COMMUN. STATIST.—SIMULA., 30(4), 949–962 (2001)

SHIFTED HILL'S ESTIMATOR FOR HEAVY TAILS

Inmaculada B. Aban* **and Mark M. Meerschaert**†

Department of Mathematics, University of Nevada,
1664 N. Virginia St., Reno, NV 89557-0045, USA

ABSTRACT

Hill's estimator is a popular method for estimating the thickness of heavy tails. In this paper we modify Hill's estimator to make it shift-invariant as well as scale-invariant. The resulting shifted Hill's estimator is a more robust method of estimating tail thickness.

Key Words: Maximum likelihood estimation; Heavy tails; Order statistics

1. INTRODUCTION

We say that a random variable X has heavy tails if $P(|X| > x) \to 0$ at a polynomial rate. In this case, some of the moments of X will be undefined. Heavy tail random variables are important in applications to finance, electrical engineering, physics and hydrology, see for example (1, 2, 3, 4, 5, 6).

*Corresponding author. E-mail: aban@unr.edu. Partially supported by UNR Junior Faculty Research Grant.
†Partially supported by NSF-EAR grant 9980484.

Hill (7) developed an estimator that is commonly used to determine the rate at which the tails diminish, and this is very useful in practice, see for example (8, 9, 10). Hill's estimator is the conditional maximum likelihood estimator (MLE) for the Pareto distribution $P(X > x) = Cx^{-\alpha}$ conditional on $X \geq d$ for some fixed $d > 0$. This estimator can be applied to a wide variety of distributions, such as the stable and type II extreme value distributions, whose tails are approximately Pareto (11, 12, 13, 14). Hill's estimator is more robust than an MLE based on one of these distributions because it only assumes that the distributional tail has a given form, not the entire distribution. Other tail estimators have also been proposed, for example (15, 16, 17, 18, 19, 20), but Hill's estimator continues to be the most popular in practice.

In this paper we propose a practical modification to Hill's estimator that extends its utility. Hill's estimator is scale-invariant, so that a multiplicative factor in the data will not affect the estimate. But it is not shift-invariant, so that an additive factor will distort the estimates. In this paper we extend Hill's estimator to include information about the shift. Using the same methods as Hill, we compute the conditional MLE for a shifted Pareto $P(X > x) = C(x - s)^{-\alpha}$ which gives estimates of $\alpha > 0$, $C > 0$, and $s \in \mathbb{R}^1$. The resulting estimator is both scale-invariant and shift-invariant. Then we investigate robustness by applying the new estimator to simulated stable data. As observed by McCulloch (21), Resnick (22) and Fofack and Nolan (23), Hill's estimator performs poorly for stable data with $1.5 < \alpha < 2$. The shifted Hill's estimator performs much better, indicating that the problem with Hill's estimator in this case is resolved by accounting for a shift.

2. SHIFTED HILL'S ESTIMATOR

Hill's estimator provides a robust method for measuring the thickness of heavy tails by approximating the distributional tail with a power function. In practice it is often true that $P(X > x) \approx Cx^{-\alpha}$ for $x > 0$ sufficiently large. Then the idea is to estimate the parameters $C > 0$ and $\alpha > 0$ by a conditional maximum likelihood estimate based on the $r + 1$ $(0 \leq r < n)$ largest order statistics, which represent only the portion of the tail for which the power law approximation holds. For the shifted Hill's estimator which is the subject of this paper, we rely on a slightly more general approximation scheme $P(X > x) \approx C(x - s)^{-\alpha}$ where s is an arbitrary shift. In practical applications of the theorem of this section, one should choose r as large as possible, but small enough so that the largest $r + 1$ order statistics lie within the portion of the distributional tail where this approximation is valid.

Our shifted Hill's estimator is the conditional maximum likelihood estimator for the parameters of a shifted Pareto distribution

$$F(x) = 1 - C(x-s)^{-\alpha} \quad x > s + C^{1/\alpha} \tag{1}$$

based on the $r+1$ largest order statistics. Our result extends the usual Hill's estimator, and in fact if $\hat{s} = 0$ we obtain an identical formula. Given a data set $X_1, \ldots, X_n$, we define the order statistics $X_{(1)} \geq \cdots \geq X_{(n)}$ where ties are broken arbitrarily.

Theorem 1. *When $X_{(r)} > X_{(r+1)}$ the conditional maximum likelihood estimator for the parameters in* (1) *based on the $r+1$ largest order statistics is given by*

$$\hat{\alpha} = \left[r^{-1} \sum_{i=1}^{r} \{\ln(X_{(i)} - \hat{s}) - \ln(X_{(r+1)} - \hat{s})\} \right]^{-1} \tag{2}$$

$$\hat{C} = (r/n)(X_{(r+1)} - \hat{s})^{\hat{\alpha}} \tag{3}$$

where the optimal shift $\hat{s}$ satisfies the equation

$$\hat{\alpha}(X_{(r+1)} - \hat{s})^{-1} = (\hat{\alpha}+1)\, r^{-1} \sum_{i=1}^{r} (X_{(i)} - \hat{s})^{-1} \tag{4}$$

for some $\hat{s} < X_{(r+1)}$.

