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

Marta Ferreira

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
Marta Ferreira

Joe ([17] 1997), usually denoted by $\lambda$, 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

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

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 $\eta$ in Ledford and Tawn ([18] 1996). More precisely, it is assumed that

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

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 $\eta$ 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

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

and thus $\eta$ 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 $T_{n-k} \geq \ldots \geq T_{n-k:n}$, i.e.,

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

where $\log$ is the natural logarithm.
Consistency requires that $k$ must be intermediate, that is, a sequence of integers $k \equiv k_n$, $1 \leq k < n$, such that

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

There is no definite formula to obtain $k$ and it must be chosen not too small to avoid high variance but also not too 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

$$F^{-1}(1 - 1/x) = C x^\gamma (1 + \frac{\beta x^\rho}{\rho + o(x^\rho)}), \ x \rightarrow \infty, \quad (1.5)$$

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

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

where $\hat{\beta}$ and $\hat{\rho}$ are suitable estimators of $\beta$ and $\rho$, 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 $\eta$. We also consider the graphical and pure heuristic procedure presented in Frahm et al. ([13] 2005) in the context of estimating the TDC $\lambda$ 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 $\eta$ given in (1.3) and therefore corresponding to the tail index of

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

Coefficient $\eta$ 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
a unit Fréchet 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 Fréchet 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,

\[
T_i^{(n)} := \min\left(\frac{n + 1}{n + 1 - R_i^X}, \frac{n + 1}{n + 1 - R_i^Y}\right), \quad i = 1, \ldots, n
\]

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 \(\eta\) 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 \(H(k)(\tilde{\beta}(n/k)\tilde{\rho})/(1 - \tilde{\rho})\).

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 \(\eta\) (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
\[
Y_i := (i + 1) \log \frac{T_{n-i:n}^{(n)} H(i)}{T_{n-(i+1):n}^{(n)} H(i + 1)} = \eta + b(n/k) \left( \frac{i}{k} \right)^{-\rho} + \epsilon_i, \quad i = 1, \ldots, k, \tag{2.2}
\]
where the error term $\epsilon_i$ 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 $\rho$ is considered fixed, even if misspecified. Matthys and Beirlant ([19] 2000) suggest $\rho = -1$. From model (2.2), the resulting least squares estimators of $\eta$ and $b(n/k)$ are given by
\[
\tilde{Y}_{k,n}^{LS} = \sum_{i=1}^{k} \left( \frac{1}{k} \right)^{-\rho} - \frac{1}{1-\rho} Y_i. \tag{2.3}
\]
Thus, by replacing these estimates in the Hill’s asymptotic mean squared error (AMSE)
\[
\text{AMSE}(H(k)) = \frac{v_2}{k} + \left( \frac{b(n/k)}{1-\rho} \right)^2,
\]
we are able to compute $\tilde{k}_1^{opt}$ as the value of $k$ that minimizes the obtained estimates of the AMSE and estimate $\eta$ as $H(\tilde{k}_1^{opt})$.

On the other hand, we can compute the approximate value of $k$ that minimizes the AMSE, given by
\[
k_{opt} \sim b(n/k) - 2/(1-2\rho) k^{-2(1-2\rho)/(1-2\rho)} \left( \frac{v_2}{2} \right)^{1/(1-2\rho)}.	ag{2.4}
\]
See, e.g., Beirlant et al. ([1] 2002). Replacing again $\eta$ and $b(n/k)$ by the respective least squares estimates in (2.3) with fixed $\rho = -1$, we derive $\tilde{k}_{opt,k}$, for $k = 3, \ldots, n$, using (2.4). Then compute $\tilde{k}_2^{opt} = \text{median} \{ \tilde{k}_{opt,k}, k = 3, \ldots, \left\lfloor \frac{n}{2} \right\rfloor \}$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding $x$ and consider $\eta$ estimated by $H(\tilde{k}_2^{opt})$.

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 $\sqrt{\hat{\eta}}(H(i) - \eta)$, $i = 1, \ldots, k - 1$, with $k \equiv k_n$ an intermediate sequence, is of order $\sqrt{\log \log n}$. More precisely, for $\rho$ fixed at $-1$, we have:

1. Consider $r_n = 2.5 \times \tilde{\eta} \times n^{0.25}$, with $\tilde{\eta} = \hat{\eta}_2^{\sqrt{n},n}$.

2. Calculate $\tilde{k}(r_n) := \min \{ k = 1, \ldots, n - 1 : \max_{i<k} \sqrt{i} |H(i) - H(k)| > r_n \}$. 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
\[
\tilde{k}_{DK} = \left[ \frac{1}{3} (2\tilde{\eta}^2)^{1/3} \left( \frac{\tilde{k}(r_n^*)}{(\tilde{k}(r_n^*))^\varepsilon} \right)^{1/(1-\varepsilon)} \right].
\]
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 \(\eta\). 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 \(y = (S_1, ..., S_k)', X = [1]_{i=1}^k\) and \(\delta\) 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, \(y = [y_1]_{i=1}^k\) (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 \(k_{\text{new}} = \max(0, \{ j : TS < F_{1, k-2,1-\alpha} \})\). If \(k_{\text{new}} = 0\), no new points are added to \(y\) and thus move forward to step 4. Return to step 1. if \(k_{\text{new}} > 0\) by considering \(k = k_{\text{new}}\).

