
Extreme Value Theory as a Theoretical Background for Power Law Behavior

Simone Alfarano¹ and Thomas Lux²

¹Department of Economics, University of Kiel, alfarano@bwl.uni-kiel.de

²Department of Economics, University of Kiel, lux@bwl.uni-kiel.de

Prepared for *Power Laws in the Social Sciences: Discovering Complexity and Non-Equilibrium Dynamics in the Social Universe* Claudio Cioffi-Revilla, editor.

Summary. Power law behavior has been recognized to be a pervasive feature of many phenomena in natural and social sciences. While immense research efforts have been devoted to the analysis of behavioral mechanisms responsible for the ubiquity of power-law scaling, the strong theoretical foundation of power laws as a very general type of limiting behavior of large realizations of stochastic processes is less well known. In this chapter, we briefly present some of the key results of extreme value theory, which provide a statistical justification for the emergence of power laws as limiting behavior for extreme fluctuations. The remarkable generality of the theory allows to abstract from the details of the system under investigation, and therefore allows its application in many diverse fields. Moreover, this theory offers new powerful techniques for the estimation of the Pareto index, detailed in the second part of this chapter.

1 Extreme Value Theory

In modelling the extreme events of a random variable, Extreme Value Theory is the counterpart of the Central Limit Theorem (EVT and CLT in the following) for sums. However, while the CLT is concerned with “small” fluctuations around the mean resulting from an aggregation process, the EVT provides results on the asymptotic behavior of the extreme realizations (maxima and minima).

The CLT states that a sum of N random variables independently drawn from a common distribution function¹ $F(x)$ with finite variance, $X = x_1 + x_2 + \dots + x_N$, converges to the Normal distribution as N goes to infinity². The form of the original distribution $F(x)$ does not affect the asymptotic behavior (that in any case is Gaussian), but only the speed of convergence. It is worth emphasizing that the CLT only concerns the *central* part of the distribution of the sum X , but it does not provide information on the behavior of its extremes. In particular, the CLT does **not** imply that the laws governing the extreme realizations will be those governing the extremes of the limiting Normal distribution. Rather the tail behavior of the original distribution $F(x)$ might be preserved under certain conditions despite the overall attraction towards a Gaussian shape. This means that, in the aggregation process, the region around the mean is gradually better approximated by the Gaussian, while the outmost part might still follow a different type of extreme value distribution. Since this extreme region will be expelled towards large values of X in the aggregation process, it will become less and less ‘visible’ for the sums of random variables although it is still effective

¹We use the same notation as in the second chapter of the book, $F(x)$ denotes the cumulative distribution function $F(x) = Pr(X < x)$ and $f(x)$ the associated probability density function.

²See Feller [7] for a generalization of the CLT to dependent and non-identical distributed variables, and to distributions with infinite variance.

asymptotically³.

The EVT deals with these extreme events, providing a classification of *continuous* distributions according to the behavior of the *tail region* or their *extreme realizations*. The theory distinguishes three limiting stable distributions for the maximum values of a random variable, called Generalized Extreme Value Distributions (GEV), and the three associated Generalized Pareto Distributions (GPD), which are the limiting distributions for the tail region of the pertinent distribution.

1.1 GEV: Limiting Distributions for Extrema

Let us consider a stationary sequence of iid variables $\{x_i\}_{i=1}^N$ with a common distribution function $F(x)$. By dividing the entire data-set into L non-overlapping sub-samples, and taking the maximum M_j from every sub-sample, we will end up with a subset of maxima $\{M_j\}_{j=1}^L$ (so-called block maxima). The limit law of this sequence $\{M_j\}_{j=1}^L \equiv M(L)$ is given by the following theorem⁴:

Theorem 1 (Fisher and Tippett, Gnedenko [1], [16]). *If there exist two normalizing constants $a_L > 0$ and $b_L \in \mathfrak{R}$, and a non-degenerate distribution H such that*

$$\frac{M(L) - b_L}{a_L} \xrightarrow{d} H,$$

where the subscript d indicates convergence in distribution, then H belongs to one of the following extreme value distributions:

$$(1a) \quad \text{Fréchet: } G_{1,\alpha}(x) = \begin{cases} 0 & x < 0 \\ \exp[-x^{-\alpha}] & x \geq 0, \end{cases}$$

$$(1b) \quad \text{Weibull: } G_{2,\alpha}(x) = \begin{cases} \exp[-(-x)^\alpha] & x \leq 0 \\ 1 & x > 0, \end{cases}$$

$$(1c) \quad \text{Gumbel: } G_3(x) = \exp[-e^{-x}] \quad x \in \mathfrak{R}.$$

Based on the previous theorem, distributions can be classified into three categories: (i) *heavy-tailed* distributions, whose extremes follow the first type of law; (ii) *short-tailed* distributions with finite end-point, whose extremes obey the Weibull's type; and (iii) *medium tailed* distributions, whose extremes are governed by the third category. Note that in cases (i) and (ii) we have a one-parameter family of distributions, parametrized by the shape coefficient α . Representative members of the three groups are respectively: the Student-t, the uniform distribution and the Normal. The von Mises representation of the GEV provides a unified formula for the previous three limiting distributions (1a), (1b) and (1c):

$$G_\gamma = \exp[-(1 + \gamma x)^{-\frac{1}{\gamma}}].$$

For a positive γ we recover the Fréchet distribution, negative γ corresponds to the Weibull type, and the limit case $\gamma \rightarrow 0$ describes the Gumbel formula. The shape parameters of the two representations are related to each other by the formula $\alpha = \frac{1}{\gamma}$ for the distribution (1a), and $\alpha = -\frac{1}{\gamma}$ for the type (1b). The Von Mises formula turns out to be very useful in that it nests all these types of limiting behavior in a unified framework and, via estimation of γ , allows inference on the relevant limit laws.

³An excellent reference for a detailed description of this aspect of CLT, in a financial perspective, is the book of Cont and Bouchaud [2].

⁴It is also possible to derive equivalent results for the minimum values.

1.2 GPD: Limiting Distributions for the Tails

The second set of results focuses on the tails of the distributions instead of maxima; the selected events are, in this case, those events that exceed a given threshold u . Let us first introduce the Generalized Pareto Distributions (GPD in the following) using the so-called α -parameterization:

$$(2a) \quad W_{1,\alpha} = 1 - x^{-\alpha}, \quad x \geq 1,$$

$$(2b) \quad W_{2,\alpha} = 1 - (-x)^\alpha, \quad -1 \leq x \leq 0,$$

$$(2c) \quad W_3 = 1 - \exp(-x), \quad x \geq 0.$$

All three distributions (2a) to (2c) assume the value zero outside the pertinent intervals. For the GPD a similar one-parameter representation exists as with the extreme value distributions:

$$(3) \quad W_\gamma = 1 - (1 + \gamma x)^{-\frac{1}{\gamma}},$$

where for $\gamma > 0$, $\gamma < 0$ and $\gamma \rightarrow 0$ we recover the first, second and third group, respectively. γ is called the shape parameter. The relations between α and γ are again $\alpha = \frac{1}{\gamma}$ for the first type, and $\alpha = -\frac{1}{\gamma}$ for the second type.

The basic result for the limiting behavior of the tail region of a distribution is:

Theorem 2 (Pickands, Balkema and de Haan [1], [16]). *Let us consider a continuous distribution function $F(x)$ and a threshold u smaller than the right end-point $\omega(F)$ ⁵. We define the associated distribution function of exceedances over the threshold u of the distribution $F(x)$ as:*

$$F^{[u]}(x) = Pr(X \leq x | X > u).$$

If $F^{[u]}(a_u x + b_u)$ has a non-degenerate continuous limiting distribution function as u goes to the right-end point $\omega(F)$ then, with appropriate adjustment of scale and location, $F^{[u]}(x)$ converges to one of the GPD distributions:

$$|F^{[u]}(x) - W_{\gamma,u,\sigma_u}(x)| \rightarrow 0, \quad u \rightarrow \omega(F),$$

for shape, location and scale parameters⁶ γ , u and $\sigma_u > 0$, respectively. The theorem can also be formulated in terms of the α -parameterization.

The previous limiting theorem allows for a classification of distributions according to the behavior of their tails: hyperbolic decline, if the distribution converges to the first type of GPD, a finite end-point, if it converges to the second type, or exponential decline, if its limiting distribution belongs to the third type.

