \;\sim\; Cp^{-\gamma}, \quad \text{as} \quad p \to 0.$$

An obvious estimator of $\chi_{1-p}$ is thus $\widehat{C}p^{-\hat\gamma}$, with $\widehat{C}$ and $\hat\gamma$ any consistent estimators of $C$ and $\gamma$, respectively.

Given a sample $(X_1, X_2, \ldots, X_n)$, let us denote $(X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n})$ the set of associated ascending o.s. Denoting $Y$ a standard Pareto model, i.e., a model such that $F_Y(y) = 1 - 1/y$, $y > 1$, the universal uniform transformation and the fact that $Y_{n-k:n} \overset{p}{\sim} (n/k)$ for intermediate $k$, enables us to write $X_{n-k:n} \overset{p}{\sim} C(n/k)^\gamma$, as $n \to \infty$. Consequently, an obvious estimator of $C$, proposed in Hall (1982), is

$$\widehat{C} \equiv C_{k,n,\hat{\gamma}} := X_{n-k:n}(k/n)^{\hat{\gamma}}$$

and

$$Q_{k,p_n,\hat{\gamma}} = \widehat{C}\ p_n^{-\hat{\gamma}} = X_{n-k:n}\,(k/np_n)^{\hat{\gamma}}$$

is the obvious quantile-estimator at the level $p$ (Weissman, 1978). The semi-parametric estimation of high quantiles depends thus strongly on the estimation of the tail index $\gamma$, the primary parameter of extreme events.

In the classical approach, we often consider for $\hat{\gamma}$ either the Hill estimator (Hill, 1975) or the moment estimator (Dekkers *et al.*, 1989), both based on the $k+1$ top o.s., denoted $\underline{X}_k := (X_{n-k:n}, \cdots, X_{n:n})$. The Hill estimator is the average of the log-excesses,

$$H_{k,n} \equiv H_n(\underline{X}_k) \equiv \widehat{\gamma}_{k,n,H} := \frac{1}{k}\sum_{i=1}^{k}\left(\ln X_{n-i+1:n} - \ln X_{n-k:n}\right), \tag{1.6}$$

and the moment estimator has the functional expression,

$$M_{k,n} \equiv M_n(\underline{X}_k) \equiv \widehat{\gamma}_{k,n,M} := M_{k,n}^{(1)} + 1 - \frac{1}{2}\left\{1 - \left(M_{k,n}^{(1)}\right)^2/M_{k,n}^{(2)}\right\}^{-1}, \tag{1.7}$$

with $M_{k,n}^{(\alpha)}$ defined by

$$M_{k,n}^{(\alpha)} \equiv M_n^{(\alpha)}(\underline{X}_k) := \frac{1}{k}\sum_{i=1}^{k}\left(\ln X_{n-i+1:n} - \ln X_{n-k:n}\right)^{\alpha}, \quad \alpha = 1, 2. \tag{1.8}$$

Under the second order framework in (1.4) and for any intermediate sequence $k$, i.e. whenever (1.3) holds, we have for the Hill estimator $H$, in (1.6), and for the moment estimator $M$, in (1.7), generally denoted by $T$, the validity of following asymptotic distributional representation,

$$\widehat{\gamma}_{k,n,T} \overset{d}{=} \gamma + \frac{\sigma_T\ P_{k,T}}{\sqrt{k}} + b_T\ A(n/k)(1 + o_p(1)), \tag{1.9}$$

where $P_{k,T}$ is asymptotically standard normal and

$$\sigma_{_H}^2 := \gamma^2, \quad b_{_H} := \frac{1}{1-\rho}, \quad \sigma_{_M}^2 := 1+\gamma^2 \quad \text{and} \quad b_{_M} := \frac{\gamma(1-\rho)+\rho}{\gamma(1-\rho)^2}. \tag{1.10}$$

Most of the semi-parametric quantile estimators in the literature, like the ones in Gomes and Figueiredo (2006), Gomes and Pestana (2007), Beirlant *et al.* (2008), Caeiro and Gomes (2008), as well as in other papers on semi-parametric quantile estimation prior to 2005, do not enjoy the adequate behaviour in the presence of linear transformations of the data, a behaviour related with the fact that for any quantile $\chi_{1-p}$ we have

$$\chi_{1-p}(\lambda+\delta X) = \lambda + \delta\chi_{1-p}(X) \tag{1.11}$$

for any model $X$, real $\lambda$ and positive $\delta$.

Recently, and for $\gamma > 0$, Araújo Santos *et al.* (2006) provided quantile estimators with the linear property in (1.11), based upon a sample of excesses over a random threshold $X_{n_q:n}$, denoted

$$\underline{X}^{(q)} := \big(X_{n:n} - X_{n_q:n}, \cdots, X_{n_q+1:n} - X_{n_q:n}\big), \tag{1.12}$$

where $X_{i:n}$, $1 \le i \le n$, is, as mentioned before, the sample of ascending o.s. associated to the random sample $(X_1, X_2, \ldots, X_n)$ and $n_q := [nq]+1$, with:

- $0 < q < 1$, for d.f.'s with finite or infinite left endpoint $x_{_F} := \inf\{x : F(x) > 0\}$ (*the random threshold $X_{n_q:n}$ is an empirical quantile*);
- $q = 0$, for d.f.'s with finite left endpoint $x_{_F}$ (*the random threshold is the minimum,* $X_{1:n}$).

Such estimators were named PORT-quantile estimators, with PORT standing for *peaks over random threshold*, and are based on the PORT-Hill and PORT-moment estimators, generically denoted $T(q) \equiv T_{k,n}(q) := T_n(\underline{X}^{(q)})$ for $T = H$ or $M$, $k < n - n_q$, and with $H_n(\underline{X}_k)$ and $M_n(\underline{X}_k)$ provided in (1.6) and (1.7), respectively. They are given by

$$\widetilde{Q}_{k,p_n,T(q)} := \big(X_{n-k:n} - X_{n_q:n}\big)\left(\frac{k}{np_n}\right)^{T(q)} + X_{n_q:n}, \tag{1.13}$$

where $T(q)$ can more generally be any consistent estimator of the tail index $\gamma$, made location/scale invariant by using any of the transformed samples $\underline{X}^{(q)}$, in (1.12). The PORT-Hill and the PORT-moment estimators have been studied by simulation in Gomes *et al.* (2008).

