Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®
Journal of Statistical Planning and Inference 119 (2004) 341–352

journal of statistical planning and inference
www.elsevier.com/locate/jspi

# Generalized least-squares estimators for the thickness of heavy tails

Inmaculada B. Aban[*,1], Mark M. Meerschaert[1,2]

*Department of Mathematics, College of Arts and Sciences, University of Nevada, Ansari Business Building, Reno, NV 89557-0045, USA*

Received 25 October 2001; accepted 4 August 2002

**Abstract**

For a probability distribution with power law tails, a log–log transformation makes the tails of the empirical distribution function resemble a straight line, leading to a least-squares estimate of the tail thickness. Taking into account the mean and covariance structure of the extreme order statistics leads to improved tail estimators, and a surprising connection with Hill's estimator.


*Keywords:* Least-squares; Linear regression; Heavy tails; Order statistics

## 1. Introduction

A random variable $X$ has heavy tails if the fractional absolute moments $E|X|^\rho = \infty$ for some real $\rho > 0$. This occurs, for example, if $y = P(X > x) = Cx^{-\alpha}$ for $x > 0$ large. In this case, $\log y = \log C - \alpha \log x$ so that a log–log plot of the distribution function for $X$ has a linear tail. Then the slope of the best fitting line through the points $(\log(X_{(n-i+1)}), \log(i/n))$ for $1 \leqslant i \leqslant r$, where $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ are the order statistics (ties are broken arbitrarily) and $r \leqslant n$ counts the number of extreme order statistics, should be approximately equal to $-\alpha$. This graphical estimate of the tail thickness was pioneered by Mandelbrot (1963, 1982) in his seminal work on heavy tails and fractals. Fig. 1 illustrates this method for a typical financial data set. The right portion of the graph is nearly linear, indicating that this data set is heavy tailed.

* Corresponding author. Tel.: +17757846773; fax: +17757846378.
*E-mail addresses:* aban@unr.edu (I.B. Aban), mcubed@unr.edu (M.M. Meerschaert).
1 Prof. Aban and Prof. Meerschaert were partially supported by NSF-DMS grant 0139927.
2 Prof. Meerschaert was partially supported by NSF-DES grant 9980484.

0378-3758/02/$ - see front matter 
PII: S0378-3758(02)00419-6

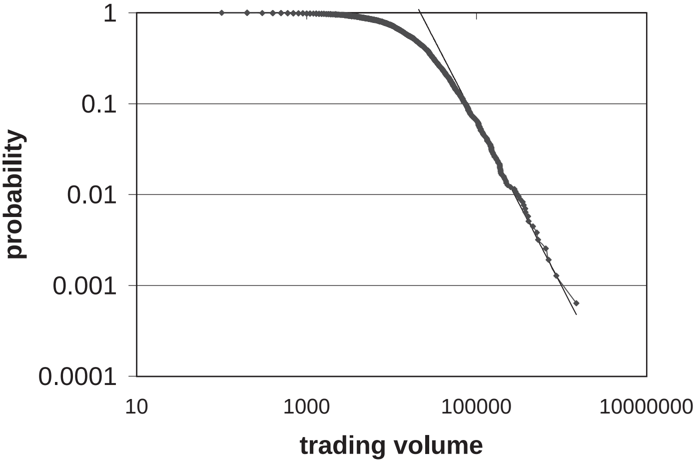


Fig. 1. Daily trading volume of Amazon, Inc. stock from 20 March 1995 to 18 June 2001.

In linear regression models it is standard to place the random response variables on the vertical axis. The slope of the best fitting line through the points $(\log(i/n), \log(X_{(n-i+1)}))$ for $1 \leqslant i \leqslant r$ should be approximately equal to $-1/\alpha$. This method of estimating $\alpha$ was proposed independently by Kratz and Resnick (1996), who call it a $qq$-estimator, and Schultze and Steinebach (1996), who consider the equivalent problem of least-squares estimation for data with exponential tails. This equivalence holds because a random variable $X$ has a Pareto distribution with $P(X > x) = Cx^{-\alpha}$ if and only if $Y = \ln X$ has a shifted exponential distribution with $P(Y > y) = e^{-\alpha(y-d)}$ where $d = \alpha^{-1}\ln C$. Csörgő and Viharos (1997) establish asymptotic normality of this estimator for $P(|X| > x)$ regularly varying with index $-\alpha$, see also Csörgő and Viharos (1998) and Viharos (1999) for some extensions, and a comparison of different tail estimators. The most popular of these is Hill's estimator (Hill, 1975)

$$\hat{\alpha}_{\mathrm{H}}(r) = \left[\frac{1}{r-1}\sum_{i=1}^{r-1}(\ln X_{(n-i+1)} - \ln X_{(n-r+1)})\right]^{-1} \tag{1.1}$$

which is the conditional maximum likelihood estimator for $\alpha$ based on the $r$ largest order statistics for nonnegative data with a Pareto tail. [3] Asymptotic normality for Hill's estimator was addressed in Hall (1982), Hall and Welsh (1985), Haeusler and Teugels (1985), Csörgő and Mason (1985), Beirlant and Teugels (1989) and other references cited there.

As a problem in linear regression, tail estimation is complicated by the fact that deviations of the extreme order statistics from their respective means are neither independent nor identically distributed. Furthermore, if $P(X > x) = Cx^{-\alpha}$ then the mean of the random variable $\ln(X_{(n-i+1)})$ is not equal to $\ln(i/n)$. In this paper, we develop new linear regression estimators of $\alpha$ and $C$, taking into account both the mean and covariance structure of the tail data. We compute the best linear unbiased estimator

[3] Hill's original estimator divides by $r$ instead of $r-1$ but the unbiased version (1.1) is now standard.

(BLUE) of $\alpha^{-1}$ and $d=\alpha^{-1}\ln C$, based on the $r$ largest order statistics, and we show that the resulting estimator is also the uniformly minimum variance unbiased estimator (UMVUE). Surprisingly, this estimator for $\alpha$ also turns out to equal Hill's estimator, establishing an equivalence between the MLE and BLUE estimators, and placing Hill's estimator in the context of linear regression.

