The Annals of Statistics,
1983, Vol. 11, No. 4, 1019–1031

SPECIAL INVITED PAPER

ESTIMATING THE STABLE INDEX α IN ORDER TO MEASURE TAIL THICKNESS: A CRITIQUE[1]

BY WILLIAM H. DUMOUCHEL

Massachusetts Institute of Technology

Stable laws are often fit to outlier-prone data and, if the index α is estimated to be much less than two, then the normal law is rejected in favor of an infinite-variance stable law. This paper derives the theoretical properties of such a procedure. When the true distribution is stable, the distribution of the m.l.e. of α is non-regular if $\alpha = 2$. When the true distribution is not stable, the estimate of α is not a robust measure of the rate of decrease of the tail probabilities. A more robust procedure is developed, and a statistic for describing and comparing the tail-shapes of arbitrary samples is proposed.

1. Introduction and example. The independent identically distributed variables $X_1, \cdots, X_n$ are said to have a stable distribution with index α if $X_1 + \cdots + X_n$ has the same distribution as $\delta_n + n^{1/\alpha}X_1$, $0 < \alpha \le 2$. If $\alpha = 2$ the distribution of X is normal and, of course, has moments of all orders. If $\alpha < 2$, the distribution is called *Stable Paretian*, since the tail probabilites are approximately like those of a Pareto distribution, $P(X > x) \sim kx^{-\alpha}$, $x \to \infty$, and only moments of order less than α exist. See Feller (1966) and DuMouchel (1973b) for more of the mathematical properties of stable laws.

Since the work of Mendlebrot (1960, 1963, 1969) the use of stable distributions to model data suspected to have heavy tailed distributions, especially certain economic variables such as stock price changes, has become popular. Some recent references are Press (1975), Samuelson (1976), Koutrouvelis (1980), McCullough (1978), Feuerverger and McDonough (1981), and Paulson et al. (1981).

Besides the stability property, stable laws are attractive because only stable laws have domains of attraction. Thus, if an observed quantity can be thought of as the sum or result of very many independent identically distributed effects, then it may have a stable distribution. Also, the family of stable laws is reasonably flexible, since, besides α, parameters regulating location, scale, and skewness are available.

There are practical difficulties with stable laws, primarily stemming from the fact that no closed form expression exists for most stable density functions. Tables of the stable distributions and/or percentiles can be found in Fama and Roll (1968), DuMouchel (1971), Holt and Crow (1973), and Brothers et al (1982). Also, except for the asymptotic theory of the maximum likelihood estimate, developed by DuMouchel (1973b), there is not much theory of how to estimate the parameters.

Other heavy tailed distributions like the Student's t, log normal, or Pareto families have also been used to model data with high kurtosis. See, e.g. Blattberg and Gonides (1974). But this paper focuses on the use of the stable model as proposed by Mandelbrot, Fama (1965), Roll (1968) and others. Within this model, the choice between $\alpha = 2$ and

Received October 1981; revised June 1983.

[1] Prepared under National Science Foundation Grant MCS82-01732. Reproduction in whole or in part is permitted for any purpose of the United States Government.

AMS 1980 *subject classifications.* 62F25, 62G99.

Key words and phrases. Stable distributions, infinite variance, Pareto distributions, super efficient estimation.

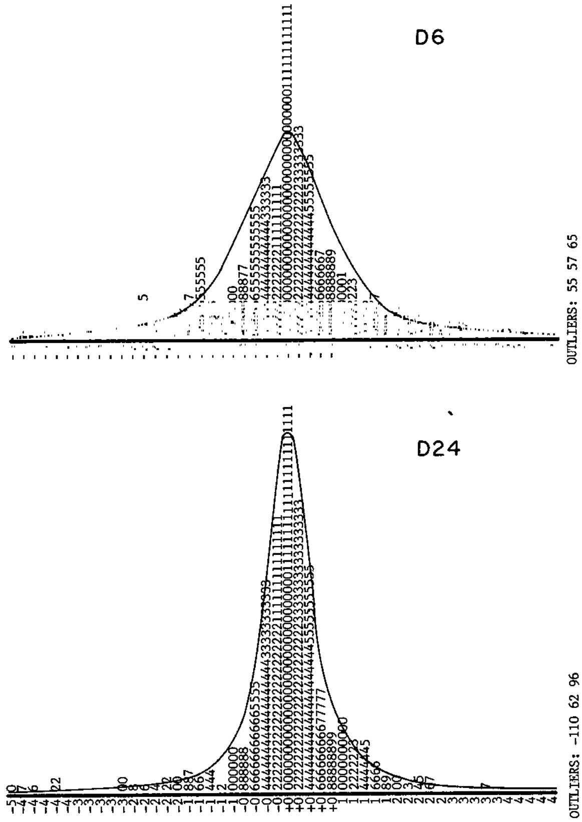

FIG. 1. *Stem and leaf plots for 304 weekly differences in treasury bill interest rates, March 1959 through December 1964. The labels D6 and D24 refer to differences for bills with 6 and 24 weeks to maturity, respectively. The units are 1/100 of one percent. Thus 45 is a .45% increase in interest rates. For each sample, three observations were too extreme to be plotted: they are listed next to the plot. Superimposed over each stem and leaf plot is the symmetric stable density estimated by maximum likelihood.*

$\alpha < 2$ is sometimes referred to as a test for infinite variance. If a stable distribution with $\alpha < 2$ seems to fit the data well, then the property $P(X > x) \sim kx^{-\alpha}$ is used to estimate the probability of extreme deviations, while the estimated value of α may be used to compare the tail behavior of this data to that of other data or other distributions.

