Estimating the Coefficient of Asymptotic Tail Independence: a Comparison of Methods

Marta Ferreira[1]

Abstract

Many multivariate analyses require the account of extreme events. Correlation is an insufficient measure to quantify tail dependence. The most common tail dependence coefficients are based on the probability of simultaneous exceedances. The coefficient of asymptotic tail independence introduced in Ledford and Tawn ([18] 1996) is a bivariate measure often used in the tail modeling of data in finance, environment, insurance, among other fields of applications. It can be estimated as the tail index of the minimum component of a random pair with transformed unit Pareto marginals. The literature regarding the estimation of the tail index is extensive. Semi-parametric inference requires the choice of the number k of the largest order statistics that lead to the best estimate, where there is a tricky trade-off between variance and bias. Many methodologies have been developed to undertake this choice, most of them applied to the Hill estimator (Hill, [16] 1975). We are going to analyze, through simulation, some of these methods within the estimation of the coefficient of asymptotic tail independence. We also compare with a minimum-variance reduced-bias Hill estimator presented in Caeiro *et al.* ([3] 2005). A pure heuristic procedure adapted from Frahm *et al.* ([13] 2005), used in a different context but with a resembling framework, will also be implemented. We will see that some of these simple tools should not be discarded in this context. Our study will be complemented by applications to real datasets.

1   Introduction

It is undeniable that extreme events have been occurring in areas like environment (e.g. climate changes due to pollution and global heating), finance (e.g., market crashes due to less regulation and globalization), telecommunications (e.g., growing traffic due to a high expanding technological development), among others. Extreme values are therefore the subject of concern of many analysts and researchers, who have come to realize that they should be dealt with some care, requiring their own treatment. For instance, the classical linear correlation is not a suitable dependence measure if the dependence characteristics in the tail differ from the remaining realizations in the sample. An illustration is addressed in Embrechts *et al.* ([9] 2002). To this end, the tail dependence coefficient (TDC) defined in

---

[1] Center of Mathematics of University of Minho, Center for Computational and Stochastic Mathematics of University of Lisbon and Center of Statistics and Applications of University of Lisbon, Portugal; msferreira@math.uminho.pt

Joe ([17] 1997), usually denoted by λ, is more appropriate. More precisely, for a random pair (X, Y) with respective marginal distribution functions (dfs) F_X and F_Y, the TDC is given by

$$\lambda = \lim_{t \downarrow 0} P(F_Y(Y) > 1 - t | F_X(X) > 1 - t), \qquad (1.1)$$

whenever the limit exists. Roughly speaking, the TDC evaluates the probability of one variable exceeding a large value given that the other also exceeds it. A positive TDC means that X and Y are tail dependent and whenever null we conclude the random pair is tail independent. In this latter case, the rate of convergence towards zero is a kind of residual tail dependence that, once ignored, may lead to an under-estimation of the risk underlying the simultaneous exceedance of a large value. On the other hand, by considering that the random variables (rv's) X and Y are tail dependent when they are actually asymptotically independent, it will result in an over-estimation of such risk. The degree of misspecification depends on the degree of asymptotic independence given by the mentioned rate of convergence, denoted η in Ledford and Tawn ([18] 1996). More precisely, it is assumed that

$$P(F_X(X) > 1 - t, F_Y(Y) > 1 - t) = t^{1/\eta} L(t), \ \eta \in (0, 1], \qquad (1.2)$$

where $L(t)$ is a slowly varying function at zero, i.e., $L(tx)/L(t) \to 1$ as $t \downarrow 0$ for all $x > 0$. We call the parameter η the coefficient of asymptotic tail independence. Whenever $\eta < 1$, X and Y are asymptotically independent and, if $\eta = 1$, asymptotic dependence holds if $L(t) \to c > 0$, as $t \downarrow 0$. In case X and Y are exactly independent then $\eta = 1/2$ and we can also discern between asymptotically vanishing negative dependence and asymptotically vanishing positive dependence if, respectively, $\eta \in (0, 1/2)$ and $\eta \in (1/2, 1)$. Observe that we can state (1.2) as

$$P\left(\min\left(\frac{1}{1 - F_X(X)}, \frac{1}{1 - F_Y(Y)}\right) > t\right) = t^{-1/\eta} L(1/t), \qquad (1.3)$$

and thus η corresponds to the tail index of the minimum of the two marginals standardized as unit Pareto. The tail index, also denoted extreme value index, quantifies the "weight" of the tail of a univariate distribution: whenever negative, null or positive it means that the tail of the underlying model is, respectively, "light", "exponential" or "heavy". In what concerns univariate extreme values, it is the primary parameter as it is implicated in all other extremal parameters, such as, extremal quantiles, right end-point of distributions, probability of exceedance of large levels, as well as return periods, among others. Therefore, the estimation of the tail index is a crucial issue, with numerous contributions in the literature. A survey on this topic can be seen, for instance, in Beirlant *et al.* ([2] 2004).

Under a semi-parametric framework in the domain of heavy tails, the Hill estimator, introduced in Hill ([16] 1975), have proved to possess good properties, being an essential tool in any application on this topic. For a random sample $(T_1, \ldots, T_n)$, the Hill estimator corresponds to the sample mean of the log-excesses of the $k + 1$ larger order statistics $T_{n:n} \geq \ldots \geq T_{n-k:n}$, i.e.,

$$H_n(k) \equiv H(k) := \frac{1}{k} \sum_{i=1}^{k} \log \frac{T_{n-i+1:n}}{T_{n-k:n}}, \ 1 \leq k < n, \qquad (1.4)$$

Consistency requires that k must be intermediate, that is, a sequence of integers $k \equiv k_n$, $1 \le k < n$, such that

$$k_n \to \infty \text{ and } k_n/n \to 0, \text{ as } n \to \infty.$$

There is no definite formula to obtain k and it must be chosen not too small to avoid high variance but also not to large to prevent high bias. Figure 1 illustrates this issue, particularly the dashed line corresponding to a unit Frchet model where the tail index is 1. Observe also that there is a kind of stable area of the sample path around the true value of the tail index, where the variance is no longer high and the bias haven't started to increase. This disadvantage is transversal to the semi-parametric tools concerning extreme values inference. In the particular case of the Hill estimator, many efforts have been made to minimize the problem, ranging from bias-corrected versions to the implementation of procedures to compute k. The minimum-variance reduced-bias (MVRB) Hill estimator presented in Caeiro *et al.* ([3] 2005; see also Neves *et al.* [21] 2015) was developed for the Hall-Welsh class (within Generalized Pareto distributions), with reciprocal quantile function

$$F^{-1}(1 - 1/x) = Cx^{\gamma} \left(1 + \gamma \beta x^{\rho}/\rho + o(x^{\rho})\right), \, x \to \infty, \tag{1.5}$$