## 2. Best linear unbiased estimator

In this section, we compute the best linear unbiased estimator for a random sample with a Pareto tail. It is mathematically simpler to begin with the equivalent case of a shifted exponential, which is related to the Pareto by a logarithmic transformation. Consider a random sample $X_1,\ldots,X_n$ from a population with cumulative distribution function of the form $H(x)=G((x-\mu)/\sigma)$, $\sigma>0$. If we define $Y=(X-\mu)/\sigma$, then $Y$ has distribution function $G(y)$ free of the parameters $\mu$ and $\sigma$. Let $X_{(1)}<X_{(2)}<\cdots<X_{(n)}$ and $Y_{(1)}<Y_{(2)}<\cdots<Y_{(n)}$ be the order statistics associated with $X_i$ and $Y_i$, $i=1,\ldots,n$. The distribution of $Y$ does not depend on the parameters, hence the means and covariances of the order statistics of $Y$, denoted by $E\{Y_{(i)}\}=v_i$ and $\mathrm{Cov}\{Y_{(i)},Y_{(j)}\}=\beta_{i,j}$, where $i,j=1,\ldots,n$, also do not depend on $\mu$ and $\sigma$. It follows from the relation $Y_{(i)}=(X_{(i)}-\mu)/\sigma$ that $E\{X_{(i)}\}=\mu+\sigma v_i$ and $\mathrm{Cov}\{X_{(i)},X_{(j)}\}=\sigma^2\beta_{i,j}$ implying that $E\{X_{(i)}\}$ is linear in the parameters $\mu$ and $\sigma$ while $\mathrm{Cov}\{X_{(i)},X_{(j)}\}$ only depends on $\sigma$.

For any matrix $\mathbf{W}$, we denote its transpose, inverse and determinant as $\mathbf{W}'$, $\mathbf{W}^{-1}$, and $|\mathbf{W}|$, respectively. Let $\mathbf{1}$ denote the $n\times 1$ vector of 1's, $\boldsymbol{v}'=(v_1,\ldots,v_n)$ denote the $n\times 1$ vector of means, let $\mathbf{A}=(\mathbf{1},\boldsymbol{v})$ denote the design matrix, and let $\mathbf{B}$ denote an $n\times n$ matrix with $ij$ element equal to $\beta_{i,j}$. Let $\mathbf{X}'=(X_{(1)},\ldots,X_{(n)})$ and $\boldsymbol{\theta}'=(\mu,\sigma)$. Our linear regression model is $\mathbf{X}=\mathbf{A}\boldsymbol{\theta}+\mathbf{e}$ where the vector of errors $\mathbf{e}$ has mean zero and covariance matrix $\sigma^2\mathbf{B}$. Following Dillon and Goldstein (1984), we transform this generalized linear regression problem to a standard form. Using the Cholesky decomposition $\sigma^2\mathbf{B}=\mathbf{R}\mathbf{R}'$, define $\mathbf{f}=\mathbf{R}^{-1}\mathbf{e}$. The vectors of errors $\mathbf{f}$ are uncorrelated, and we consider the equivalent model $\mathbf{Z}=\mathbf{D}\boldsymbol{\theta}+\mathbf{f}$ with $\mathbf{Z}=\mathbf{R}^{-1}\mathbf{X}$ and $\mathbf{D}=\mathbf{R}^{-1}\mathbf{A}$. Now the Gauss–Markov Theorem (see, e.g., Bickel and Doksum, 1977) implies that the BLUE for $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_B=(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{Z}=(\mathbf{A}'\mathbf{B}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{B}^{-1}\mathbf{X} \tag{2.1}$$

which minimizes the mean squared error $(\mathbf{Z}-\mathbf{D}\boldsymbol{\theta})'(\mathbf{Z}-\mathbf{D}\boldsymbol{\theta})=(\mathbf{X}-\mathbf{A}\boldsymbol{\theta})'\sigma^{-2}\mathbf{B}^{-1}(\mathbf{X}-\mathbf{A}\boldsymbol{\theta})$. Its mean is $\boldsymbol{\theta}$, and its covariance matrix is $\sigma^2(\mathbf{A}'\mathbf{B}^{-1}\mathbf{A})^{-1}$.

If $X_1,\ldots,X_n$ is a random sample from a shifted exponential population with cumulative distribution function $H(x)=\exp\{(x-\mu)/\sigma\}$, then $Y_1,\ldots,Y_n$ are i.i.d. unit exponential random variables. Defining the constants $a_i\equiv\sum_{k=1}^{i}(n-k+1)^{-1}$ and $b_i\equiv\sum_{k=1}^{i}(n-k+1)^{-2}$, for $i=1,\ldots,n$, it is known (see for instance, Barlow and Proschan, 1981) that $E\{Y_{(i)}\}=v_i=a_i$, and $\mathrm{Cov}\{Y_{(i)},Y_{(j)}\}=\beta_{i,j}=b_i$, $i\leqslant j$. Since the matrices $\mathbf{A}$ and $\mathbf{B}$ are completely known, we can easily compute the BLUE for $\boldsymbol{\theta}$ and its covariance matrix.

