Second order reduced bias tail index estimators under a third order framework

M. Ivette Gomes

Universidade de Lisboa and CEAUL

M. João Martins and Manuela Neves

Universidade Técnica de Lisboa, ISA

- Classical tail index estimators are known to be quite sensitive to the number k of top o.s. used in the estimation.
- The recently developed 2nd order reduced bias' estimators show less sensitivity to changes in k. We are here interested in this type of tail index estimation, based on an exponential 2nd order regression model for the scaled top log-spacings.
- The estimation of the 2nd order parameters in the bias, at a level k_1 of a larger order than that of the level k used for the tail index estimation, enables us to keep the asymptotic variance of the new estimators equal to the asymptotic variance of the Hill estimator, the ML estimator of γ, under a strict Pareto model.
- To enhance the performance of this type of estimators, we also consider the estimation of the scale second order parameter only, and of all unknown parameters, at the same level k.
- The asymptotic distributional properties of the proposed class of γ-estimators are derived under 2nd and 3rd order frameworks and the estimators are compared with other similar alternative estimators of γ, not only asymptotically, but also for finite samples through Monte Carlo techniques.
- A case-study in the field of finance will illustrate the performance of these new second order reduced bias' tail index estimators.

Introduction and motivation for the new class of tail index estimators. Heavy-tailed models are quite useful in diversified fields, like telecommunication networks and finance. In the area of *EVT*, with $U(t) = F^{\leftarrow}(1-1/t),\ t \geq 1$,

$$\boxed{F \text{ is heavy-tailed} \quad \Longleftrightarrow \quad U \in RV_\gamma.}$$

Then, and with $\gamma > 0$, we are in the domain of attraction for maxima of

$$\boxed{EV_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right),\ 1+\gamma x \geq 0 & \text{if} \quad \gamma \neq 0 \\ \exp\left(-\exp(-x)\right),\ x \in \mathbb{R} & \text{if} \quad \gamma = 0 \end{cases}.}$$

The *tail index* γ is indeed the primary parameter of extreme events.

The *second order parameter*, ρ (≤ 0), rules the rate of convergence in the 1st order condition, and is the parameter appearing in the limit

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

with $|A(t)| \in RV_\rho$.

This condition has been widely accepted as an appropriate condition to specify the tail of a Pareto-type distribution in a semi-parametric way, and it holds true for most common Pareto-type models, like the *Fréchet*, the *Generalized Pareto* and the *Student*'s t. We assume everywhere that $\rho < 0$.

To obtain information on the asymptotic bias of 2nd order reduced bias' estimators, we need further assuming a 3rd order condition, ruling now the rate of convergence in the 2nd order condition. We write such a 3rd order condition as,

$$\boxed{\lim_{t\to\infty} \frac{\frac{\ln U(tx)-\ln U(t)-\gamma \ln x}{A(t)} - \frac{x^\rho - 1}{\rho}}{B(t)} = \frac{x^{\rho+\rho'}-1}{\rho+\rho'},}$$

with $|B(t)| \in RV_{\rho'}$ and $\rho' < 0$. We have $\rho' = \rho$ for most of the common heavy-tailed d.f.'s. We shall assume to be in a class of models where, for $\beta,\ \beta' \neq 0,\ \rho,\ \rho' < 0$, we may choose

$$\boxed{A(t) = \alpha\ t^\rho =: \gamma\ \beta\ t^\rho,\ B(t) =\ \beta'\ t^{\rho'}.}$$

Basic statistics in this study ($1 \leq i \leq k < n$):

$$\boxed{V_{ik} := \ln \frac{X_{n-i+1:n}}{X_{n-k:n}}, \qquad U_i := i \left\{ \ln \frac{X_{n-i+1:n}}{X_{n-i:n}} \right\}.}$$

As usual, k is a sequence of intermediate integers:

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

and Hill's estimator of γ [Hill, 1975],

$$\boxed{H_n(k) = \sum_{i=1}^{k} V_{ik}/k \; = \; \sum_{i=1}^{k} U_i/k,}$$

is consistent for the estimation of γ.

The adequate accommodation of the bias of Hill's estimator has been extensively addressed in recent years. Beirlant *et al.* (1999) and Feuerverger and Hall (1999) consider exponential regression techniques, based on the approximations:

$$\boxed{U_i \approx \gamma \left(1 + b(n/k)(k/i)^{\rho}\right) E_i \quad \text{and} \quad U_i \approx \gamma e^{\beta \ (n/i)^{\rho}} E_i,}$$

respectively, $1 \leq i \leq k$. They then proceed to the joint estimation of the 3 unknown parameters or functionals at the same k.

Working with this second approximation, Gomes and Martins (2002) advance with the "external" estimation of the 2nd order parameter ρ, together with a 1st order approximation for the ML β-estimator. We then obtain "quasi-ML" explicit estimators of γ and β, and through the "external" estimation of ρ, we are able to reduce the asymptotic variance of the proposed tail index estimator. Such a tail index estimator is

$$\boxed{\widehat{\gamma}_n^{ML}(k) = S_0(k) - \widehat{\beta}_{\widehat{\rho}}(k)\left(\frac{n}{k}\right)^{\widehat{\rho}} S_{\widehat{\rho}}(k), \;\; S_\rho(k) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho} U_i,}$$

with

$$\boxed{\widehat{\beta}_{\widehat{\rho}}(k) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{s_{\widehat{\rho}}(k)\;\, S_0(k) - S_{\widehat{\rho}}(k)}{s_{\widehat{\rho}}(k)\;\, S_{\widehat{\rho}}(k) - S_{2\widehat{\rho}}(k)}, \qquad s_\rho(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}.}$$

The β-estimator is plugged in $\widehat{\gamma}_n^{ML}(k)$, after being computed at the same level k. We here propose an "external" estimation of both β and ρ through $\widehat{\beta}$ and $\widehat{\rho}$, both using a number of top o.s. of a larger order than the one used for the tail index estimation.

We shall thus consider, for an adequate consistent estimator $\left(\widehat{\beta},\ \widehat{\rho}\right)$ of $(\beta,\ \rho)$:

$$\boxed{ML_{\widehat{\beta},\ \widehat{\rho}}(k) := S_0(k) - \widehat{\beta}\ \left(\frac{n}{k}\right)^{\widehat{\rho}}\ S_{\widehat{\rho}}(k),}$$

Remark 1. *This estimator has been inspired in the recent papers of Gomes et al. (2004b) and Caeiro et al. (2004, 2005). These authors consider, in different ways, the "external" estimation of both the "scale" and the "shape" parameter in the A function, being able to reduce the bias without increasing the asymptotic variance, which is kept at the value γ^2 for moderate k levels. The tail index estimator in Gomes et al. (2004b) is*

$$\boxed{WH_{\widehat{\beta},\ \widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ ((i/k)^{-\widehat{\rho}}-1)/(\widehat{\rho}\ln(i/k))}\ V_{ik},}$$

with the notation WH standing for Weighted Hill estimator. Caeiro et al. (2004, 2005) consider the estimator

$$\boxed{\overline{H}_{\widehat{\beta},\ \widehat{\rho}}(k) := H(k)\left(1 - \frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right).}$$

Remark 2. *Note that $\widehat{\gamma}_n^{ML}(k) = ML_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$, when both γ and β are estimated at the same level k.*

- Whenever there is no distinction between the three *"Unbiased Hill"* estimators, or the corresponding r.v.'s, we shall often use the notation UH, generically denoting either ML or WH or $\overline{H}$.

Asymptotic behaviour of the reduced bias' tail index estimators under a third order framework. Denoting $\{E_i\}$ a sequence of i.i.d. standard exponential r.v.'s, let us denote $Z_k^{(1)} = \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}\ E_i - 1\right)$. Let us assume that only γ is unknown:

Theorem 1. *Under the 2nd order framework, further assuming that $A(t)$ may be chosen as mentioned before, and for intermediate levels k, we get, for the r.v. $ML_{\beta,\ \rho}(k)$, an asymptotic distributional representation of the type*

$$\boxed{ML_{\beta,\ \rho}(k) \ \stackrel{d}{=}\ \gamma + \frac{\gamma}{\sqrt{k}}\ Z_k^{(1)} + o_p(A(n/k)).}$$

Then, $\sqrt{k}\ \left(ML_{\beta,\ \rho}(k) - \gamma\right)$ is AN with variance equal to γ^2, and with a null mean value not only when $\sqrt{k}\ A(n/k) \longrightarrow 0$, but also when $\sqrt{k}\ A(n/k) \longrightarrow \lambda \neq 0$, finite, as $n \to \infty$.