4. Estimate \(\eta\) 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 \(\ell := k_{\text{max}} - k_{\text{min}}\). Take all the estimates \(H(k; j_0 + 2)\) with \(k_{\text{max}} \leq k \leq k_{\text{min}}\), i.e., the estimates with two additional decimal points and calculate the mode. Consider \(K\) the set of \(k\)-values corresponding to the mode. Take \(H(\hat{k})\), with \(\hat{k}\) being the maximum of \(K\). 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 \(\lambda\) 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, \(H(1), ..., H(n - 2b)\), with bandwidth \(b = \lfloor w \times n \rfloor\).

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

\[
\sum_{i=k+1}^{k+m-1} |\overline{H}(i) - \overline{H}(k)| \leq 2s,
\]

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 \(\eta\) as the mean of the values of \(p_k^*\) (consider the estimate zero if no plane region fulfills the stopping condition).

The estimation of \(\eta\) 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 \(\eta\) 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 \(t_\nu\) with correlation coefficient given by \(\rho \neq -1\) \((\lambda = 2F_{\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 an asymmetric-logistic dependence function \(\gamma(x, y) = (1 - a_1)x + (1 - a_2)y + ((a_1x)^{1/\alpha} + (a_2y)^{1/\alpha})^\alpha\), with \(x, y \geq 0, \alpha > 0\).
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 ($\beta$ and $\rho$). To this end, we have followed the indications in Caeiro et al. ([4] 2009). For the $\rho$ 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 $\rho$.

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, totaling 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 $Fr(2)$.

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 $\eta$ 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.

![scatter-plots](image1.png)

**Figure 4:** From left to right: scatter-plots of datasets I, II and III.

![sample paths](image2.png)

**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

The author wishes to thank the reviewers for their constructive and valuable comments that have improved this work. This research was financed by Portuguese Funds through FCT - Fundação para a Ciência e a Tecnologia, within the Project UID/MAT/00013/2013 and by the research centre CEMAT (Instituto Superior Técnico, Universidade de Lisboa) through the Project UID/Mult/04621/2013.

References


efficient: properties and pitfalls. *Insurance: Mathematics & Economics*, **37(1)**, 80-


(2013): Adaptive PORT-MVRB estimation: an empirical comparison of two heuris-
tic algorithms. *Journal of Statistical Computation and Simulation*, **83(6)**, 1129-


timation. In: *Extremes and Integrated Risk Management*, (Edited by P. Embrechts),


TABLE 1: Simulation results from Hill estimator, where abias denotes the absolute bias. \( k \) is the number of fails and \( r \) corresponds to the mean of the values obtained in the 100 runs.

<table>
<thead>
<tr>
<th>( F )</th>
<th>( \alpha )</th>
<th>( N )</th>
<th>( abias )</th>
<th>( rmse )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( NFGM(0) )</td>
<td>0.3511 0.3534</td>
<td>0.0261 0.0330</td>
<td>0.0384 0.0391</td>
<td>0.0288 0.0346</td>
</tr>
<tr>
<td>( NFGM(1) )</td>
<td>0.0868 0.0961</td>
<td>0.0479 0.1042</td>
<td>0.1923 0.3481</td>
<td>0.3007 0.5849</td>
</tr>
<tr>
<td>( NFGM(2) )</td>
<td>0.3114 0.3172</td>
<td>0.0302 0.1052</td>
<td>0.2318 0.2344</td>
<td>0.4309 0.4362</td>
</tr>
<tr>
<td>( NFGM(3) )</td>
<td>0.0716 0.0784</td>
<td>0.0508 0.3649</td>
<td>0.0163 0.1117</td>
<td>0.1686 0.2364</td>
</tr>
<tr>
<td>( NFGM(4) )</td>
<td>0.1133 0.1440</td>
<td>0.0388 0.4878</td>
<td>0.0387 0.1232</td>
<td>0.1881 0.2726</td>
</tr>
<tr>
<td>( NFGM(5) )</td>
<td>0.0399 0.0406</td>
<td>0.0378 0.0474</td>
<td>0.0059 0.0515</td>
<td>0.0370 0.0687</td>
</tr>
<tr>
<td>( NFGM(6) )</td>
<td>0.0764 0.1293</td>
<td>0.0320 0.1265</td>
<td>0.0191 0.3459</td>
<td>0.0383 0.0578</td>
</tr>
<tr>
<td>( NFGM(7) )</td>
<td>0.1991 0.3459</td>
<td>0.0242 0.3225</td>
<td>0.0437 0.0455</td>
<td>0.0437 0.0455</td>
</tr>
<tr>
<td>( NFGM(8) )</td>
<td>0.0111 0.0780</td>
<td>0.0350 0.2883</td>
<td>0.0532 0.0714</td>
<td>0.0878 0.1224</td>
</tr>
</tbody>
</table>

The \( abias \) and \( rmse \) are for the mean of the values obtained in the 100 runs.
Table 2: Simulation results from corrected Hill estimator, where abias denotes the absolute bias, NF the number of fails and T correspond to the mean of the k values obtained in the 100 runs.