As an example, Figure 1 displays the stem and leaf plots of two samples. The values of D6 are 304 weekly changes in interest rates for U.S. Treasury Bills with six weeks to maturity, covering the period from March 1959 through December 1964. The variable D24 is the same quantity for bills with 24 weeks to maturity. Roll (1968) found no significant serial association in these series. The sample coefficients of kurtosis are 3.1 for D6 and 23.2 for D24.

Superimposed over the histograms in Figure 1 are the maximum likelihood fits of symmetric stable densities. The estimates of α are $\hat{\alpha} = 1.37 \pm .11$ and $\hat{\alpha} = 1.23 \pm .08$ for D6 and D24 respectively. The preceding standard errors are based on the curvature of the log-likelihood function. For more details on the computation of the maximum likelihood estimate, see DuMouchel (1971). A Pearson Chi squared goodness of fit test based on 23 intervals of D6 and 19 intervals of D24 yields $\chi^2 = 36$ ($P = .013$, 19 df) and $\chi^2 = 20$ ($P = .2$, 15 df) respectively. Thus a stable law fits D24 quite well, but not D6. If $\hat{\alpha} + 3(\text{SE}) < 2$ is interpreted as "evidence of infinite variance" then both series show such evidence.

The remainder of this article derives the theoretical properties of such a standard of evidence and also provides insight into the problem of estimating the tail behavior of a

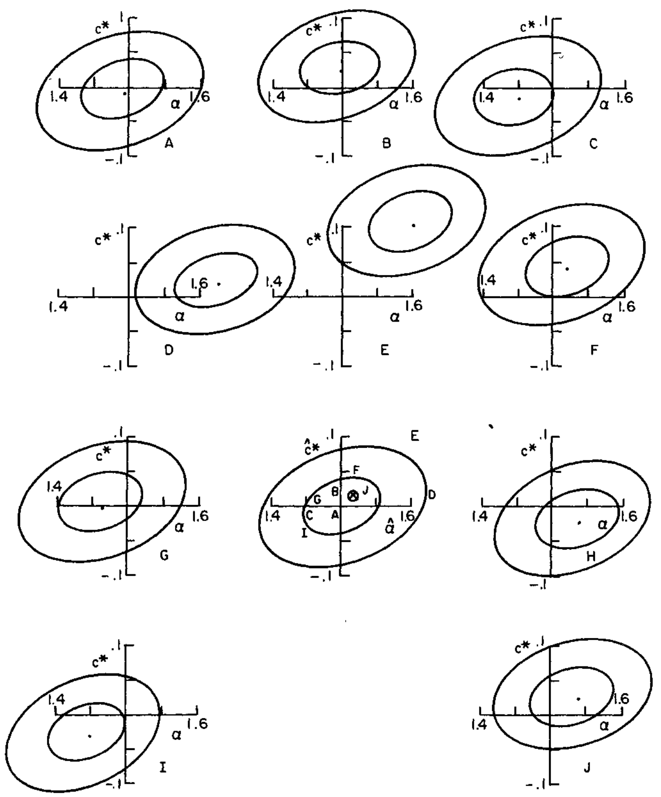

FIG. 2. *Sample 50% and 5% contours for the 10 simulated likelihood function (graphs A through J), and the 50% and 5% concentration ellipses (center graph, with the values of $(\hat{\alpha}, \hat{c}^*)$ and their mean indicated) for the sampling situation having $\alpha_0 = 1.5$, $n = 1000$.*

distribution, with special emphasis on the use of stable models. Section 2 shows that when the true distribution is normal, the asymptotic distribution of $\hat{\alpha}$ is nonregular, and if the true distribution is stable with index α less than but near two, the moderate sample distribution of $\hat{\alpha}$ is not well approximated by its asymptotically normal limit. Section 3 shows that the use of $\hat{\alpha}$ to measure tail behavior is not robust to the assumption of stability. Section 4 suggests an alternative statistic for describing and comparing the tails of distributions, and presents an alternative analysis of the treasury bill data. Section 5 contains a concluding discussion.

2. Inferences about α when α is near or equal to 2. This section assumes that a stable law model is correct and discusses the problem of estimating the index α when α equals 2 (i.e., the data are normally distributed) or α is slightly less than 2 (in which case the model has infinite variance, but the density function is quite similar to the normal density everywhere but in the extreme tails). The principal result is that nonregularities in the asymptotic theory make it easier to distinguish $\alpha = 2$ from $\alpha < 2$.