Under the third order framework we may further specify the term $o_p(A(n/k))$, which is given by

$$\boxed{A(n/k)\left(\frac{B(n/k)}{1-\rho-\rho'}-\frac{A(n/k)}{\gamma(1-2\rho)}\right)(1+o_p(1)).}$$

Consequently, even if $\sqrt{k}\ A(n/k) \to \infty$, but with λ_A and λ_B finite, $\sqrt{k}\ A^2(n/k) \to \lambda_A$ and $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B$, $\sqrt{k}\ \left(ML_{\beta,\ \rho}(k)-\gamma\right)$ is asymptotically normal with variance equal to γ^2. The asymptotic bias of $ML_{\beta,\ \rho}(k)$ is equal to

$$\boxed{b_{ML} := \frac{\lambda_B}{1-\rho-\rho'}-\frac{\lambda_A}{\gamma(1-2\rho)}.}$$

Remark 3. *If $\rho=\rho'$, $b_{ML}=(\lambda_B-\lambda_A/\gamma)/(1-2\rho)$. Since for the Burr model, with d.f. $F(x)=1-\left(1+x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, we may choose $B(t)=A(t)/\gamma$, we have $\lambda_B=\lambda_A/\gamma$ and $b_{ML}=0$.*

Remark 4. *In Caeiro et al. (2005) have been proved results similar to those of Theorem 1 for WH and $\overline{H}$. For $WH_{\beta,\ \rho}(k)$, we have got an asymptotic bias given by*

$$\boxed{b_{WH} := \frac{\lambda_B}{1-\rho-\rho'}-\frac{\lambda_A\ a_2}{2\ \gamma}, \quad a_2=\frac{-\ln\left((1-2\rho)/(1-\rho)^2\right)}{\rho^2}}.$$

For $\overline{H}_{\beta,\ \rho}(k)$, the asymptotic bias is given by

$$\boxed{b_{\overline{H}} := \frac{\lambda_{B}}{1-\rho-\rho'} - \frac{\lambda_{A}}{\gamma(1-\rho)^2}.}$$

Since $\lambda_{A} \geq 0$ and $2/a_2(\rho) > (1-\rho)^2 > 1-2\rho$ for any $\rho < 0$, we have $b_{WH} \geq b_{\overline{H}} \geq b_{ML}$.

How far is it possible to replace $(\beta,\ \rho)$ by $\left(\widehat{\beta}, \widehat{\rho}\right)$ and still get the same results as in Theorem 1?

It is possible to prove that

$$\sqrt{k}\left\{UH_{\widehat{\beta},\ \widehat{\rho}}(k) - UH_{\beta,\ \rho}(k)\right\}$$
$$\overset{p}{\sim}\ \sqrt{k}\ A(n/k)\left(\widehat{\rho}-\rho\right)\left\{a^*_{UH} \ln\left(\frac{k}{k_1}\right) + b^*_{UH}\right\} =: \widehat{W}_{k,\ k_1}$$

Now

$$\boxed{\widehat{W}_{k,\ k_1} = \begin{cases} \sqrt{k}\ A(n/k) b_{UH}\left(\widehat{\rho}(k)-\rho\right) & \text{if} \quad k = k_1 \\ o_p(1) & \text{if} \quad \sqrt{k}\ A(n/k) \to \lambda \\ & \text{and} \quad \widehat{\rho}-\rho = o_p(1/\ln n) \\ \frac{\sqrt{k}\ A(n/k)}{\sqrt{k_1}\ A(n/k_1)}\left\{a_{UH}\ln\left(\frac{k}{k_1}\right)+b_{UH}\right\} & \text{if} \quad \sqrt{k_1}\ AB(n/k_1) \to \lambda_{B_1} \\ & \text{and} \quad \sqrt{k_1}\ A^2(n/k_1) \to \lambda_{A_1} \\ \frac{\sqrt{k}\ A(n/k)B(n/k_1)}{}\left\{a_{UH}\ln\left(\frac{k}{k_1}\right)+b_{UH}\right\} & \quad \text{otherwise} \end{cases}}$$

A brief review of the second order parameters' estimators. We have nowadays a general class of ρ-estimators which work well in practice [Fraga Alves *et al.* (2003)]. Under general conditions, they are semi-parametric asymptotically normal estimators of ρ, whenever $\rho < 0$. Such a class of estimators is based on the statistics

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

parameterised in a tuning parameter $\tau \in \mathbb{R}$, where,

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right\}^j, \ j \geq 1 \quad \left[M_n^{(1)} \equiv H\right].$$

Usually for $\tau = 0$ and 1, we work here with

$$\hat{\rho}_\tau(k) \equiv \hat{\rho}_n^{(\tau)}(k) := -\left|\frac{3(T_n^{(\tau)}(k) - 1)}{T_n^{(\tau)}(k) - 3}\right|,$$

and compute it at a level k_1 of a larger order than that of the level k on which we base the tail index estimation.

Proposition 1 (ρ-estimation). *Under the second order framework, with $\rho < 0$, if k is intermediate, and if $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$, $\widehat{\rho}_n^{(\tau)}(k)$ converges in probability towards ρ, as $n \to \infty$, for any $\tau \in \mathbb{R}$. Under the 3rd order framework, if $\sqrt{k}\ A^2(n/k) \longrightarrow \lambda_{_A}$, finite, and $\sqrt{k}\ A(n/k)\ B(n/k) \longrightarrow \lambda_{_B}$, also finite, $\sqrt{k}\ A(n/k) \left(\widehat{\rho}_n^{(\tau)}(k) - \rho\right)$ is AN with asymptotic variance*

$$\boxed{\sigma_\rho^2 \equiv \sigma_\rho^2(\gamma) = \left(\frac{\gamma(1-\rho)^3}{\rho}\right)^2 \left(2\rho^2 - 2\rho + 1\right).}$$

There is moreover a possibly non-null asymptotic bias given by $\{\lambda_{_A} u_\rho + \lambda_{_B} v_\rho\}$, where

$$\boxed{u_\rho = \frac{\rho\left(\tau(1-2\rho)^2(3-\rho)(3-2\rho) - 6\rho\left(4\rho^3 - 16\rho^2 + 20\rho - 7\right)\right)}{12\ \gamma\left((1-\rho)(1-2\rho)\right)^2},}$$

$$\boxed{v_\rho \equiv v_\rho(\rho') = \left(1 + \frac{\rho'}{\rho}\right)\left(\frac{1-\rho}{1-\rho-\rho'}\right)^3.}$$

Estimation of β based on the scaled log-spacings. We have considered the estimator of β obtained in Gomes and Martins (2002), already defined, and based on the scaled log-spacings U_i, $1 \le i \le k$. When estimating β "externally", we have used $\widehat{\beta}_j = \widehat{\beta}_{\widehat{\rho}_j}$, $j = 0,\ 1$. We thus need the distributional behaviour of $\widehat{\beta}_{\widehat{\rho}(k)}(k)$:

Proposition 2. *If the 2nd order condition holds, the rate of convergence of $\widehat{\beta}_{\widehat{\rho}(k)}(k)$ is of the order of $\left\{\ln(n/k)/(\sqrt{k}\ A(n/k))\right\}$, which must converge towards zero, so that $\widehat{\beta}_{\widehat{\rho}(k)}(k)$ is consistent for the estimation of β, and*

$$\boxed{\frac{\sqrt{k}\ A(n/k)}{\ln(n/k)}\left(\frac{\widehat{\beta}_{\widehat{\rho}(k)}(k)-\beta}{\beta}\right) \overset{p}{\sim} -\sqrt{k}\ A(n/k)\left(\widehat{\rho}(k)-\rho\right).}$$

If apart from $\sqrt{k}\ A(n/k)/\ln(n/k) \to \infty$, we assume that $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$ and $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_{_B}$, both finite,

$$\boxed{\frac{\sqrt{k}\ A(n/k)}{\ln(n/k)}\left(\frac{\beta-\beta_{\widehat{\rho}(k)}(k)}{\beta}\right)}$$

is AN, with asymptotic variance σ^2_ρ and a bias $\{\lambda_{_A}\ u_\rho + \lambda_{_B}\ v_\rho\}$, with u_ρ and v_ρ given before.