where $\gamma > 0$ is the tail index of model F, $C > 0$, and $\beta \neq 0$ and $\rho < 0$ are second order parameters. The MVRB Hill estimator is given by

$$CH_n(k) \equiv CH(k) := H(k) \left(1 - \frac{\widehat{\beta}(n/k)^{\widehat{\rho}}}{1 - \widehat{\rho}}\right), \, 1 \le k < n, \tag{1.6}$$

where $\widehat{\beta}$ and $\widehat{\rho}$ are suitable estimators of β and ρ, respectively. Details about these latter are addressed in Caeiro *et al.* ([4] 2009) and references therein. We will denote it "corrected Hill" (CH). Our aim is to compare, through simulation, several methods regarding the Hill and corrected Hill estimators applied to the estimation of η. We also consider the graphical and pure heuristic procedure presented in Frahm *et al.* ([13] 2005) in the context of estimating the TDC λ in (1.1), also relying on the choice of k upper order statistics with the same bias/variance controversy. All the estimation procedures are described in Section 2. The simulation study is conducted in Section 3 and applications to real datasets appear in Section 4. A small discussion ends this work in Section 5.

2 Estimation methods

In this section we describe the procedures that we are going to consider in the estimation of the coefficient of asymptotic tail independence η given in (1.3) and therefore corresponding to the tail index of

$$T = \min((1 - F_X(X))^{-1}, (1 - F_Y(Y))^{-1}). \tag{2.1}$$

Coefficient η is positive and we can use positive tail index estimators such as Hill. Observe that T is the minimum between two unit Pareto r.v.'s Alternatively, we can also undertake
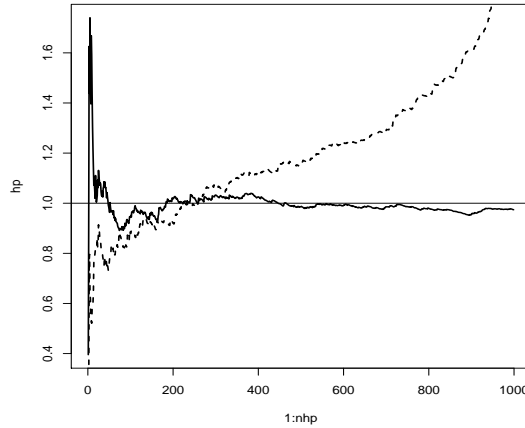
Figure 1: Hill plots of 1000 realizations of a unit Pareto (full line) and a unit Fréchet (dashed line), both with tail index equal to 1 (horizontal line).

a unit Frchet marginal transformation since $1 - F_X(X) \sim - \log F_X(X)$. However, in the sequel, we prosecute with unit Pareto marginals, since the Hill estimator has smaller bias in the Pareto models than in the Frchet ones (see Figure 1; see also Draisma *et al.* [6] 2004 and references therein). In order to estimate the unknown marginal df's F_X and F_Y we consider their empirical counterparts (ranks of the components), more precisely,

$$T_i^{(n)} := \min((n+1)/(n+1-R_i^X), (n+1)/(n+1-R_i^Y)), \ i = 1, \ldots, n$$

where R_i^X denotes the rank of X_i among $(X_1, \ldots, X_n)$ and R_i^Y denotes the rank of Y_i among $(Y_1, \ldots, Y_n)$.

The estimation of η through the tail index estimators Hill and maximum likelihood (Smith, [24] 1987) was addressed in Draisma *et al.* ([6] 2004). Other estimators were also considered in Poon *et al.* ([23] 2003; see also references therein) and more recently in Goegebeur and Guillou ([14] 2013) and Dutang *et al.* ([8] 2014). However, no method was analyzed in order to attain the best choice of k in estimation.

In the domain of positive tail indexes, the Hill estimator is the most widely studied and many developments have been appearing around it. The main topics concern methods to obtain the value of k related to the number of tail observations to use in estimation and procedures to control the bias without increasing the variance. The corrected Hill version in (1.6), for instance, removes from Hill its dominant bias component estimated by $H(k)(\widehat{\beta}(n/k)^{\widehat{\rho}})/(1 - \widehat{\rho})$.

In the following, we describe the methods developed in literature for the Hill estimator to compute the value of k, that will be used to estimate η (the tail index of rv T in (2.1)) in our simulation study.

Based on Beirlant *et al.* ([1] 2002) and little restrictive conditions on the underlying

model, we have

$$Y_i := (i+1) \log \frac{T_{n-i:n}^{(n)} H(i)}{T_{n-(i+1):n}^{(n)} H(i+1)} = \eta + b(n/k) \left(\frac{i}{k}\right)^{-\rho} + \epsilon_i, \ i = 1, ..., k, \quad (2.2)$$

where the error term ϵ_i is zero-centered and b is a positive function such that $b(x) \to 0$, as $x \to \infty$. Extensive simulation studies conclude that the results tend to be better when ρ is considered fixed, even if misspecified. Matthys and Beirlant ([19] 2000) suggest $\rho = -1$. From model (2.2), the resulting least squares estimators of η and $b(n/k)$ are given by

$$\widetilde{\eta}_{k,n}^{LS} = \overline{Y}_k - \widetilde{b}_{k,n}^{LS}/(1-\rho) \quad \text{and} \quad \widetilde{b}_{k,n}^{LS} = \frac{(1-\rho)^2(1-2\rho)}{\rho^2} \frac{1}{k} \sum_{i=1}^{k} \left(\left(\frac{i}{k}\right)^{-\rho} - \frac{1}{1-\rho} \right) Y_i. \quad (2.3)$$

Thus, by replacing these estimates in the Hill's asymptotic mean squared error (AMSE)

$$\text{AMSE}(H(k)) = \frac{\eta^2}{k} + \left(\frac{b(n/k)}{1-\rho} \right)^2,$$

we are able to compute $\widehat{k}_{opt}^1$ as the value of k that minimizes the obtained estimates of the AMSE and estimate η as $H(\widehat{k}_{opt}^1)$.

On the other hand, we can compute the approximate value of k that minimizes the AMSE, given by

$$k_{opt} \sim b(n/k)^{-2/(1-2\rho)} k^{-2\rho/(1-2\rho)} \left(\frac{\eta^2(1-\rho)^2}{-2\rho} \right)^{1/(1-2\rho)}. \quad (2.4)$$

See, e.g., Beirlant *et al.* ([1] 2002). Replacing again η and $b(n/k)$ by the respective least squares estimates in (2.3) with fixed $\rho = -1$, we derive $\widehat{k}_{opt,k}$, for $k = 3, ..., n$, using (2.4). Then compute $\widehat{k}_{opt}^2 = median\{\widehat{k}_{opt,k}, \ k = 3, ..., \lfloor \frac{n}{2} \rfloor\}$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x and consider η estimated by $H(\widehat{k}_{opt}^2)$.