Now, suppose we assume that only the largest observations fit the shifted exponential model. In this case, we can compute the BLUE for $\boldsymbol{\theta}$ based on the mean and covariance of just these observations. Let $\mathbf{1}_r$ be a $r\times 1$ vector of 1's, $\mathbf{X}_r'=(X_{(n-r+1)},\ldots,X_{(n)})$,

$\boldsymbol{v}_r = (v_{(n-r+1)}, \ldots, v_{(n)})$, $\mathbf{A}_r = (\mathbf{1}_r, \boldsymbol{v}_r)$, and $\mathbf{B}_r = (\beta_{i,j}\colon i,j = n-r+1,\ldots,n)$. Now the BLUE for $\boldsymbol{\theta}$ based on the $r$ largest order statistics is given by $\hat{\boldsymbol{\theta}}_{B_r} = (\mathbf{A}_r'\mathbf{B}_r^{-1}\mathbf{A}_r)^{-1}\mathbf{A}_r'\mathbf{B}_r^{-1}\mathbf{X}_r$. This estimator, which minimizes the mean squared error $(\mathbf{X}_r - \mathbf{A}_r\boldsymbol{\theta})'\sigma^{-2}\mathbf{B}_r^{-1}(\mathbf{X}_r - \mathbf{A}_r\boldsymbol{\theta})$, has mean $\boldsymbol{\theta}$ and covariance matrix $\sigma^2(\mathbf{A}_r'\mathbf{B}_r^{-1}\mathbf{A}_r)^{-1}$. Next we explicitly compute $\hat{\boldsymbol{\theta}}_{B_r} = (\hat{\mu}_{B_r}, \hat{\sigma}_{B_r})'$. This requires some preliminary matrix calculations.

**Lemma 2.1.** *The Cholesky decomposition for* $\mathbf{B}_r$ *is given by the matrix* $\mathbf{R}$ *with elements*

$$\mathbf{R}_{i,j} = \begin{cases} b_{n-r+1}^{1/2} & \text{when } j = 1, \\ (r-j+1)^{-1} & \text{when } i \geqslant j \text{ and } j > 1, \\ 0 & \text{otherwise.} \end{cases}$$

**Proof.** Using the above definition of $\mathbf{R}$, it can be shown that $\mathbf{R}\mathbf{R}' = \mathbf{B}_r$. Since $\mathbf{R}$ is lower triangular with positive diagonal entries, this is the unique Cholesky decomposition of $\mathbf{B}_r$. □

**Lemma 2.2.** *The matrix* $\mathbf{R}^{-1}$ *has the following elements*:

$$\mathbf{R}_{i,j}^{-1} = \begin{cases} b_{n-r+1}^{-1/2} & \text{when } i = j = 1, \\ (r-i+1) & \text{when } i = j \text{ and } j > 1, \\ -(r-i+1) & \text{when } i = j+1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

**Proof.** To prove the result, it can be shown that $\mathbf{R}^{-1}\mathbf{R} = \mathbf{I}$. □

**Lemma 2.3.** *The matrix* $\mathbf{R}^{-1}\mathbf{A}_r$ *has elements given by*

$$(\mathbf{R}^{-1}\mathbf{A}_r)_{i,j} = \begin{cases} b_{n-r+1}^{-1/2} & \text{when } i = 1, j = 1, \\ b_{n-r+1}^{-1/2}a_{n-r+1} & \text{when } i = 1, j = 2, \\ 1 & \text{when } i > 1, j = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

**Proof.** Using Lemma (2.2) and the definition of $\mathbf{A}_r$, the result follows. □

**Lemma 2.4.** *Define* $\mathbf{M} = (\mathbf{A}_r'\mathbf{B}_r^{-1}\mathbf{A}_r)^{-1}\mathbf{A}_r'\mathbf{B}_r^{-1}$ *where* $\mathbf{A}_r$ *is a* $2\times r$ *matrix with* $(\mathbf{A}_r)_{i,1} = 1$ *and* $(\mathbf{A}_r)_{i,2} = a_{n-r+i}$ *for all* $1 \leqslant i \leqslant r$. *We claim that*

$$\mathbf{M}_{i,j} = \begin{cases} 1 + a_{n-r+1} & \text{when } i = 1, j = 1, \\ -1 & \text{when } i = 2, j = 1, \\ -a_{n-r+1}/(r-1) & \text{when } i = 1, j > 1, \\ 1/(r-1) & \text{when } i = 2, j > 1. \end{cases} \tag{2.4}$$

**Proof.** We will verify this in several steps. Since $\mathbf{RR}' = \mathbf{B}_r$ we have $\mathbf{B}_r^{-1} = (\mathbf{R}^{-1})'\mathbf{R}^{-1}$ so that

$$\mathbf{M} = (\mathbf{A}_r'(\mathbf{R}^{-1})'\mathbf{R}^{-1}\mathbf{A}_r)^{-1}\mathbf{A}_r'(\mathbf{R}^{-1})'\mathbf{R}^{-1} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{R}^{-1}$$

where $\mathbf{Q} = \mathbf{R}^{-1}\mathbf{A}_r$. Let $\mathbf{C} = \mathbf{Q}'\mathbf{Q}$. After some simplifications, it can be shown that

$$\mathbf{C} = \begin{pmatrix} b_{n-r+1}^{-1} & b_{n-r+1}^{-1}a_{n-r+1} \\ b_{n-r+1}^{-1}a_{n-r+1} & b_{n-r+1}^{-1}a_{n-r+1}^2 + (r-1) \end{pmatrix}. \tag{2.5}$$

Note that $\Delta = |\mathbf{C}| = b_{n-r+1}^{-1}(r-1)$ and so

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{a_{n-r+1}^2}{r-1} + b_{n-r+1} & -\frac{a_{n-r+1}}{r-1} \\ -\frac{a_{n-r+1}}{r-1} & \frac{1}{r-1} \end{pmatrix}. \tag{2.6}$$

Next define $\mathbf{E} = \mathbf{C}^{-1}\mathbf{Q}'$ and note that

$$\mathbf{E} = \begin{pmatrix} b_{n-r+1}^{1/2} & -\frac{a_{n-r+1}}{r-1} & \cdots & -\frac{a_{n-r+1}}{r-1} \\ 0 & \frac{1}{r-1} & \cdots & \frac{1}{r-1} \end{pmatrix}.$$

Finally, we have $\mathbf{M} = \mathbf{ER}^{-1}$ so that

$$\begin{aligned} \mathbf{M}_{i,j} &= \sum_{k=1}^{r} \mathbf{E}_{i,k}\mathbf{R}_{k,j}^{-1} \\ &= \sum_{k=1}^{r} \left\{ b_{n-r+1}^{1/2} I_{\{i=k=1\}} - \frac{a_{n-r+1}}{r-1} I_{\{i=1,k>1\}} + \frac{1}{r-1} I_{\{i=2,k>1\}} \right\} \\ &\quad \times \{ b_{n-r+1}^{-1/2} I_{\{k=j=1\}} + (r-k+1) I_{\{k=j>1\}} - (r-k+1) I_{\{k=j+1\}} \} \end{aligned} \tag{2.7}$$

Simplifying the above expression leads us to the result. □

Next we apply the matrix calculations above to determine a simplified form of the BLUE for $\mu$ and $\sigma$, based on the $r$ largest order statistics.