In summary, through theorems (1) and (2), EVT allows to abstract from the specific distribution governing the fluctuations of the overall system (that can be a financial market or a geological structure, for instance) when investigating extremes, and to concentrate solely on the behavior of the large observations. Moreover, the theory provides the functional form for the description of the tail, given by eqs. (2a), (2b) and (2c), or, in a more compact form by eq. (3). The GPD formalization is very flexible in describing the tail behavior, although it depends on one parameter only, the *index* α (after accounting for location and scale). *EVT, therefore, provides a rationale for the widespread observation of Pareto or power law behavior, since it is the limiting behavior of large events for a whole class of probability distributions.*

⁵The right-end point $\omega(F)$ of a distribution is defined as $\omega(F) \equiv \sup\{x : F(x) < 1\}$.

⁶If a random variable X has a distribution function F , then $\sigma X + \mu$ has a distribution function $F_{\mu,\sigma} = F[(X - \mu)/\sigma]$, where μ and σ are the location and scale parameter, respectively. Eq (3) represent the standard form of the GPDs in the γ parametrization, where the location parameter is 0 and the scale parameter is 1.

2 Estimation of the Tail Index

The estimation of the index α is the central issue in empirical research dealing with extreme events, since it allows to precisely quantify the likelihood of large fluctuations.

However, any attempt of estimation of the tail index has to cope with three key problems. The first one arises from the operational definition of an ‘extreme’ realization, i.e. one has to decide which events, from the complete set of data points, belong to the subset relevant for the estimation of α . This problem is denoted as the *threshold selection problem*. Then, it is necessary to construct an estimator that provides a compromise between a potential bias, due to the only approximate validity of the power law, and the variance of the estimate which strongly depends on the size of the selected subset of extreme events. Finally, the method should ideally also provide information on the distribution of the estimates.

There are several simple heuristic regression approaches for the estimation of the tail index, often using graphical procedures: mean excess function and mean log-excess function, for instance [1], [16]. Although these methods are unsatisfactory in so far as often the distribution of estimates is unknown, they might be useful for discriminating among the three different types of tail behavior, namely power law, exponential decay or finite distribution with a end-point. A more rigorous procedure is the one proposed by Hill [11], that has become the standard tool for estimating of power law exponents in economics and other related fields.

2.1 Hill Estimator

The Hill estimator is the conditional maximum likelihood estimator for heavy-tailed distributions. If we assume that the data points exceeding a given threshold u follow a Pareto distribution with index α , the distribution of realizations exceeding u reads:

$$(4) \quad F^{[u]}(x) = 1 - \left(\frac{u}{x}\right)^\alpha, \quad x \geq u.$$

As has been shown by Hill, the (conditional) maximum likelihood estimator of the parameter α in eq. (4) assumes a particularly simple form. Denote by $\{x_i\}_{i=1}^N$ a sample of size N whose k largest values are assumed to obey to eq. (4). The parameter γ (or α) can then be estimated by:

$$(5) \quad \hat{\gamma}_{k,N} = (\hat{\alpha}_{k,N})^{-1} = \frac{1}{k} \sum_{i=1}^k [\ln x_{(N-i+1)} - \ln x_{(N-k)}],$$

with $x_{(i)}$ the order statistics of the series x , $x_{(N)} > x_{(N-1)} > \dots > x_{(1)}$, i.e. $x_{(N)}$ is the maximum of x , $x_{(N-1)}$ is the second largest value etc. As it is assumed that eq. (4) only applies to a fraction k/N of the largest values, we only consider the $x_{(i)}$ above the threshold u , $x_{(N-k)} = u > x_{(N-k-1)}$, where k is the number of selected large realizations, from the entire sample of N observations. It has been shown that under some mild additional restrictions on the behavior of the underlying distribution function, $\hat{\gamma}_{k,N}$ is asymptotically Gaussian with mean γ (i.e. the inverse of the true index) and variance $(\gamma^2 k)^{-1}$. Given the asymptotic Normality of $\hat{\gamma}$, the $(1-x)\%$ confidence intervals can be computed as:

$$(6) \quad \hat{\gamma} \pm \lambda_{x/2} \frac{\hat{\gamma}}{\sqrt{k}},$$

where $\lambda_{x/2}$ is the $(1-x/2)$ standard Normal quantile. It is important to emphasize that each different threshold value might lead to a different Hill estimator. The two subscripts N and k indicate dependence of the estimated value on the number of data points and on the chosen threshold, respectively. In the appendix a mathematical proof of the formula (5) and a detailed description of the procedure are given. The simplicity of eq. (5) together with the desirable asymptotic properties of consistency and Normality are the advantages of the Hill estimator over many other procedures.