The class of estimators suggested here is also a function of the sample of the excesses $\underline{X}^{(q)}$ in (1.12). We use the PORT methodology and incorporate Hill and moment PORT-classes of *tail index* estimators in one of the classes of quantile estimators suggested in Caeiro and Gomes (2008), slightly modified in order to satisfy the linear property in (1.11). More specifically, we shall consider quantile estimators of the type,

$$\widetilde{Q}_{k,p_n,T(q)} := \frac{X_{n-[k/2]:n} - X_{n-k:n}}{2^{T(q)} - 1} \left(\frac{k}{n\,p_n}\right)^{T(q)} + X_{n_q:n}. \tag{1.14}$$

Under convenient restrictions on the underlying model, these classes of estimators are consistent and asymptotically normal for adequate $k$, the number of top o.s. used in the semi-parametric estimation of $\chi_{1-p}$.

In Section 2, we shall present a few introductory technical details and asymptotic preliminary results associated with the PORT methodology. The asymptotic behaviour of the PORT-classes of *tail index* estimators under study, together with the asymptotic comparison of the PORT-Hill and the PORT-moment estimators at optimal levels, will be derived in Section 3. Finally, in Section 4, we derive the asymptotic behaviour of the new classes of PORT-quantile estimators.

# 2 Technical preliminaries related with the PORT methodology

## 2.1 The second order framework for heavy-tailed models with a real shift

If we introduce a deterministic shift, i.e. a new location, $\lambda$, in the underlying model $X$, with quantile function $U_X(t)$, the transformed random variable (r.v.) $Y = X + \lambda$ has an associated quantile function given by $U_s(t) \equiv U_Y(t) = U_X(t) + \lambda$ and condition (1.4) can be rewritten as

$$\lim_{t\to\infty} \frac{\ln U_s(tx) - \ln U_s(t) - \gamma \ln x}{A_s(t)} = \frac{x^{\rho_s} - 1}{\rho_s}, \tag{2.1}$$

for all $x > 0$, with $|A_s| \in RV_{\rho_s}$.

Let $F$ be a model with quantile function $U(t) \equiv U_X(t)$, given in (1.5). Then

$$U_Y(t) = Ct^\gamma \left(1 + \frac{A(t)}{\rho} + \lambda C^{-1} t^{-\gamma} + o(t^\rho)\right), \quad \text{as} \quad t \to \infty.$$

Therefore both $U_s(t) = U_Y(t)$ and $U(t) = U_X(t)$ are asymptotically equivalent to $Ct^\gamma$, but

$$\rho_s = \begin{cases} \rho & \text{if} \quad \rho > -\gamma \\ -\gamma & \text{if} \quad \rho \le -\gamma. \end{cases}$$

The function $A_s(t)$ in (2.1) can be chosen as

$$A_s(t) := \begin{cases} -\frac{\gamma\lambda}{U(t)}, & \text{if} \quad \rho < -\gamma \\ A(t) - \frac{\gamma\lambda}{U(t)}, & \text{if} \quad \rho = -\gamma \\ A(t), & \text{if} \quad \rho > -\gamma. \end{cases}$$

## 2.2 Asymptotic preliminary results in the PORT methodology

In this subsection we begin with the presentation of the asymptotic results for the statistics $M_{k,n}^{(\alpha,q)} \equiv M_n^{(\alpha,q)} = M_n^{(\alpha)}\big(\underline{X}^{(q)}\big)$, $k < n - n_q$, based on the sample of excesses $\underline{X}^{(q)}$, $0 \le q < 1$, in (1.12) and with $M_n^{(\alpha)}(\underline{X}_k)$ provided in (1.8).

In the following, $\chi_q^*$ denotes the $q$-quantile of $F$: $F(\chi_q^*) = q$ (by convention $\chi_0^* = x_F$) so that,

$$X_{n_q:n} \underset{n\to\infty}{\overset{p}{\longrightarrow}} \chi_q^* \qquad \text{for} \qquad 0 \le q < 1. \tag{2.2}$$

We present, without proof, the following Lemma:

**Lemma 2.1** (Araújo Santos *et al.*, 2006)**.** *If the second order condition* (1.4) *holds, if* $k = k_n$ *is an intermediate sequence, i.e.* (1.3) *holds, for any real* $q$*,* $0 \le q < 1$*, with* $F(\chi_q^*) = q$ $(\chi_0^* = x_F$*, finite), and for* $\alpha = 1,\ 2$*,*

$$M_{k,n}^{(\alpha,q)} - \frac{1}{k}\sum_{i=1}^{k}\Big(\ln \frac{X_{n-i+1:n} - \chi_q^*}{X_{n-k:n} - \chi_q^*}\Big)^\alpha = o_p\Big(\frac{1}{U(n/k)}\Big).$$

**Remark 2.1.** *Note that if* $q \in (0,1)$, $X_{n_q:n} - \chi_q^* = O_p(1/\sqrt{n}\ )$ *and we can assure that* $\sqrt{k}\left[M_{k,n}^{(\alpha,q)} - \frac{1}{k}\sum_{i=1}^{k}\left(\ln \frac{X_{n-i+1:n}-\chi_q^*}{X_{n-k:n}-\chi_q^*}\right)^{\alpha}\right] = O_p\left(\sqrt{k/n}/U(n/k)\right) = o_p(1)$, *for* $\alpha = 1,\ 2$.

**Proposition 2.1.** *If the second order condition* (1.4) *holds and* $k = k_n$ *is an intermediate sequence, i.e.* (1.3) *holds, the statistics* $M_{k,n}^{(\alpha,q)} = M_n^{(\alpha)}(\underline{X}^{(q)})$, *with* $k < n - n_q$, $M_n^{(\alpha)}(\underline{X}_k)$ *given in* (1.8) *and* $\underline{X}^{(q)}$ *given in* (1.12), *satisfy for* $\alpha = 1, 2$,

$$M_{k,n}^{(1,q)} \stackrel{d}{=} M_{k,n}^{(1)} + \frac{\gamma\ \chi_q^*}{(1+\gamma)\ U(n/k)}\ (1 + o_p(1)), \tag{2.3}$$

$$M_{k,n}^{(2,q)} \stackrel{d}{=} M_{k,n}^{(2)} + \frac{2\gamma^2\ (2+\gamma)\ \chi_q^*}{(1+\gamma)^2\ U(n/k)}\ (1 + o_p(1)). \tag{2.4}$$

*Proof.* The first moment of the log-excesses can be rewritten as

$$M_{k,n}^{(1,q)} = \frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{X_{n-i+1:n} - X_{n_q:n}}{X_{n-k:n} - X_{n_q:n}}\right) = M_{k,n}^{(1)} + \frac{1}{k}\sum_{i=1}^{k}\ln\left(\frac{1 - \frac{X_{n_q:n}}{X_{n-i+1:n}}}{1 - \frac{X_{n_q:n}}{X_{n-k:n}}}\right).$$

