ESTIMATION OF THE MAXIMAL MOMENT EXPONENT WITH CENSORED DATA

Nuno Crato

Department of Mathematics, ISEG – Instituto Superior de Economia e Gestão,
R. do Quelhas 6, 1200–781 Lisbon, Portugal

Key words: Extreme event, Heavy tail, Hill estimator, Pareto distribution, Stable distribution.

ABSTRACT

Heavy-tailed distributions have been used to model phenomena in which extreme events occur with high probability. In these type of occurrences, it is likely that extreme events are not observable after a certain threshold. Appropriate estimators are needed to deal with this type of censored data. We show that the well-known Hill-Hall estimator is unable to deal with censored data and yields highly biased estimates. We propose and study an unbiased modified maximum likelihood estimator, as well as a truncated tail regression estimator. We assess the expected value and the variance of these estimators in the cases of stable- and Pareto-distributed data.

1. INTRODUCTION

Heavy-tailed distributions such as the Pareto and the non-Gaussian stable distributions have been extensively used to model situations in which extreme values are observed with a relatively high probability.

The interest on these type of probability laws can be traced back to the work of Vilfredo Pareto on income distribution and to the work of Paul Lévy on the properties of stable random variables. Later on, Mandelbrot's work on fractal behavior and self-similar laws was instrumental in establishing the use of heavy-tailed distributions for modeling various real-world phenomena. In two influential papers, Mandelbrot (1960, 1963) introduced stable laws for modeling the stock market price changes. Areas as diverse as statistical physics, weather forecasts, earthquake prediction, economics, and risk theory have been using heavy-tailed distributions (e.g., Mandelbrot 1983, Embrechts, Klüppelberg, and Mikosh 1997). More recently, these distributions have been applied for modeling time delays on the World Wide Web (Willinger *et al.* 1994) and computing costs of random algorithms (Gomes, Selman, and Crato 1997).

Various methods have been proposed for estimating the parameters of stable and Pareto distributions. Huhey (1991) provides a survey and a comparison of a number of methods. Adler, Feldman, and Taqqu (1998) provide a recent and thorough overview.

Many times, the single most important parameter to estimate on heavy-tailed data is the characteristic exponent α of the assumed distribution. This parameter determines the rate of decay of the tail which determines the probability of occurrence of extreme events and the existence of moments for the distribution. In order to estimate this parameter α, it is natural to pay a particular attention to extreme events. In many cases, the available methods are semiparametric: they estimate α and are not otherwise concerned with the characteristics of the distribution. The maximum-likelihood estimator introduced by Hill (1975) and studied by Hall (1982), henceforth the Hill-Hall estimator, is a parametric estimator for observations generated from a Pareto distribution. This estimator, however, has been used as a semiparametric estimator for other heavy-tail distributions. Once truncated for non-extreme observations, it provides an asymptotically unbiased estimator for the tail behavior. The Hill-Hall estimator is arguably the most natural and the most used of such estimators (see Embrechts, Klüppelberg, and Mikosh, 1997, p. 331).

It is possible, however, to face censored observations and to be unable to observe the most extreme data. Consider, for instance, the measurement of physical phenomena, such as wind speed or earthquake intensity. Non-Gaussian stable distributions have been deemed appropriate for this type of phenomena. However, in many extreme situations no measurements are available, since extreme hurricanes or very destructive earthquakes can damage the gauges. Consider also financial data, such as stock market returns for which non-Gaussian stable models have been used (de Lima 1997). In moments of high volatility, exactly when extreme data appear, many stock exchange markets have rules for limiting the transactions or even for closing the market, in order to avoid extreme oscillations. Consider, finally, the study of random algorithms. In many cases, the computing costs of some instances are so high that the algorithms have to stop and run with different starting points—the exact computing costs are not observable after a certain threshold (Gomes, Selman, and Crato 1998).

A truncation of extreme values constitutes a major problem for the available estimators. To the best of our knowledge, no estimator for the α parameter has been developed for explicitly dealing with censoring and it is worthwhile to develop appropriate methods.