Nevertheless, it is important to emphasize its potential sensitivity to the choice of the tail size. For better understanding of this problem, let us consider a simple example. In Figure (1), a typical ‘Hill plot’ is exhibited, i.e. the tail index estimated using eq. (5) as a function of the tail size, for a Student-t

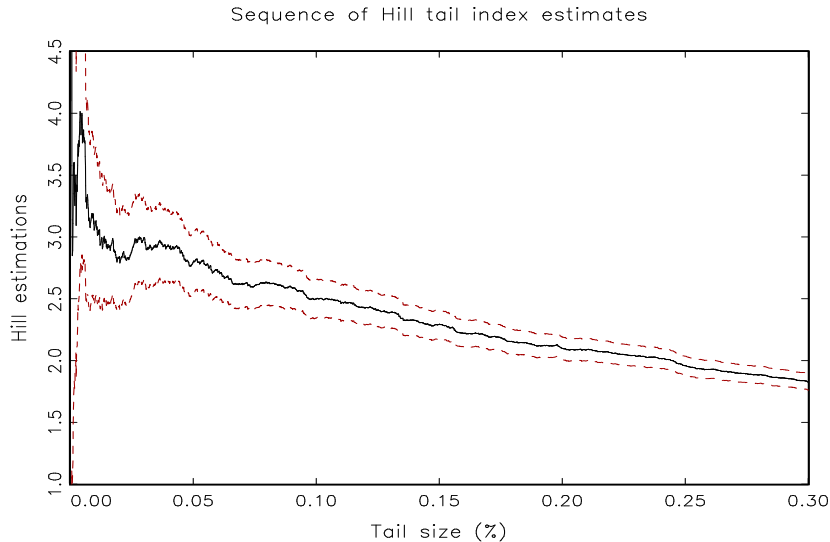


Fig. 1. Hill plot for 10 000 realizations from a random variable drawn from a Student- t distribution, with 3 degrees of freedom. The dashed lines represent the 95% confidence interval, computed using eq. (6). Note that the Hill estimator reported in this plot is $\hat{\alpha} = \frac{1}{\gamma}$.

random variable with 3 degrees of freedom. The theoretical value of the index of the tail for this distribution is exactly its degree of freedom [7]. Nevertheless, we can observe a monotonic increase of the estimated value, starting from a low $\hat{\alpha} = 1.83$ at 30% tail size, to $\hat{\alpha} = 3.09$ at the 1% tail size⁷. The underlying assumption of power law tails is the more accurate the further one goes to the outer part of the distribution. The large downward bias at, e.g., 30% tail sizes is, then, not too surprising, since the true power law behavior is strongly contaminated by the entries from more central parts of the distribution.

For this reason, it is not immediately obvious what the appropriate tail fraction should be that would give the “best” estimator for the ‘true’ parameter α . A possible practical approach could be an “eyeball” method, searching for a region in the Hill plot where the estimated values are approximately constant. However, there is no clear evidence of the existence of such a plateau, and, moreover, this approach has all the drawbacks of a subjective graphical data analysis.

A data-driven criterium for the selection of the tail size was suggested by Hill himself. It is known that the logarithm of a variable following a Pareto distribution with index α obeys an Exponential distribution with scale parameter $1/\alpha$ and location parameter 0. More precisely, let us consider a random variable x distributed according to a Pareto distribution as in eq. (4), the log-transformed variable $\nu = \ln(u/x)$ follows the Exponential distribution:

$$\text{Exp}(\nu, 0, \alpha) \equiv 1 - \exp(-\alpha\nu).$$

One can exploit the previous relation and test the appropriateness of the considered tail size, simply performing a goodness-of-fit test of ν against the pertinent Exponential distribution, which gives information on the convergence of the original sample to a Pareto distribution. In other words, rejection of the Exponential distribution for the transformed data implies a rejection of the approximated power law decay for the original sample, at the considered tail size. Operationally, we transform the descending order statistics of the original data into new random variables $\nu_i = i\alpha_{k,N} \ln[x_{(i)}/x_{(i+1)}]$ for $i < N - k$, which should follow an Exponential distribution $\text{exp}(\nu_i, 0, 1)$, if k is suitably chosen. Note the dependence of α

⁷This phenomenon is not limited to simulated data, but it is also a frequent empirical fact observed, for example, in financial data, as shown by Lux [14] among others.