Proof. We adapt the proof of Hill (7). It is convenient to transform the data, taking $Z_{(i)} = (X_{(n-i+1)} - s)^{-1}$ so that $G(z) = P(Z \leq z) = Cz^{\alpha}$ and $Z_{(1)} \geq \cdots \geq Z_{(n)}$. Since $U_{(i)} = G(Z_{(i)})$ are (decreasing) order statistics from a uniform distribution, $E_{(i)} = -\ln U_{(i)}$ are (increasing) order statistics from a unit exponential. Following [24] pp. 20–21, we let $Y_i = (n-i+1)(E_{(i)} - E_{(i-1)})$, and hence, one can easily check $\{Y_i, i = 1, \ldots, n\}$ are independent and identically distributed unit exponential. Define $Y^* = nE_{(n-r+1)} = n(Y_1/n + \cdots + Y_{n-r+1}/r)$ so that $Y_n, \ldots, Y_{n-r+2}, Y^*$ are mutually independent with joint density $\exp(-y_{(n)} - \cdots - y_{(n-r+2)})p(y^*)$, where $p(y^*)$ is the density of Y^*. Since $U_{(n-r+1)}$ has density $K_1 u^{r-1}(1-u)^{n-r}$ it follows that $Y^* = -n \ln U_{(n-r+1)}$ has density $p(y) = K_2 \exp(-y/n)^r (1 - \exp(y/n))^{n-r}$ where $\{K_j, j = 1, 2\}$ are positive constants. Use the fact that $Y_{(i)} = (n-i+1)\ [-\ln G(Z_{(i)}) + \ln G(Z_{(i-1)})] = \alpha(n-i+1)(\ln Z_{(i-1)} - \ln Z_{(i)})$, for $i = 1, \ldots, r-1$, and

$\exp(-Y^*/n) = CZ^{\alpha}_{(n-r+1)}$ to obtain the likelihood

$$K\left(\prod_{i=1}^{r}\frac{\alpha}{z_{(n-i+1)}}\right)\exp\left(-\alpha\sum_{i=1}^{r-1} i\left[\ln z_{(n-i)} - \ln z_{(n-i+1)}\right]\right)$$
$$\times\left(Cz^{\alpha}_{(n-r+1)}\right)^{r}\left(1 - Cz^{\alpha}_{(n-r+1)}\right)^{n-r},$$

conditional on the values of the $r+1$ smallest order statistics, $Z_{(n)} = z_{(n)}, \ldots, Z_{(n-r+1)} = z_{(n-r+1)}$, where $K > 0$ does not depend on the data or the parameters. Next, condition on $Z_{(n-r+1)} \le d < Z_{(n-r)}$, which multiplies the conditional likelihood by a factor of $(1 - Cd^{\alpha})^{n-r}/(1 - Cz^{\alpha}_{(n-r+1)})^{n-r}$. Then simplify the sum in the exponential to obtain

$$K\left(\prod_{i=1}^{r}\frac{\alpha}{z_{(n-i+1)}}\right)\exp\left(-(r-1)\alpha\ln z_{(n-r+1)} + \alpha\sum_{i=1}^{r-1}\ln z_{(n-i+1)}\right)$$
$$\times\left(Cz^{\alpha}_{(n-r+1)}\right)^{r}\left(1 - Cd^{\alpha}\right)^{n-r}$$

and collect terms to get

$$K\alpha^{r}\exp\left((\alpha-1)\sum_{i=1}^{r}\ln z_{(n-i+1)}\right)C^{r}\left(1 - Cd^{\alpha}\right)^{n-r}.$$

Substitute $\beta = Cd^{\alpha}$ to obtain the conditional likelihood function

$$K\alpha^{r}\beta^{r}(1-\beta)^{n-r}\exp\left(-\alpha\, r\ln d + (\alpha-1)\sum_{i=1}^{r}\ln z_{(n-i+1)}\right)$$

which is similar to (2.7) in (7). In terms of the original data, this conditional likelihood becomes

$$L = K\alpha^{r}\beta^{r}(1-\beta)^{n-r}\exp\left(\alpha\, r\ln(D-s) - (\alpha-1)\sum_{i=1}^{r}\ln(X_{(i)} - s)\right)$$
$$\times\prod_{i=1}^{r}(X_{(i)} - s)^{-2}, \qquad (5)$$

where $d = (D-s)^{-1}$ and the product term comes from the change of variable formula for a multivariable probability density. Consequently, the conditional log-likelihood