Remark 5. *The theoretical and simulated results in Fraga Alves et al. (2003), together with the use of these ρ-estimators in 2nd order reduced bias tail index estimators, has led us to advise in practice the consideration of the level*

$$\boxed{k_1 = \min\left(n-1, [2n/\ln\ln n]\right)}$$

(*not chosen in any optimal way*). *Indeed, practitioners should not choose blindly the value of τ. It is sensible to draw a few sample paths of $\widehat{\rho}_\tau(k)$, as functions of k, electing the τ providing the highest stability for large k, by means of any stability criterion. Anyway, in all the Monte Carlo simulations we have considered this level k_1 and the ρ-estimators $\widehat{\rho}_0$ if $\rho \geq -1$ and $\widehat{\rho}_1$ if $\rho < -1$.*

Remark 6. *The maintenance of part of Theorem 1 could be achieved only under:*

Condition *U: There exists a $\tau = \tau_{_U}$ such that $\widehat{\rho}_{\tau_{_U}}(k)$ is unbiased for the estimation of ρ.*

We may then work with $\widehat{\rho}_{_U} = \widehat{\rho}_{\tau_{_U}}(k_1)$ and $\widehat{\beta}_{_U} = \widehat{\beta}_{\widehat{\rho}_{_U}}(k_1)$, k_1 any level such that $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \to \infty$ and still get $\widehat{\rho}_{_U} - \rho = O_p\left(1/\left(\sqrt{k_1}\ A(n/k_1)\right)\right)$

Tail index estimation based on the estimation of ρ (only) at a lower threshold. Let us assume first that we estimate both β and ρ "externally" at the level k_1. We may state the following:

Theorem 2. *Under the conditions of Th. 1, let us consider $UH_{\hat\beta,\ \hat\rho}(k)$, for any of the proposed estimators $\hat\beta$ and $\hat\rho$, computed at a level k_1 of a larger order than k. Then, $\sqrt{k}\left\{UH_{\hat\beta,\ \hat\rho}(k)-\gamma\right\}$ is asymptotically normal with variance γ^2 and null mean value, not only when $\sqrt{k}\ A(n/k) \to 0$, but also whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite.*
Under the third order framework and Condition U, if $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$ and $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_{_B}$, with $\lambda_{_A}$ and $\lambda_{_B}$ finite, $\sqrt{k}\ \left(UH_{\hat\beta_{U},\ \hat\rho_{U}}(k)-\gamma\right)$ is AN with variance equal to γ^2. The asymptotic bias of $UH_{\hat\beta_{U},\ \hat\rho_{U}}(k)$ is equal to b_{UH}.

If we consider γ and β estimated at the same level, we have an increase in the asymptotic variance:

Theorem 3. *If the 3rd order condition holds, if $k = k_n$ is a sequence of intermediate integers, and if $\sqrt{k}\ A(n/k) \to \infty$, with $\sqrt{k}\ A^2(n/k)$ and $\sqrt{k}\ A(n/k)\ B(n/k)$ converging both towards zero, as $n \to \infty$, the asymptotic variance of $UH_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}(k)$ increases of a factor $((1-\rho)/\rho)^2 > 1,\ \forall \rho \leq 0$.*

If $\sqrt{k}\ A^2(n/k) \to \lambda_A$ & $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B$, finite, the asymptotic variances of all the $UH_{\widehat{\beta}_{\widehat{\rho}_U}(k),\ \widehat{\rho}_U}(k)$ statistics are kept equal to $(\gamma(1-\rho)/\rho)^2$, and the asymptotic bias of $ML_{\widehat{\beta}_{\rho}(k),\ \rho}(k)$ is given by

$$\boxed{b^*_{ML} = \frac{\lambda_A(1-\rho)}{\gamma(1-2\rho)(1-3\rho)} - \frac{\lambda_B\ \rho'(1-\rho)}{\rho(1-\rho-\rho')(1-2\rho-\rho')}.}$$

Remark 7. *Again, $b^*_{WH} \geq b^*_{\overline{H}} \geq b^*_{ML}$.*

Estimation of γ, β and ρ at the same level k.

If we estimate the 3 parameters γ, β and ρ at the same k:

Theorem 4. *If the 3rd order condition holds, if $k = k_n$ is a sequence of intermediate integers, and if $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, with $\sqrt{k}\ A^2(n/k) \to 0$ and $\sqrt{k}\ A(n/k)\ B(n/k) \to 0$, then,*

$$\boxed{\sqrt{k}\left(UH_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} Normal\left(0,\ \sigma_3^2\right),}$$

where $\boxed{\sigma_3^2 = \gamma^2\left(1 + \left(\frac{1-\rho}{\rho}\right)^2 - \frac{2\rho(1-\rho)^3}{\rho^2}\right).}$

If $\sqrt{k}\ A^2(n/k) \to \lambda_A$ and $\sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B$, both finite, the asymptotic bias of the $UH_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\ \widehat{\rho}(k)}(k)$ statistics are given by

$$b^{**}_{ML} = \lambda_B \left(\frac{1}{1-\rho-\rho'} - \frac{v_\rho}{(1-\rho)^2} \right) - \frac{\lambda_A}{(1-\rho)^2} \left(\frac{1}{\gamma} + u_\rho \right),$$

$$b^{**}_{WH} = \lambda_B \left(\frac{1}{1-\rho-\rho'} - \frac{v_\rho}{(1-\rho)^2} \right) - \lambda_A \left(\frac{a_2(\rho)}{2\gamma} + \frac{u_\rho}{(1-\rho)^2} \right)$$

and

$$b^{**}_{\overline{H}} = \lambda_B \left(\frac{1}{1-\rho-\rho'} - \frac{v_\rho}{(1-\rho)^2} \right) - \lambda_A \left(\frac{1}{\gamma(1-2\rho)} + \frac{u_\rho}{(1-\rho)^2} \right),$$

respectively, with $a_2(\rho)$, u_ρ and v_ρ given before.

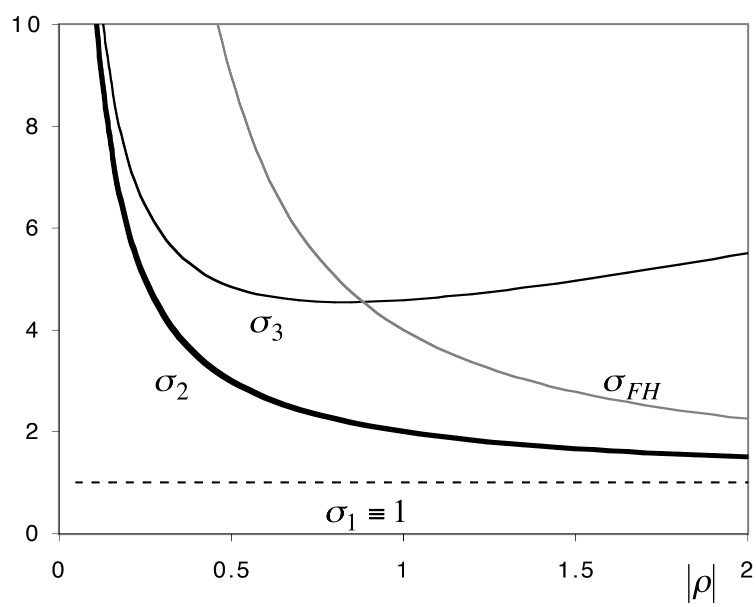
Remark 8. *Again,* $b^{**}_{WH} \geq b^{**}_{\overline{H}} \geq b^{**}_{ML}$.

Remark 9. *If we compare Theorems 2, 3 and 4, we see that it seems convenient to estimate the two second order parameters β and ρ at a level k_1 of a larger order than the level k used for the tail index estimation. Note however that we are not able to guarantee the asymptotic variances in Theorems 2 and 3, when we base the tail index estimation on optimal levels for ρ. To attain those variances we had to assume Condition U.*

Remark 10. *The asymptotic variance of the estimator in* Feuerverger and Hall (1999) (*where also the 3 parameters are computed at the same* k) *is given by* $\sigma^2_{FH} := \gamma^2\left((1-\rho)/\rho\right)^4$. *We have*

$$\boxed{\begin{array}{ll} \sigma_1 < \sigma_2 < \sigma_3 < \sigma_{FH} & \textit{if} \quad |\rho| < 0.8832 \\ \sigma_1 < \sigma_2 < \sigma_{FH} < \sigma_3 & \textit{if} \quad |\rho| > 0.8832. \end{array}}$$

In the following figure, we provide both a picture and some values of $\sigma_1/\gamma \equiv 1$, σ_2/γ, σ_3/γ *and* σ_{FH}/γ, *as functions of* $|\rho|$.