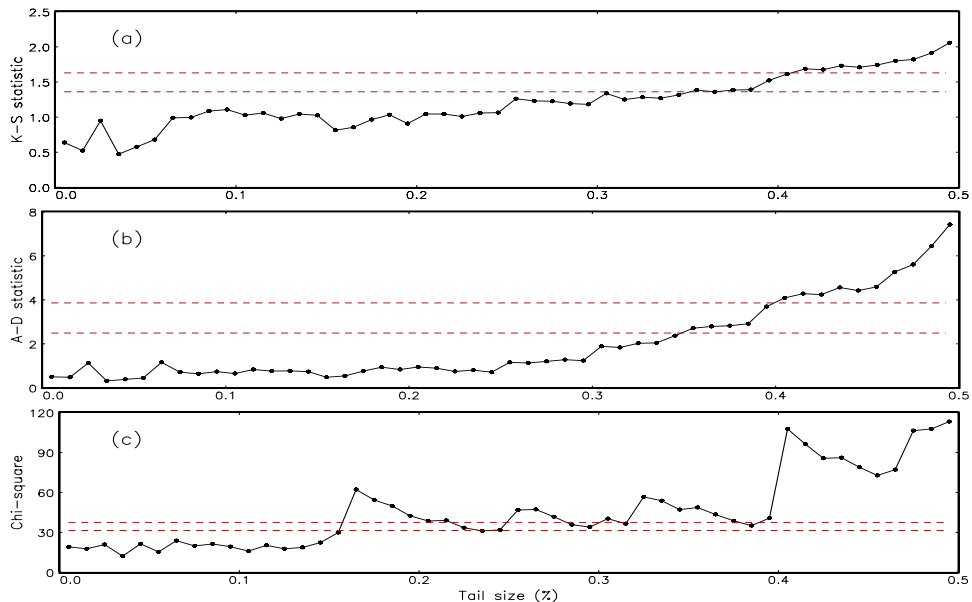


Fig. 2. Panel (a) shows the sequence of Kolmogorov-Smirnov statistics as a function of the points located in the tail of the distribution, expressed as a fraction of the entire sample. Panel (b) refers to the Andersen-Darling statistics. Panel (c) shows the sequence of Chi-square tests computed with 20 bins. In all cases, the dashed lines refer to the appropriate critical values at 5% and 1% confidence levels.

on the considered threshold $u = x_{(N-k)}$. One may now use various goodness-of-fit tests to investigate the appropriateness of the hypothesized Exponential distribution for the new random variable ν_i (one might, of course, equivalently apply goodness-of-fit tests for the originally estimated Pareto distribution). Figure 2 shows the outcome of a complete sequence of goodness-of-fit tests using different statistics, namely the Chi-square, Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) statistics⁸, applied to the same sample of Student-t random variables used to compute the Hill estimator in Figure 1.

The appropriate tail sizes, suggested by the K-S or A-D goodness-of-fit tests, turn out to be too large - higher than 30% of the entire sample, with a corresponding Hill estimator smaller than 2, therefore far from the true value 3, which would not even be included in the associated confidence intervals (see Figure 1). The Chi-squared test seems to perform better, suggesting a tail size of about 15%. However, once again the estimated value results to be smaller than the ‘true’ value 3. All in all, the goodness-of-fit tests, especially the K-S and A-D, seem to be inappropriate for the choice of an optimal tail size. The reason can be found in the strong downward bias of the Hill estimator, which artificially enlarges the region of validity of the Pareto approximation (4). Note that although, as in our example, a certain estimate of α might be quite different from the ‘true’ limiting index (e.g. at the sizes of 20% to 30%), the Pareto distribution with this index might have a shape quite close to that of the 30% tail (after all, α has been estimated to provide the best fit to exactly this sub-sample). As can be seen from our example, goodness-of-fit tests are not too successful when trying to figure out what index from a set of estimates for different tail sizes is the one closest to the asymptotic scaling behavior.