Further reading of the methods is referred to Beirlant *et al.* ([1] 2002), Matthys and Beirlant ([19] 2000) and references therein. In the sequel, they are shortly denoted, respectively, AMSE and KOPT.

The adaptive procedure of Drees and Kaufmann ([6] 1998) looks for the optimum k under which the bias starts to dominate the variance. The method is developed for the Hall-Welsh class of models defined in (1.5), for which it is proved that the maximum random fluctuation of $\sqrt{i}(H(i) - \eta)$, $i = 1, ..., k - 1$, with $k \equiv k_n$ an intermediate sequence, is of order $\sqrt{\log \log n}$. More precisely, for ρ fixed at -1, we have:

1. Consider $r_n = 2.5 \times \widetilde{\eta} \times n^{0.25}$, with $\widetilde{\eta} = \widehat{\eta}_{2\sqrt{n},n}$.

2. Calculate $\widetilde{k}(r_n) := \min\{k = 1, ..., n - 1 : \max_{i<k} \sqrt{i}|H(i) - H(k)| > r_n\}$. If $\sqrt{i}|H(i) - H(k)| > r_n$ doesn't hold for any k, consider $0.9 \times r_n$ to r_n and repeat step 2, otherwise move to step 3.

3. For $\varepsilon \in (0, 1)$, usually $\varepsilon = 0.7$, obtain

$$\widehat{k}_{\text{DK}} = \left\lfloor \frac{1}{3}(2\widetilde{\eta}^2)^{1/3} \left(\frac{\widetilde{k}(r_n^\varepsilon)}{(\widetilde{k}(r_n))^\varepsilon} \right)^{1/(1-\varepsilon)} \right\rfloor$$

This method will be shortly referred DK.

Sousa and Michailidis (2004) method is based on the Hill sum plot, (k, S_k), $k = 1, ..., n-1$, where $S_k = kH(k)$. We have $E(S_k) = k\eta$, an thus the sumplot must be approximately linear for the values of k where $H(k) \approx \eta$, with the respective slope being an estimator of η. The method essentially seeks the breakdown of linearity. Their approach is based on a sequential testing procedure implemented in McGee and Carleton ([20] 1970), leaning over approximately Pareto tail models. More precisely, consider the regression model $y = X\eta + \delta$, with $y = (S_1, ..., S_k)'$, $X = [1\,i]_{i=1}^{k}$ and δ the error term. It is checked the null hypothesis that a new point y_0 is adjacent to the left or to the right of the set of points $y = (y_1, ..., y_k)$, through the statistics

$$TS = s^{-2}\left((y_0 - \widehat{y}_0^*)^2 + \sum_{i=1}^{k}(\widehat{y}_i - \widehat{y}_i^*)^2\right),$$

where $*$ denotes the predictions based on $k+1$ and $s^2 = (k-2)^{-1}(y'y - \widehat{\eta}X'y)$. Since TS is approximately distributed by $F_{1,k-2}$, the null hypothesis is rejected if TS is larger than the $(1-\alpha)$-quantile, $F_{1,k-2;1-\alpha}$. The method, shortly denoted SP from now on, is described in the following algorithm:

1. Fit a least-squares regression line to the initial $k = \nu n$ upper observations, $y = [y_i]_{i=1}^{k}$ (usually $\nu = 0.02$).

2. Using the test statistic TS, determine if a new point $y_0 = y_j$ for $j > k$, belongs to the original set of points y. Go adding points until the null hypothesis is rejected.

3. Consider $k_{\texttt{new}} = \max(0, \{j : TS < F_{1,k-2;1-\alpha}\})$. If $k_{\texttt{new}} = 0$, no new points are added to y and thus move forward to step 4. Return to step 1. if $k_{\texttt{new}} > 0$ by considering $k = k_{\texttt{new}}$.

4. Estimate η by considering the obtained k.

The heuristic procedure introduced in Gomes *et al.* ([15] 2013), searches for the supposed stable region encompassing the best k to be estimated. More precisely, we need first to obtain the minimum value j_0, such that the rounded values to j decimal places of $H(k)$, $1 \leq k < n$, denoted $H(k; j)$ are not all equal. Identify the set of values of k associated to equal consecutive values of $H(k; j_0)$. Consider the set with largest range $\ell := k_{\max} - k_{\min}$. Take all the estimates $H(k; j_0 + 2)$ with $k_{\max} \leq k \leq k_{\min}$, i.e., the estimates with two additional decimal points and calculate the mode. Consider $\mathcal{K}$ the set of k-values corresponding to the mode. Take $H(\widehat{k})$, with $\widehat{k}$ being the maximum of $\mathcal{K}$. Since it was specially designed for reduced-bias estimators, we shortly referred it as RB method hereinafter.

Frahm *et al.* ([13] 2005) also presented a heuristic procedure that can be applied to all estimators depending on a number k of rv's whose choice bears the mentioned trade-off between bias and variance. Indeed is was developed within the estimation of the TDC λ defined in (1.1). It was adapted to the Hill estimator in Ferreira ([11, 12] 2014, 2015) as follows:

1. Smooth the Hill plot $(k, H(k))$ by taking the means of $2b + 1$ successive points, $\overline{H}(1), ..., \overline{H}(n - 2b)$, with bandwidth $b = \lfloor w \times n \rfloor$.

2. Define the regions $p_k = (\overline{H}(k), ..., \overline{H}(k + m - 1))$, $k = 1, ..., n - 2b - m + 1$, with length $m = \lfloor \sqrt{n - 2b} \rfloor$. The algorithm stops at the first region satisfying

$$\sum_{i=k+1}^{k+m-1} \left| \overline{H}(i) - \overline{H}(k) \right| \le 2s,$$

   where s is the empirical standard-deviation of $\overline{H}(1), ..., \overline{H}(n - 2b)$.

3. Consider the chosen plateau region p_k* and estimate η as the mean of the values of p_k* (consider the estimate zero if no plane region fulfills the stopping condition).

The estimation of η through the plateau method was analyzed in Ferreira and Silva ([10] 2014) with respect to the sensibility of the bandwidth. The value $w = 0.005$ seems a reasonable choice (thus each moving average in step 1. consists in 1% of the data), also suggested in Frahm *et al.* ([13] 2005). In the sequel it will be referred as plateau method (in short PLAT).

Both RB and PLAT are simultaneously graphical and free-assumption methods since they are based on the search of a plane region of the estimator's plot that presumably contains the best sample fraction k to be estimated through a totally "ad-hoc" procedure. The sumplot is also a graphical method and the remaining procedures are neither graphical nor free-assumption.