**Theorem 2.5.** *Let $X_1, \ldots, X_n$ be a random sample from a shifted exponential population with cumulative distribution function $H(x;\mu,\sigma) = 1 - \exp[(x-\mu)/\sigma]$. The best linear unbiased estimator for $\mu$ and $\sigma$ based on the $r$ largest order statistics is*

$$\hat{\mu}_{B_r} = X_{(n-r+1)} - (r-1)^{-1} a_{n-r+1} \sum_{i=1}^{r} \{X_{(n-i+1)} - X_{(n-r+1)}\}, \tag{2.8}$$

$$\hat{\sigma}_{B_r} = (r-1)^{-1} \sum_{i=1}^{r} \{X_{(n-i+1)} - X_{(n-r+1)}\}. \tag{2.9}$$

**Proof.** Since $\hat{\boldsymbol{\theta}}_{B_r} = (\mathbf{A}_r'\mathbf{B}_r^{-1}\mathbf{A}_r)^{-1}\mathbf{A}_r'\mathbf{B}_r^{-1}\mathbf{X}_r = \mathbf{M}\mathbf{X}_r$, the results of the theorem follow immediately from Lemma 2.4. □

**Corollary 2.6.** *Let $X_1,\ldots,X_n$ be a random sample from a Pareto population with cumulative distribution function $H(x;\alpha,C)=1-Cx^{-\alpha}$. The best linear unbiased estimator for $\alpha^{-1}$ based on the r largest order statistics is the same as Hill's estimator* (1.1).

**Proof.** Recall that $\ln X$ has a shifted exponential distribution with parameters $\sigma=\alpha^{-1}$ and $\mu=\alpha^{-1}\ln C$ and apply Theorem 2.5, noting that the $i=r$ term is zero in the sum. □

**Remark 2.7.** Theorem 2.5 also yields an estimator $\hat{C}_{B_r}=\exp(\hat{\mu}_{B_r}/\hat{\sigma}_{B_r})$ for $C$ which is different from Hill's estimator $\hat{C}_{\mathrm{H}}=(r/n)X_{(n-r+1)}^{\hat{\alpha}_{\mathrm{H}}}$. Of course neither of these is the BLUE for $C$ since they are not linear. Alternatively, the BLUE (2.8) can be used along with the parameterization $P(X>x)=(x/\mathrm{e}^{\mu})^{-\alpha}$.

**Remark 2.8.** Since the BLUE for $\alpha^{-1}$ is the same as Hill's estimator, this estimator is asymptotically normal for certain classes of distributions satisfying $P(X>x)=x^{-\alpha}L(x)$ with $L(x)$ slowly varying. See Csörgő and Viharos (1997) for a survey of these results.

## 3. Uniformly minimum variance unbiased estimator

In this section, we show that the best linear unbiased estimators for the parameters of a shifted exponential distribution, given in Theorem 2.5, are also the UMVUE. Then it follows that Hill's estimator, the BLUE for the tail thickness parameter $\alpha^{-1}$ of a heavy tailed distribution, is also the UMVUE.

**Theorem 3.1.** *The best linear unbiased estimators given by Eqs.* (2.8) *and* (2.9) *in Theorem* 2.5 *are also the unique UMVUE for $\mu$ and $\sigma$.*

**Proof.** From theory of order statistics (see, e.g., David, 1981), the joint density function of the $r$ largest order statistics $(X_{(n-r+1)},X_{(n-r+2)},\ldots,X_{(n)})$ from a shifted exponential distribution with cumulative distribution function $H(x;\mu,\sigma)=1-\exp[(x-\mu)/\sigma]$ is given by

$$\frac{n!}{(n-r)!}\sigma^{-r}\left\{\exp\left[-\left(\frac{\sum_{i=1}^{r}x_{(n-r+i)}-r\mu}{\sigma}\right)\right]\right\}$$
$$\left\{1-\exp\left[-\left(\frac{x_{(n-r+1)}-\mu}{\sigma}\right)\right]\right\}^{n-r} I\{x_{(n-r+1)}\geqslant\mu\}.$$

By the Factorization Theorem (see, e.g., Theorem 6.5 in Lehmann and Casella, 1998), $(X_{(n-r+1)},\sum_{i=1}^{r}X_{(n-r+i)})$ are jointly sufficient statistics for $(\mu,\sigma)$. Consequently $(T_1,T_2)$, where $T_1=X_{(n-r+1)}$ and $T_2=\sum_{i=1}^{r}[X_{(n-r+i)}-X_{(n-r+1)}]$, are also jointly sufficient for $(\mu,\sigma)$ since they are 1–1 transformations of $(X_{(n-r+1)},\sum_{i=1}^{r} X_{(n-r+i)})$ (see, e.g., Lehmann and Casella, 1998, pp. 36–37).