Since $\ln(1+x) \sim x$, as $x \to 0$,

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k}\ln\left(\frac{1 - \frac{X_{n_q:n}}{X_{n-i+1:n}}}{1 - \frac{X_{n_q:n}}{X_{n-k:n}}}\right) &\sim \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n_q:n}}{X_{n-k:n}} - \frac{X_{n_q:n}}{X_{n-i+1:n}}\right) \\ &= \frac{1}{k}\sum_{i=1}^{k}\left[\frac{X_{n_q:n}}{X_{n-k:n}}\left(1 - \frac{X_{n-k:n}}{X_{n-i+1:n}}\right)\right].\end{aligned}$$

If $k = k_n$ is intermediate, i.e. (1.3) holds, and $\{Y_i\}_{i=1,\ldots,k}$ is a sequence of independent and identically distributed (i.i.d.) standard Pareto r.v.'s, then $Y_{n-k:n} \stackrel{p}{\sim} (n/k)$ and,

$$\begin{aligned}M_{k,n}^{(1,q)} &= M_{k,n}^{(1)} + \frac{\chi_q^*}{U(n/k)}\frac{1}{k}\sum_{i=1}^{k}\left(1 - \frac{X_{n-k:n}}{X_{n-i+1:n}}\right) \\ &\stackrel{d}{=} M_{k,n}^{(1)} + \frac{\chi_q^*}{U(n/k)}\frac{1}{k}\sum_{i=1}^{k}\left(1 - Y_i^{-\gamma}\right)(1 + o_p(1)).\end{aligned}$$

Given that $\mathbb{E}[Y^{-\gamma}] = 1/(1+\gamma)$ and by the weak law of large numbers we get (2.3).

For $\alpha = 2$ and using similar developments, we have

$$M_{k,n}^{(2,q)} \stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\left[\ln\frac{X_{n-i+1:n}}{X_{n-k:n}} + \frac{\chi_q^*}{U(n/k)}\left(1 - Y_i^{-\gamma}\right)(1+o_p(1))\right]^2$$

$$\stackrel{d}{=} M_{k,n}^{(2)} + \frac{2\chi_q^*}{U(n/k)}\frac{1}{k}\sum_{i=1}^{k}\left(\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)\left(1 - Y_i^{-\gamma}\right)(1+o_p(1))$$

$$\stackrel{d}{=} M_{k,n}^{(2)} + \frac{2\chi_q^*}{U(n/k)}\Big[M_{k,n}^{(1)} - \frac{1}{k}\sum_{i=1}^{k}\big(\gamma\ \ln Y_i\ Y_i^{-\gamma}\big)(1+o_p(1))\Big].$$

Since $\mathbb{E}[\ln Y\ Y^{-\gamma}] = 1/(1+\gamma)^2$ and by the weak law of large numbers we get (2.4). □

It has been proved in Martins (2000) (see also, Gomes and Martins, 2001) that, under the second order framework, (1.4), and for levels $k$ such that (1.3) holds, we get for $M_{k,n}^{(\alpha)} = M_n^{(\alpha)}(\underline{X}_k)$, in (1.8), an asymptotic distributional representation of the type,

$$M_{k,n}^{(1)} \stackrel{d}{=} \gamma + \frac{\gamma Z_k^{(1)}}{\sqrt{k}} + \frac{A(n/k)}{1-\rho}(1+o_p(1)), \tag{2.5}$$

$$M_{k,n}^{(2)} \stackrel{d}{=} 2\gamma^2 + \frac{2\sqrt{5}\gamma^2 Z_k^{(2)}}{\sqrt{k}} + \frac{2\gamma\left(1-(1-\rho)^2\right)}{\rho(1-\rho)^2}A(n/k)(1+o_p(1)), \tag{2.6}$$

where, with $\{E_i\}_{i\geq 1}$ a sequence of i.i.d. standard exponencial r.v.'s,

$$Z_k^{(\alpha)} = \frac{\sqrt{k}}{\sqrt{\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)}}\Big(\frac{1}{k}\sum_{i=1}^{k}E_i^\alpha - \Gamma(\alpha+1)\Big), \quad \alpha = 1, 2, \tag{2.7}$$

is asymptotically standard normal. The covariance structure of $Z_k^{(\alpha)}$ is given by

$$\mathbb{C}ov\left(Z_k^\alpha, Z_k^\beta\right) = \frac{\Gamma(\alpha+\beta+1) - \Gamma(\alpha+1)\Gamma(\beta+1)}{\sqrt{\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)}\sqrt{\Gamma(2\beta+1) - \Gamma^2(\beta+1)}}. \tag{2.8}$$

# 3 Asymptotic behaviour of the PORT-classes of tail index estimators

In this section we present, under the validity of the second order condition in (1.4), the asymptotic distributional representations of the PORT-Hill estimators, $H_{k,n}(q) :=$

$H_n(\underline{X}^{(q)})$, and the PORT-moment estimators, $M_{k,n}(q) := M_n(\underline{X}^{(q)})$, with functional expressions given by

$$
\begin{aligned}
H_{k,n}(q) \equiv \hat{\gamma}_{k,n,H(q)} &= \frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{X_{n-i+1:n} - X_{n_q:n}}{X_{n-k:n} - X_{n_q:n}}\right) \\
\text{and} & \\
M_{k,n}(q) \equiv \hat{\gamma}_{k,n,M(q)} &= M_{k,n}^{(1,q)} + 1 - \frac{1}{2}\left\{1 - \left(M_{k,n}^{(1,q)}\right)^2 / M_{k,n}^{(2,q)}\right\}^{-1},
\end{aligned}
$$

respectively, $k < n - n_q$, $M_{k,n}^{(\alpha,q)} = M_n^{(\alpha)}\big(\underline{X}^{(q)}\big)$, $M_n^{(\alpha)}(\underline{X}_k)$ and $\underline{X}^{(q)}$ provided in (1.8) and (1.12), respectively.

The following theorem has been proved in Araújo Santos *et al.* (2006).