A more rigorous approach is to use a criterion for endogenous selection of the *optimal cut-off value*, based on the statistical properties of the Hill estimator itself, which will be introduced in the next paragraph.

⁸For an exhaustive description of the goodness-of-fit tests and the tabulated critical values see [3], in particular Table 4.2, p.105.

2.2 Optimal Tail Size

We now turn to alternative data-driven criteria for an endogenous selection of the threshold value. A natural choice is the minimization of the mean squared error (MSE in the following) of the estimated $\hat{\gamma}(= \frac{1}{\alpha})$, since this approach allows to mediate between the bias, which increases with the tail size, and the variance, which, on the contrary, decreases when extending the tail fraction. The optimal value is, then, defined as:

$$k_N^{opt} = \min_k E[(\hat{\gamma}_{k,N} - \gamma)^2] = \min_k (Var[\hat{\gamma}_{k,N}] + Bias^2[\hat{\gamma}_{k,N}]) .$$

It is relatively easy to evaluate the variance, given the asymptotic normality of the estimator (see [8, 9]). In order to assess the contribution of the bias, it is necessary to introduce a second order expansion of the tail⁹:

$$(7) \quad F(x|x > u) = 1 - ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})) .$$

Note that the first and the second term have the same functional form. This is a crucial assumption, since a slower decay term, such as a $\log x$, would prevent the asymptotic convergence towards the power law, while an exponential term would have so rapid convergence that it would not affect the behavior of the tail. The expansion (7) holds exactly for many textbook distributions, such as the Student-t. Dropping terms of third and higher orders in eq. (7), the first and the second moments of the Hill estimator can be computed, which allows to derive the asymptotic MSE as a function of the underlying parameters of the expansion (7) [9]. Hall has shown that the optimal tail size is given by:

$$(8) \quad k_N^{opt} = s(a, b; \alpha, \beta) N^{\frac{2\alpha}{2\alpha+\beta}} ,$$

where $s(\cdot)$ is a function of the underlying parameters of the expansion (7) which is independent of N . The major practical problem in using (8) is that it requests a preliminary estimate of α , as well as of the parameter β of the second-order approximation.

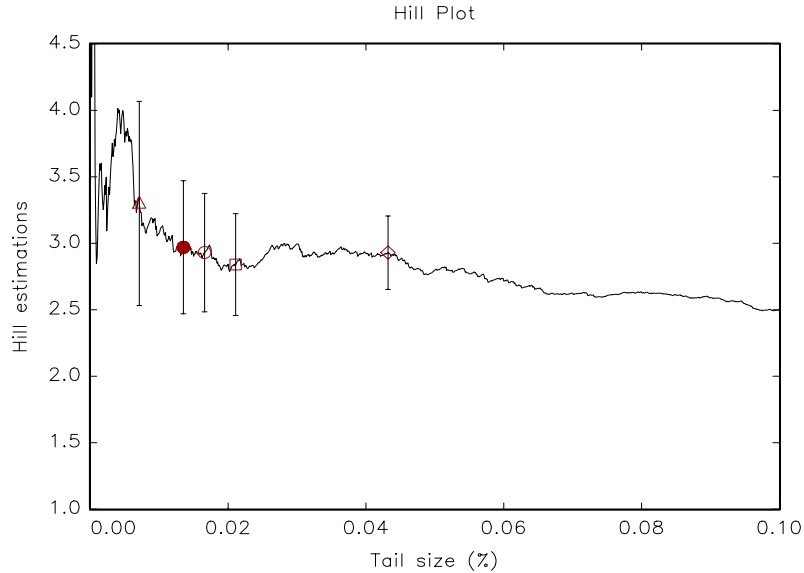


Fig. 3. The four points mark the estimates of α chosen by methods (1) thorough (4). The empty triangle, circle, diamond and square refer to the method (1), (2), (3) and (4), respectively. The full circle marks the estimate associated to the theoretical value k_N^{opt} . The error bars show the 95% confidence interval, computed from the asymptotic distribution of α , using eq. (6).