We have studied two methods. First, we have used a variant of the Hill-Hall maximum likelihood estimator, which is a widely used estimator. Second, we have applied a crude regression method to the log empirical quantiles, using only observations resulting from a lower and upper truncation. By means of simulation, we show that these estimators deal appropriately with the upper truncation, while the original Hill-Hall estimator is seriously biased.

The plan for the rest of this paper is as follows. Section 2 presents heavy-tail distributions and discusses estimators for the characteristic exponent α. Section 3 presents a simulation study comparing the non-corrected Hill-Hall, the modified maximum likelihood, and the regression estimators. Serious biases on the non-corrected Hill-Hall estimator are detected and the finite-sample properties of the new estimator are investigated. Section 4 concludes.

2. Estimating the characteristic exponent α

We will consider probability laws, such as the stable and the Pareto laws, which asymptotically have tails of the Pareto-Lévy form, *viz.*

$$\Pr\{|X| > x\} \sim C\, x^{-\alpha}, \tag{1}$$

where α is a constant in the interval $(0, 2)$. One or both tails of these distributions have a *hyperbolic decay*. Without loss of generality, we will discuss the right tail behavior and assume that the distribution has support on the positive half line only.

The constant α is called the *characteristic exponent* or *index of stability* of the distribution. Since the existence or nonexistence of moments is completely determined by the tail behavior, α is also called the *maximal moment exponent* of the distribution. Moments of X of order less than α are finite while higher order moments are infinite, i.e., $\alpha = \sup\{a > 0 : \mathrm{E}|X|^a < \infty\}$.

As in many real-life applications, we will consider only the asymptotic tail behavior of the distributions and the estimation of the parameter α.

Stable distributions with $\alpha < 2$—also called *non-Gaussian stable* distributions—are an important example of probability laws with heavy tails of the Pareto-Lévy type. Distributions with tails of the form (1) are in the *domain of attraction* of stable distributions, *i.e.*, properly normalized sums of variables with tails of the Pareto-Lévy type converge in distribution to an α-stable random variable.

Stable laws

Stable laws are defined either by their characteristic functions or by their closure to normalized convolutions. More precisely, we say the random variable X follows a *stable law* if there is an $\alpha \in (0, 2]$ and a real number D_n such that

$$X_1 + X_2 + \ldots + X_n \overset{\mathcal{D}}{=} n^{1/\alpha} X + D_n \tag{2}$$

where $X_1, X_2, \ldots, X_n$ are independent copies of X and $\overset{\mathcal{D}}{=}$ means equality in distribution.

Only in a few special cases the form of a stable distribution is known in analytic form. These cases include the Gaussian distribution, which corresponds to a stable distribution with $\alpha = 2$. For $\alpha < 2$, this parameter is the maximal moment exponent of the stable law. A modern treatment of stable distributions can be found in Samorodnitsky and Taqqu (1994).

Pareto laws

We will consider Pareto random variables X with probability density functions of the form

$$f(x) = \alpha C^{\alpha} x^{-\alpha-1}, \quad \text{for } 0 < C \le x < \infty,\ 0 < \alpha. \tag{3}$$

This is a Pareto distribution of the first of three kinds. Although the estimators we are going to discuss could be applied to Pareto laws without this restriction, we will consider the case $\alpha \le 2$.

The Hill-Hall maximum likelihood estimator

Let $X_{n1} \le X_{n2} \le \ldots \le X_{nn}$ be the order statistics, i.e., the ordered values of the sample $X_1, X_2 \ldots, X_n$. Set $r < n$ as a truncation value which allows to consider only the extreme observations. The Hill estimator is

$$\hat{\alpha}_r = \left(r^{-1} \sum_{j=1}^{r} \ln X_{n,n-j+1} - \ln X_{n,n-r} \right)^{-1}. \tag{4}$$

Hall (1982) has established the asymptotic normality of this estimator and determined the optimal choice of the truncation parameter r. However, since this parameter is a function of the *unknown* parameters of the distribution, it is common to use a set of different truncation values.