Whenever $\alpha < 2$, the maximum likelihood estimate will have an asymptotically normal distribution with mean α and variance determined by the Fisher information. See DuMouchel (1973b, 1975) for details. However, for fixed sample size, the accuracy of the normal approximation becomes worse as $\alpha \to 2$, and the standard asymptotic theory fails at $\alpha = 2$.

Figures 2 and 3 show the results of a simulation experiment in which 10 samples of size 1000 were drawn from each of 2 symmetric stable distributions. Figure 2 shows the contours of the resulting likelihood functions with respect to the index parameter α and

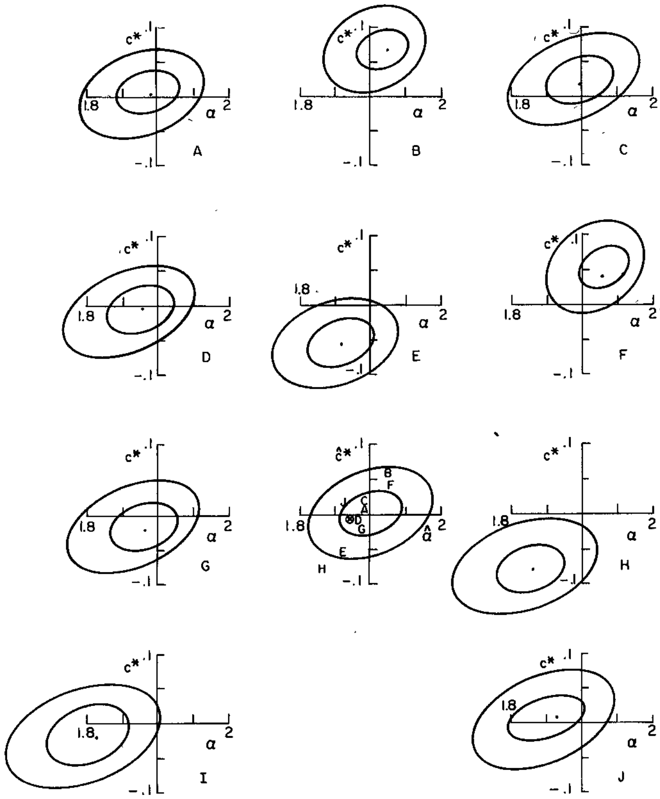

FIG. 3. *Sample 50% and 5% contours for the 10 simulated likelihood functions (graphs* A *through* J*), and the 50% and 5% concentration ellipses (center graph, with the values of* $(\hat{\alpha}, \hat{c}^*)$ *and their mean indicated) for the sampling situation having* $\alpha_0 = 1.9$, $n = 1000$.

scale parameter c (actually $c^* = \log c$ is presented) when the true value of α is 1.5. The same information when $\alpha = 1.9$ is displayed in Figure 3. (The location parameter δ was assumed known and equal to 0 in these simulations. When the distribution is known to be symmetric, $\hat{\delta}$ is asymptotically independent of $(\hat{\alpha}, \hat{c})$ and simple robust estimates of location will perform about as well as the m.l.e. See DuMouchel (1975) for details.) The center graph in each figure depicts the 10 maximum likelihood estimates with elliptical contours which include 50% and 95% of the asymptotic distribution of the m.l.e.'s. Note that Figure 2, depicting the $\alpha = 1.5$ case, shows smoother, more symmetrical contours than does Figure 3. The latter figure shows only 2 of the 10 values of $\hat{\alpha}$ greater than $\alpha = 1.9$ and these two simulations show likelihood functions especially skewed away from the boundary $\alpha = 2$ of the parameter space.

The mean and standard deviations of the 10 values of $\hat{\alpha}$ are 1.876 and .037 respectively, so that a test of $E(\hat{\alpha}) = 1.9$ would yield $t = 2.05$ with nine degrees of freedom, (2 sided $P = .07$) so that there is some evidence that $\hat{\alpha}$ is biased and the likelihood functions are far from normally shaped, even when the sample size is as large as 1000. See DuMouchel (1971) for more details of the generation of the stable variates and computation of the maximum likelihood estimates.

Next consider the situation when the true distribution is normal ($\alpha = 2$). Since observable asymmetry is obvious ground for rejecting normality, we consider only symmetric stable alternatives, and will also assume the center is known to be 0 (see previous parenthetical discussion). Thus we consider the hypothesis $H_0: \alpha = 2$ versus $H_1: \alpha < 2$ with c as a nuisance parameter.

TABLE 1
Proportion of times $\hat{\alpha} < 2$ when the true value of $\alpha = 2$ and the sample size is n.

# of samples	n	Proportion ± Standard error	
		known scale	unknown scale
500	100	.18 ± .02	.16 ± .02
534	1000	.16 ± .02	.13 ± .02
205	10000	.08 ± .02	.08 ± .02

Note: Each sample consisted of n pseudo-random normally distributed variates using the IMSL function ggnorm.