**Theorem 3.1** (Araújo Santos *et al.*, 2006)**.** *If the second order condition* (1.4) *holds, $k = k_n$ is an intermediate sequence of positive integers, i.e.* (1.3) *holds, and for any real q, $0 \le q < 1$, we have for $T_{k,n}(q)$, with $T$ denoting either $H$ or $M$, an asymptotic distributional representation of the type,*

$$T_{k,n}(q) \stackrel{d}{=} \gamma + \frac{\sigma_T \; P_{k,T}}{\sqrt{k}} + \left(b_T \; A(n/k) + c_T \; \frac{\chi_q^*}{U(n/k)}\right)(1 + o_p(1)), \tag{3.1}$$

*where $P_{k,T}$ is asymptotically standard normal, $P_{k,H} = Z_k^{(1)}$ defined in* (2.7)*, with $\sigma_T^2$ and $b_T$ provided in* (1.10)*, and*

$$c_H := \frac{\gamma}{1+\gamma}, \qquad c_M := \frac{\gamma^2}{(1+\gamma)^2}. \tag{3.2}$$

**Corollary 3.1.** *Under the conditions of Theorem* 3.1*, the following results hold:*

- *For values of $\gamma + \rho < 0$ and $\chi_q^* \neq 0$,*

  $$T_{k,n}(q) \stackrel{d}{=} \gamma + \frac{\sigma_T \; P_{k,T}}{\sqrt{k}} + c_T \; \frac{\chi_q^*}{U(n/k)}(1 + o_p(1)).$$

  *If $\sqrt{k}/U(n/k) \to \lambda_1$ finite, then*

  $$\sqrt{k}\,(T_{k,n}(q) - \gamma) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal(\lambda_1 c_T \; \chi_q^*, \sigma_T^2).$$

- *For values of* $\gamma+\rho>0$ *or* $\gamma+\rho\le 0$ *and* $\chi_q^*=0$,

$$T_{k,n}(q) \;\overset{d}{=}\; \gamma + \frac{\sigma_{_T}\ P_{k,T}}{\sqrt{k}} + b_{_T}\ A(n/k)(1+o_p(1)).$$

*If* $\sqrt{k}A(n/k)\to\lambda_2$ *finite, then*

$$\sqrt{k}\,(T_{k,n}(q)-\gamma) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal(\lambda_2 b_{_T}, \sigma_{_T}^2).$$

- *For values of* $\gamma+\rho=0\ \wedge\ \chi_q^*\neq 0$,

$$T_{k,n}(q) \;\overset{d}{=}\; \gamma + \frac{\sigma_{_T}\ P_{k,T}}{\sqrt{k}} + \left(b_{_T}\ A(n/k) + c_{_T}\ \frac{\chi_q^*}{U(n/k)}\right)(1+o_p(1)).$$

*If* $\sqrt{k}/U(n/k)\to\lambda_1$ *and* $\sqrt{k}A(n/k)\to\lambda_2$*, with* $\lambda_1$ *and* $\lambda_2$ *both finite, then*

$$\sqrt{k}\,(T_{k,n}(q)-\gamma) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal(\lambda_1\ c_{_T}\ \chi_q^* + \lambda_2 b_{_T}, \sigma_{_T}^2).$$

### 3.1 Asymptotic comparison at optimal levels

We now proceed to an asymptotic comparison of the estimators at their optimal levels in the lines of the Haan and Peng (1998), Gomes and Martins (2001), Gomes *et al.* (2005, 2007) and Gomes and Neves (2008). Suppose that $\widehat{\gamma}_{k,n,\bullet(q)}$, now denoted $\widehat{\gamma}_{\bullet(q)}(k)$, is a general semi-parametric PORT-tail index estimator, with distributional representation,

$$\widehat{\gamma}_{\bullet(q)}(k) = \gamma + \frac{\sigma_\bullet}{\sqrt{k}}\ P_{k,\bullet} + \left(b_\bullet\ A(n/k) + c_\bullet\ \frac{\chi_q^*}{U(n/k)}\right)(1+o_p(1)), \tag{3.3}$$

which hold for any intermediate $k$, and where $P_{k,\bullet}$ is an asymptotically standard normal r.v. Given the results presented in Corollary 3.1, the Asymptotic Mean Square Error (AMSE) of $\widehat{\gamma}_{\bullet(q)}(k)$ is

$$AMSE[\widehat{\gamma}_{\bullet(q)}(k)] := \begin{cases} \frac{\sigma_\bullet^2}{k} + c_\bullet^2\ \frac{(\chi_q^*)^2}{U^2(n/k)}, & \text{if}\quad \rho<-\gamma\ \wedge\ \chi_q^*\neq 0 \\ \frac{\sigma_\bullet^2}{k} + b_\bullet^2\ A^2(n/k), & \text{if}\quad \rho>-\gamma\ \vee\ (\rho\le-\gamma\wedge\chi_q^*=0) \\ \frac{\sigma_\bullet^2}{k} + \left(b_\bullet + c_\bullet \frac{\chi_q^*}{C\gamma\beta}\right)^2 A^2(n/k), & \text{if}\quad \rho=-\gamma\ \wedge\ \chi_q^*\neq 0, \end{cases} \tag{3.4}$$

where $Var_\infty[\widehat{\gamma}_{\bullet(q)}(k)] := \sigma_\bullet^2/k$ and

$$Bias_\infty[\widehat{\gamma}_{\bullet(q)}(k)] :=$$
$$\begin{cases} c_\bullet \, \frac{\chi_q^*}{U(n/k)} =: d_\bullet^{(1)}/U(n/k), & \text{if} \quad \rho < -\gamma \ \wedge \ \chi_q^* \neq 0 \\ b_\bullet \ A(n/k) =: d_\bullet^{(2)} \ A(n/k), & \text{if} \quad \rho > -\gamma \ \vee \ (\rho \leq -\gamma \wedge \chi_q^* = 0) \\ \left(b_\bullet + c_\bullet \frac{\chi_q^*}{C\gamma\beta}\right) A(n/k) =: d_\bullet^{(3)} \ A(n/k), & \text{if} \quad \rho = -\gamma \ \wedge \ \chi_q^* \neq 0, \end{cases}$$

with $d_\bullet^{(1)} = c_\bullet \ \chi_q^*$, $d_\bullet^{(2)} = b_\bullet$ and $d_\bullet^{(3)} = b_\bullet + c_\bullet \chi_q^*/(C\gamma\beta)$.