**Lemma 3.2.** *Let $X_{(n-r+1)} < X_{(n-r+2)} < \cdots < X_{(n)}$ be the r largest order statistics from a shifted exponential with shift parameter $\mu$ and scale parameter $\sigma$. Then $T_1 = X_{(n-r+1)}$ and $T_2 = \sum_{i=1}^{r-1}[X_{(n-r+i)} - X_{(n-r+1)}]$ are independently distributed where $T_1$ has marginal density*

$$h_1(t_1;\mu,\sigma) = \frac{n!}{(n-r)!(r-1)!}\sigma^{-1}\exp\left[-\frac{(n-r+1)(t_1-\mu)}{\sigma}\right] \times\left\{1-\exp\left[-\left(\frac{t_1-\mu}{\sigma}\right)\right]\right\}^{r-1} I\{t_1 \geqslant \mu\} \tag{3.1}$$

*and $T_2$ has a Gamma $(r-1,\sigma^{-1})$ density given by $h_2(t_2;\sigma)=[(r-2)!]^{-1}\sigma^{-(r-1)}t_2^{r-2} e^{-t_2/\sigma}$.*

**Proof.** Using standard results on order statistics (see, e.g., David, 1981), the marginal density function of the $j$th smallest order statistic is given by

$$\frac{n!}{(n-j)!(j-1)!}[H(x_{(j)};\mu,\sigma)]^{n-j}h(x_{(j)};\mu,\sigma)[1-H(x_{(n-r+1)};\mu,\sigma)]^{j-1}.$$

Letting $j=n-r+1$, substituting the density and cumulative distribution functions of the shifted exponential, and simplifying, we obtain (3.1). To derive the density of $T_2$, note that if

$$Z_i = (n-i+1)\frac{(X_{(i)}-X_{(i-1)})}{\sigma}, \quad X_{(0)}=0, \quad i=1,\dots,n, \tag{3.2}$$

then $Z_1,\dots,Z_n$ are iid unit exponential random variables (see David, 1981, pp. 20–21 for details). Hence

$$\sum_{i=n-r+2}^{n} Z_i = \sigma^{-1}\sum_{i=n-r+2}^{n}(n-i+1)(X_{(i)}-X_{(i-1)}) = \sigma^{-1}\sum_{i=1}^{r-1}(X_{(n-r+i)}-X_{(n-r+1)}). \tag{3.3}$$

Since $\sum_{i=n-r+2}^{n} Z_i$ is distributed as Gamma$(r-1,1)$, then $T_2=\sum_{i=1}^{r-1}(X_{(n-r+i)} - X_{(n-r+1)})$ is distributed as Gamma$(r-1,\sigma^{-1})$. The independence of $T_1$ and $T_2$ follows from the fact that

$$T_1 = X_{(n-r+1)} = \sigma\sum_{i=1}^{n-r+1}\frac{Z_i}{n-i+1} \quad \text{and}$$

$$T_2 = \sum_{i=1}^{r-1}(X_{(n-r+i)}-X_{(n-r+1)}) = \sigma\sum_{i=n-r+2}^{n} Z_i$$

in view of (3.2). □

Next we show that $(T_1, T_2)$ are complete statistics. This means that if $E_{\mu,\sigma}[f(T_1, T_2)]=0$, for all $\mu, \sigma$ for some measurable real-valued integrable function $f$ then $f(t_1, t_2) = 0$ a.e.-$(\mathrm{d}t_1 \times \mathrm{d}t_2)$. We adapt the proof found in Lehmann and Casella (1998, p. 43). Note that

$$0 = E_{\mu,\sigma}[f(T_1, T_2)] = \int_\mu^\infty \int_0^\infty f(t_1, t_2) h_2(t_2; \sigma)\, \mathrm{d}t_2 h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1$$

$$= \int_\mu^\infty g(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1$$

$$= \int_\mu^\infty g^+(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1 - \int_\mu^\infty g^-(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1,$$

where

$$g(t_1, \sigma) = E_\sigma[f(t_1, T_2)] \tag{3.4}$$

and $g^+$ and $g^-$ are the positive and negative parts of $g$, respectively, implying that

$$\int_\mu^\infty g^+(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1 = \int_\mu^\infty g^-(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1$$

for all $-\infty < \mu < \infty$, and for a fixed $\sigma > 0$. Consequently,

$$\int_A g^+(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1 = \int_A g^-(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1 \tag{3.5}$$

for any Borel set, $A$. If we let $A = \{t_1\colon g(t_1, \sigma) > 0\}$, then $\int_A g^+(t_1, \sigma) h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1 = 0$ which follows from (3.5). Similarly, if we take $A = \{t_1 : g(t_1, \sigma) < 0\}$, then $\int_A g^-(t_1, \sigma) \times h_1(t_1; \mu, \sigma)\, \mathrm{d}t_1 = 0$. Thus, for any fixed $\sigma$, $g(t_1, \sigma) = 0$ a.e.-$\mathrm{d}t_1$. It then follows that $\int_\mu^\infty g^+(t_1, \sigma)\, \mathrm{d}t_1 = 0$. Since $f, h_1$ are jointly measurable, $g^+$ is jointly measurable, and the Fubini theorem yields

$$0 = \int_0^\infty \int_\mu^\infty g^+(t_1, \sigma)\, \mathrm{d}t_1\, \mathrm{d}\sigma = \int_\mu^\infty \left( \int_0^\infty g^+(t_1, \sigma)\, \mathrm{d}\sigma \right) \mathrm{d}t_1,$$

so that $\int_0^\infty g^+(t_1, \sigma)\, \mathrm{d}\sigma = 0$ a.e.-$\mathrm{d}t_1 I_{\{t_1 > \mu\}}$ for any $\mu$ and hence a.e.-$\mathrm{d}t_1$. In other words, there is a measurable set $B \subset (-\infty, \infty)$ such that $\int_{t_1 \notin B} \mathrm{d}t_1 = 0$ and $\int_0^\infty g^+(t_1, \sigma)\, \mathrm{d}\sigma = 0$ for all $t_1 \in B$. Then for any $t_1 \in B$ we have $g^+(t_1, \sigma) = 0$ a.e.-$\mathrm{d}\sigma$. Using a similar argument, we can show that $g^-(t_1, \sigma) = 0$ a.e.-$\mathrm{d}\sigma$ so that $g(t_1, \sigma) = 0$ a.e.-$\mathrm{d}\sigma$.