The maximum likelihood estimator for censored data

If the most extreme data are not observed, then we have to modify the Hill-Hall maximum likelihood estimator, by conditioning both on the lower (arbitrary) truncation and the upper censoring value. By following the arguments of Hill (1975) with the appropriate changes, it is found that a third term has to be added to (4). Details are available from the author upon request.

The resulting estimator is

$$\hat{\alpha}_{r,u} = \left(\frac{1}{r} \sum_{j=1}^{r-1} \ln X_{n,n-r+j} + \frac{u+1}{r} \ln X_{n,n} - \frac{u+r}{r} \ln X_{n,n-r} \right)^{-1} . \tag{5}$$

The variance of this estimator can be computed as in Hill (1975), obtaining

$$\widehat{\mathrm{Var}}(\hat{\alpha}) = \frac{\hat{\alpha}^2 (r+1)^2}{r^2 (r-1)}. \tag{6}$$

With this notation, n is the number of observed variables, $r+1$ is the number of selected upper order statistics, and u is the number of unobserved extreme values. If all variables are observable and there is no extreme truncation, then $u = 0$ and (5) is just the usual Hill-Hall estimator.

The regression estimator

If a Pareto-Lévy tail is observed, then the rate of decrease of the estimated density is hyperbolic — *i.e.*, slower than the exponential rate. The complement to one of the cumulative distribution, $\overline{F}(\cdot)$, also displays a hyperbolic decay:

$$\overline{F}(x) = 1 - F(x) = \Pr\{X > x\} \sim C\, x^{-\alpha}. \tag{7}$$

Then, for a heavy-tailed random variable, a log-log plot of the frequency of observed values after x should show an approximate linear decrease at the tail. Moreover, the slope of the observed linear decrease provides an estimate of the index α. In contrast, for a distribution with an exponentially decreasing tail, the log-log plot should show a faster-than-linear decrease of the tail.

A natural regression estimator stemming from (7) is the ordinary least squares (OLS) estimator, discussed in Adler, Feldman, and Taqqu (1998, p. 10). This estimator can be readily expressed in terms of a selected number of extreme statistics. The censored data case is handled without any special problems. The only practical point to stress is the need to record the number u of unobserved data.

Assume a sample of $k = n + u$ iid random variables is drawn. Assume also that only the n smallest values of the random variable X are observed and we constitute the order statistics $X_{n1} \leq X_{n2} \leq \ldots \leq X_{nn}$. Assume that, for $X_{n,n-r} \leq X \leq X_{nn}$, the tail distribution is of the Pareto-Lévy type. Then, it is easy to verify that the OLS regression estimator for the maximal moment exponent α is

$$\hat{\alpha} = -\frac{\sum l_i \log X_{ni} - \sum l_i \sum \log X_{ni} / (r+1)}{\sum (\log X_{ni})^2 - \left(\sum \log X_{ni}\right)^2 / (r+1)} \tag{8}$$

where $l_i = \log \frac{n+u-i}{n+u}$ and the sums range from $i = n - r$ to $i = n$. This estimator is simply the regression estimator properly considering the quantiles in face of censured data. If all $k = n + u$ values of the sample are observed, then $u = 0$ and $k = n$.

3. Finite sample properties of the estimators

In order to study the finite sample performance of the estimators, a large size simulation experiment was conducted. We considered the following options and performed 1,000 replications for each resulting case.

Distribution. We generated random samples from both Pareto laws of the first kind (equation 3) and Stable laws with maximal positive skewness. Pareto variates X were simply generated by a direct transformation of pseudo-random uniform variates U: $X = U^{-1/\alpha}$. Stable variates were generated by the method of Chambers, Mallows and Stuck (1976).

Maximal Moment Exponent. We used the values $\alpha = 0.50$, 0.75, 1.00, 1.25, 1.50, 1.75, and 2.00. For the stable distribution, this last case does not correspond to a Pareto-Lévy tail and was studied for reference.

Sample Size. We used sample sizes $k = 500$, 1,000, 5,000, and 10,000.