$\lvert\rho\rvert$	σ_1	σ_2	σ_3	σ_{FH}
0.1	1.00	11.00	12.19	121.00
0.2	1.00	6.00	7.37	36.00
0.3	1.00	4.33	5.87	18.78
0.4	1.00	3.50	5.19	12.25
0.5	1.00	3.00	4.85	9.00
1.0	1.00	2.00	4.58	4.00
1.5	1.00	1.67	4.96	2.78
2.0	1.00	1.50	5.50	2.25
2.5	1.00	1.40	6.10	1.96
3.0	1.00	1.33	6.74	1.78

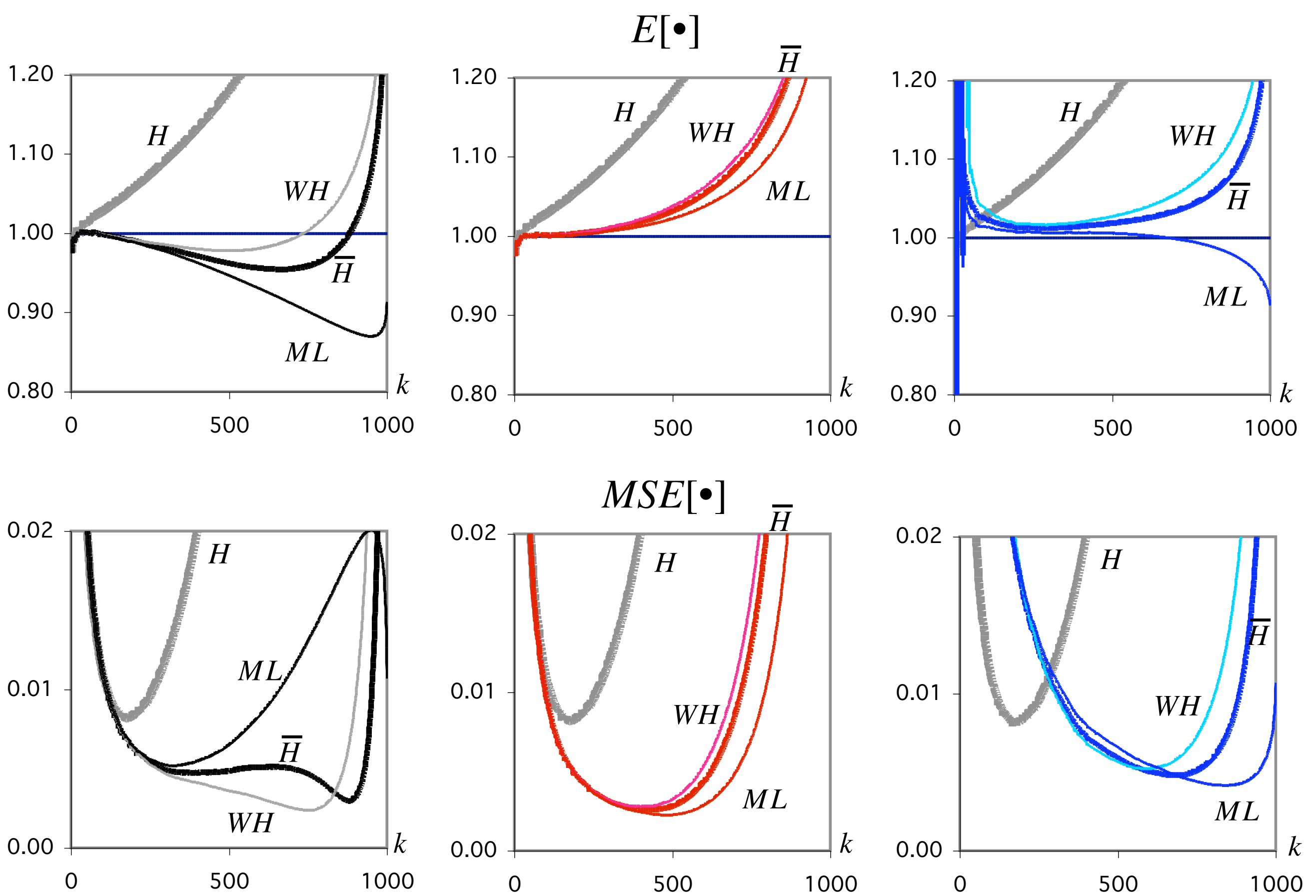
Finite sample behaviour of the estimators. In the simulations we have considered the following models:

- the *Fréchet* model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, $\gamma > 0$, for which $\rho' = \rho = -1$, $\beta = 0.5$;
- the GP model, with d.f. $F(x) = 1-(1+\gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho' = \rho = -\gamma$, $\beta = 1$;
- the *Burr* model, with d.f. $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$, $x \geq 0$, $\gamma > 0$, $\rho' = \rho < 0$, $\beta = 1$;
- the *Student*'s t_ν-model with ν degrees of freedom, for which $\gamma = 1/\nu$ and $\rho' = \rho = -2/\nu$.

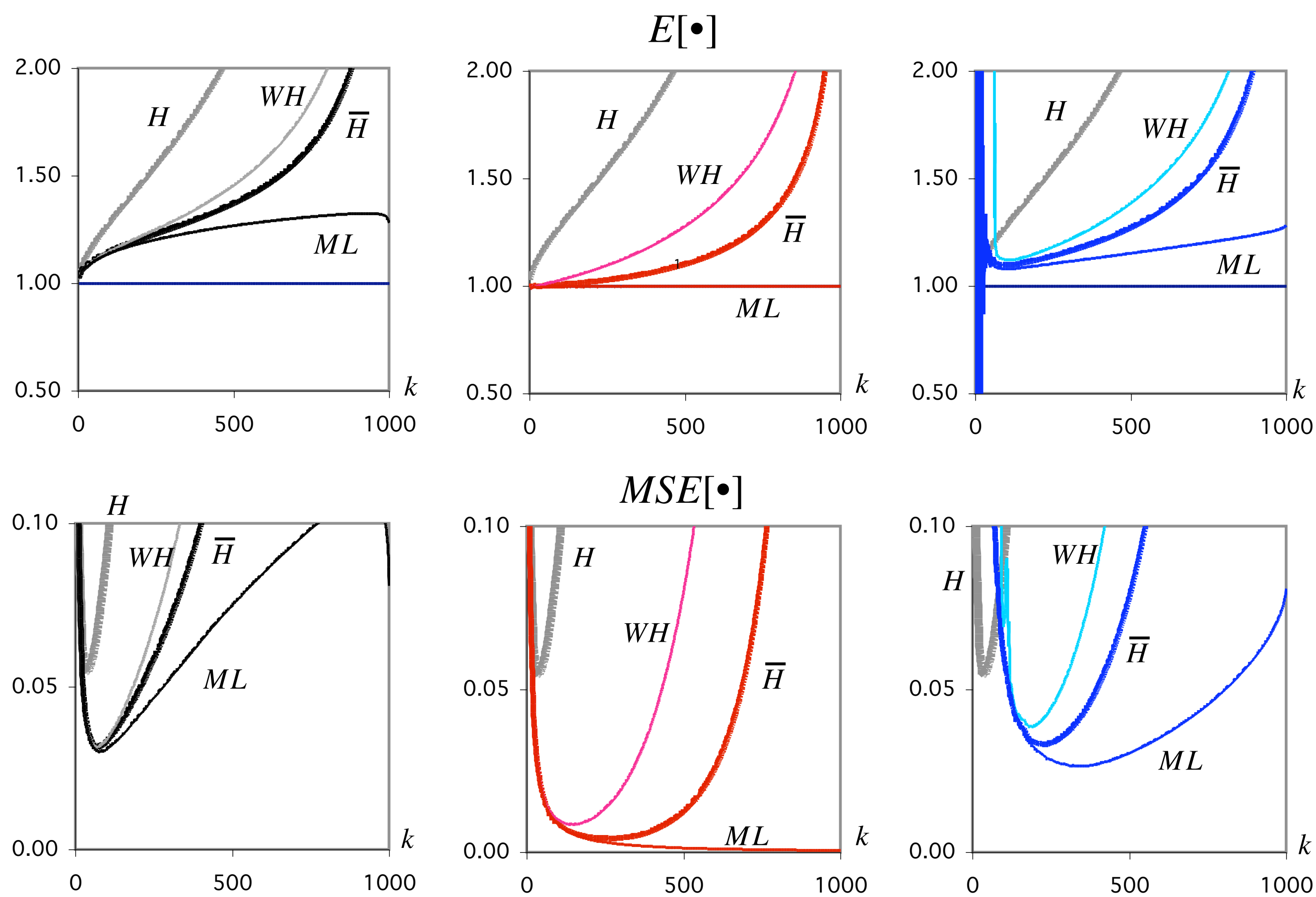
Mean values and MSE patterns. We have implemented simulation experiments with 5000 runs, estimating β at the same level $k_1 = \min\left(n-1, [2n/\ln_2 n]\right)$ we have used for the estimation of ρ and at the level k used for the tail index estimation. These estimators of ρ and β have been incorporated in the *"Unbiased Hill"*-estimators.