$$\ln L = K_0 + r\ln\alpha + r\ln\beta + (n-r)\ln(1-\beta) + \alpha r\ln(D-s) - (\alpha+1)$$
$$\times\sum_{i=1}^{r}\ln(X_{(i)} - s),$$

and we seek the global maximum over the parameter space consisting of all (α, β, s) for which $\alpha > 0$, $0 < \beta < 1$, $s < D$ and $s + C^{1/\alpha} = s + \beta^{1/\alpha}(D - s) < X_{(r)}$. At an interior maximum

$$\begin{aligned} 0 &= \frac{\partial \ln L}{\partial \alpha} = \frac{r}{\alpha} + r\ln(D - s) - \sum_{i=1}^{r} \ln(X_{(i)} - s) \\ 0 &= \frac{\partial \ln L}{\partial \beta} = \frac{r}{\beta} - \frac{n - r}{1 - \beta} \\ 0 &= \frac{\partial \ln L}{\partial s} = \frac{-\alpha r}{D - s} + (\alpha + 1) \sum_{i=1}^{r} (X_{(i)} - s)^{-1} \end{aligned} \tag{6}$$

must hold. The solution to (6) depends on the value of D, which will be unknown in practice. Since we assume $X_{(r)} \geq D > X_{(r+1)}$ we could estimate D by $X_{(r)}$ or $X_{(r+1)}$ or some point in between. We use the estimate $D = X_{(r+1)}$ to be consistent with standard usage for the original Hill estimator (8–11, 13, 22, 23). Now the theorem follows easily, using $C = \beta(D - s)^{\alpha}$.

Remarks. The conditional likelihood (5) can also be obtained using the joint density of the order statistics (e.g., see equation (2.2.2) in (24). Note that the formulas for $\hat{\alpha}$ and $\hat{C}$ are the same as for the usual Hill's estimator, once the data is shifted by $\hat{s}$. Therefore any implementation of Hill's estimator may be used once an estimate of the true shift is computed and applied. Although the proof of the original Hill's estimator also requires $X_{(r)} > X_{(r+1)}$, this fact is commonly ignored, and the resulting formulas are used even in the case of a tie. Various refinements of Hill's estimator have been proposed to smooth or sharpen the estimate of α, see for example (11, 22, 25). These refinements can also be applied to the shifted Hill's estimator, and we believe that the resulting improvements would make a worthwhile subject for future research.

3. NUMERICAL IMPLEMENTATION

The shifted Hill's estimator is computed by solving a three dimensional optimization problem. The estimator maximizes the conditional likelihood function (5) over the parameter space consisting of all (α, β, s) for which $\alpha > 0$, $0 < \beta < 1$, $s < D$ and $s + C^{1/\alpha} = s + \beta^{1/\alpha}(D - s) < X_{(r)}$. At an interior critical point, the equations (6) hold. Since $s < D$ it follows that

$$\begin{aligned} s + \beta^{1/\alpha}(D - s) &= s(1 - \beta^{1/\alpha}) + \beta^{1/\alpha} D < D(1 - \beta^{1/\alpha}) + \beta^{1/\alpha} D \\ &= D \leq X_{(r)} \end{aligned}$$

is automatically satisfied, so that it suffices to solve (6) over the space of all (α, β, s) for which $\alpha > 0$, $0 < \beta < 1$, and $s < D$.

Lemma. *The conditional likelihood function L in* (5) *tends to zero as* $\alpha \to 0+$, $\alpha \to \infty$, $\beta \to 0+$, $\beta \to 1-$, $s \to D-$, or $s \to -\infty$.

Proof. Note that $M = D - s > 0$ and $C_i = X_{(i)} - D > 0$ for $i = 1, \ldots, r$. Then

$$\ln L = K_1 + r\ln\alpha + \alpha\sum_{i=1}^{r}\ln\left(\frac{D-s}{X_{(i)}-s}\right) = K_1 + r\ln\alpha + \alpha\sum_{i=1}^{r}\ln\left(\frac{M}{M+C_i}\right)$$

where K_1 does not depend on α. Then it follows easily that $\ln L \to -\infty$ as $\alpha \to 0+$ or as $\alpha \to \infty$. Next write $\ln L = K_2 + r\ln\beta + (n-r)\ln(1-\beta)$ where K_2 does not depend on β. Then it follows that $\ln L \to -\infty$ as $\beta \to 0+$ or $\beta \to 1-$. Finally write

$$\ln L = K_3 + \alpha\sum_{i=1}^{r}\ln\left(\frac{M}{M+C_i}\right) - \sum_{i=1}^{r}\ln(M + C_i)$$

where K_3 does not depend on s. Then $\ln L \to -\infty$ as $s \to D-$ (so that $M \to 0+$) or $s \to -\infty$ (so that $M \to \infty$). This completes the proof.