The limiting distribution of $\hat{\alpha}$ under H_0 is not regular. If it were, we would expect that $\hat{\alpha} = 2$ about 50% of the time, and that conditionally on $\hat{\alpha} < 2$, $\hat{\alpha}$ would have an approximate half-normal distribution when the sample size n is large. Instead, it is shown in the Appendix that $P(\hat{\alpha} = 2) \to 1$ as $n \to \infty$. Thus the fixed test, "Reject H_0 if $\hat{\alpha} < 2$, otherwise accept," will have error probabilities of both types approaching 0 as $n \to \infty$.

Although the appendix merely shows that if $\alpha = 2$, as $n \to \infty$, $P(\hat{\alpha} < 2) \to 0$, we conjecture further that, for some $K > 0$,

$$P(\hat{\alpha} < 2) \sim K/\log n.$$

As a check, simulations were performed and are reported in Table 1. Random samples of size 100, 1000, and 10000 were drawn from a normal distribution and the proportion of times $\hat{\alpha} < 2$ was .19, .16, and .08 respectively, when the scale parameter c was assumed known. When c was estimated also, the corresponding proportions were .16, .13, and .08.

Thus we see that $P(\hat{\alpha} < 2)$ does decrease as n becomes large. If the conjecture that $P(\hat{\alpha} < 2) \sim K/\log n$ is true, these simulations suggest the simple rule of thumb $P(\hat{\alpha} < 2) \approx (3 \log_{10} n)^{-1}$. This rule requires a few million observations before the critical region "reject if $\hat{\alpha} < 2$" acquires the coveted .05 level of significance.

No theoretical results are available for the distribution of $\hat{\alpha}$ conditional on $\hat{\alpha} < 2$. No obvious pattern showed up in the simulations. The first percentiles of $\hat{\alpha}$ (fifth lowest values of $\hat{\alpha}$ out of about 500 simulations) were 1.83 and 1.97 for $n = 100$ and 1000 respectively, suggesting that perhaps the scale of $2 - \hat{\alpha}$ decreases faster than rate $n^{-1/2}$.

3. Robustness to the assumption of stability. This section shows that inferences about tail behavior based on estimating the stable index are not robust. The lack of robustness is demonstrated by constructing a family of distributions which are relatively hard to distinguish from stable laws, but which give rise to misleading estimates of tail behavior when stable laws are fit to them.

The alternative densities have Pareto tails but match the normal density in the center. They are denoted $f_\gamma(x)$, for $\gamma \geq 0$:

$$f_\gamma(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \text{for} \quad |x| < 1$$

$$= \frac{\Phi(-1)}{\sigma} \left[\frac{1 + \gamma(|x| - 1)}{\sigma}\right]^{-\gamma^{-1}-1} \qquad \text{for} \quad |x| \geq 1.$$

(When $\gamma = 0$, the second factor is taken as the limit, $\exp-(|X| - 1)/\sigma$. This parameterization allows exponential tail behavior to be a limiting case of Pareto tail behavior. See the next section. Note also that for $\gamma > .5$, the tail probabilities of f_γ are asymptotically proportional to those of s_α, if $\alpha = 1/\gamma$.)

The scale parameter in the Pareto piece is chosen to make $f_\gamma(x)$ continuous, so that $\sigma = \Phi(-1)\sqrt{2\pi e}$, where $\Phi(-1) = .1587 \cdots$ is a normal tail probability. It is not claimed that

TABLE 2
Estimates and goodness-of-fit measures when $n = 1000$ points which fit the distribution f_γ exactly are fitted to a stable law.

Pareto γ	$\frac{1}{\gamma}$	Fitted $\alpha \pm$ SE	Noncentrality	P	Fitted c	Pareto $x_{.995}$	Stable $x_{.995}$	Pareto $x_{.999}$	Stable $x_{.999}$
0	∞	1.86 ± .04	5.6	.11	.711	3.3	3.4	4.3	7.0
.2	5	1.71 ± .05	4.6	.15	.694	4.3	4.9	6.8	12.0
.5	2	1.54 ± .05	1.9	.29	.674	6.1	7.3	16.2	20.2
.67	1.5	1.46 ± .05	2.9	.23	.667	9.9	9.4	28.9	28.1
1.0	1.0	1.32 ± .04	11.1	.02	.656	21.2	12.8	104.4	42.8

any particular data are generated exactly according to the densities f_γ, only that the a priori plausibility of the densities f_γ is not much lower than that of the stable densities s_α, for most data sets. (Fisk (1961) fitted U.S. income data to a model with a Pareto upper tail and a log-logistic distribution for the rest.) If data which follow the distribution f_γ are instead fit to the family s_α, what estimate $\hat{\alpha}$ will result, and how good will the fit look?