Since $f$ is integrable and $h_2(t_2; \sigma) = [(r-2)!]^{-1} \sigma^{-(r-1)} t_2^{r-2} \mathrm{e}^{-t_2/\sigma}$, if $\sigma_n \to \sigma > 0$ then $\sigma_n > \sigma/2$ for all $n$ large, and then a straightforward application of the dominated convergence theorem implies that

$$g(t_1, \sigma_n) = \int_0^\infty f(t_1, t_2) h_2(t_2; \sigma_n)\, \mathrm{d}t_2 \to \int_0^\infty f(t_1, t_2) h_2(t_2; \sigma)\, \mathrm{d}t_2 = g(t_1, \sigma),$$

so that $g(t_1, \sigma)$ is a continuous function of $\sigma$ for any fixed $t_1$. Then for any $t_1 \in B$ we have $g(t_1, \sigma) = 0$ for all $\sigma > 0$. Since the densities $h_2$ constitute an exponential family, $T_2$ is a complete statistic for $\sigma$, and hence it follows from $0 = g(t_1,\sigma) = E_\sigma[f(t_1,T_2)]$ that $f(t_1,t_2) = 0$ a.e.-d$t_2$. In other words, for any $t_1 \in B$ there exists a measurable set $A(t_1) \subset (0,\infty)$ such that $\int_{t_2 \notin A(t_1)} \mathrm{d}t_2 = 0$ and $f(t_1,t_2) = 0$ for all $t_1 \in B, t_2 \in A(t_1)$. Since $t_1 \in B$ and $t_2 \in A(t_1)$ imply $f(t_1,t_2) = 0$, $f(t_1,t_2) \neq 0$ implies that either $t_1 \notin B$ or $t_1 \in B, t_2 \notin A(t_1)$. Hence

$$\begin{aligned}\int\int I_{\{f(t_1,t_2)\neq 0\}} \mathrm{d}t_2\,\mathrm{d}t_1 &\leqslant \int\int (I_{\{t_1\notin B\}} + I_{\{t_1 \in B, t_2 \notin A(t_1)\}})\,\mathrm{d}t_2\,\mathrm{d}t_1\\ &= \int\int I_{\{t_1\notin B\}} \mathrm{d}t_1\,\mathrm{d}t_2 + \int_{t_1\in B}\int I_{\{t_2\notin A(t_1)\}} \mathrm{d}t_2\,\mathrm{d}t_1 = 0+0\end{aligned}$$

so that $f(t_1,t_2) = 0$ a.e.-d$t_1 \times$ d$t_2$. Hence $(T_1,T_2)$ are jointly complete statistics for $(\mu,\sigma)$.

Finally, since $\hat\mu_{B_r}$ and $\hat\sigma_{B_r}$ are unbiased estimators of $\mu$ and $\sigma$ and are functions of the jointly complete sufficient statistics $(T_1,T_2)$, $(\hat\mu_{B_r},\hat\sigma_{B_r})$ is the unique UMVUE for $(\mu,\sigma)$ (see, e.g., Theorem 1.11 of Lehmann and Casella, 1998) based on the $r$ largest order statistics. □

**Remark 3.3.** The fact that Hill's estimator is at once the BLUE, UMVUE, and conditional MLE for $\alpha^{-1}$ argues strongly for its use in the case of data with Pareto tails. However, real data are usually not exactly Pareto, even in the tail, so the question of robustness remains open. For example, although it is common to employ Hill's estimator for financial data which many believe to be at least approximately stable, McCulloch (1997) points out that Hill's estimator performs poorly for such data when the stable index $1.5 < \alpha < 2$ (see also Fofack and Nolan, 1999). Meerschaert and Scheffler (1998) suggest a different estimator based on the asymptotics of the sample variance, which is more robust in this situation. Aban and Meerschaert (2001) develop a shift-invariant version of Hill's estimator, which is another way to increase robustness. Beirlant et al. (1999) use an exponential regression model to obtain tail estimates assuming a second-order regular variation condition, which adjusts for the bias in Hill's estimator in this case. A closely related estimator to this latter estimator was obtained by Feuerverger and Hall (1999) which is also based on a nonlinear regression, under similar model assumptions, to obtain a different bias correction.

## 4. Other least-squares estimators

In Section 2, we showed that Hill's estimator for the tail thickness parameter $\alpha^{-1}$ is actually a linear regression estimator, taking into account the mean and covariance structure of the largest order statistics. Other estimators can be obtained in the same manner, using a different set of assumptions about the data. If we ignore the covariance matrix $\mathbf{B}_r$ of the $r$ largest order statistics and instead minimize $(\mathbf{X}_r - \mathbf{A}_r\boldsymbol{\theta})'(\mathbf{X}_r - \mathbf{A}_r\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, the resulting least-squares estimator for $\boldsymbol{\theta}$ based on the $r$ largest order statistics is $\hat{\boldsymbol{\theta}}_{S_r} = (\mathbf{A}_r'\mathbf{A}_r)^{-1}\mathbf{A}_r'\mathbf{X}_r$ with covariance matrix $\sigma^2(\mathbf{A}_r'\mathbf{A}_r)^{-1}(\mathbf{A}_r'\mathbf{B}_r\mathbf{A}_r)(\mathbf{A}_r'\mathbf{A}_r)^{-1}$. This simplified version of BLUE (SBLUE) is typically used when the covariance matrix of the data is unknown (see, e.g., David, 1981). This method leads to the tail thickness estimator

$$\hat{\alpha}_{S_r}^{-1} = \sum_{i=1}^{r} k_i \ln X_{(n-i+1)}$$

with

$$k_i = \frac{\bar{a}_r(a_{n-i+1} - \bar{a}_r)}{\sum_{i=1}^{r}(a_{n-i+1} - \bar{a}_r)^2}$$

and

$$\bar{a}_r = \frac{1}{r}\sum_{i=1}^{r} a_{n-i+1}.$$

This estimator is still unbiased for $\alpha^{-1}$ but has a larger variance than the BLUE.