Substituting $D = X_{(r+1)}$ into (6) leads to the formulas (2) and (3) along with the constraint equation (4) that identifies the optimal shift. Equation (2) gives the optimal $\hat{\alpha}$ as a function of $\hat{s}$. Substituting into (4) yields one equation

$$G(\hat{s}, \hat{\alpha}(\hat{s})) = -r\hat{\alpha}(\hat{s})(X_{(r+1)} - \hat{s})^{-1} + (\hat{\alpha}(\hat{s}) + 1)\sum_{i=1}^{r}(X_{(i)} - \hat{s})^{-1} = 0 \qquad (7)$$

in one variable $\hat{s}$. After solving equation (7) for the optimal shift, we can substitute back into (2) and (3) to obtain $\hat{\alpha}$ and $\hat{C}$. In most cases, equation (7) has a unique solution on the interval $-\infty < \hat{s} < X_{(r+1)}$. Figure 1 displays typical graphs of the constraint function in (7) as a function of $\hat{s}$ for varying values of r based on a sample of size $n = 20,000$ from a symmetric stable distribution with tail parameter $\alpha = 1.8$. When r is very small, (7) may have no solution (see for instance Figure 1 when $r = 5$.) In practical applications, we consider the parameter estimates as functions of the number of order statistics used, so this is not a problem. However, it can be annoying when running repeated simulations. Therefore we have developed a simple test to ensure that the equation (7) has a solution.

Given an ordered sample $X_{(1)} \geq \cdots \geq X_{(n)}$ let $C_i = X_{(i)} - X_{(r+1)}$ denote the exceedences past the $(r+1)$st order statistic for $i = 1, \ldots, r$. The mean and variance of these data are given by

$$\mu_r = r^{-1}\sum_{i=1}^{r} C_i \quad \text{and} \quad \sigma_r^2 = r^{-1}\sum_{i=1}^{r}(C_i - \mu_r)^2. \qquad (8)$$

For heavy tailed data it is commonly observed that $\sigma_r > \mu_r$.

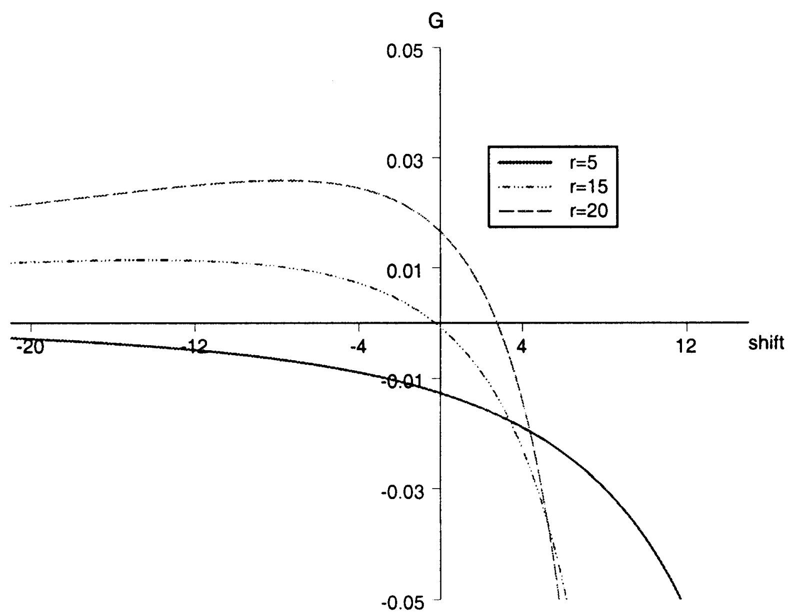

Figure 1. Constraint function G as a function of the shift estimate for varying values of r based on a sample of size $n = 20,000$ from a symmetric stable distribution with tail parameter $\alpha = 1.8$.

Theorem 2. *Suppose that* (8) *holds with* $C_i = X_{(i)} - X_{(r+1)}$ *and that* $\sigma_r > \mu_r$. *Then* (7) *holds for some* $-\infty < \hat{s} < X_{(r+1)}$.

Proof. Write $M = X_{(r+1)} - \hat{s}$ so that $X_{(i)} - \hat{s} = C_i + M$ for $i = 1, \dots, r$. Now we develop a property of our constraint function. Recall that Theorem 1 requires the condition $X_{(r)} > X_{(r+1)}$, just like the original Hill's estimator.

Lemma. *If* $X_{(r)} > X_{(r+1)}$ *then* $G(\hat{s}, \hat{\alpha}(\hat{s})) \to -\infty$ as $\hat{s} \uparrow X_{(r+1)}$.