⁹Other equivalent parameterizations also exist in the literature (see [1, 16]).

Several methods have been developed recently for estimating the second-order tail parameter β and the optimal tail fraction. In the following, a brief description of the procedures is given, leaving the technical details, often quite involved, to the pertinent literature¹⁰:

1. The first method, introduced by Hall [10], is based on bootstrapping procedure using sub-samples smaller than the entire data set. The contribution of the bias is extracted via the minimization of the empirical MSE of the sub-samples, computed using an initial estimate of α from the overall sample. Hall assumes $\alpha = \beta$, which applies to some distributions but, in general, turns out to be a quite restrictive assumption. Danielson and de Vries [5] generalize this approach by adding a moment estimator for β .
2. A more general sub-sample bootstrap procedure is implemented in the method developed by Danielson *et al.* [4]. They show that the MSE of a suitable combination of first and second log moments leads to an equivalent and more convenient minimization problem than the MSE of the Hill estimator itself. This auxiliary statistic is, indeed, a consistent estimator of α , and, moreover, no initial estimate is needed in order to compute the bootstrap MSE. The optimal tail is, then, found by comparing two estimates from sub-samples of different sizes (with arbitrary n_1 and $n_2 = \frac{n_1^2}{N}$).
3. Drees and Kaufmann [6] construct a “stopping rule” for the Hill estimates to extract the bias. Their starting point is the observation that large deviations of the estimates from the expected order of magnitude, given by $\ln(\ln N)$, are attributable to the bias term. The optimal tail is, then, derived by using two different “stopping times”, after computing a consistent moment estimator for β from the pertinent Hill estimators.
4. Beirlant *et al.* [1] use as a starting point the relationship between the Hill estimator and Pareto quantile plots. The optimal value for the tail size is estimated via a weighted least squared regression derived from this relationship. Weights are, then, iteratively adjusted until convergence of both the weights and the tail fractions derived from them is obtained.
5. Finally, Beirlant *et al.* [1] also generalize the previous method by applying the same idea to the von Mises γ -parameterization. Their method, therefore, does not only allow for data-driven selection of the cut-off value k of data with a limiting Pareto distribution, but also nests the possibilities of exponential decline and finite endpoint.

Estimates based on the first four methods, applied to the same sample previously considered in Figure 1, are exhibited in Figure 3. Estimated optimal tail sizes vary between about 1% and 4% tail regions. The resulting estimates of α are all very close to the true value, and, on the basis of 95% confidence intervals, the ‘true’ value of 3 cannot be rejected for any of these estimates. Interesting, it is possible to compute the value k_N^{opt} of eq. (8) for the Student-t distribution for a given sample size N , since the expansion (7) holds exactly for this parametric family. Figure 3 shows this theoretical value -which turns out as $k_{10\,000}^{opt} = 135$ (note the dependence on sample size) - together with the estimates from the four considered procedures.

3 Final remarks

Some words of cautions in the application of the optimal cut-off procedures are in order here. For example, violation of the assumption of iid-ness of the data could lead to a huge inflation of the ‘true’ 95% confidence intervals compared to those of the asymptotic distribution in eq. (6). While the Hill estimator remains a consistent estimator for dependent data [17], it might, in fact, not be asymptotically Normally distributed any more for certain stochastic processes. As an example, Kearns and Pagan [12] have shown that for the IGARCH process the confidence intervals of the Hill’s estimator, obtained with a bootstrap method, can be seven or eight times higher than the standard deviation computed under the Normality assumption. Moreover, note that for an exact power law function, the second order term in the expansion (7) is undetermined, and the procedures of the endogenous selection of the tail size would produce unsystematic output over the whole range of the available data set.

As a final remark, it is worth emphasizing that since the Hill estimator is the conditional maximum likelihood estimator for the case of power-law tails (2a), it always yields a positive and finite value of the index of the tail. We, therefore, cannot expect this procedure to detect exponential tails or finite endpoints. As can be seen via Monte Carlo simulations, one would obtain high values of estimated tail indices when

¹⁰For a more detailed description of the methods see [13].

applying the Hill method to data drawn from, e.g., the Normal or an Exponential distribution. This is to be expected as in this case we have limiting behavior according to eq. (2c), and therefore, $\gamma = 0$ or $\alpha \rightarrow \infty$. Since the 'true' limit is not attainable, the exponential decay would be indicated by high empirical estimates of α . It, therefore, seems wise to not attribute too much credibility to estimates beyond, say, $\alpha = 10$ as -at least with typical sizes of data set available in the social sciences- a power-law tail behavior with such an index would be virtually indistinguishable from exponential decay.