The tail index estimators $UH_{\hat{\beta}_{j1},\ \hat{\rho}_j}(k)$, $j = 0$ or 1, according as $|\rho| \leq 1$ or $|\rho| > 1$, seem to work reasonably well, as illustrated in the following figures, where we picture, for different underlying models, and a sample size $n = 1000$, mean values ($E[\bullet]$) and the MSEs ($MSE[\bullet]$).

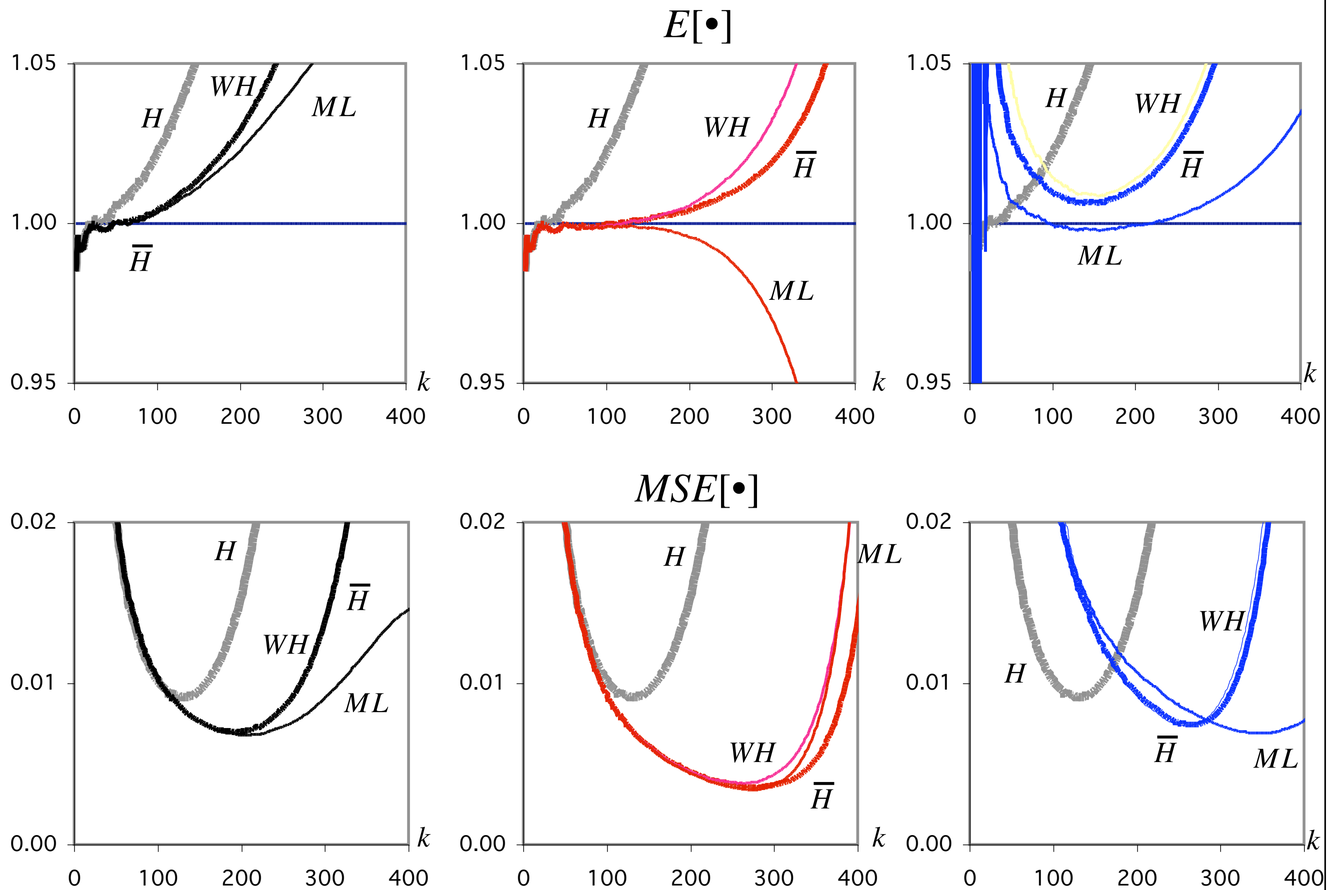
Finite sample behaviour of the estimators.



Underlying *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).



Underlying *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

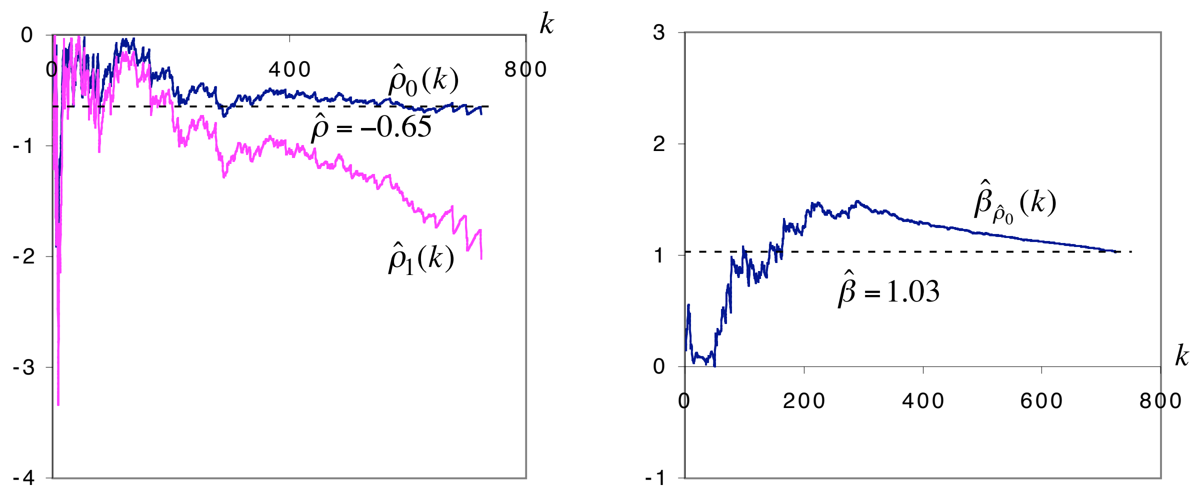


Underlying *Student* parent with with $\nu = 1$ degrees of freedom ($\gamma = 1$ and $\rho = -2$).

An overall conclusion. The main advantage of these estimators lies on the fact that we may estimate β and ρ adequately through $\widehat{\beta}$ and $\widehat{\rho}$ so that the MSE of the new estimator is smaller than the MSE of Hill's estimator for all k, even when $|\rho| > 1$, a region where has been difficult to find alternatives for the Hill estimator. And this happens together with a higher stability of the sample paths around the target value γ. These new estimators work indeed better than the Hill estimator for all values of k, contrarily to the alternatives so far available in the literature.