Table 2 answers these questions for the five values $\gamma = 0, .2, .5, .67, 1$, and sample size $N = 1000$. For each value of γ, the stable parameters (α, c) for which $(1/c)s_\alpha(x/c)$ is "closest" to f_γ is found by maximizing the quantity

$$L(\alpha, c) = \int_0^\infty f_\gamma(x)\log\left[\frac{1}{c}\, s_\alpha\left(\frac{x}{c}\right)\right] dx$$

with respect to α and c. (Actually, a finite discrete sum was used to approximate $L(\alpha, c)$ in the computations.) These values of (α, c) will be the limit of the maximum likelihood estimates of α and c if a stable distribution is fit to large samples taken from f_γ. The fitted value of α is shown in column 3 of Table 2, plus or minus its supposed standard error if the sample size were $N = 1000$. The standard errors are computed as described in DuMouchel (1975).

Looking at row one of Table 2, if $\gamma = 0$ and the data have exponential tails, the estimated α will tend to 1.86 and, if $N = 1000$, the theoretical standard error of .04 will lead to a firm conclusion that $\alpha < 2$.

In order to see how well the data from f_γ can be expected to fit a stable law, the asymptotic behavior of a Pearson Chi squared goodness of fit test is computed. The positive line $X > 0$ was broken somewhat arbitrarily into 13 intervals, and the approximate non-centrality parameter

$$\lambda = 2000 \textstyle\sum_{i=1}^{13} (F_i - S_i)^2/S_i$$

is computed, where F_i and S_i are the probabilities that f_γ and s_α assign to the ith interval, respectively. If a sample of size 1000 is drawn from f_γ, the Pearson Chi squared statistic would have approximately a noncentral Chi squared distribution with 10 degrees of freedom and noncentrality λ. (This is not quite true, since the result of Chernoff and Lehmann (1952) states that the asymptotic distribution is not quite that of Chi squared if the parameter estimates are based on ungrouped data. But the true situation should be close enough for the present purpose.)

When the true model has exponential tails, the non-centrality is 5.6, so that the expected value of the Chi squared test statistic is about 15.6, to be compared to the distribution with ten degrees of freedom. Since the non-centrality will be proportional to the sample size, if $N = 200$, the expected value of the goodness-of-fit statistic is just 10 + .2(5.6) = 11.1. Clearly the power of the goodness of fit test would be very small. Even for the larger sample size, the probability of rejecting the stable assumption is not large. Column five shows $P(\chi^2_{10} > 10 + \text{non-centrality})$, or the attained level of significance of the goodness-of-fit test if the test statistic equaled its expected value. For the row corresponding to $\gamma = 0$ it is .11, not even significant at the 10% level for $N = 1000$. I think most statisticians would be satisfied with the fit.

Looking at the other rows of Table 2, when γ is .2 or .5, a similar phenomenon occurs. The corresponding true distributions are in the domain of attraction of the normal distribution, but the estimated α's are 1.71 ± .05 and 1.54 ± .05 respectively, and the goodness-of-fit based on 1000 observations is not likely to reject. When $\gamma = .67$ the "true" $\hat{\alpha} = 1/\gamma$ is 1.5 and the estimated is 1.46 ± .05, so relatively little bias occurs. When $\gamma = 1$ the bias occurs in the other direction and $\hat{\alpha}$ is 1.32 ± .04, although in this case, columns 4 and 5 show that f_1 and $s_{1.32}$ are different enough that the test of fit should catch it when $N = 1000$.

The remaining columns of Table 3 provide more comparisons of f_γ and its "best fitting" stable distribution $c^{-1}s_\alpha(x/c)$. Column 6 is the best fitting scale parameter, c, and the last 4 columns compare the 99.5 and 99.9 percentiles of the two distributions. In all five cases, the error is greater when estimating the more extreme percentile. If a sample of size 1000 is available, then it would seem that the 99.9th percentile is a logical choice for defining the "extreme" tails, since the largest observation should be about that large.

Of course, it is always possible to develop more sophisticated tests of fit to the stable model, especially for specific alternatives like the densities f_γ, but these calculations show that the stable maximum likelihood estimates of tail behavior are not robust to the stability assumption, even when alternatives which are relatively hard to distinguish from stable laws are considered. (Paulson et al. (1981) perform a direct test for the stability property. After estimating the index α based on all n observations, they then reestimate α based on $n/2$ sums of 2 observations, $n/4$ sums of 4 observations, $n/8$ sums of 8 observations, etc., and see whether this sequence of estimates of α show any upward trend, as would be expected if the parent distribution had finite variance.)

4. Letting the tails speak for themselves. The rationales for using stable laws are usually that the stability property is useful and that the generalized central limit theorem makes a stable law plausible. However, DuMouchel (1973a) showed that the latter argument is not very persuasive, especially when $1.5 < \alpha < 2$, because of the slow convergence of convolutions to their limiting stable law. Blattberg and Gonedes (1974) reported success in fitting t-distributions to stock price data.

A more natural way of modeling the tail behavior of data is to let the tails "speak for themselves" by basing the inferences on the extreme observations without making any assumptions about the center of the distribution. This paper does not attempt a thorough analysis of how to estimate the tails of a distribution. DuMouchel and Olshen (1975), Hill (1975), Weissman (1979), Breiman et al. (1978, 1981) have discussed the problem of estimating tail behavior. A common family of distributions used in such cases is the Pareto family, $P(X > x) = x^{-\alpha}$ for $x > 1$, $\alpha > 0$.