Let $k_{0,\bullet(q)}$:=arg inf$_k$ $AMSE[\widehat{\gamma}_{\bullet(q)}(k)]$ be the so-called optimal level for the estimation of $\gamma$ through $\widehat{\gamma}_{\bullet(q)}(k)$, i.e., the level associated with a minimum asymptotic mean square error, and let us denote $\widehat{\gamma}_{n0,\bullet(q)} := \widehat{\gamma}_{\bullet(q)}(k_{0,\bullet(q)})$, the estimator computed at its optimal level. The use of regular variation theory enables us to prove that, whenever $d_\bullet^{(i)} \neq 0$, $i = 1, 2, 3$, there exists a function $\varphi(n) = \varphi(n; \rho, \gamma)$, dependent only on the underlying model, and not on the estimator, such that

$$\begin{aligned} LMSE[\widehat{\gamma}_{n0,\bullet(q)}] \ &:= \ \lim_{n\to\infty} \varphi(n) AMSE[\widehat{\gamma}_{n0,\bullet(q)}] \\ &= \begin{cases} \frac{1+2\gamma}{2\gamma} \left(\sigma_\bullet^2\right)^{\frac{2\gamma}{1+2\gamma}} \left((d_\bullet^{(1)})^2\right)^{\frac{1}{1+2\gamma}}, & \text{if} \quad \rho < -\gamma \ \wedge \ \chi_q^* \neq 0 \\ \frac{2\rho-1}{2\rho} \left(\sigma_\bullet^2\right)^{-\frac{2\rho}{1-2\rho}} \left((d_\bullet^{(2)})^2\right)^{\frac{1}{1-2\rho}}, & \text{if} \quad \rho > -\gamma \ \vee \ (\rho \leq -\gamma \wedge \chi_q^* = 0) \\ \frac{2\rho-1}{2\rho} \left(\sigma_\bullet^2\right)^{-\frac{2\rho}{1-2\rho}} \left((d_\bullet^{(3)})^2\right)^{\frac{1}{1-2\rho}}, & \text{if} \quad \rho = -\gamma \ \wedge \ \chi_q^* \neq 0 \end{cases} \end{aligned} \tag{3.5}$$

where LMSE stands for *limiting mean-squared error.*

It is then sensible to consider the following:

**Definition 3.1.** *Given* $\widehat{\gamma}_{n0,T_1(q)} = \widehat{\gamma}_{T_{1(q)}}(k_{0,T_1(q)})$ *and* $\widehat{\gamma}_{n0,T_2(q)} = \widehat{\gamma}_{T_{2(q)}}(k_{0,T_2(q)})$*, two biased PORT-estimators* $\widehat{\gamma}_{T_{1(q)}}$ *and* $\widehat{\gamma}_{T_{2(q)}}$ *for which distributional representations of the type* (3.3) *hold with constants* $(\sigma_{T_1}, d_{T_1})$ *and* $(\sigma_{T_2}, d_{T_2})$*,* $d_{T_1}$*,* $d_{T_2} \neq 0$*, respectively, both computed at their optimal levels, the Asymptotic Root Efficiency* $(AREFF)$ *of* $\widehat{\gamma}_{T_{1(q)}}$ *relatively to* $\widehat{\gamma}_{T_{2(q)}}$ *is*

$$AREFF_{T_1(q)|T_2(q)} \equiv AREFF_{\widehat{\gamma}_{T_{1(q)}}|\widehat{\gamma}_{T_{2(q)}}} := \sqrt{LMSE\left[\widehat{\gamma}_{n0,T_2(q)}\right]/LMSE\left[\widehat{\gamma}_{n0,T_1(q)}\right]},$$

*with LMSE given in* (3.5).

**Remark 3.1.** *Note that this measure was devised so that the higher the AREFF measure, the better the first estimator is.*

**Remark 3.2.** *The optimal levels* $k_{0,T(q)}$ *for the estimation of* $\gamma$ *through* $\widehat{\gamma}_{T(q)}(k)$*, with* $T$ *denoting either* $H$ *or* $M$ *are denoted by* $k_{0,H(q)}$ *and* $k_{0,M(q)}$ *and are given in Table* 1.

Table 1: Optimal levels for the estimation of $\gamma$ through PORT-Hill and PORT-moment estimators

| | $k_{0,H(q)}$ | $k_{0,M(q)}$ |
|---|---|---|
| $\gamma+\rho<0 \ \wedge \ \chi_q^*\neq 0$ | $\left(\frac{C\ (1+\gamma)\ n^{\gamma}}{\lvert\chi_q^*\rvert\sqrt{2\gamma}}\right)^{2/(1+2\gamma)}$ | $\left(\frac{C\ \sqrt{1+\gamma^2}(1+\gamma)^2\ n^{\gamma}}{\gamma^2\ \lvert\chi_q^*\rvert\sqrt{2\gamma}}\right)^{2/(1+2\gamma)}$ |
| $\gamma+\rho>0 \ \vee \ (\rho\leq-\gamma\wedge\chi_q^*=0)$ | $\left(\frac{(1-\rho)\ n^{-\rho}}{\lvert\beta\rvert\sqrt{-2\rho}}\right)^{2/(1-2\rho)}$ | $\left(\frac{\sqrt{1+\gamma^2}(1-\rho)^2\ n^{-\rho}}{\lvert\gamma(1-\rho)+\rho\rvert\ \lvert\beta\rvert\sqrt{-2\rho}}\right)^{2/(1-2\rho)}$ |
| $\gamma+\rho=0 \ \wedge \ \chi_q^*\neq 0$ | $\left(\frac{C\ (1-\rho)\ n^{-\rho}}{\lvert\beta\ C\ +\chi_q^*\rvert\sqrt{-2\rho}}\right)^{2/(1-2\rho)}$ | $\left(\frac{C\ \sqrt{1+\rho^2}(1-\rho)^2\ n^{-\rho}}{\rho^2\lvert\beta\ C+\chi_q^*\rvert\sqrt{-2\rho}}\right)^{2/(1-2\rho)}$ |

**Proposition 3.1.** *The AREFF-indicator of* $\widehat{\gamma}_{M(q)}$ *relatively to* $\widehat{\gamma}_{H(q)}$ *is:*

$$AREFF_{M(q)|H(q)}=\begin{cases}\left(\frac{\gamma^2}{1+\gamma^2}\right)^{\frac{\gamma}{1+2\gamma}}\left(\frac{1+\gamma}{\gamma}\right)^{\frac{1}{1+2\gamma}}, & \text{if } \ \rho<-\gamma \ \wedge \ \chi_q^*\neq 0\\ \left(\frac{\gamma^2}{1+\gamma^2}\right)^{\frac{-\rho}{1-2\rho}}\left(\frac{\gamma(1-\rho)}{\gamma(1-\rho)+\rho}\right)^{\frac{1}{1-2\rho}}, & \text{if } \ \rho>-\gamma \ \vee \ (\rho<-\gamma\wedge\chi_q^*=0)\\ \left(\frac{\rho^2}{1+\rho^2}\right)^{\frac{-\rho}{1-2\rho}}\left(\frac{1-\rho}{|\rho|}\right)^{\frac{1}{1-2\rho}}, & \text{if } \ \rho=-\gamma.\end{cases}$$