Proof. As $\hat{s} \uparrow X_{(r+1)}$ we have $M \to 0$. Then

$$\alpha(M) = \hat{\alpha}(\hat{s}) = \left(-\ln M + r^{-1} \sum_{i=1}^{r} \ln(C_i + M)\right)^{-1}$$

where $\ln(C_i + M) \geq \ln(C_r)$ for all $1 \leq i \leq r$ and all $M > 0$. Then $\alpha(M)^{-1} \geq \ln(C_r/M) \to \infty$, so $\alpha(M) \to 0$ as $M \to 0$. Now write $G(\hat{s}, \hat{\alpha}(\hat{s})) = I_1 + I_2$

where

$$I_1 = \frac{-r\alpha(M)}{X_{(r+1)} - \hat{s}} = \frac{-r}{M\alpha(M)^{-1}} \leq \frac{-r}{M\ln(C_r/M)} \to -\infty$$

as $M \to 0$ and

$$I_2 = (\alpha(M) + 1)\sum_{i=1}^{r}\frac{1}{C_i + M} \leq (\alpha(M) + 1)\frac{r}{C_r}$$

so that $I_1 + I_2 \to -\infty$ as $M \to 0$. This completes the proof of the Lemma.

Now in order to show that $G(\hat{s}, \hat{\alpha}(\hat{s})) = 0$ for some $-\infty < \hat{s} < X_{(r+1)}$ it will suffice to show that this function is positive for some $\hat{s}$. Recall that $\mu_r = r^{-1}\sum_{i=1}^{r} C_i$ and let $\mu_r^{(2)} = r^{-1}\sum_{i=1}^{r} C_i^2$. Since $\ln(1 + C_i/M) = C_i/M - 1/2(C_i/M)^2 + O(M^{-3})$ we have

$$\frac{1}{r}\sum_{i=1}^{r}\ln(1 + C_i/M) = \frac{1}{rM}\sum_{i=1}^{r} C_i - \frac{1}{2rM^2}\sum_{i=1}^{r} C_i^2 + O(M^{-3})$$

and so for $\alpha = \alpha(M)$ we have

$$\begin{aligned}\left(\frac{\alpha C_i}{M}\right)^{-1} &= \frac{M}{C_i}\left(\frac{1}{rM}\sum_{i=1}^{r} C_i - \frac{1}{2rM^2}\sum_{i=1}^{r} C_i^2 + O(M^{-3})\right)\\ &= \frac{\mu_r}{C_i} - \frac{\mu_r^{(2)}}{2C_iM} + O(M^{-2})\\ &= \frac{\mu_r}{C_i}\left(1 - \frac{\mu_r^{(2)}}{2M\mu_r} + O(M^{-2})\right).\end{aligned}$$

Since $(1 + K_0/M)^{-1} \sim (1 + K_0/M)$ as $M \to \infty$ for any real constant K_0 it follows that

$$\left(\frac{\alpha C_i}{M}\right) = \frac{C_i}{\mu_r}\left(1 + \frac{\mu_r^{(2)}}{2M\mu_r} + O\big(M^{-2}\big)\right)$$

so that

$$1 - \left(\frac{\alpha C_i}{M}\right) = 1 - \frac{C_i}{\mu_r} - \frac{C_i\mu_r^{(2)}}{2M\mu_r^2} + O(M^{-2})$$

as $M \to \infty$. Now write

$$\begin{aligned} G(\hat{s}, \hat{\alpha}(\hat{s})) &= \sum_{i=1}^{r} \left(\frac{M - \alpha C_i}{M(C_i + M)} \right) \\ &= \sum_{i=1}^{r} \left(\frac{1}{C_i + M} \right) \left(1 - \frac{\alpha C_i}{M} \right) \\ &= \sum_{i=1}^{r} \left(\frac{1}{C_i + M} \right) \left(\frac{\mu_r - C_i}{\mu_r} - \frac{C_i \mu_r^{(2)}}{2M\mu_r^2} + O(M^{-2}) \right) \\ &= K(M) \sum_{i=1}^{r} \left(\mu_r - C_i - \frac{C_i \mu_r^{(2)}}{2M\mu_r} + O(M^{-2}) \right) \prod_{j \neq i} (M + C_j) \end{aligned}$$

where

$$K(M)^{-1} = \mu_r \prod_{j=1}^{r} (M + C_j)$$

and the remaining product is taken over all $1 \le j \le r$ such that $j \neq i$. Expanding this product term shows that the last expression above equals

$$K(M) \left(\sum_{i=1}^{r} (\mu_r - C_i) M^{r-1} + \sum_{i=1}^{r} (\mu_r - C_i) M^{r-2} \sum_{j \neq i} C_j - \frac{\mu_r^{(2)}}{2\mu_r} \times \sum_{i=1}^{r} C_i M^{r-2} + O(M^{r-3}) \right).$$