3   Simulation study

In this section we compare through simulation the performance of the methods described above within the estimation of η through the under study estimators Hill in (1.4) and corrected Hill in (1.6).

We have generated 100 runs of samples of sizes $n = 100, 1000, 5000$ from the following models:

- Bivariate Normal distribution ($\eta = (1+\rho)/2$; see, e.g., Draisma *et al.* [6] 2004); we consider correlation $\rho = -0.2$ ($\eta = 0.4$), $\rho = 0.2$ ($\eta = 0.6$) and $\rho = 0.8$ ($\eta = 0.9$); we use notation, respectively, $N(-0.2)$, $N(0.2)$ and $N(0.8)$.

- Bivariate t-Student distribution t_ν with correlation coefficient given by $\rho \ne -1$ ($\lambda = 2F_{t_{\nu+1}}\left(-\sqrt{(\nu + 1)(1 - \rho)/(1 + \rho)}\right)$, see Embrechts *et al.* [9] 2002; we have $\lambda > 0$ and thus $\eta = 1$); we consider $\nu = 4$ and $\rho = 0.25$ ($\lambda = 0.1438$) and $\nu = 1$ and $\rho = 0.75$ ($\lambda = 0.6464$); we use notation, respectively, t_4 and t_1.

- Bivariate extreme value distribution with a asymmetric-logistic dependence function $\ell(x, y) = (1 - a_1)x + (1 - a_2)y + ((a_1 x)^{1/\alpha} + (a_2 y)^{1/\alpha})^\alpha$, with $x, y \ge 0$,

dependence parameter $\alpha \in (0,1]$ and asymmetric parameters $a_1, a_2 \in (0,1]$ ($\lambda = 2 - l(1,1)$, see Beirlant *et al.* [1] 2004; we have $\lambda > 0$ and thus $\eta = 1$); we consider $\alpha = 0.7$ and $a_1 = 0.4$, $a_2 = 0.2$ ($\lambda = 0.1010$) and $\alpha = 0.3$ and $a_1 = 0.6$, $a_2 = 0.8$ ($\lambda = 0.5182$); we use notation, respectively, $AL(0.7)$ and $AL(0.3)$.

- Farlie-Gumbel-Morgenstern distribution with dependence 0.5 ($\eta = 0.5$, see Dutang *et al.* [8] 2014); we use notation $FGM(0.5)$.

- Frank distribution with dependence 2 ($\eta = 0.5$, see Dutang *et al.* [8] 2014); we use notation $Fr(2)$.

Observe that the case $N(0.8)$ is an asymptotic tail independent model close to tail dependence since $\eta = 0.9 \approx 1$. On the other hand, the cases t_4 and $AL(0.7)$ are tail dependent cases ($\eta = 1$) near asymptotic tail independence since $\lambda = 0.1438 \approx 0$ and $\lambda = 0.1010 \approx 0$, respectively. We consider these examples in order to assess the robustness of the methods in border cases.

In Figures 2 and 3 are plotted, respectively, the results of the simulated values of the absolute bias and root mean squared error (rmse), for the Hill and corrected Hill estimators, in the case $n = 1000$. All the results are presented in Table 1 concerning the Hill estimator and Table 2 with respect to the corrected Hill. Observe that this latter case requires the estimation of additional second order parameters (β and ρ). To this end, we have followed the indications in Caeiro *et al.* ([4] 2009). For the ρ estimation, there was an overall best performance whenever it was taken fixed at value -1, leading to the reported results.

The largest differences between Hill and corrected Hill can be noticed in the above mentioned border cases, with the corrected one presenting lower absolute bias and rmse. The other models also show this difference but in a small amount. We remark that we are working with the minimum of Pareto rv's and the Hill estimator is unbiased in the Pareto case. The FGM and Frank models behave otherwise with a little lower absolute bias and rmse within the Hill estimator, for either estimated or several fixed values tried for ρ.

The failure cases in the DK method (column "NF" of Tables 1 and 2) correspond to an estimate of k out of the range $\{1, \ldots, n-1\}$, which were ignored in the results. It sets up the worst performance, which may be justified by the fact that the class of models underlying the scope of application of this method excludes the simple Pareto law.

The corrected Hill exhibits better results in general, particularly for methods KOPT, PLAT and AMSE, followed by SP and RB, in large sample sizes (n¿=1000). The PLAT procedure also performs well with the Hill estimator unlike the SP.

For $n = 100$, we have good results within RB and SP based on corrected Hill. Once again, the PLAT method behaves well in both estimators.

The border cases of weak tail dependence (t_4 and $AL(0.7)$) are critical throughout all evaluated procedures and estimators. On the other hand, the methods are robust in the border case of tail independence near dependence expressed in model $N(0.8)$.

4   Applications

In this section we illustrate the methods with three datasets analyzed in literature:

Figure 2: Simulated results of the absolute bias of Hill (full) and corrected Hill (dashed), for $n = 1000$, of the models (left-to-right and top-to-down): $N(-0.2)$, $N(0.2)$, $N(0.8)$, t_4, t_1, $AL(0.3)$, $AL(0.7)$, $FGM(0.5)$ and $Fr(2)$.

- I: The data consists of closing stock index levels of S&P 500 from the US and FTSE 100 from the UK, over the period 11 December 1989 to 31 May 2000, totalizing 2733 observed pairs (see, e.g., Poon *et al.* ([23] 2003)).

- II: The wave-surge data corresponding to 2894 paired observations collected during 1971-77 in Cornwall (England); it was analyzed in Coles and Tawn ([5] 1994) and later also in Ramos and Ledford ([22] 2009) under a parametric view.