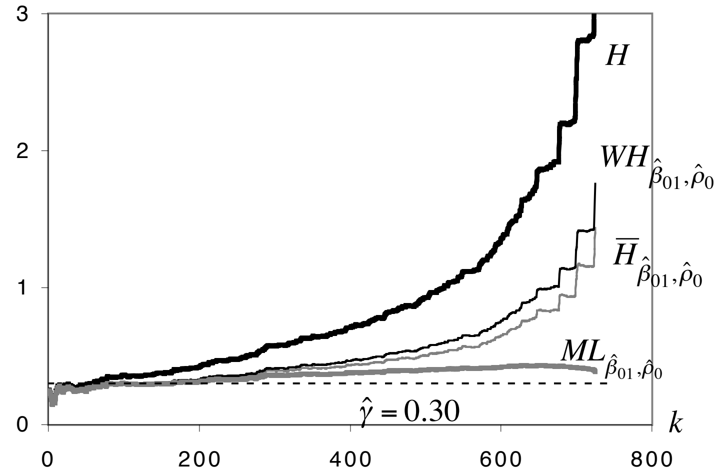
A case-study. We shall here consider the performance of the above mentioned estimators in the analysis of the Euro-UK Pound daily exchange rates from January 4, 1999 until December 14, 2004.

In the following figure, working with the $n_0 = 725$ positive log-returns, we picture the sample paths of $\widehat{\rho}_\tau(k)$, $\tau = 0$, and 1 (*left*), together with the sample paths of $\widehat{\beta}_{\widehat{\rho}_0}(k)$, for $\tau = 0$, as functions of k, together with the estimates, $\widehat{\rho}_0 = \widehat{\rho}_0(725) = -0.65$ and $\widehat{\beta}_0 = \widehat{\beta}_{\widehat{\rho}_0}(725) = 1.03$.



Remark 11. *The sample paths of the ρ-estimates associated to $\tau = 0$ and $\tau = 1$ lead us to choose, on the basis of any stability criterion, the estimate associated to $\tau = 0$. From the experience we have with this class of estimates, this means that $|\rho| \le 1$, and indeed, $\widehat{\rho}_0 = \widehat{\rho}_0(725) = -0.65$. The use of $\widehat{\beta}_{\widehat{\rho}_0}(k)$, computed at the level k_1, leads then us to the estimate $\widehat{\beta}_0 = 1.03$.*

Tail index estimates of the Log-returns are next presented:



Remark 12. *The Hill estimator exhibits a relevant bias, as may be seen from this figure. We are for sure a long way from the strict Pareto model. The ML statistic is the one with smallest bias, among the statistics considered.*

How to estimate γ? We have obtained estimates of the second order parameters β and ρ, and we may thus proceed to the estimation of the optimal k for the Hill estimator:

$$\boxed{\widehat{k}_0^H = 56 \implies \widehat{\gamma} = 0.2986.}$$

For these UH-reduced bias' estimators, we have not yet ways to estimate the optimal levels.

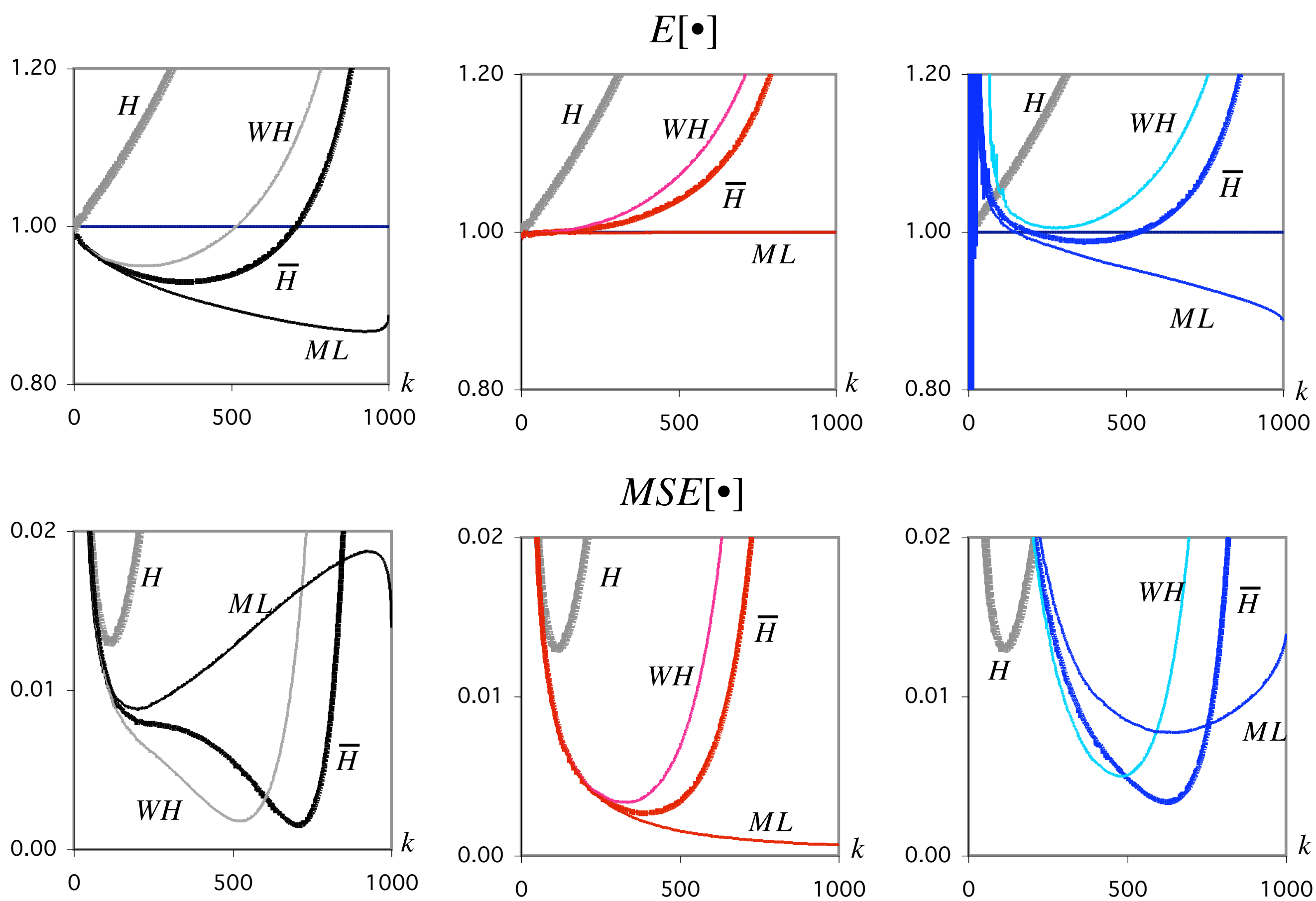
But we know that any estimate considered on the basis of $ML_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$ (or any of the other two reduced bias' statistics) performs for sure better than the estimate based on $H(k)$ for any level k. Here, we represent the estimate $\widehat{\gamma} \equiv \widehat{\gamma}_{_{ML}} =$ 0.30, the median of the ML estimates, for an adequate region of thresholds. If we use this same criterion on the estimates WH and $\overline{H}$ we are also led to the same estimate, $\widehat{\gamma}_{_{WH}} \equiv \widehat{\gamma}_{_{\overline{H}}} = 0.30$.