Unobserved Data Points. We censored the u most extreme observations on the range $u = 1$, 5, 10, 50, 100, and 500. However, we never censored more than 10% of the data, since this is not a realistic situation.

Truncation. The data points actually used for the estimations are the $r + 1$ most extreme observed values. This parameter was set as a fraction of the number of observed values: 1%, 2.5%, 5%, 10%, 15%, and 20%.

Estimators. We used (i) the uncorrected Hill-Hall estimator (Hill), as displayed in equation (4); (ii) the corrected for truncation Hill-Hall estimator (T. Hill), as displayed in equation (5); and (iii) the regression estimator (Regression), as displayed in equation (8). In all cases, the standard deviation (s.d.) of the obtained estimates was computed. In the T. Hill case, we also show the theoretical standard deviation (Theor. s.d.) as computed from equation (6).

The complete results of the simulations are available from the author upon request. In the discussion and the tables presented below, we highlight the main results. Table I displays a typical situation for Pareto distributed data. Table II displays the same situation for stable distributed data.

Firstly, we note a fact that persists across all studied cases: *the Pareto distributed data allows for better estimates of α than the stable data*. As McCulloch (1997) has noted in the context of non-censured samples, only asymptotically does the stable distribution display a hyperbolic tail. However, the Hill-Hall estimator (4) assumes a perfectly hyperbolic tail. As a result, estimates for α are positively biased for this distribution. In contrast, the Pareto distribution has a tail which displays a hyperbolic decay across its complete support. As a result, the corresponding estimates are less biased.

This can have serious consequences when assessing the existence of moments is crucial for later data modeling (see, e.g., de Lima 1997). Consider, for instance, the case of $k = 1000$ data points, noncensored data, $u = 0$ (Hill and the T. Hill estimators coincide), and $r = 0.05(k - u) = 50$ points used for the estimation. This corresponds to the first blocks of the tables I and II. In the case $\alpha = 1.75$, the estimates for the Pareto data correctly indicate that data was generated from a distribution with no finite variance ($\alpha < 2$). The estimates for the corresponding stable data, however, indicate that data was generated from a distribution with finite variance ($\alpha \geq 2$). This type of positive bias for the stable data can again be very problematic on a neighborhood of 1, making data distributed from a stable law with infinite mean ($\alpha \leq 1$) look as if generated by a distribution with a first moment ($\alpha > 1$).

Secondly, we note that the *all estimators display marked biases*. These biases are systematically positive for the Hill-Hall estimator and systematically negative for the regression estimator. As tables I and II show, the positive bias of the Hill-Hall estimator is more serious for the stable data, while the negative bias of the regression estimator is less serious for the same data. This fact is related with the bias above noted: the stable tail shape forces the values of the estimates to increase; consequently, the Hill-Hall positive bias is worsened and the regression negative bias is weakened.

Thirdly, restricting ourselves to the situation to which the estimators were designed, i.e., noncensored data from a perfect hyperbolic tail, we note that *the Hill-Hall estimator behaves better than the regression estimator*. This is not surprising, since the Hill-Hall is the maximum likelihood estimator.If we analyze the first block of Table I, we note that both the bias and the variance of the Hill-Hall estimator are smaller than the corresponding values on the regression estimator.

TABLE I: Average and standard deviation of the $\hat{\alpha}$ estimates for samples with size $k = 1000$ resulting from a Pareto distribution with tail parameter α; sample truncation set as $r = 0.05(k - u)$ with u censored values.