- III: The Loss-ALAE data analyzed in Beirlant *et al.* ([2] 2004; see also references therein) consisting of 1500 pairs of registered claims (in USD) corresponding to an
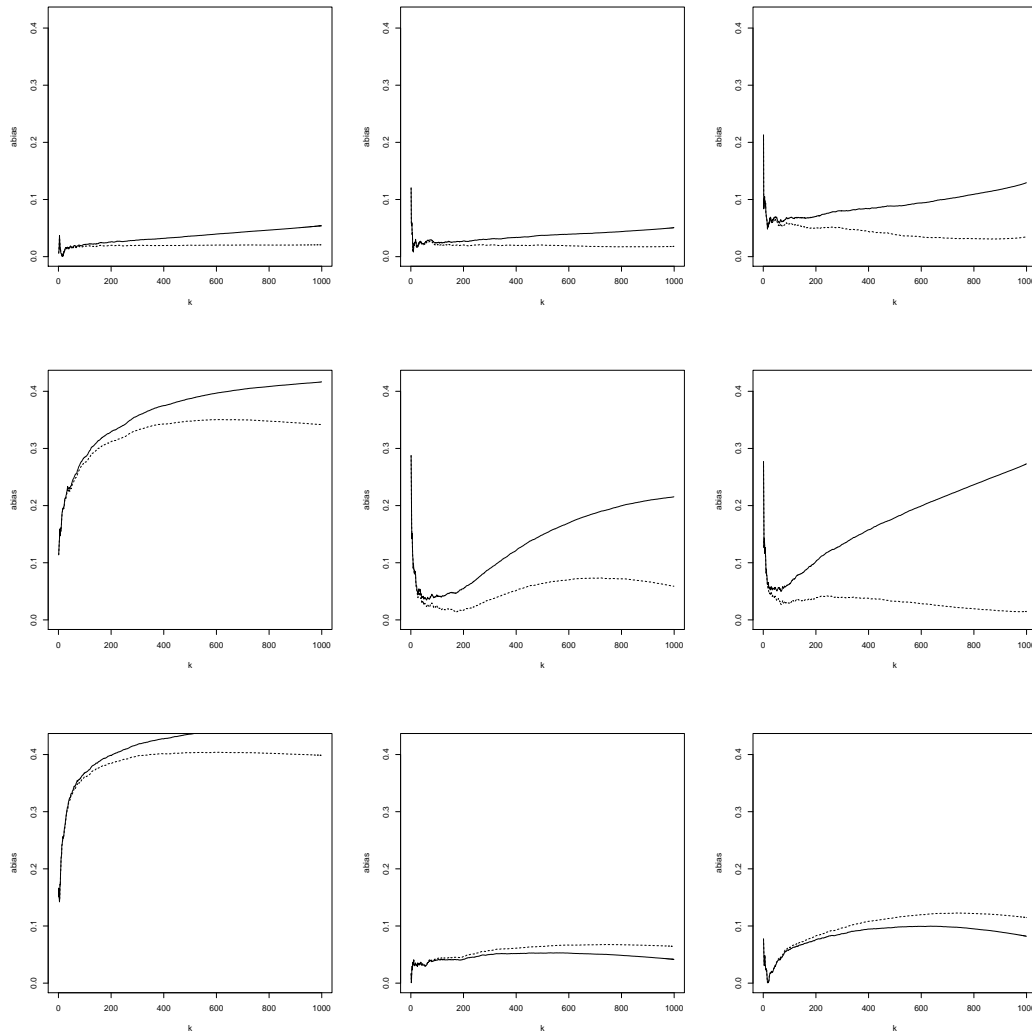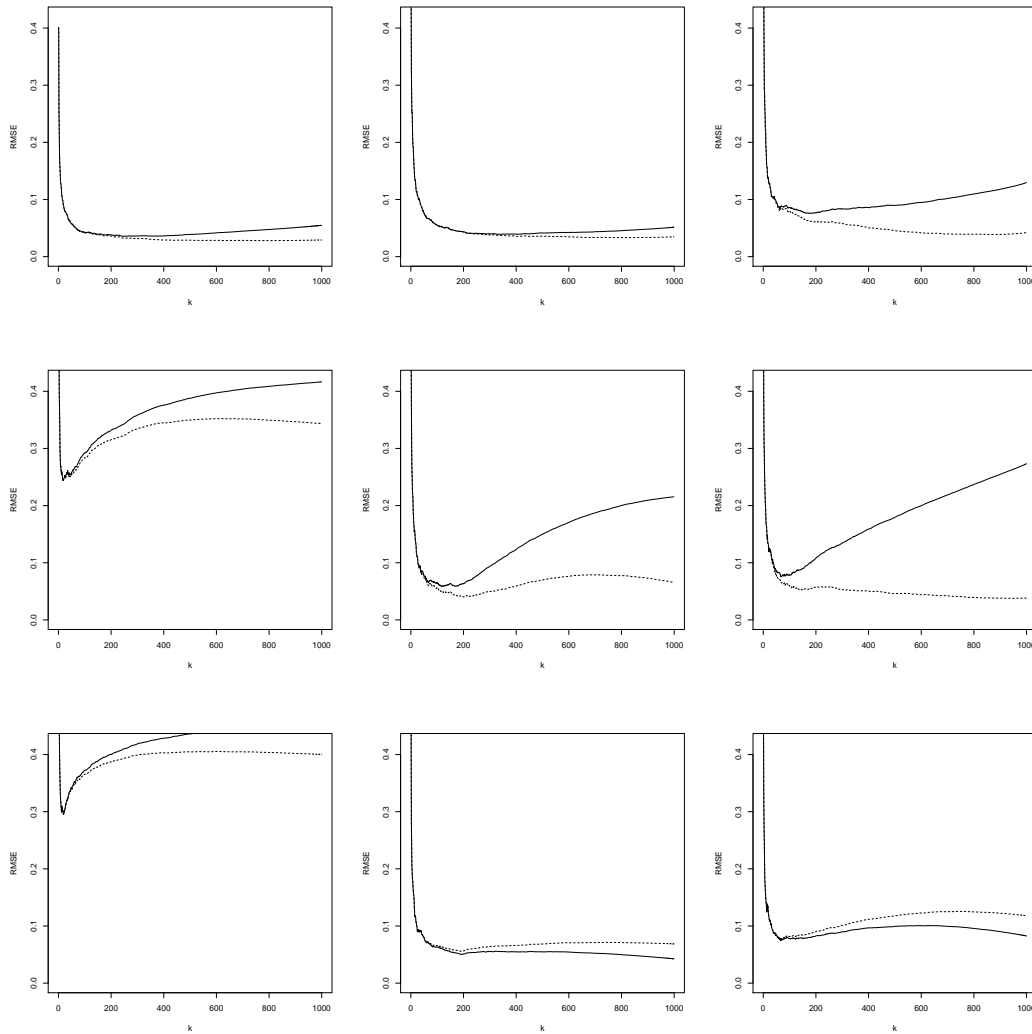
Figure 3: Simulated results of the rmse of Hill (full) and corrected Hill (dashed), for $n = 1000$, of the models (left-to-right and top-to-down): $N(-0.2)$, $N(0.2)$, $N(0.8)$, t_4, t_1, $AL(0.3)$, $AL(0.7)$, $FGM(0.5)$ and $Fr(2)$.

indemnity payment (loss) and an allocated loss adjustment expense (ALAE).