Another least-squares estimator for tail thickness is the $qq$-estimator of Kratz and Resnick (1996), or equivalently the exponential tail estimator of Schultze and Steinebach (1996). Given $X_1,\ldots,X_n$ i.i.d. with $P(X > x) = Cx^{-\alpha}$ for $x$ large, the $qq$-estimator is based on the regression equation $\ln X_{(n-i+1)} = \alpha^{-1}\ln C - \alpha^{-1}\ln(i/n)$, $i = 1,\ldots,r$. Let $\mathbf{X}_r' = (\ln X_{(n-r+1)},\ldots,\ln X_{(n)})$ and let $\mathbf{1}_r$ be a $r \times 1$ vector of 1's. If we define $\boldsymbol{\eta}_r' = (\eta_r,\ldots,\eta_1)$ where $\eta_i = \ln(i/n)$, then the design matrix for this least-squares problem is $\mathbf{E}_r = (\mathbf{1}_r, \boldsymbol{\eta}_r)$, and the estimator, $\hat{\boldsymbol{\theta}}_{qq_r} = (\mathbf{E}_r'\mathbf{E}_r)^{-1}\mathbf{E}_r'\mathbf{X}_r$, minimizes $(\mathbf{X}_r - \mathbf{E}_r\boldsymbol{\theta})'(\mathbf{X}_r - \mathbf{E}_r\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Unlike the BLUE and SBLUE, this estimator is biased with mean vector $(\mathbf{E}_r'\mathbf{E}_r)^{-1}\mathbf{E}_r'\mathbf{A}_r\boldsymbol{\theta}$ and covariance matrix $\sigma^2(\mathbf{E}_r'\mathbf{E}_r)^{-1}(\mathbf{E}_r'\mathbf{B}_r\mathbf{E}_r)(\mathbf{E}_r'\mathbf{E}_r)^{-1}$.

One way to sharpen the $qq$-estimator is to take into account the variance–covariance matrix of the log-transformed data. This is similar to the BLUE but with a different design matrix. Minimizing $(\mathbf{X}_r - \mathbf{E}_r\boldsymbol{\theta})'\mathbf{B}_r^{-1}(\mathbf{X}_r - \mathbf{E}_r\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, we obtain the empirical-based least-squares estimator (ELSE), $\hat{\boldsymbol{\theta}}_{E_r} = (\mathbf{E}_r'\mathbf{B}_r^{-1}\mathbf{E}_r)^{-1}\mathbf{E}_r'\mathbf{B}_r^{-1}\mathbf{X}_r$. This estimator is biased with mean $(\mathbf{E}_r'\mathbf{B}_r^{-1}\mathbf{E}_r)^{-1}\mathbf{E}_r'\mathbf{B}_r^{-1}\mathbf{A}_r\boldsymbol{\theta}$ and covariance matrix $\sigma^2(\mathbf{E}_r'\mathbf{B}_r^{-1}\mathbf{E}_r)^{-1}$. Here the linear regression model is $\mathbf{X}_r = \mathbf{E}_r\boldsymbol{\theta} + \mathbf{e}$ where $E\{\mathbf{e}\} = \mathbf{a} \neq 0$. Subtracting $\mathbf{a}$ from both sides gives a generalized linear regression model with mean zero errors, and then the same argument that lead to (2.1) shows that the ELSE has minimum variance among all linear estimators with this bias.

Fig. 2 compares the theoretical variances of all four least-squares estimators of $\alpha^{-1}$ in the case $\alpha = 1$ and $n = 10{,}000$. Typically the variance of the $qq$-estimator and SBLUE, which do not take the covariance structure of the order statistics into account, are about twice as large as BLUE and ELSE. This is in agreement with known results: Hall (1982) showed that Hill's estimator of $\alpha$ is asymptotically normal with variance $\alpha^2/r$, while Csörgő and Viharos (1997) showed that the $qq$-estimator is asymptotically normal with variance $2\alpha^2/r$.

Fig. 3 applies each of these estimators to the data set ($n = 1570$ daily trading volumes) from Fig. 1. Since the estimators with the smaller theoretical variance (BLUE and ELSE) are considerably less smooth, some practitioners may prefer the smoother estimators (SBLUE and $qq$) despite their theoretical shortcomings. Alternatively, one could smooth the BLUE or ELSE estimators (e.g., see Resnick and Stărică (1997) and recall that Hill's estimator for $\alpha^{-1}$ is the BLUE). Hill's estimator $\hat{C}_{\mathrm{H}} = (r/n)X_{(n-r+1)}^{\hat{\alpha}_{\mathrm{H}}}$

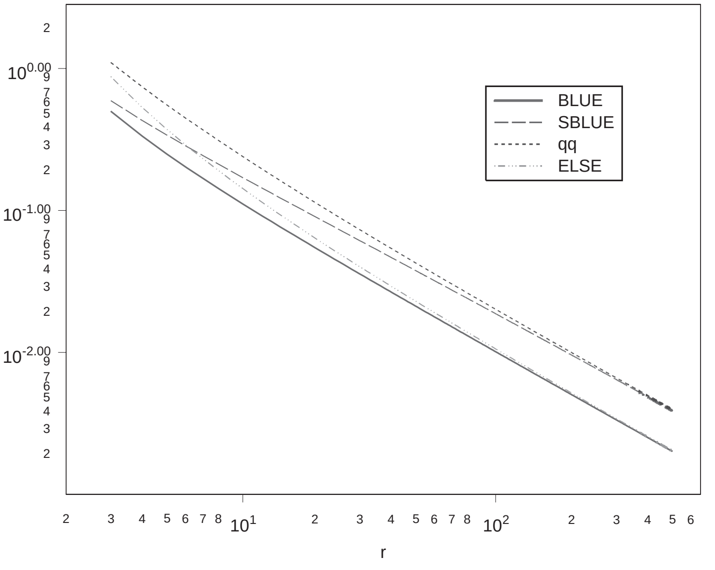


Fig. 2. Theoretical variances for least-squares estimators of $\hat{\alpha}^{-1}$ based on the $r$ largest order statistics, in the case $\alpha = 1$ and $n = 10,000$.