This section suggests fitting the largest and/or the smallest ten percent of the sample by a generalization of the Pareto distribution. The generalized Pareto family to be proposed here has attractive theoretical and practical properties, and leads naturally to a descriptive tail-behavior statistic for comparison to other samples or other distributions. Finally, application of this more robust method to the Treasury Bill data will lead to conclusions different from those based on the stable model.

DEFINITION. The random variable Z is said to have the generalized Pareto distribution with parameters γ and σ [$Z \sim \text{GP}(\gamma, \sigma)$] if $P(Z > z) = (1 + \gamma z/\sigma)^{-1/\gamma}$, $-\infty < \gamma < \infty$, $0 < \sigma < \infty$, $z > 0$, $\gamma z > -\sigma$. (The distributions with $\gamma = 0$ are defined to be the exponential distributions with mean σ.)

This family of distributions has several nice properties. All manner of tail behaviors are represented. When $\gamma > 0$, $P(Z > z) \sim kz^{-1/\gamma}$, so that $1/\gamma$ is seen to be comparable to the α of the Pareto and Stable Paretian families. At $\gamma = 0$ is the exponential distribution.

TABLE 3
Tail-shape parameters, γ, describing the upper ten percent of several common distributions.

Distribution	γ
Uniform	−1.000
Triangular	−.5000
Normal[a]	−.151
Exponential	0.000
Student (5 DF)[b]	.099
Log Normal	.259
Student (2 DF)	.452
Cauchy	.988

[a] Although the normal distribution has infinite range, the choice of the 99.995 percentile as the largest percentile used in fitting γ is not crucial. If 10000 z's are chosen so that

$$\Phi(x_{.9} + z_k) = .900005 + .00001\,k, \quad k = 0, 1, \cdots, 9999;$$

then the fitted γ becomes −.145, rather than the value −.151 shown.
[b] The value $\gamma = .099$ differs from the value $\mathrm{DF}^{-1} = .2$ that would describe the very extreme tail of this Student distribution. Some such difference should be expected, since the Student distributions approach normality as $\mathrm{DF}^{-1} \to 0$, while the GP distribution with $\gamma = 0$ is exponential.

Anscombe (1961) mentions that the distributions with $\gamma > 0$ can be generated by gamma mixtures of exponential distributions. When $\gamma < 0$ the distribution has shorter tails than the exponential, in fact, even a finite range, since $0 < z < -\sigma/\gamma$. However, if γ is near 0, say $\gamma = -.1$ and $\sigma = .1$, the density is $(1 - z)^9$, which has a very smooth contact with the z-axis and so can serve well as a model for a short-tailed distribution with possibly infinite range. When $\gamma = -1$, the uniform distribution over $(0, \sigma)$ results, while $\gamma = -.5$ corresponds to a triangular distribution.

The distributions with $\gamma > -.5$ obey the regular large sample theory, (as can be shown using the results of Hall, 1982) with an easily evaluated information matrix. Thus, the asymptotic distribution of the m.l.e. based on a sample of size n has mean (γ, σ) and covariance matrix

$$n^{-1}\begin{pmatrix} (1+\gamma)^2 & \sigma(1+\gamma) \\ \sigma(1+\gamma) & 2\sigma^2(1+\gamma) \end{pmatrix}.$$

To use this generalized Pareto family to estimate the tail behavior from sample observations on a variable X, it would be necessary to choose a value x_0, say, and then let $Z = X - x_0$ for all $X > x_0$. (Or let $Z = x_0 - X$ for $X < x_0$ to estimate the left tail.) The choice of x_0 is somewhat arbitrary. Since all the GP(γ, σ) densities are convex, one suggestion is to look at the sample histogram and choose x_0 to be near the point of inflection of each tail. If the GP model fits for $X > x_0$, then choosing a cutoff point $x_1 > x_0$, larger than necessary, results in a GP distribution with the value of γ unchanged, but with σ replaced by $\sigma + \gamma(x_1 - x_0)$.

To be less subjective, we propose choosing x_0 to be the 90th (or 10th) percentile of the sample. The maximum likelihood estimate of γ applied to the resulting sample of Z's is defined as the descriptive measure of tailshape. By convention, if the upper tail of a *continuous* distribution F is to be described, first the values of $z_0, z_1, \cdots, z_{999}$ are defined by

$$F(x_{.9} + z_k) = .90005 + .0001k, \quad k = 0, 1, \cdots, 999;$$

where $F(x_{.9}) = .9$. Then the quantity

$$1000 \log \sigma + (1 + \gamma^{-1}) \sum_0^{999} \log(1 + \gamma z_k/\sigma)$$

is minimized with respect to (γ, σ); the resulting value of γ being the descriptive tail-shape parameter. Table 3 shows the γ's which describe several common distributions.