Remark 13. *Another possible way to find an adequate estimate of γ is to consider the largest run criterion suggested in Gomes et al. (2004a): let us consider a set of tail index estimates $\widehat{\gamma}_i(k)$, $1 \le k < n$, $i \in \mathcal{I}$, based on the observed sample of size n. Consider those estimates with a small number r of decimal figures, and denote them $\widehat{\gamma}_{i|r}(k)$. For any value $i \in \mathcal{I}$ and for any possible value a in the domain of $\widehat{\gamma}_{i|r}(k)$, consider the largest run associated with a, i.e., $R_i(a)$, the maximum number of consecutive k values such that $\widehat{\gamma}_{i|r}(k) = a$. Compute $a_i^{^M} := \arg\max_a R_i(a)$. Consider as a data-driven estimate of the tail index γ, $\widehat{\gamma} = a_{i_0}^{^M}$ with $i_0 := \arg\max_i a_i^{^M}$.*

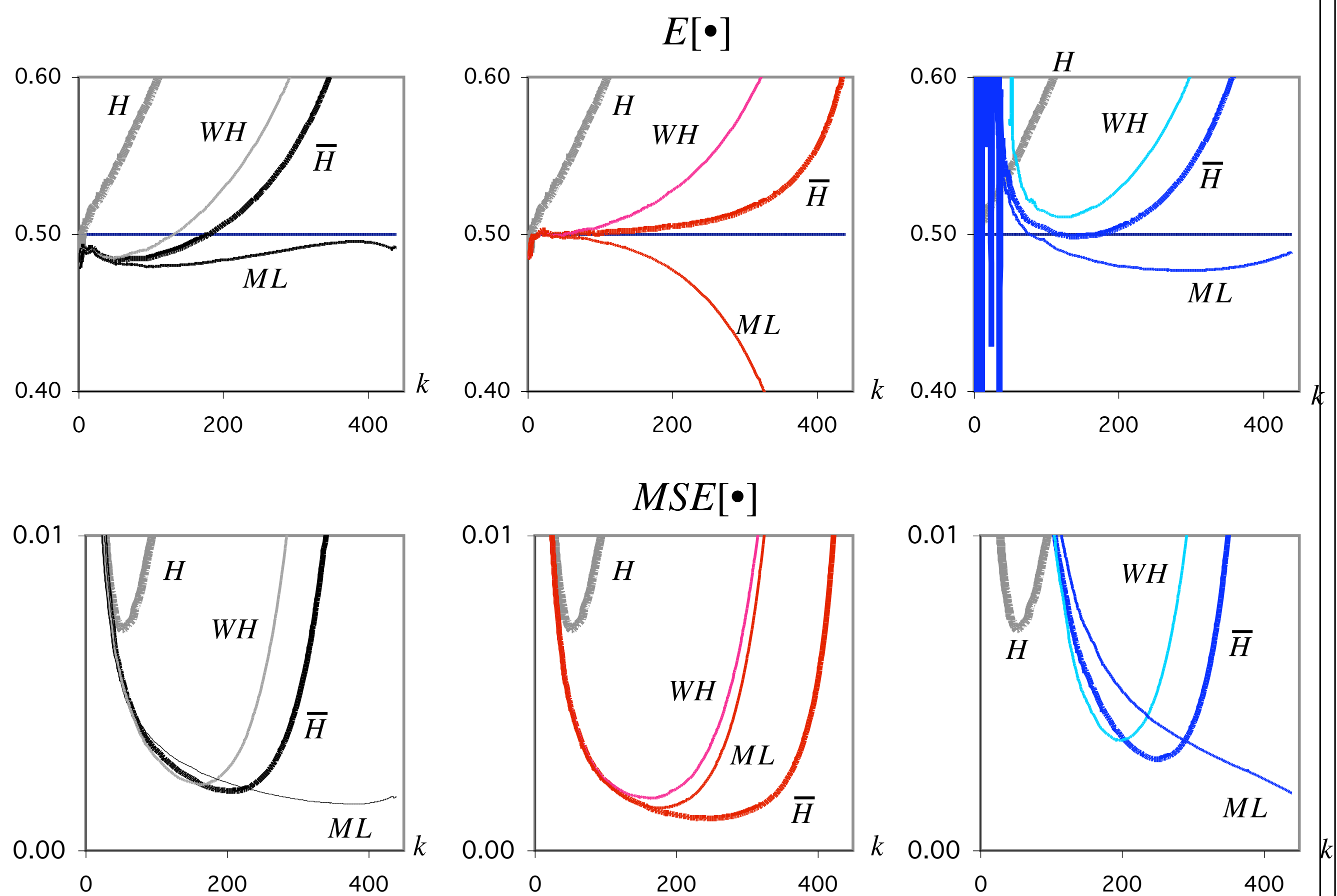
If we consider this same criterion, but the estimates with one decimal figure only, and if we go to concordance regions, related to runs of the reduced bias' estimates with one decimal figure, we are led to the estimate $\widehat{\gamma} = 0.3$ in the region $288 \leq k \leq 380$. If, in this region we now consider the estimates with two decimal figures, the $\overline{H}$-estimates provide an estimate equal to 0.30, with a run of size 20 ($145 \leq k \leq 166$), the WH-estimates provide an estimate equal to 0.32, with a run of size 32 ($169 \leq k \leq 198$) and the ML-estimates provide an estimate equal to 0.29, with a run of size 30 ($113 \leq k \leq 142$). But the three reduced bias estimates are equal to 0.30, between $k = 80$ and $k = 97$, i.e., provides us with a joint run of size 18. We have thus decided for the choice $\widehat{\gamma} = 0.3$, the one pictured in the previous figure.

References

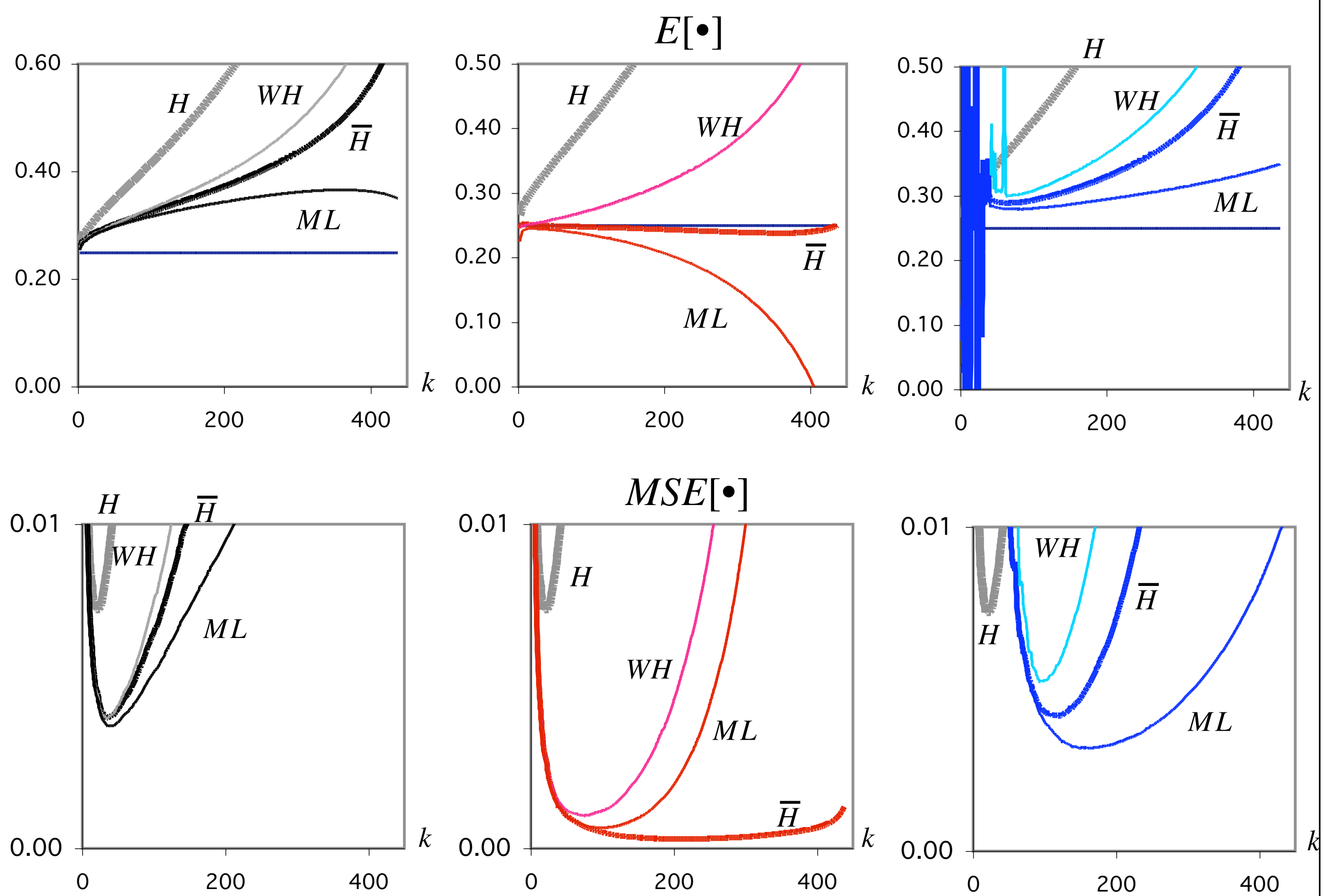
1. Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2**, 177-200.
2. Caeiro, F., Gomes, M. I. and Rodrigues, L. (2005). *A comparative study of two classes of bias reduced estimators under a third order framework*. Notas e Comunicações CEAUL 04/05.
3. Caeiro, F. Gomes, M. I. and Pestana, D. D. (2004). *Direct reduction of bias of the classical Hill estimator.* Notas e Comunicações CEAUL 16/04. Submitted.
4. Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* **27**, 760-781.
5. Fraga Alves, M. I., Gomes, M. I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**:1, 193-213.
6. Gomes, M. I., Caeiro, F. and Figueiredo, F. (2004a). Bias reduction of a tail index estimator trough an external estimation of the 2nd order parameter. *Statistics* **38**(6), 497-510.
7. Gomes, M. I. Haan, L. de and Rodrigues, L. (2004b). *Tail index estimation through accommodation of bias in the weighted log-excesses*. Notas e Comunicações C.E.A.U.L. 14/2004. *Submitted*.
8. Gomes, M. I. and Martins, M. J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**:1, 5-31.
9. Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.
10. Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4**, 191-231.