This $AREFF$-measure is presented in Figure 1, where we can see that the gain in efficiency for the PORT-moment estimator happens for a large region of values of $(\gamma,\rho)$.

| $\rho$ \ $\gamma$ | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 | 1.10 | 1.20 | 1.30 | 1.40 | 1.50 | 1.60 | 1.70 | 1.80 | 1.90 | 2.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.10 | 5.02 | 1.26 | 1.10 | 1.05 | 1.03 | 1.03 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| -0.20 | 5.02 | 2.26 | 1.25 | 1.11 | 1.06 | 1.04 | 1.04 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 |
| -0.30 | 5.02 | 2.26 | 1.57 | 1.18 | 1.09 | 1.06 | 1.04 | 1.04 | 1.04 | 1.03 | 1.03 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
| -0.40 | 5.02 | 2.26 | 1.57 | 1.29 | 1.12 | 1.07 | 1.05 | 1.04 | 1.03 | 1.03 | 1.03 | 1.03 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
| -0.50 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.08 | 1.05 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.04 | 1.04 | 1.04 | 1.04 |
| -0.60 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.03 | 1.03 | 1.02 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 |
| -0.70 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.03 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 |
| -0.80 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.03 | 1.03 | 1.03 | 1.03 |
| -0.90 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.03 |
| -1.00 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| -1.10 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 |
| -1.20 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 | 1.01 | 1.02 |
| -1.30 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 |
| -1.40 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 |
| -1.50 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.01 |
| -1.60 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 |
| -1.70 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 |
| -1.80 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 |
| -1.90 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| -2.00 | 5.02 | 2.26 | 1.57 | 1.29 | 1.16 | 1.09 | 1.05 | 1.02 | 1.01 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |

$1.00 < AREFF_{M(q)|H(q)} \le 1.10$ | $1.10 < AREFF_{M(q)|H(q)} \le 1.30$ | $1.30 < AREFF_{M(q)|H(q)} \le 2.30$ | $AREFF_{M(q)|H(q)} > 2.30$

Figure 1: Asymptotic efficiency of $\widehat{\gamma}_{M(q)}$ relatively to $\widehat{\gamma}_{H(q)}$ in the $(\gamma, \rho)$-plane whenever $\chi_q^* \neq 0$.

# 4 Asymptotic behaviour of the PORT-quantile estimators

We first present the following result, proved in Ferreira *et al.* (2003), on the asymptotic behaviour of intermediate o.s.:

**Proposition 4.1** (Ferreira *et al.*, 2003). *Under the second order framework in* (1.4) *and for intermediate sequences of positive integers $k$, i.e. if* (1.3) *holds,*

$$X_{n-k:n} \stackrel{d}{=} U(n/k)\Big(1 + \frac{\gamma\, B_k}{\sqrt{k}} + o_p(A(n/k))\Big), \tag{4.1}$$

*where $B_k$ is asymptotically standard normal, and*

$$\mathbb{C}ov(B_i, B_j) = \sqrt{i\ j}\ \Big(\frac{1 - j/n}{j-1}\Big), \quad i < j.$$

We shall now consider and study the new PORT-quantile estimator, in (1.14), given by

$$\widetilde{Q}_{k,p_n,T(q)} = \frac{X_{n-[k/2]:n} - X_{n-k:n}}{2^{T(q)} - 1}\Big(\frac{k}{n\,p_n}\Big)^{T(q)} + X_{n_q:n}.$$

We can state the following result:

**Theorem 4.1.** *Let us assume that the second order condition* (1.4) *holds, with* $A(t) = \gamma\beta t^\rho$, *that* $k$ *is an intermediate sequence of integers, i.e.* (1.3) *holds, and that* $\ln(np_n)/\sqrt{k} \to 0$, *as* $n \to \infty$, *with* $p_n$ *given in* (1.2). *Then, for any real* $q$, $0 \le q < 1$ *and with* $T$ *denoting either* $H$ *or* $M$,

$$\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\left(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}} - 1\right) \stackrel{d}{=} \sigma_{_T}\;P_{k,T} + \sqrt{k}\left(b_{_T}\;A(n/k) + c_{_T}\;\frac{\chi^*_q}{U(n/k)}\right)(1+o_p(1))$$
$$-\left\{\frac{\sqrt{k}\;A(n/k)}{\ln\left(\frac{k}{np_n}\right)}\left(b_{_T}\frac{2^\gamma\ln 2}{2^\gamma - 1} - \frac{2^{\gamma+\rho}-1}{\rho(2^\gamma-1)}\right) + \frac{\sqrt{k}\;\chi^*_q}{U(n/k)\ln\left(\frac{k}{np_n}\right)}\;\frac{2^\gamma\ln 2}{2^\gamma - 1}\right\}(1+o_p(1)), \quad (4.2)$$

*with* $(b_{_T}, \sigma_{_T})$ *and* $c_{_T}$ *given in* (1.10) *and* (3.2), *respectively, and where* $P_k^T$ *is asymptotically standard normal.*

*Proof.* The PORT-quantile estimator in (1.14) can be written as

$$\widetilde{Q}_{k,p_n,T(q)} := X_{n-k:n}\left[\left(\frac{X_{n-[k/2]:n}}{X_{n-k:n}} - 1\right)\frac{1}{2^{T(q)}-1}\Big(\frac{k}{n\,p_n}\Big)^{T(q)} + \frac{X_{n_q:n}}{X_{n-k:n}}\right].$$

Therefore,

$$\widetilde{Q}_{k,p_n,T(q)} - \chi_{1-p_n} = X_{n-k:n}\left[\left(\frac{X_{n-[k/2]:n}}{X_{n-k:n}} - 1\right)\frac{1}{2^{T(q)}-1}\Big(\frac{k}{n\,p_n}\Big)^{T(q)} + \frac{X_{n_q:n}}{X_{n-k:n}} - \frac{\chi_{1-p_n}}{X_{n-k:n}}\right].$$

As (2.2) holds, we can say that $X_{n_q:n}/X_{n-k:n} = o_p(1)$, and using the second order condition in (1.4), we can guarantee that

$$\frac{X_{n-[k/2]:n}}{X_{n-k:n}} = \frac{U\left(2\frac{n}{k}\right)}{U\left(\frac{n}{k}\right)} \stackrel{d}{=} 2^\gamma\left(1 + \frac{2^\rho - 1}{\rho}A(n/k)(1+o(1))\right).$$