The choice of the upper ten percent, rather than some other fraction of the sample, is a compromise between the practical need for enough observations to reliably estimate γ and the theoretical desire to describe the behavior of $F(x)$ as $x \to \infty$. If the tail of F is actually in the generalized Pareto family, the standard error of $\hat{\gamma}$ is approximately $(1 + \gamma)/\sqrt{m}$, where m is the number of observations used to estimate γ. This leads to an approximate standard error of .3 if the upper ten percent of 1000 exponential variates are used. Referring to Table 3, this sample size cannot reliably distinguish the tails of the normal, exponential, and log-normal distributions. (The log normal distribution used in Table 3 is such that log X has variance 1.) Using an even smaller fraction of the observations would restrict profitable use of the statistic to much larger sample sizes. On the other hand, to use more than the upper one-tenth of a sample would seem to allow too much dependence on the central part of the distribution.

Breiman et al. (1979, 1981) also propose a measure of tail heaviness of a continuous distribution F. It is

$$H(p) = l''(x_{1-p})/[l'(x_{1-p})]^2,$$

where $l(x) = -\log(1 - F(x))$ and x_p satisfies $F(x_p) = p$. They show that $H(p)$ has desirable theoretical properties but they do not discuss its estimation from samples.

Returning now to the Treasury Bills data, Table 4 shows the results of fitting the GP model to the samples of size 60 formed by subtracting the 30 smallest observations from the 31st smallest, and by subtracting the 31st largest from the largest 30, then merging the two groups.

Reference to the estimate of γ and their approximate standard errors in Table 4 shows that the D6 series is not at all consistent with an infinite variance model, since $\hat{\gamma} = -.081$ indicates tail behavior lighter than the exponential distribution. For the D24 series $\hat{\gamma} = .227$ and $\hat{\gamma}/\text{SE} = 1.4$, so that even here an exponential tail model cannot be ruled out with much confidence.

On the other hand this analysis does rule out the tail behaviors which the stable model predicted. Table 4 provides one-sided 95% lower confidence limits for γ^{-1} and for comparison, the 95% upper confidence limits for the stable index α, from the analysis of Section one. If the stable model holds for these data, the estimates of γ^{-1} should be near those of α. On the contrary, the lower limit for γ^{-1} is greater than the upper limit for α in

TABLE 4

Maximum likelihood estimates of the generalized Pareto Laws for the variables D6 *and* D24, *applied to the outer* 20% *of each data set. The confidence regions for* γ^{-1} *assume a normal distribution for* $\hat{\gamma}$, *while the confidence regions for* α *are based on the stable model analysis of Section* 1.

Statistics	**6 week bills**	**24 week bills**
no. of cases used	60	60
$\hat{\gamma}$(SE)	−.081(.119)	.227(.158)
$\hat{\sigma}$(SE)	12.7(2.2)	11.1(2.2)
95% lower confidence bound for γ^{-1}	$\gamma^{-1} > 8.73$	$\gamma^{-1} > 2.05$
95% upper confidence bound for α	$\alpha < 1.55$	$\alpha < 1.36$

both data sets. We conclude that these data are much less outlier-prone than a stable law analysis would lead us to believe.

5. Discussion. The good news on estimating the tail behavior of possibly long tailed distributions is that, if a stable law is assumed, the problem of distinguishing $\alpha = 2$ from $\alpha < 2$ is relatively easy. The Fisher information per observation is large, and the results of Section 2 show that the asymptotic behavior of a test of $\alpha = 2$ is non-regular in a way that favors making a correct decision. The bad news is that this very feature of the stable family, namely that the distributions having $\alpha < 2$ are so different from the normal distribution, leads to a biased measure of tail behavior if the true distribution is not stable. For example, a distribution with exponential tails could easily be diagnosed as having infinite variance.

To let the tails speak for themselves we suggest fitting a generalized Pareto model to the data outside the 10th and 90th percentiles. (It is interesting to note that when the simple Pareto model, $P(X > x \mid X > x_0) = (x/x_0)^{-\alpha}$, was fit to the Treasury Bill data, the same biases occurred as in the stable law analyses.)

Actually describing the tail behavior of data, rather than merely screening for outliers, is becoming more frequent, especially in the analysis of large data sets. As techniques for doing so become more standardized and sophisticated, their use may become almost as common as those for describing location and scale. The computer programs of Hoaglin and Peters (1979) are one example of this trend. A short computer program to find the maximum likelihood estimates of (γ, σ) for a sample from $\mathrm{GP}(\gamma, \sigma)$, written in the language APL, is available from the author.

Although the advantages of robustness are obvious, the proposed GP procedure is very inefficient if the true distribution is in fact stable. As an example, suppose it is symmetric stable with $\alpha = 1.5$. Then calculations presented in DuMouchel (1975) show that $V(\hat{\alpha}) \sim (1.54)^2/n$. On the other hand, if the outer $n/5$ observations are fit to the Pareto family, $V(\hat{\gamma}) \sim (1 + \gamma)^2/(n/5)$, and the asymptotic variance of $1/\hat{\gamma}$ is $\gamma^{-4}V(\hat{\gamma}) = 5(1 + \gamma)^2/n\gamma^4$ which equals $(8.39)^2/n$ if $\gamma = 1/\alpha = 2/3$. Thus the efficiency of the robust procedure is only $(1.54/8.39)^2 = .034$, a large price to pay for robustness.