Since $\sum_{i=1}^{r} (\mu_r - C_i) = 0$ it follows that $G(\hat{s}, \hat{\alpha}(\hat{s})) > 0$ for all M sufficiently large (i.e., for all sufficiently large negative numbers $\hat{s}$) if and only if

$$\sum_{i=1}^{r} (\mu_r - C_i) \sum_{j \neq i} C_j > \frac{\mu_r^{(2)}}{2\mu_r} \sum_{i=1}^{r} C_i = \frac{r}{2} \mu_r^{(2)}.$$

Since

$$\sum_{i=1}^{r} \sum_{j=1}^{r} (\mu_r - C_i)(\mu_r - C_j) = \left(\sum_{i=1}^{r} (\mu_r - C_i) \right)^2 = 0$$

it follows that

$$\sum_{i=1}^{r} (\mu_r - C_i) \sum_{j \neq i} C_j = - \sum_{i=1}^{r} \sum_{j \neq i} (\mu_r - C_i)(\mu_r - C_j) = \sum_{i=1}^{r} (\mu_r - C_i)^2 = r\sigma_r^2$$

so the condition reduces to $2\sigma_r^2 > \mu_r^{(2)}$. Now use the well known formula $\sigma_r^2 = \mu_r^{(2)} - \mu_r^2$ to reduce this to $\sigma_r^2 > \mu_r^2$.

4. ROBUSTNESS

The main argument for the use of Hill's estimator is robustness. Since it only depends on the shape of the probability tails, it can be applied in situations where the form of the distribution is unknown. This is typically the case in applications to finance, where heavy tails are common. Jansen and de Vries (9) and Loretan and Phillips (10) analyze a large number of financial data sets using Hill's estimator. Typically they find values of α around 3.0 on the basis of the largest $r = 50$ order statistics from a sample of size $n = 3,000$. McCulloch (21) criticizes these findings, arguing that Hill's estimator may significantly inflate estimates of α for data which approximately follow a stable distribution. Another method of estimating α for these data sets is to use the stable MLE recently developed by Nolan (26). If probability plots show an adequate fit to a stable distribution, then this method is preferred. If not, then Hill's estimator may still apply because it only assumes that the distributional tail has a given form, not the entire distribution.

The use of stable distributions for financial data is motivated by the extended central limit theorem. Sums of independent and identically distributed random variables can only converge to a stable distribution, see for example Feller (2). Hence if price changes result from the additive effect of many small independent random shocks, their distribution must be approximately stable. In the special case where the constituent shocks have a finite variance, we obtain a normal limit, which is a special case of a stable limit. A random variable X is said to have a nondegenerate stable distribution if its characteristic function is given by

$$\mathrm{E}\{\exp i\theta X\} = \begin{cases} \exp\{-\sigma^\alpha|\theta|^\alpha(1 - i\beta(\operatorname{sign}\theta)\tan(\pi\alpha/2)) + i\mu\theta\} & \text{for } \alpha \neq 1, \\ \exp\{-\sigma|\theta|(1 + i\beta(2/\pi))(\operatorname{sign}\theta)\ln\theta + i\mu\theta\} & \text{for } \alpha = 1, \end{cases}$$

where the tail index $0 < \alpha \le 2$, the scale $\sigma > 0$, center $-\infty < \mu < \infty$, and skewness $-1 \le \beta \le 1$. When $\alpha = 2$, we obtain the special case of a normal distribution. When $\alpha < 2$ the stable distribution has heavy tails. In this case, $P(X > x) \sim Cx^{-\alpha}$ as $x \to \infty$ where

$$C = \left(\frac{1+\beta}{2}\right)\frac{\sigma^\alpha(1-\alpha)}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}, \tag{9}$$

see for example Samorodnitsky and Taqqu (14). Hence a stable distribution is approximately Pareto for large $x > 0$. See Fofack and Nolan (23) for a more detailed discussion on fitting a Pareto to the tail of a stable distribution.

We investigated the observed sampling distribution of the shifted Hill's estimator when applied to stable data. We generated $m = 1,000$ random samples of size $n = 20,000$ from a stable distribution with $\alpha = 1.8$, $\sigma = 1$, $\mu = 0$, and $\beta = 0$, using the stable random variable generator from the IMSL Fortran subroutine library (27). Then we computed the shifted Hill's estimator from these data as a function of the tail sample size r. For purposes of comparison we also computed the usual Hill's estimator. In view of formula (2) we chose to examine $\hat{\alpha}^{-1}$ rather than $\hat{\alpha}$. Figure 2 and Table 1 summarize the results of our simulation. The observed sampling distributions are approximately normal with a standard deviation roughly proportional to $r^{-1/2}$. The original Hill's estimator for α^{-1} is significantly below the true value of $\alpha^{-1} = 0.55\bar{5}$, thus overestimating α, while the shifted Hill's estimator is centered near $\alpha^{-1} = 0.55\bar{5}$. The spread of the simulated sampling distribution for the shifted Hill's estimator is quite a bit larger than the original Hill's estimator, but this is the price of robustness. We also note that this stable distribution is symmetric, so that its mean, median and mode are all equal to zero. Fofack and Nolan (23) suggest centering the data to its sample median or mode before using Hill's estimator. Our data are already centered, and the shifted Hill's estimator still performs better than the original. In this case, shifting to the median is far from optimal.