α	.50	.75	1.00	1.25	1.50	1.75	2.00
$u = 0$							
Hill	0.511	0.765	1.019	1.282	1.532	1.800	2.035
(s.d.)	(0.076)	(0.114)	(0.150)	(0.183)	(0.212)	(0.267)	(0.313)
T. Hill	0.511	0.765	1.019	1.282	1.532	1.800	2.035
(s.d)	(0.076)	(0.114)	(0.150)	(0.183)	(0.212)	(0.267)	(0.313)
(Theor. s.d.)	(0.073)	(0.109)	(0.146)	(0.182)	(0.219)	(0.255)	(0.291)
Regression	0.462	0.697	0.919	1.160	1.379	1.633	1.847
(s.d.)	(0.092)	(0.135)	(0.184)	(0.224)	(0.268)	(0.329)	(0.381)
$u = 1$							
Hill	0.547	0.827	1.093	1.378	1.643	1.931	2.199
(s.d.)	(0.080)	(0.122)	(0.162)	(0.200)	(0.241)	(0.287)	(0.319)
T. Hill	0.508	0.768	1.015	1.279	1.526	1.793	2.043
(s.d)	(0.074)	(0.113)	(0.149)	(0.184)	(0.223)	(0.265)	(0.295)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.474	0.719	0.944	1.192	1.428	1.671	1.910
(s.d.)	(0.084)	(0.132)	(0.172)	(0.213)	(0.253)	(0.306)	(0.340)
$u = 5$							
Hill	0.668	1.001	1.338	1.681	2.000	2.338	2.676
(s.d.)	(0.105)	(0.145)	(0.201)	(0.247)	(0.300)	(0.369)	(0.413)
T. Hill	0.511	0.766	1.022	1.287	1.528	1.790	2.048
(s.d)	(0.079)	(0.110)	(0.151)	(0.184)	(0.226)	(0.273)	(0.311)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.491	0.735	0.981	1.234	1.464	1.716	1.954
(s.d.)	(0.088)	(0.123)	(0.166)	(0.202)	(0.251)	(0.297)	(0.334)
$u = 10$							
Hill	0.787	1.190	1.575	1.979	2.374	2.794	3.152
(s.d.)	(0.117)	(0.179)	(0.233)	(0.306)	(0.376)	(0.408)	(0.471)
T. Hill	0.508	0.768	1.015	1.275	1.530	1.796	2.034
(s.d)	(0.073)	(0.112)	(0.144)	(0.186)	(0.231)	(0.250)	(0.292)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.493	0.744	0.983	1.238	1.480	1.733	1.969
(s.d.)	(0.081)	(0.125)	(0.156)	(0.209)	(0.249)	(0.269)	(0.314)
$u = 50$							
Hill	1.702	2.553	3.433	4.264	5.124	6.013	6.871
(s.d.)	(0.278)	(0.393)	(0.587)	(0.694)	(0.820)	(1.008)	(1.097)
T. Hill	0.509	0.761	1.021	1.275	1.533	1.805	2.048
(s.d)	(0.077)	(0.108)	(0.155)	(0.187)	(0.223)	(0.278)	(0.298)
(Theor. s.d.)	(0.075)	(0.113)	(0.151)	(0.188)	(0.226)	(0.264)	(0.301)
Regression	0.498	0.747	0.999	1.249	1.508	1.772	2.005
(s.d.)	(0.082)	(0.116)	(0.164)	(0.196)	(0.240)	(0.291)	(0.323)

TABLE II: Average and standard deviation of the $\hat{\alpha}$ estimates for samples with size $k = 1000$ resulting from a stable distribution with tail parameter α; sample truncation set as $r = 0.05(k - u)$ with u censored values.