Since $\chi_{1-p_n} = U(1/p_n)$, the result in Proposition 4.1 enables us to write

$$\frac{\chi_{1-p_n}}{X_{n-k:n}} = \frac{U\left(\frac{n}{k}\,\frac{k}{np_n}\right)}{U\left(\frac{n}{k}\right)} \times \frac{U\left(\frac{n}{k}\right)}{X_{n-k:n}} \stackrel{d}{=} \left(\frac{k}{np_n}\right)^\gamma\Big(1 - \frac{A(n/k)}{\rho}(1+o(1))\Big)$$
$$\times\Big(1 - \frac{\gamma\;B_k}{\sqrt{k}} + o_p(A(n/k))\Big) = \left(\frac{k}{np_n}\right)^\gamma\Big(1 - \frac{\gamma\;B_k}{\sqrt{k}} - \frac{A(n/k)}{\rho}(1+o_p(1))\Big),$$

and the use of the delta method leads us to

$$\frac{\left(\frac{k}{np_n}\right)^{T(q)}}{2^{T(q)}-1} \stackrel{d}{=} \frac{\left(\frac{k}{np_n}\right)^{\gamma}}{2^{\gamma}-1}\Big(1+(T_k(q)-\gamma)\left(\ln\left(\frac{k}{np_n}\right)-\frac{2^\gamma\ln 2}{2^\gamma-1}\right)(1+o_p(1))\Big).$$

Therefore

$$\begin{aligned}\widetilde{Q}_{k,p_n,T(q)}-\chi_{1-p_n} &\stackrel{d}{=} \left(\frac{k}{np_n}\right)^{\gamma} X_{n-k:n}\Big[(T_k(q)-\gamma)\left(\ln\left(\frac{k}{np_n}\right)-\frac{2^\gamma\ln 2}{2^\gamma-1}\right)(1+o_p(1)) \\ &\qquad +\frac{\gamma\ B_k}{\sqrt{k}}+\frac{A(n/k)(2^{\gamma+\rho}-1)}{\rho(2^\gamma-1)}\ (1+o_p(1))\Big] \\ &\stackrel{d}{=} \left(\frac{k}{np_n}\right)^{\gamma} U(n/k)\Big(1+\frac{\gamma\ B_k}{\sqrt{k}}+\frac{A(n/k)}{\rho}\ (1+o_p(1))\Big) \\ &\quad\times\Big[(T_k(q)-\gamma)\left(\ln\left(\frac{k}{np_n}\right)-\frac{2^\gamma\ln 2}{2^\gamma-1}\right)\ (1+o_p(1)) \\ &\qquad\qquad +\frac{\gamma\ B_k}{\sqrt{k}}+\frac{A(n/k)(2^{\gamma+\rho}-1)}{\rho(2^\gamma-1)}\ (1+o_p(1))\Big],\end{aligned}$$

using also the result presented in Proposition 4.1. Since $\big(\gamma\ B_k/\sqrt{k}+(A(n/k)/\rho)(1+o_p(1))\big)=o_p\big(1/\sqrt{k}\big)$, then

$$\begin{aligned}\widetilde{Q}_{k,p_n,T(q)}-\chi_{1-p_n} &\stackrel{d}{=} \left(\frac{k}{np_n}\right)^{\gamma} U(n/k)\Big[(T_k(q)-\gamma)\left(\ln\left(\frac{k}{np_n}\right)-\frac{2^\gamma\ln 2}{2^\gamma-1}\right)(1+o_p(1)) \\ &\qquad +\frac{\gamma\ B_k}{\sqrt{k}}+\frac{A(n/k)(2^{\gamma+\rho}-1)}{\rho(2^\gamma-1)}\ (1+o_p(1))\Big].\end{aligned}$$

Notice that $\chi_{1-p_n}=\big(k/(np_n)\big)^{\gamma}U(n/k)$, and then

$$\begin{aligned}\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\left(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\right) &\stackrel{d}{=} \sqrt{k}\,(T_k(q)-\gamma)-\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\,(T_k(q)-\gamma)\,\frac{2^\gamma\ln 2}{2^\gamma-1} \\ &\qquad +\frac{\gamma\ B_k}{\ln\left(\frac{k}{np_n}\right)}+\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\ \frac{A(n/k)(2^{\gamma+\rho}-1)}{\rho(2^\gamma-1)}.\end{aligned}$$

Using the distributional representation of $T_k(q)$ in (3.1) and since $\ln\left(k/(np_n)\right)\to\infty$, (4.2) follows. □

**Corollary 4.1.** *Under the conditions of Theorem 4.1, the following results hold:*

- *For values of* $\gamma+\rho<0$ *and* $\chi_q^*\neq 0$,

$$\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big) \overset{d}{=} \sigma_{_T}\ P_{k,T}+\sqrt{k}\Big(c_{_T}\ \frac{\chi_q^*}{U(n/k)}\Big)\ (1+o_p(1)) - \frac{\sqrt{k}\ \chi_q^*}{U(n/k)\ln\left(\frac{k}{np_n}\right)}\frac{2^\gamma\ln 2}{2^\gamma-1}\ (1+o_p(1)),$$

*If* $\sqrt{k}/U(n/k)\to\lambda_1$ *finite, then*

$$\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal(\lambda_1 c_{_T}\ \chi_q^*,\sigma_{_T}^2).$$

- *For values of* $\gamma+\rho>0$ *or* $\gamma+\rho\leq 0$ *and* $\chi_q^*=0$,

$$\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big) \overset{d}{=} \sigma_{_T}\ P_{k,T}+\sqrt{k}\,(b_{_T}\ A(n/k))\,(1+o_p(1)) - \frac{\sqrt{k}A(n/k)}{\ln\left(\frac{k}{np_n}\right)}\Big(b_{_T}\frac{2^\gamma\ln 2}{2^\gamma-1}-\frac{1}{\rho}\Big)(1+o_p(1)),$$

*If* $\sqrt{k}A(n/k)\to\lambda_2$ *finite, then*

$$\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal(\lambda_2 b_{_T},\sigma_{_T}^2).$$

- *For values of* $\gamma+\rho=0\ \wedge\ \chi_q^*\neq 0$,

$$\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big) \overset{d}{=} \sigma_{_T}\ P_{k,T}+\sqrt{k}\Big(b_{_T}\ A(n/k)+c_{_T}\ \frac{\chi_q^*}{U(n/k)}\Big)(1+o_p(1)) - \Big\{\frac{\sqrt{k}A(n/k)}{\ln\left(\frac{k}{np_n}\right)}\left(b_{_T}\frac{2^\gamma\ln 2}{2^\gamma-1}-\frac{1}{\rho}\right)+\frac{\sqrt{k}\ \chi_q^*}{U(n/k)\ln\left(\frac{k}{np_n}\right)}\frac{2^\gamma\ln 2}{2^\gamma-1}\Big\}(1+o_p(1)),$$