Summary statistics for the shift and dispersion are given in Table 2. The observed sampling distributions are significantly skewed with numerous outliers. Both the optimal shift and dispersion vary significantly depending on the number r of upper order statistics used. This is because the best fitting shifted Pareto density depends on how much of the stable density tail we are trying to fit. It is also interesting to note that the optimal shift s cannot be estimated independent of the optimal dispersion C. Simulation results for the

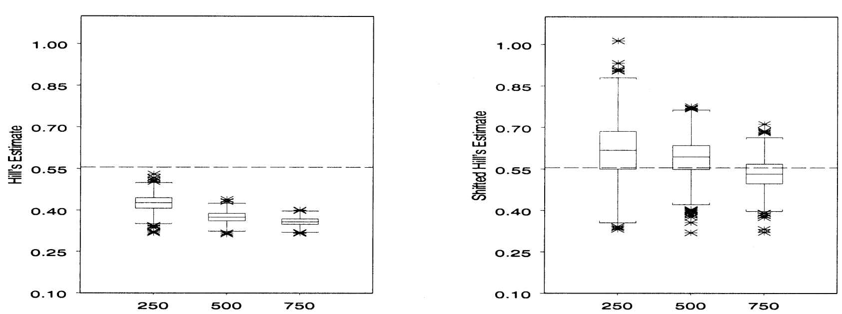

Figure 2. Observed sampling distribution of $\hat{\alpha}^{-1}$ based on $m = 1,000$ simulated data sets of size $n = 20,000$ from a stable distribution with $\sigma = 1$, $\mu = 0$, $\beta = 0$ and $\alpha = 1.8$ ($\alpha^{-1} = 0.55\bar{5}$ denoted by the broken horizontal line).

Table 1. Summary Statistics for $\hat{\alpha}^{-1}$ Based on $m = 1{,}000$ Data Sets of Size $n = 20{,}000$ from a Stable Distribution with $\sigma = 1$, $\mu = 0$, $\beta = 0$ and $\alpha = 1.8$ ($\alpha^{-1} = 0.55\bar{5}$)

	$r = 250$		$r = 500$		$r = 750$	
Statistic	Hill's	Shifted Hill's	Hill's	Shifted Hill's	Hill's	Shifted Hill's
Minimum	0.317	0.332	0.313	0.320	0.316	0.321
Lower quartile	0.407	0.550	0.361	0.549	0.349	0.498
Median	0.427	0.618	0.374	0.595	0.358	0.533
Upper quartile	0.445	0.686	0.388	0.636	0.368	0.568
Maximum	0.530	1.013	0.439	0.776	0.400	0.712
Mean	0.426	0.619	0.375	0.592	0.358	0.532
Std dev	0.030	0.100	0.019	0.068	0.015	0.055

Table 2. Summary Statistics for $\hat{s}$ and $\hat{C}$ Based on $m = 1{,}000$ Simulated Data Sets of Size $n = 20{,}000$ from a Stable Distribution with $\alpha = 1.8$, $\sigma = 1$, $\mu = 0$ and $\beta = 0$

	$\hat{s}$			$\hat{C}$		
Statistic	$r = 250$	$r = 500$	$r = 750$	$r = 250$	$r = 500$	$r = 750$
Minimum	−1.152	−0.078	−0.078	0.012	0.024	0.037
Lower quartile	1.310	1.402	1.057	0.032	0.044	0.069
Median	1.728	1.576	1.193	0.046	0.054	0.085
Upper quartile	2.052	1.734	1.359	0.073	0.069	0.109
Maximum	2.912	2.154	1.776	1.430	0.990	0.976
Median abs dev	0.362	0.164	0.148	0.017	0.012	0.018

stable distribution considered with $r = 500$ (the upper 2.5% of the data) suggest an optimal shift of around $s = 1.5$ and an optimal dispersion of around $C = 0.05$.

We also studied the behavior of the shifted Hill's estimator for stable data with other values of $0 < \alpha \leq 2$ and $-1 \leq \beta \leq 1$. For smaller values of α, the original Hill's estimator performs better than the shifted Hill's estimator, while for $1.5 < \alpha < 2$ the shifted Hill's estimator performs better. We lose resolution in both estimators as β decreases, since our estimates are based on the right tail, but if we apply both estimators to absolute values of the observations, then both estimators are insensitive to the skewness parameter β. The other two parameters of a stable distribution are the scale σ and center μ. The shifted Hill's estimator is both scale-invariant and shift-invariant, so it is unaffected by changes in these parameters.