α	.50	.75	1.00	1.25	1.50	1.75	2.00
$u = 0$							
Hill	0.513	0.804	1.068	1.116	1.512	2.440	6.080
(s.d.)	(0.076)	(0.119)	(0.148)	(0.153)	(0.229)	(0.405)	(0.773)
T. Hill	0.513	0.804	1.068	1.116	1.512	2.440	6.080
(s.d)	(0.076)	(0.119)	(0.148)	(0.153)	(0.229)	(0.405)	(0.773)
(Theor. s.d.)	(0.073)	(0.109)	(0.146)	(0.182)	(0.219)	(0.255)	(0.291)
Regression	0.464	0.715	0.960	1.055	1.347	1.895	6.604
(s.d.)	(0.091)	(0.147)	(0.189)	(0.188)	(0.259)	(0.436)	(0.883)
$u = 1$							
Hill	0.546	0.862	1.150	1.195	1.653	2.687	6.348
(s.d.)	(0.080)	(0.124)	(0.173)	(0.170)	(0.258)	(0.453)	(0.824)
T. Hill	0.508	0.799	1.067	1.114	1.533	2.462	5.978
(s.d)	(0.074)	(0.114)	(0.159)	(0.157)	(0.238)	(0.415)	(0.759)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.473	0.735	0.994	1.084	1.403	2.012	6.374
(s.d.)	(0.086)	(0.133)	(0.184)	(0.178)	(0.252)	(0.424)	(0.824)
$u = 5$							
Hill	0.668	1.057	1.395	1.463	2.067	3.530	7.161
(s.d.)	(0.096)	(0.156)	(0.201)	(0.253)	(0.340)	(0.642)	(0.946)
T. Hill	0.511	0.805	1.066	1.125	1.574	2.601	5.660
(s.d)	(0.072)	(0.117)	(0.151)	(0.182)	(0.246)	(0.466)	(0.706)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.491	0.763	1.023	1.075	1.463	2.282	5.853
(s.d.)	(0.081)	(0.127)	(0.171)	(0.171)	(0.243)	(0.473)	(0.796)
$u = 10$							
Hill	0.782	1.279	1.650	1.815	2.540	4.405	8.041
(s.d.)	(0.120)	(0.202)	(0.263)	(0.368)	(0.432)	(0.818)	(1.129)
T. Hill	0.504	0.819	1.063	1.160	1.608	2.721	5.380
(s.d)	(0.074)	(0.125)	(0.163)	(0.204)	(0.255)	(0.508)	(0.701)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.487	0.786	1.030	1.067	1.481	2.507	5.492
(s.d.)	(0.079)	(0.131)	(0.176)	(0.171)	(0.248)	(0.553)	(0.773)
$u = 50$							
Hill	1.697	2.881	3.487	12.026	8.950	10.443	12.872
(s.d.)	(0.281)	(0.470)	(0.560)	(3.456)	(1.730)	(1.691)	(2.051)
T. Hill	0.507	0.853	1.047	3.021	2.516	3.137	3.986
(s.d)	(0.077)	(0.127)	(0.151)	(1.083)	(0.516)	(0.473)	(0.557)
(Theor. s.d.)	(0.075)	(0.113)	(0.151)	(0.188)	(0.226)	(0.264)	(0.301)
Regression	0.498	0.834	1.032	2.927	2.452	3.076	3.962
(s.d.)	(0.081)	(0.135)	(0.162)	(1.137)	(0.536)	(0.494)	(0.601)

Finally, we analyze the behavior of the estimators in the presence of censored data. As tables I and II perfectly show, *even a very small truncation, say* $u = 5$ *or even* $u = 1$*, is enough to seriously bias the Hill-Hall estimator*. Consider, for instance, the Pareto distributed data for sample size $k = 1000$, a 5% truncation r, and 5 censored data points. The positive bias attains values of 0.3 and 0.5, which can seriously mislead the practitioner. The magnitude of the bias increases with the actual values of α.

For the stable distributed data, the bias due to censoring adds to the existing positive bias. With the parameters considered above, data generated by a stable distribution with $\alpha = 1.5$ is likely to be mistaken by a distribution with finite moments.

The bias increases as the sample size decreases, as expected. The bias also increases as the number of censored data points increases. However, even for very large samples, say $k = 10000$, and a very extreme truncation, say $r = 0.01(k - u)$, the biases are still apparent. Table III reveals these biases for large sample sizes.

These results show how unreliable the Hill-Hall estimator is in the presence of censoring. We now analyze how the new corrected estimator and the regression method are able to cope with censoring.

Firstly, we notice that *the biases do not deteriorate with censored data*, provided censoring is taken into account. The truncated maximum likelihood estimator (T. Hill) reveals a positive bias, which does not change with censoring, and the regression estimator reveals a negative bias, which does not change either. This means that, whatever drawbacks these estimators have, these drawbacks are not exacerbated by censoring, provided this fact is correctly taken into account.