<table>
<thead>
<tr>
<th>SP</th>
<th>KOPT</th>
<th>PLAT</th>
<th>AMSE</th>
<th>FGM</th>
<th>Fr</th>
<th>n = 100</th>
<th>KOPT</th>
<th>PLAT</th>
<th>AMSE</th>
<th>FGM</th>
<th>Fr</th>
<th>n = 5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = (0.2)</td>
<td>0.986</td>
<td>0.119</td>
<td>0.127</td>
<td>0.122</td>
<td>0.132</td>
<td>0.119</td>
<td>0.127</td>
<td>0.122</td>
<td>0.132</td>
<td>0.119</td>
<td>0.127</td>
<td>0.122</td>
</tr>
<tr>
<td>0.0186</td>
<td>0.0994</td>
<td>0.0982</td>
<td>0.0957</td>
<td>0.0977</td>
<td>0.1085</td>
<td>0.1072</td>
<td>0.1065</td>
<td>0.1085</td>
<td>0.1072</td>
<td>0.1065</td>
<td>0.1085</td>
<td>0.1072</td>
</tr>
<tr>
<td>0.0653</td>
<td>0.0659</td>
<td>0.0666</td>
<td>0.0669</td>
<td>0.0671</td>
<td>0.0677</td>
<td>0.0677</td>
<td>0.0671</td>
<td>0.0677</td>
<td>0.0671</td>
<td>0.0677</td>
<td>0.0671</td>
<td>0.0677</td>
</tr>
<tr>
<td>91</td>
<td>90</td>
<td>93</td>
<td>96</td>
<td>96</td>
<td>92</td>
<td>91</td>
<td>90</td>
<td>93</td>
<td>96</td>
<td>96</td>
<td>92</td>
<td>91</td>
</tr>
<tr>
<td>91</td>
<td>90</td>
<td>93</td>
<td>96</td>
<td>96</td>
<td>92</td>
<td>91</td>
<td>90</td>
<td>93</td>
<td>96</td>
<td>96</td>
<td>92</td>
<td>91</td>
</tr>
<tr>
<td>Table 2: Simulation results from corrected Hill estimator, where abias denotes the absolute bias, NF the number of fails and T correspond to the mean of the k values obtained in the 100 runs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Estimates of $\eta$ and respective values $k$, of datasets I, II and III.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>k</th>
<th>II</th>
<th>k</th>
<th>III</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>DK</td>
<td>0.6510</td>
<td>21</td>
<td>0.8255</td>
<td>83</td>
<td>0.7827</td>
<td>78</td>
</tr>
<tr>
<td>SP</td>
<td>0.6025</td>
<td>2592</td>
<td>0.5922</td>
<td>2893</td>
<td>0.6584</td>
<td>1499</td>
</tr>
<tr>
<td>KOPT</td>
<td>0.6733</td>
<td>744</td>
<td>0.9137</td>
<td>738</td>
<td>0.8444</td>
<td>135</td>
</tr>
<tr>
<td>AMSE</td>
<td>0.6494</td>
<td>955</td>
<td>0.7076</td>
<td>1244</td>
<td>0.6850</td>
<td>1172</td>
</tr>
<tr>
<td>RB</td>
<td>0.6041</td>
<td>2477</td>
<td>0.5967</td>
<td>2772</td>
<td>0.7428</td>
<td>708</td>
</tr>
<tr>
<td>PLAT</td>
<td>0.7148</td>
<td>–</td>
<td>0.8755</td>
<td>–</td>
<td>0.8110</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>k</th>
<th>II</th>
<th>k</th>
<th>III</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>DK</td>
<td>0.7654</td>
<td>5</td>
<td>0.4521</td>
<td>1</td>
<td>0.7044</td>
<td>27</td>
</tr>
<tr>
<td>SP</td>
<td>0.6725</td>
<td>2592</td>
<td>0.8581</td>
<td>2893</td>
<td>0.8671</td>
<td>1499</td>
</tr>
<tr>
<td>KOPT</td>
<td>0.7070</td>
<td>585</td>
<td>0.8991</td>
<td>412</td>
<td>0.8661</td>
<td>176</td>
</tr>
<tr>
<td>AMSE</td>
<td>0.6925</td>
<td>726</td>
<td>0.8997</td>
<td>596</td>
<td>0.8386</td>
<td>678</td>
</tr>
<tr>
<td>RB</td>
<td>0.6652</td>
<td>2264</td>
<td>0.8300</td>
<td>2040</td>
<td>0.8671</td>
<td>1499</td>
</tr>
<tr>
<td>PLAT</td>
<td>0.7261</td>
<td>–</td>
<td>0.8908</td>
<td>–</td>
<td>0.8524</td>
<td>–</td>
</tr>
</tbody>
</table>