The original Hill's estimator is scale-invariant, but of course the shift will affect this estimator, which is the entire point of this paper.

5. CONCLUSIONS

In this paper we compute a shifted Hill's estimator as the conditional maximum likelihood estimator based on the upper order statistics of a shifted Pareto distribution. Accounting for the shift is important because Hill's estimator is not shift-invariant. The resulting shifted Hill's estimator has a wider sampling distribution than the original Hill's estimator, but it is more robust. McCulloch (21) reports that the usual Hill's estimator greatly overestimates the tail parameter α for simulated stable data with $1.5 < \alpha < 2$. The shifted Hill's estimator does a much better job of estimating α, showing that the problem with Hill's estimator in this case is nothing more than a failure to account for a shift. Stable or nearly stable random variables are commonly observed in applications to finance. Hill's estimator is often used to provide a robust estimator of the tail parameter α for these data sets. We recommend computing the optimal shift for Hill's estimator in all such applications.

REFERENCES

1. Anderson, P.; Meerschaert, M. Modeling river flows with heavy tails. Water Resources Research, **1998**, *34*, 2271–2280.
2. Feller, W. *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd Ed., Wiley, New York, 1971.
3. Janicki, A.; Weron, A. *Simulation and Chaotic Behavior of α Stable Stochastic Processes*, Marcel Dekker, New York, 1994.
4. Mandelbrot, B. *The Fractal Geometry of Nature*, W.H. Freeman, San Francisco, 1982.
5. Mittnik, S.; Rachev, S. *Asset and Option Pricing with Alternative Stable Models*, Wiley, New York, 1998.
6. Nikias, C.; Shao, M. *Signal Processing with Alpha Stable Distributions and Applications*, Wiley, New York, 1995.
7. Hill, B. A simple general approach to inference about the tail of a distribution. Ann. Statist., **1975**, *3* (5), 1163–1173.
8. Anderson, P.; Meerschaert, M. Periodic moving averages of random variables with regularly varying tails. Ann. Statist., **1997**, *25*, 771–785.
9. Jansen, D.; de Vries, C. On the frequency of large stock market returns: Putting booms and busts into perspective. Review of Econ. and Statist., **1991**, *73*, 18–24.

10. Loretan, M.; Phillips, P. Testing the covariance stationarity of heavy tailed time series. J. Empirical Finance, **1994**, *1*, 211–248.
11. Hall, P. On some simple estimates of an exponent of regular variation. J. Royal Statist. Soc. B, **1982**, *44*, 37–42.
12. Leadbetter, M.; Lindgren, G.; Rootzén, H. *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York, 1980.
13. Resnick, S.; Stărică, C. Consistency of Hill's estimator for dependent data. J. Appl. Probab., **1995**, *32*, 139–167.
14. Samorodnitsky, G.; Taqqu, M. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, London, 1994.
15. Dekkers, A.; Einmahl, J.; de Haan, L. A moment estimator for the index of an extreme value distribution. Ann. Statist., **1989**, *17*, 1833–1855.
16. Drees, H. Refined Pickands estimators of the extreme value index. Ann. Statist., **1995**, *23*, 2059–2080.
17. Falk, M. On testing the extreme value index via the POT-method. Ann. Statist., **1995**, *23*, 2013–2035.
18. Kratz, M.; Resnick, S. The qq estimator and heavy tails. Stochastic Models, **1996**, *12*, 699–724.
19. Meerschaert, M.; Scheffler, H. P. A simple robust estimator for the thickness of heavy tails. J. Statist. Plann. Inf., **1998**, *71*, 19–34.
20. Wei, X. Asymptotically efficient estimation of the index of regular variation. Ann. Statist., **1995**, *23*, 2036–2058.
21. McCulloch, J. Measuring tail thickness to estimate the stable index α: A critique. J. Business Econ. Statist., **1997**, *15*, 74–81.
22. Resnick, S. Heavy tail modeling and teletraffic data. Ann. Statist., **1997**, *25*, 1805–1869.
23. Fofack, H.; Nolan, J. P. Tail Behavior, Modes and Other Characteristics of Stable Distribution. Extremes, **1999**, *2* (1), 39–58.
24. David, H. *Order Statistics*, 2nd Ed., Wiley, New York, 1981.
25. Resnick, S.; Stărică, C. Smoothing the Hill's estimator. Adv. Appl. Probab., **1997**, *29*, 271–293.
26. Nolan, J. Maximum likelihood estimation and diagnostics for stable distributions, **1998**, preprint.
27. IMSL Math/Stat Library. Visual Numerics, Inc., Houston, 1994.

Received April 2000
Revised April 2001