References

1. J. Beirlant, P. Vynckier, and J.L. Teugels. *Practical Analysis of Extreme Values*. University Press, Leuven, 1996.
2. J-P Bouchaud and M. Potters. *Theory of Financial Risks. From Statistical Physics to Risk Management*. University press, Cambridge, 2000.
3. R.B. D'Agostino and M.A. Stephens. *Goodness-of-Fit Techniques*. Marcel-Dekker, inc., New York, 1986.
4. J. Danielsson, L. de Hann, Peng L., and C.G. de Vries. Using a bootstrap method to choose the optimal sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76:226–248, 2001.
5. J. Danielsson and C.G. de Vries. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4:241–257, 1997.
6. H. Drees and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and Their Applications*, 75:149–172, 1998.
7. W. Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley and Sons, New York, 1971.
8. C.M. Goldie and R.L. Smith. Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics*, 38:45–71, 1987.
9. P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B*, 44:37–42, 1982.
10. P. Hall. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32:177–203, 1990.
11. B. M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1173, 1975.
12. P. Kearns and A. Pagan. Estimating the density tail index for financial time series. *Review of Economics and Statistics*, 79:171–175, 1997.
13. T. Lux. The limiting extremal behavior of speculative returns: an analysis of intra-daily data from the Frankfurt stock exchange. *Applied Financial Economics*, 11:299–315, 1996.
14. T. Lux. On moment condition failure in german stock returns: an application of recent advances in extreme value statistics. *Empirical Economics*, 20:641–652, 2000.
15. D.M. Mason. Laws of large numbers for sums of extreme values. *Annals of Probability*, 10:754–764, 1982.
16. R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser, Basle, 1997.
17. H. Rootzen, M.R. Leadbetter, and L. de Hann. On the distribution of tail array sums for strongly mixing stationary sequences. *Annals of Applied Probability*, 8:868–885, 1998.

Appendix

In this appendix we derive the maximum likelihood estimator of the index α of a Pareto law, developed by Hill [11].

Assume we are interested in the tail behavior of a sequence of T iid variables x_i , which we assume are distributed approximatively according to a Pareto distribution with index α :

$$F^{[u]}(x) = 1 - \left(\frac{u}{x}\right)^\alpha, \quad x \geq u,$$

where u is a scaling parameter that regulates the amplitude of the tail. The corresponding density is given by:

$$f(x) = \alpha \frac{u^\alpha}{x^{\alpha+1}}.$$

The maximum of the log-likelihood function is the solution of the maximization problem:

$$L_T(\alpha) = \max_{\alpha} \sum_{i=1}^T \ln f(x_i) = \max_{\alpha} \sum_{i=1}^T [\ln \alpha + \alpha \ln u - (\alpha + 1) \ln x_i],$$

which yields

$$\frac{1}{\alpha} = \frac{1}{T} \sum_{i=1}^T [\ln x_i - \ln u].$$

The implementation of this estimation procedure for a time series is straightforward. Let us assume we have N observations of an iid random variable. The first step is the selection of a suitable subset of k sample elements that belong to the tail of the distribution. Then, taking descending order statistics, i.e. we arrange the data in the order $x_{(N)} > x_{(N-1)} \dots > x_{(N-k)}$, where now $u = \min(x_{(i)}) = x_{(N-k)}$. If we denote by γ the inverse of α , we end up with eq. (5) in the main text:

$$\hat{\gamma}_{k,N} = H_{k,N} = \frac{1}{k} \sum_{i=1}^k [\ln x_{(N-i+1)} - \ln x_{(N-k)}].$$

It is obvious that this procedure is optimal for data strictly following a Pareto distribution. Asymptotic consistency of the estimator for fat-tailed distributions has been demonstrated by Mason [15], for $k \rightarrow \infty$ and $k/N \rightarrow 0$; moreover, for IID sequences, asymptotic Normality has been demonstrated by Hall [9] and Goldie *et al.* [8].