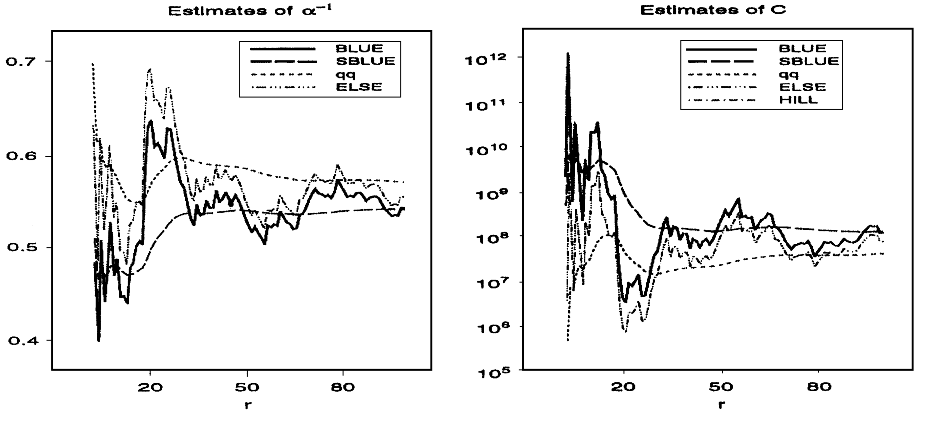


Fig. 3. Estimates of $\alpha^{-1}$ (left) and $C$ (right) with varying values of $r$ for daily trading volume of Amazon, Inc. stock from 20 March 1995 to 18 June 2001. Hill's estimates for $C$ are indistinguishable from BLUE.

differs from the BLUE $\hat{C} = \exp(\hat{\mu}/\hat{\sigma})$ but the difference is too small to see on this graph. The fitted line on Fig. 1 uses the values $\alpha = 1.808$ and $C = 6.973 \times 10^7$ obtained from the BLUE with $r = 75$. Hill's estimator with $r = 75$ yields the same $\alpha$, as expected in view of Corollary 2.6, and $C = 7.022 \times 10^7$, very close to the BLUE.

**Remark 4.1.** Another way to improve the $qq$-estimator is to make a continuity correction. Using the regression equation

$$\ln X_{(n-i+1)} = \alpha^{-1}\ln C - \alpha^{-1}\ln\left(\frac{i-1/2}{n}\right), \quad i = 1,\ldots,r,$$

significantly reduces the bias of the $qq$-estimator, as well as the ELSE.

## References


Aban, I., Meerschaert, M., 2001. Shifted Hill's estimator for heavy tails. Comm. Statist. Simulation Comput. 30, 949–962.

Barlow, R., Proschan, F., 1981. Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Winston, New York.

Beirlant, J., Teugels, J.L., 1989. Asymptotic normality of Hill's estimator. Extreme Value Theory. Lecture Notes in Statistics, Vol. 51, Springer, Berlin, pp. 148–155.

Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression. Extremes 2, 177–200.

Bickel, P., Doksum, K., 1977. Mathematical Statistics. Holden-Day, San Francisco.

Csörgő, S., Mason, D.M., 1985. Central limit theorems for sums of extreme values. Math. Proc. Cambridge Philos. Soc. 98, 547–558.

Csörgő, S., Viharos, L., 1997. Asymptotic normality of least-squares estimators of tail indices. Bernoulli 3, 351–370.

Csörgő, S., Viharos, L., 1998. Estimating the tail index. In: Szyszkowicz, B. (Ed.), Asymptotic Methods in Probability and Statistics: a Volume in Honour of Miklós Csörgő. North-Holland, Amsterdam, pp. 833–881.

David, H., 1981. Order Statistics, 2nd Edition. Wiley, New York.

Dillon, W., Goldstein, M., 1984. Multivariate Analysis. Wiley, New York.

Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. Ann. Statist. 27, 760–781.

Fofack, H., Nolan, J., 1999. Tail behavior, modes and other characteristics of stable distribution. Extremes 2, 39–58.

Haeusler, E., Teugels, J.L., 1985. On the asymptotic normality of Hill's estimator for the index of regular variation. Ann. Statist. 13, 743–756.

Hall, P., 1982. On some simple estimates of an exponent of regular variation. J. Roy. Statist. Soc. B 44, 37–42.

Hall, P., Welsh, A.H., 1985. Adaptive estimators of parameters of regular variation. Ann. Statist. 13, 331–341.

Hill, B., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1173.

Kratz, M., Resnick, S., 1996. The $qq$-estimator and heavy tails. Comm. Statist. Stochastic Models 12, 699–724.

Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation, 2nd Edition. Springer, New York.

Mandelbrot, B., 1963. The variation of certain speculative prices. J. Business 36, 394–419.

Mandelbrot, B., 1982. The Fractal Geometry of Nature. W.H. Freeman, San Francisco.

McCulloch, J., 1997. Measuring tail thickness to estimate the stable index $\alpha$: a critique. J. Business Econom. Statist. 15, 74–81.

Meerschaert, M., Scheffler, H.P., 1998. A simple robust estimator for the thickness of heavy tails. J. Statist. Plann. Inference 71, 19–34.

Resnick, S., Stărică, C., 1997. Smoothing the Hill's estimator. Adv. Appl. Probab. 29, 271–293.

Schultze, J., Steinebach, J., 1996. On least-squares estimation of an exponential tail coefficient. Statist. Decisions 14, 353–372.

Viharos, L., 1999. Weighted least-squares estimators of tail indices. Probab. Math. Statist. 19, 249–265.