The respective scatter-plots are placed in Figure 4. For the US and UK stock market returns, the largest values in each tail for one variable correspond to reasonably large values of the same sign for the other variable, hinting an asymptotic independence but not exactly independence. In the wave-surge data, the dependence seems a bit more persistent within large values, as well as in Loss-ALAE data. The Hill and corrected Hill sample paths of η estimates are pictured in Figure 5. Table 3 reproduces the estimates obtained with each method and estimators under study. The estimation results found in literature for the financial (I), environmental (II) and insurance datasets (III) are respec-

tively approximated by 0.731, 0.85 and 0.9. The results seem to be in accordance with the simulation study.
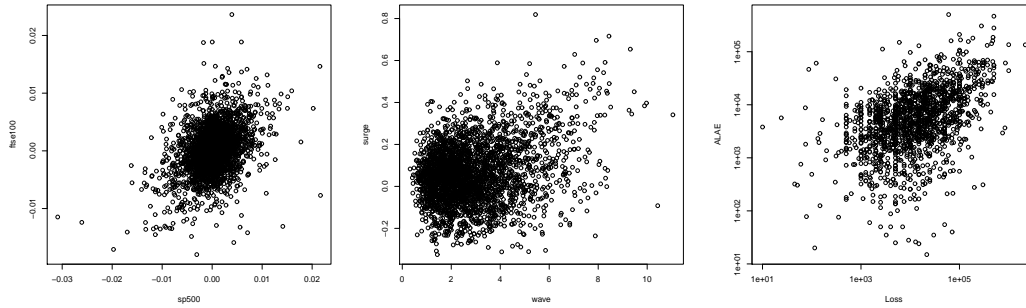
Figure 4: From left to right: scatter-plots of datasets I, II and III.
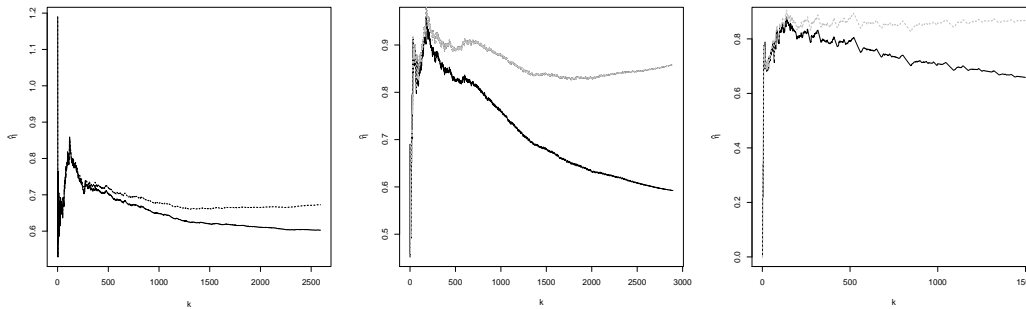
Figure 5: From left to right: sample paths of Hill (full;black) corrected Hill (dashed;grey) of datasets I, II and III.

5 Discussion

In this paper we have analyzed some simple estimation methods for the coefficient of asymptotic tail independence, with some of them revealing promising results. However, the choice of the estimator is not completely straightforward. It can be seen from simulation results that the ordinary Hill estimator may be still preferred over the corrected one in some situations. Also in boundary cases of tail dependence near independence, there are still some worrying errors to correct. These will be topics of a future research.

Acknowledgment

References

[1] Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002): On Exponential Representation of Log-Spacings of Extreme Order Statistics. *Extremes*, **5**, 157-180.

[2] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004): *Statistics of Extremes: Theory and Applications*. J. Wiley & Sons.

[3] Caeiro, F., Gomes, M.I. and Pestana, D.D. (2005): Direct reduction of bias of the classical Hill estimator. *Revstat*, **3(2)**, 111-136.

[4] Caeiro, F., Gomes, M.I. and Henriques-Rodrigues, L. (2009): Reduced-Bias Tail Index Estimators Under a Third-Order Framework. *Communications in Statistics - Theory and Methods*, **38(7)**, 1019-1040.

[5] Coles, S.G. and Tawn, J.A. (1994): Statistical methods for multivariate extremes: an application to structural design (with discussion). *Appl. Statist.*, **43**, 1-48.

[6] Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2004): Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, **10(2)**, 251-280.

[7] Drees, H. and Kaufmann, E. (1998): Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process Appl.*, **75**, 149-172.

[8] Dutang, C., Goegebeur, Y. and Guillou, A. (2014): Robust and bias-corrected estimation of the coefficient of tail dependence. *Insurance: Mathematics and Economics*, **57**, 46-57.

[9] Embrechts, P., McNeil, A. and Straumann, D. (2002): Correlation and dependency in risk management: properties and pitfalls. In: *Risk Management: Value at Risk and Beyond*, M.A.H. Dempster, Ed. Cambridge University Press, 176–223.

[10] Ferreira, M. and Silva, S. (2014): An Analysis of a Heuristic Procedure to Evaluate Tail (in)dependence. *Journal of Probability and Statistics*, Vol. 2014, Article ID 913621, 15 pages.

[11] Ferreira, M. (2014): A Heuristic Procedure to Estimate the Tail Index. Proceedings of the 14th International Conference in Computational Science and Its Applications - ICCSA 2014, June 30 - July 3 (2014), Guimares, Portugal, *IEEE-Computer Society*, 4264a241, 241-245.

[12] Ferreira, M. (2015): Estimating the tail index: Another algorithmic method. *Prob-Stat Forum*, **08**, 45-53.

[13] Frahm, G., Junker, M. and Schmidt R. (2005): Estimating the tail-dependence co-efficient: properties and pitfalls. *Insurance: Mathematics & Economics*, **37(1)**, 80-100.

[14] Goegebeur, Y. and Guillou, A. (2013): Asymptotically unbiased estimation of the coefficient of tail dependence. *Scand. J. Stat.*, **40**, 174-189

[15] Gomes, M.I., Henriques-Rodrigues, L., Fraga Alves, M.I. and Manjunath, B.G. (2013): Adaptive PORT-MVRB estimation: an empirical comparison of two heuristic algorithms. *Journal of Statistical Computation and Simulation*, **83(6)**, 1129-1144.

[16] Hill, B.M. (1975): A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Stat.*, **3**, 1163-1174.

[17] Joe, H. (1997): *Multivariate Models and Dependence Concepts*. Harry Joe, Chapman & Hall.

[18] Ledford, A. and Tawn, J. (1996): Statistics for near independence in multivariate extreme values. *Biometrika*, **83(1)**, 169-187.

[19] Matthys, G. and Beirlant, J. (2000): Adaptive Threshold Selection in Tail Index Estimation. In: *Extremes and Integrated Risk Management*, (Edited by P. Embrechts), 37-49. Risk Books, London.

[20] McGee, V.E. and Carleton, W.T. (1970): Piecewise Regression. *Journal of the American Statistical Association*, **65**, 1109-1124.