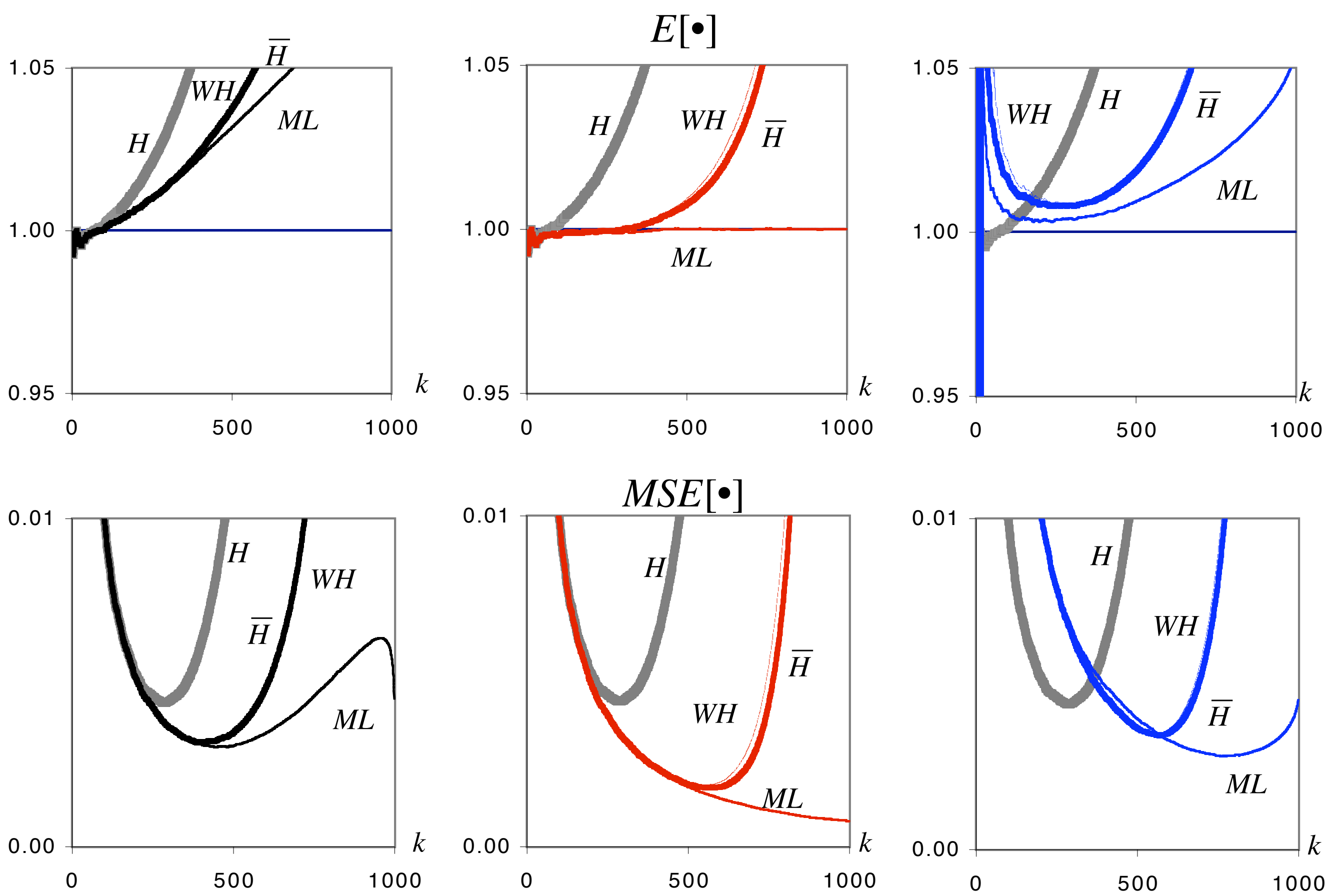
Underlying *Burr* parent with $\gamma = 1$ and $\rho = -1$.



Underlying *Student* parent with $\nu = 2$ degrees of freedom ($\gamma = 0.5$ and $\rho = -1$).



Underlying *Student* parent with $\nu = 4$ degrees of freedom ($\gamma = 0.25$ and $\rho = -0.5$).



Underlying *Burr* parent with $\gamma = 1$ and $\rho = -2$.

Remark 14. *Note that the comment done in Remark 4 is coherent with the pictures of the mean values, the one of $\overline{H}$ staying in between that of WH (above) and that of ML (below).*

Remark 15. *For the Fréchet model, and among the $UH_{\widehat{\beta},\ \widehat{\rho}}$ estimators, the $ML_{\widehat{\beta},\ \widehat{\rho}}$ statistic is the one exhibiting the worst performance in terms of bias and minimum MSE. The $WH_{\widehat{\beta},\ \widehat{\rho}}$ estimator exhibits the best performance among the three statistics considered. Things work the other way round, either with the r.v.'s $UH_{\beta,\ \rho}$ or with the statistics $UH_{\widehat{\beta}_{\widehat{\rho}}(k),\ \widehat{\rho}}$.*

Remark 16. *For a Burr model and for any of the estimators considered, $BIAS/\gamma$ and MSE/γ^2 are independent of γ, for every ρ. We may further draw the following comments, whenever we work with a Burr underlying model:*

- *The ML statistic behaves as a really unbiased estimator of γ, should we get to know the true value of β and ρ. Indeed $b_{_{ML}} = 0$ (see Remark 3).*
- *For values of $\rho > -1$, the ML-statistic is better than the $\overline{H}$-statistic, which on its turn behaves better than the WH-statistic, both regarding bias and MSE and in all situations.*

- For $\rho = -1$, the same pattern appears if we consider β and ρ known. If we estimate β and ρ through $\widehat{\beta}_{01}$ and $\widehat{\rho}_{01}$, the ML-statistic is the worst one; the $\overline{H}$ statistic is the best one regarding MSE at the optimal level, but the WH-statistic is the one with the smallest bias for not too large values of k. If we estimate only ρ through $\widehat{\rho}_0$, the $\overline{H}$ statistic is the best one, followed by the WH-statistic, being the ML-statistic the worst one, both in terms of bias, as well as minimal MSE.

- For $\rho < -1$, we need to use $\widehat{\rho}_1$. In all the simulated cases the ML-statistic is always the best one, being the $\overline{H}$ and the WH-statistics almost equivalent.

Remark 17. *For a Student model with ν degrees of freedom (Figures 4, 6 and 8), and whenever we assume β and ρ known, the most stable sample path around the target value γ is achieved by the $\overline{H}$-statistic. And such a fact leads this statistic to have the smallest mean squared error, followed by the ML and next the WH statistics, for all values of ν.*

If we need to estimate β and ρ, the ML-statistic is the one with the smallest MSE at the optimal level, also for every ν. Next comes the $\overline{H}$-statistic, quite close to the WH-statistic when $\nu < 2$, i.e. $\rho < -1$.

Remark 18. *The discrepancy, in some of the models, between the behaviour of the estimators under study, in the left figures, and the r.v.'s in the central ones, suggests that some improvement in the estimation of second order parameters β and ρ is still welcome.*