Secondly, we note that *the biases given by the corrected estimator decrease with the sample size and do not increase with censoring*. Consider Table III, we do not see any bias pattern across censoring sizes. In many cases, the bias even decreases as censoring increases, but we cannot detect any rule over the natural simulation uncertainty.

Thirdly, we note the *reasonable performance of the regression estimator*. The resulting biases are of the same order of magnitude as the ones resulting from the modified Hill-Hall estimator.

Finally, we note the *very reasonable behavior of the corrected Hill-Hall estimator*. This estimator works as well with censored as with noncensored data. The magnitude of the biases are within the limits expected for this type of parameter estimation—the problem is well-known for its difficulty.

TABLE III: Average and standard deviation of the $\hat{\alpha}$ estimates for samples with sizes k resulting from a Pareto distribution with tail parameter α; sample truncation set as $r = 0.01(k - u)$ with u censored values.

α	.50	.75	1.00	1.25	1.50	1.75	2.00
$k = 5000$							
$u = 1$							
Hill	0.551	0.824	1.090	1.373	1.634	1.913	2.218
(s.d.)	(0.084)	(0.118)	(0.156)	(0.199)	(0.237)	(0.270)	(0.312)
T. Hill	0.512	0.765	1.012	1.274	1.517	1.777	2.060
(s.d)	(0.078)	(0.110)	(0.144)	(0.184)	(0.219)	(0.250)	(0.290)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.479	0.714	0.945	1.190	1.415	1.652	1.930
(s.d.)	(0.088)	(0.128)	(0.171)	(0.209)	(0.253)	(0.283)	(0.338)
$u = 5$							
Hill	0.668	1.004	1.337	1.655	1.985	2.323	2.669
(s.d.)	(0.097)	(0.152)	(0.196)	(0.244)	(0.284)	(0.357)	(0.412)
T. Hill	0.512	0.769	1.021	1.270	1.521	1.779	2.042
(s.d)	(0.072)	(0.113)	(0.147)	(0.182)	(0.212)	(0.269)	(0.307)
(Theor. s.d.)	(0.074)	(0.110)	(0.147)	(0.184)	(0.221)	(0.258)	(0.295)
Regression	0.492	0.739	0.976	1.226	1.462	1.711	1.959
(s.d.)	(0.082)	(0.125)	(0.163)	(0.203)	(0.227)	(0.299)	(0.327)
$k = 10000$							
$u = 1$							
Hill	0.529	0.790	1.055	1.323	1.589	1.847	2.110
(s.d.)	(0.053)	(0.082)	(0.108)	(0.138)	(0.165)	(0.178)	(0.212)
T. Hill	0.506	0.756	1.010	1.268	1.522	1.769	2.020
(s.d)	(0.051)	(0.078)	(0.103)	(0.132)	(0.158)	(0.171)	(0.202)
(Theor. s.d.)	(0.051)	(0.077)	(0.102)	(0.128)	(0.153)	(0.179)	(0.204)
Regression	0.482	0.719	0.965	1.213	1.453	1.691	1.928
(s.d.)	(0.060)	(0.094)	(0.123)	(0.156)	(0.189)	(0.218)	(0.244)
$u = 5$							
Hill	0.595	0.891	1.186	1.483	1.783	2.077	2.386
(s.d.)	(0.063)	(0.089)	(0.125)	(0.153)	(0.187)	(0.221)	(0.251)
T. Hill	0.507	0.758	1.010	1.262	1.516	1.770	2.031
(s.d)	(0.053)	(0.074)	(0.105)	(0.127)	(0.155)	(0.187)	(0.209)
(Theor. s.d.)	(0.051)	(0.077)	(0.102)	(0.128)	(0.153)	(0.179)	(0.204)
Regression	0.493	0.739	0.983	1.226	1.473	1.725	1.978
(s.d.)	(0.060)	(0.088)	(0.121)	(0.141)	(0.175)	(0.214)	(0.237)

The estimator also displayed a variability with fine agreement with the one theoretically computed—the standard error computed from the simulations (s.d.) and the standard deviation resulting from (6) are very close indeed. This fact allows for a reasonable confidence in the use of the proposed estimator.