[21] Neves, M., Gomes, M.I., Figueiredo, F. and Prata-Gomes, D. (2015): Modeling Extreme Events: Sample Fraction Adaptive Choice in Parameter Estimation. *Journal of Statistical Theory and Practice*, **9(1)**, 184-199.

[22] Ramos, A. and Ledford, A. (2009): A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society, Series B*, **71**, 219-241.

[23] Poon, S.-H., Rockinger, M. and Tawn, J. (2003): Modelling extreme-value dependence in international stock markets. *Statistica Sinica*, **13**, 929-953.

[24] Smith, R.L. (1987): Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174-1207.

[25] Sousa, B. and Michailidis, G. (2004): A Diagnostic Plot for Estimating the Tail Index of a Distribution. *Journal of Computational and Graphical Statistics*, **13(4)**, 1–22.

	SP			KOPT			AMSE			RB			DK				PLAT	
$n = 100$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	NF	abias	rmse
$N(-0.2)$	0.0449	0.0590	90	0.0387	0.1232	12	0.0258	0.0579	68	0.0286	0.0470	69	0.0350	0.2883	3	4	0.0111	0.0780
$N(0.2)$	0.0574	0.0698	89	0.1202	0.2002	15	0.0878	0.1224	64	0.0532	0.0714	75	0.0388	0.4878	4	2	0.0384	0.1042
$N(0.8)$	0.1372	0.1460	93	0.1881	0.2726	16	0.1935	0.2402	77	0.1323	0.1397	75	0.1320	0.4158	8	7	0.1133	0.1440
t_4	0.4187	0.4223	96	0.4121	0.4458	20	0.4309	0.4362	79	0.4155	0.4188	76	0.3007	0.5849	3	5	0.3539	0.3734
t_1	0.2266	0.2323	96	0.1605	0.2297	14	0.2318	0.2344	95	0.2144	0.2199	76	0.1923	0.3481	12	5	0.1300	0.1507
$AL(0.7)$	0.4642	0.4658	94	0.4625	0.4895	18	0.4784	0.4863	92	0.4572	0.4594	78	0.3447	0.6026	4	3	0.4199	0.4342
$AL(0.3)$	0.2825	0.2855	98	0.1686	0.2364	17	0.2877	0.3024	73	0.2498	0.2556	74	0.1991	0.3459	14	6	0.1585	0.1864
$FGM(0.5)$	0.0383	0.0578	90	0.0507	0.1683	12	0.0163	0.1117	56	0.0362	0.0585	75	0.0508	0.3649	6	8	0.0302	0.1052
$Fr(2)$	0.0805	0.0954	88	0.2065	0.1762	13	0.0320	0.1265	61	0.0839	0.0960	77	0.0041	0.3391	5	5	0.0764	0.1293
$n = 1000$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	NF	abias	rmse
$N(-0.2)$	0.0425	0.0546	819	0.0059	0.0515	121	0.0378	0.0474	652	0.0437	0.0455	755	0.0242	0.3225	48	2	0.0247	0.0399
$N(0.2)$	0.0462	0.0642	826	0.0370	0.0687	171	0.0519	0.0690	777	0.0394	0.0432	754	0.0223	0.3651	39	0	0.0297	0.0452
$N(0.8)$	0.1178	0.1266	866	0.0832	0.0907	277	0.1231	0.1239	920	0.0926	0.0940	625	0.0991	0.3588	84	1	0.0716	0.0784
t_4	0.3921	0.4013	893	0.3303	0.3339	220	0.3703	0.3737	460	0.4056	0.4061	822	0.0431	0.6092	29	1	0.3114	0.3172
t_1	0.1975	0.2095	933	0.0777	0.0896	238	0.1530	0.1562	509	0.1886	0.1906	779	0.0479	0.1042	78	0	0.0554	0.0664
$AL(0.7)$	0.4518	0.4544	941	0.3906	0.3931	197	0.4245	0.4270	592	0.4392	0.4398	643	0.1613	0.6207	45	4	0.3827	0.3864
$AL(0.3)$	0.2369	0.2597	885	0.1282	0.1356	303	0.1821	0.1859	496	0.1940	0.1945	580	0.0800	0.1506	108	1	0.0868	0.0961
$FGM(0.5)$	0.0358	0.0430	846	0.0303	0.0525	178	0.0429	0.0600	630	0.0487	0.0516	762	0.0216	0.3347	50	0	0.0415	0.0532
$Fr(2)$	0.0630	0.0859	696	0.0305	0.0791	132	0.0409	0.1136	405	0.0952	0.0963	786	0.0380	0.3451	50	3	0.0691	0.0795
$n = 5000$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	NF	abias	rmse
$N(-0.2)$	0.0485	0.0515	4369	0.0217	0.0280	629	0.0424	0.0445	3353	0.0399	0.0406	3135	0.0920	0.3383	572	1	0.0214	0.0271
$N(0.2)$	0.0486	0.0490	4804	0.0288	0.0346	847	0.0410	0.0422	3684	0.0384	0.0391	3590	0.0601	0.4406	402	1	0.0261	0.0330
$N(0.8)$	0.1253	0.1261	4902	0.0725	0.0745	1343	0.1021	0.1043	3357	0.0907	0.0915	3052	0.0696	0.2242	737	0	0.0585	0.0625
t_4	0.4103	0.4117	4853	0.2709	0.2745	548	0.2746	0.2829	648	0.4106	0.4107	4418	0.0636	0.4472	34	1	0.2653	0.2688
t_1	0.2075	0.2090	4902	0.0499	0.0543	1062	0.0804	0.0843	1442	0.2039	0.2043	4573	0.0209	0.0393	235	0	0.0201	0.0328
$AL(0.7)$	0.4594	0.4595	4999	0.3428	0.3448	457	0.3558	0.3633	1178	0.4411	0.4413	3222	0.1898	0.5659	20	2	0.3511	0.3534
$AL(0.3)$	0.2694	0.2712	4950	0.0956	0.0989	969	0.1100	0.1137	1101	0.1989	0.1998	3024	0.0499	0.0641	298	0	0.0529	0.0642
$FGM(0.5)$	0.0391	0.0422	4562	0.0277	0.0387	705	0.0415	0.0460	2053	0.0487	0.0494	3655	0.0421	0.3120	190	0	0.0313	0.0379
$Fr(2)$	0.0831	0.0842	4854	0.0620	0.0684	617	0.0862	0.0926	1590	0.1027	0.1030	3650	0.0035	0.2501	286	0	0.0692	0.0738

Table 1: Simulation results from Hill estimator, where *abias* denotes the absolute bias, *NF* the number of fails and $\bar{k}$ correspond to the mean of the k values obtained in the 100 runs.