*If* $\sqrt{k}/U(n/k)\to\lambda_1$ *and* $\sqrt{k}A(n/k)\to\lambda_2$, *with* $\lambda_1$ *and* $\lambda_2$ *both finite, then*

$$\frac{\sqrt{k}}{\ln\left(\frac{k}{np_n}\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal(\lambda_1\ c_{_T}\ \chi_q^*+\lambda_2 b_{_T},\sigma_{_T}^2).$$

**Remark 4.1.** *Notice that, under a second order framework framework, the mean value and the variance of the r.v.* $\sqrt{k}\big(\widehat{\gamma}_{k,n,T(q)}-\gamma\big)$, *provided in* Corolary 3.1, *are equal to the ones of* $\sqrt{k}\big(\widetilde{Q}_{k,p_n,T(q)}/\chi_{1-p_n}-1\big)/\ln\left(k/(np_n)\right)$.

**Remark 4.2.** *The PORT-quantile estimator in* (1.13), *studied in Araújo Santos et al.* (2006) *is asymptotically equivalent to the new PORT-quantile estimator defined in* (1.14).

Since $\ln\left(k/(np_n)\right)$ goes to infinity very slowly, we can state a pre-asymptotic distributional representation, for moderate $k$ and $n$:

**Corollary 4.2.** *Under the conditions of Theorem 4.1 and for moderate values of $k$ and $n$, the following pre-asymptotic results hold:*

- *For values of $\gamma+\rho<0$ and $\chi_q^*\neq 0$, if $\sqrt{k}/U(n/k)\to\lambda_1$ finite, then*

$$\frac{\sqrt{k}}{\ln\left(k/(np_n)\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big)\ \overset{d}{\approx}\ Normal\Big(\mu_1,\sigma_T^2\Big(1+\frac{\gamma^2}{\sigma_T^2\ \ln^2\left(k/(np_n)\right)}\Big)\Big),$$

  *where*

$$\mu_1:=\lambda_1 c_T\ \chi_q^*\Big(1-\frac{2^\gamma\ \ln 2}{2^\gamma-1}\frac{1}{c_T\ \ln\left(k/(np_n)\right)}\Big).$$

- *For values of $\gamma+\rho>0$ or $\gamma+\rho\le 0$ and $\chi_q^*=0$, if $\sqrt{k}A(n/k)\to\lambda_2$ finite, then*

$$\frac{\sqrt{k}}{\ln\left(k/(np_n)\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big)\ \overset{d}{\approx}\ Normal\Big(\mu_2,\sigma_T^2\Big(1+\frac{\gamma^2}{\sigma_T^2\ \ln^2\left(k/(np_n)\right)}\Big)\Big),$$

  *where*

$$\mu_2:=\lambda_2 b_T\ \Big(1+\frac{1}{\ln\left(k/(np_n)\right)}\Big(\frac{2^\gamma\ \ln 2}{2^\gamma-1}-\frac{1}{\rho\ b_T}\Big)\Big).$$

- *For values of $\gamma+\rho=0\ \wedge\ \chi_q^*\neq 0$, if $\sqrt{k}/U(n/k)\to\lambda_1$ and $\sqrt{k}A(n/k)\to\lambda_2$, with $\lambda_1$ and $\lambda_2$ both finite, then*

$$\frac{\sqrt{k}}{\ln\left(k/(np_n)\right)}\Big(\frac{\widetilde{Q}_{k,p_n,T(q)}}{\chi_{1-p_n}}-1\Big)\ \overset{d}{\approx}\ Normal\Big(\mu_1+\mu_2,\sigma_T^2\Big(1+\frac{\gamma^2}{\sigma_T^2\ \ln^2\left(k/(np_n)\right)}\Big)\Big),$$

*where*

$$\mu_1 + \mu_2 = \lambda_1 c_{_T}\ \chi_q^*\Big(1 - \frac{2^\gamma\ l2}{2^\gamma - 1}\frac{1}{c_{_T}\ \ln\left(k/(np_n)\right)}\Big) + \lambda_2 b_{_T}\ \Big(1 + \frac{1}{\ln\left(k/(np_n)\right)}\Big(\frac{2^\gamma\ \ln 2}{2^\gamma - 1} - \frac{1}{\rho\ b_{_T}}\Big)\Big).$$

## References


[1] Araújo Santos, P.; Fraga Alves, M.I.; Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation. *Revstat* **4**, no. 3, 227–247.

[2] Beirlant, J.; Figueiredo, F.; Gomes, M.I.; Vandewalle, B. (2008). Improved reduced-bias tail index and quantile estimators. *J. Statist. Plann. and Inference* **138**: 6, 1851-1870.

[3] Caeiro, F.; Gomes, M.I. (2008). Semi-parametric second-order reduced-bias high quantile estimation. *Test*, in press, DOI: 10.1007/s11749-008-0108-8.

[4] Dekkers, A.; Einmahl, J.; de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* **17**, 1833–1855.

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

[6] Geluk, J.; de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems.* CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.

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

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

[9] Gomes, M.I.; Fraga Alves, M.I.; Araújo Santos, P. (2008). PORT Hill and Moment Estimators for Heavy-Tailed Models. *Commun. Statist. – Simul. & Comput.* **37**, 1281-1306.

[10] Gomes, M.I.; Martins, M.J. (2001). Generalizations of the Hill estimator: asymptotic versus finite sample behaviour. *J. Statist. Planning and Inference* **93**, 161-180.

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

[12] Gomes, M.I.; Miranda, C.; Viseu, C. (2007). Reduced bias extreme value index estimation and the Jackknife methodology, *Statistica Neerlandica* **61**:2, 243-270.

[13] Gomes, M.I.; Neves, C. (2008). Asymptotic comparison of the mixed moment and classical extreme value index estimators. *Statistics and Probability Letters* **78**:6, 643-653.

[14] Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. American Statistical Association*, Vol. **102**, Nº 477, 280-292.

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

[16] Haan, L. de; Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52**, 60-70.

[17] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. R. Statist. Soc.* **44**, no. 1, 37–42.

[18] Hall, P.; Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.

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

[20] Martins, M. J. (2000). *Estimação de Caudas Pesadas - Variantes ao Estimador de Hill.* Tese de Doutoramento, D.E.I.O., Faculdade de Ciências da Universidade de Lisboa.

[21] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*. Wiley Series in Probability and Statistics.

[22] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer.

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