4. Conclusions and future work

We have assessed the bias and variance of different estimators for the maximal moment exponent of Pareto and stable distributions. Specifically, we have considered the well-known Hill-Hall estimator, a modified Hill-Hall estimator able to deal with censored extreme data, and a regression method.

All methods worked better for the data generated by Pareto laws than for the data generated by stable laws, confirming recent findings of McCulloch (1997). The stable laws only asymptotically have tails of the Pareto-Lévy type, while Pareto laws have this type of tail behavior throughout all their support.

When data are uncensored and the only truncation is the one chosen by the analyst in order to select the extreme observations, the original and the modified Hill-Hall estimators coincide and the results are very similar to the ones obtained by the regression method. As it has been noted by other researchers, all methods showed some finite-sample bias, even for samples of large size. However, the maximum likelihood estimator performed generally better than the regression method.

When data are censored for the most extreme observations, this study reveals that the original Hill-Hall estimator is very seriously biased upwards, even when a very small number of data points are censored.

With censoring, both the regression method and the modified Hill-Hall estimator performed relatively well. The biases obtained were very small in both cases, although the regression methods has shown, in general, a larger variance.

We believe these findings prompt the reevaluation the empirical work that has used truncated data for estimating the tail exponent.

ACKNOWLEDGEMENTS

I thank Leslie Dowling-DaCosta and David DaCosta for expert computational help, Carla Pedro Gomes and Bart Selman for raising the problem here discussed, and Bonnie K. Ray for various helpful suggestions.

BIBLIOGRAPHY

Adler R., Feldman R., and Taqqu M.S., editors (1998). *A Practical Guide To Heavy Tails: Statistical Techniques for Analyzing Heavy-Tailed Distributions*, Birkhäuser, Boston.

Chambers, J. M., Mallows, C.L., and Stuck, B.W. (1976) A method for simulating stable random variables. *Journal of the American Statistical Association* **71**, 340–344.

de Lima, P.J.F. (1997) On the robustness of nonlinearity tests to moment condition failure. *Journal of Econometrics* 76, 251–280.

Embrechts, Paul; Klüppelberg, Claudia, and Mikosh, Thomas (1997) *Modelling Extremal Events: For Insurance and Finance*, Springer.

Gomes, C.P., Selman, B., and Crato, N. (1997) Heavy-tailed distributions in combinatorial search, in Gert Smolka (Ed.), *Principles and Practice of Constraint Programming, CP 97*, Lecture Notes in Computer Science 1330, Springer, 121–135.

Gomes, C.P., Selman, B., and Crato, N. (1998) Heavy-tailed phenomena in combinatorial search, Department of Computer Science, Cornell University, mimeo.

Gomes, C.P., Selman, B., Crato, N., and Kautz, H. (2000) Heavy-tailed phenomena in satisfiability and constraint satisfaction problems, *Journal of Automated Research*, forthcoming.

Hall, P. (1982) On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society,* **B 44**, 37–42.

Hill, B. M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174.

Huhey, Richard L. (1991) A survey and comparison of methods for estimating extreme right tail-area quantiles. *Communications in Statistics — Theory and methods* **20**, 1463–1496.

Mandelbrot, B. (1960) The Pareto-Lévy law and the distribution of income. *International Economic Review* **1**, 79–106.

Mandelbrot, B. (1963) The variation of certain speculative prices. *Journal of Business* **36**, 394–419.

Mandelbrot, B. (1983) *The fractal geometry of nature.* Freeman: New York. 1983.

McCulloch, J. Huston (1997) Measuring tail thickness to estimate the stable index α: A critique, *Journal of Business & Economic Statistics* **15**, 74–81.

Renyi, A. (1953) *On the theory of order statistics. Acta Math. Acad. Sci. Hungar*, 4, 191–232.

Samorodnitsky, G. and Taqqu, M. S. (1994) *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York.

Willinger, Walter, Taqqu, Murad S., Leland, Will E., and Wilson, Daniel V. (1994). Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements. *Statistical Science* **10**, 67–85.