	SP			KOPT			AMSE			RB			DK				PLAT	
	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	abias	rmse	$\bar{k}$	NF	abias	rmse
$n=100$																		
$N(-0.2)$	0.0186	0.0653	91	0.0427	0.1287	12	0.0032	0.0738	57	0.0137	0.0603	74	0.0416	0.2880	2	5	0.0076	0.0795
$N(0.2)$	0.0164	0.0977	90	0.1085	0.2044	15	0.0458	0.1295	58	0.0202	0.0961	74	0.0514	0.4604	3	1	0.0241	0.1130
$N(0.8)$	0.0594	0.1066	93	0.1717	0.2675	17	0.1014	0.1860	66	0.0658	0.1050	77	0.1025	0.4099	7	6	0.0959	0.1436
t_4	0.3446	0.3618	96	0.3846	0.4268	20	0.3649	0.3810	66	0.3566	0.3704	71	0.2871	0.6015	3	2	0.3361	0.3610
t_1	0.0952	0.1261	96	0.1369	0.2112	15	0.1104	0.1337	78	0.1118	0.1387	75	0.1437	0.3297	5	0	0.0850	0.1215
$AL(0.7)$	0.3995	0.4123	93	0.4528	0.4846	18	0.4245	0.4410	60	0.4122	0.4227	76	0.3313	0.5980	4	4	0.4046	0.4237
$AL(0.3)$	0.0437	0.1355	96	0.1187	0.2105	21	0.0781	0.1698	66	0.0609	0.1418	71	0.1537	0.3491	7	3	0.0865	0.1519
$FGM(0.5)$	0.0659	0.1121	89	0.0439	0.1749	13	0.0199	0.1345	55	0.0565	0.1036	72	0.0468	0.3775	3	8	0.0393	0.1170
$Fr(2)$	0.1237	0.1549	88	0.0199	0.1794	13	0.0733	0.1718	58	0.1210	0.1482	73	0.0048	0.3401	4	7	0.0912	0.1499
$n=1000$																		
$N(-0.2)$	0.0165	0.0357	819	0.0008	0.0514	120	0.0119	0.0495	463	0.0204	0.0286	948	0.0103	0.3473	9	2	0.0206	0.0367
$N(0.2)$	0.0200	0.0539	808	0.0305	0.0662	169	0.0273	0.0608	515	0.0179	0.0342	913	0.0432	0.3660	18	1	0.0222	0.0442
$N(0.8)$	0.0353	0.0552	848	0.0545	0.0674	253	0.0450	0.0505	527	0.0359	0.0438	837	0.1318	0.4158	23	7	0.0514	0.0622
t_4	0.3255	0.3343	893	0.3061	0.3109	197	0.3275	0.3317	296	0.3471	0.3489	838	0.0806	0.6097	23	0	0.3042	0.3100
t_1	0.0514	0.0680	924	0.0278	0.0474	238	0.0525	0.0617	331	0.0667	0.0731	827	0.1303	0.2741	54	2	0.0276	0.0439
$AL(0.7)$	0.3937	0.3969	941	0.3751	0.3786	183	0.3920	0.3948	365	0.4009	0.4023	935	0.1170	0.6324	14	5	0.3781	0.3817
$AL(0.3)$	0.0063	0.0538	857	0.0371	0.0572	210	0.0610	0.1185	241	0.0239	0.0409	797	0.1388	0.2870	42	2	0.0413	0.0559
$FGM(0.5)$	0.0547	0.0649	846	0.0356	0.0572	180	0.0617	0.0671	600	0.0657	0.0698	904	0.0288	0.3346	42	0	0.0446	0.0585
$Fr(2)$	0.0854	0.1104	668	0.0371	0.0841	140	0.0845	0.1253	516	0.1172	0.1200	814	0.0442	0.3355	62	1	0.0729	0.0843
$n=5000$																		
$N(-0.2)$	0.0199	0.0240	4368	0.0156	0.0248	584	0.0206	0.0254	1823	0.0208	0.0225	4686	0.0520	0.3914	59	0	0.0200	0.0248
$N(0.2)$	0.0173	0.0223	4804	0.0223	0.0304	865	0.0193	0.0232	2720	0.0159	0.0205	4766	0.0292	0.4502	71	1	0.0212	0.0289
$N(0.8)$	0.0324	0.0346	4902	0.0458	0.0495	1110	0.0481	0.0502	1291	0.0308	0.0337	4466	0.0992	0.4592	78	3	0.0475	0.0521
t_4	0.3349	0.3360	4853	0.2495	0.2535	454	0.2549	0.2620	487	0.3447	0.3451	3940	0.1093	0.5322	39	2	0.2666	0.2696
t_1	0.0446	0.0473	4901	0.0008	0.0191	747	0.0127	0.0216	979	0.0535	0.0566	3782	0.0849	0.2497	311	2	0.0032	0.0214
$AL(0.7)$	0.3967	0.3971	4999	0.3303	0.3325	397	0.3210	0.3279	465	0.4003	0.4007	4203	0.1226	0.6323	26	3	0.3509	0.3529
$AL(0.3)$	0.0157	0.0260	4902	0.0315	0.0401	550	0.0414	0.0485	633	0.0194	0.0292	4520	0.1619	0.3958	210	1	0.0355	0.0427
$FGM(0.5)$	0.0561	0.0602	4562	0.0316	0.0431	727	0.0511	0.0567	2440	0.0619	0.0634	4196	0.0484	0.2880	218	0	0.0328	0.0399
$Fr(2)$	0.1186	0.1208	4854	0.0693	0.0758	691	0.1225	0.1162	2431	0.1264	0.1270	4100	0.0333	0.2429	241	1	0.0703	0.0757

Table 2: Simulation results from corrected Hill estimator, where *abias* denotes the absolute bias, NF the number of fails and $\bar{k}$ correspond to the mean of the k values obtained in the 100 runs.

$H(k)$	I	k	II	k	III	k
DK	0.6510	21	0.8255	83	0.7827	78
SP	0.6025	2592	0.5922	2893	0.6584	1499
KOPT	0.6733	744	0.9137	738	0.8444	135
AMSE	0.6494	955	0.7076	1244	0.6850	1172
RB	0.6041	2477	0.5967	2772	0.7428	708
PLAT	0.7148	–	0.8755	–	0.8110	–
$CH(k)$	I	k	II	k	III	k
DK	0.7654	5	0.4521	1	0.7044	27
SP	0.6725	2592	0.8581	2893	0.8671	1499
KOPT	0.7070	585	0.8991	412	0.8661	176
AMSE	0.6925	726	0.8997	596	0.8386	678
RB	0.6652	2264	0.8300	2040	0.8671	1499
PLAT	0.7261	–	0.8908	–	0.8524	–

Table 3: Estimates of η and respective values k, of datasets I, II and III.