Whether the price is worth paying depends on the investigator's opinions as to the plausibility of the stable hypothesis and the possible costs of being wrong. Alternative strategies which fit t-distributions or other families of long tailed distributions to all of the data are beside the point. Presumably, any method which uses the central observations to help make inferences about tail behavior is somewhat nonrobust.

APPENDIX

Proof that $P(\hat{\alpha} < 2) \to 0$. Denote by $s_\alpha(x)$ the symmetric stable density, with scale parameter $c = 1$, and by $i(\alpha)$ the corresponding Fisher information: $i(\alpha) = \int [g_\alpha(x)]^2 s_\alpha(x)\, dx$, where

$$g_\alpha(x) = \dot{s}_\alpha(x)/s_\alpha(x)$$

and

$$\dot{s}_\alpha(x) = \partial s_\alpha(x)/\partial\alpha.$$

PROPOSITION 1. *$i(\alpha) \to \infty$ as $\alpha \to 2$.*

PROOF. Bergstrom's (1952) expansion represents $s_\alpha(x) = k(\alpha)x^{-\alpha-1} + O(x^{-2\alpha-1})$ as $x \to \infty$, and DuMouchel (1975) shows that $\dot{s}_\alpha(x) = \dot{k}(\alpha)x^{-\alpha-1} - k(\alpha)\log(x)x^{-\alpha-1} + O((\log x)^2 x^{-2\alpha-1})$. Strictly speaking, both expansions hold at $\alpha = 2$, although since $k(\alpha) = \pi^{-1}\Gamma(1 + \alpha)\sin \pi\alpha/2 \to 0$ as $\alpha \to 2$, this is trivially true for the series for $s_2(x)$, while $\dot{s}_2(x) \sim -x^{-3}$ as $x \to \infty$. Of course, $s_2(x)$ is the normal density with variance 2, so that $g_2(x) \sim -x^{-3}e^{x^2/4}\sqrt{4\pi}$ and $E(\mid g_2(X) \mid^p) = \infty$ for $p > 1$ when $\alpha = 2$.

PROPOSITION 2. (The statement and proof of Proposition 2 are due to M. Woodroofe.) *If* $X_1, \cdots, X_n$ *is a sample from the distribution with* $\alpha = 2$, *then* $\bar{g}_2 = \bar{g}_2^{(n)} = n^{-1} \sum g_2(X_k)$ *is in the domain of attraction of the stable distribution with* $\alpha = 1$, $\beta = -1$. *That is, there exists constants* a_n *and* b_n *such that* $P(a_n \bar{g}_2^{(n)} - nb_n < y) \to S_{1,-1}(y)$ *where* $S_{\alpha,\beta}(y)$ *is the distribution function of the stable law with parameters* (α, β). *The constants* a_n *and* b_n *are*:

$$a_n = (\log n)^2, \quad b_n = 2 \int_0^\infty \sin\left(\frac{a_n}{n} g_2(x)\right) s_2(x)\, dx. \tag{1}$$

OUTLINE OF PROOF. The statistic $\bar{g}_2$ is the average of n i.i.d. variables $g_2(X_k)$, where X_k is normal and $g_2(x)$ is an even function bounded from above and asymptotic to $-|x|^{-3}\exp(x^2/4)/\sqrt{4\pi}$ as $|x| \to \infty$. Therefore,

$$\begin{aligned} P(g_2(X) < -y) &\sim P(|X|^{-3}\exp(X^2/4) > \sqrt{4\pi}y) \\ &\sim P(|X| > \sqrt{4 \log y + 6 \log\log y + 12 \log 2 + 2 \log(4\pi)}). \end{aligned}$$

Since $P(|X| > R) \sim 4s_2(R)/R$,

$$\begin{aligned} P(|X| > \sqrt{(\cdot)}) &\sim \frac{4}{\sqrt{(\cdot)}} s_2(\sqrt{(\cdot)}) \\ &\sim \frac{4}{\sqrt{4 \log y}\sqrt{4\pi}} \exp\left(-\left(\log y + \frac{3}{2} \log\log y + 3 \log 2 + \frac{1}{2} \log 4\pi\right)\right) \\ &\sim \frac{1}{16\pi} \frac{(\log y)^{-2}}{y}. \end{aligned}$$

Therefore, $g_2(X)$ is in the domain of attraction of a stable law with $\alpha = 1$ since $P(g_2(X) < -y) \sim y^{-1}l(y)$ where $l(y)$ is slowly varying. Since g_2 is bounded above, only its lower tail is long, so that there exist constants a_n, b_n such that $P(a_n\bar{g}_2 - nb_n < y) \to S_{1-1}(y)$ as desired. The forms of the constants given by equation (1) follows from results given in Chapter XVII, Section 5 of Feller (1966), especially formula (5.16). (The quantity a_n of Feller is here denoted $n/a_n = n/(\log n)^2$).

PROPOSITION 3. *For* $\varepsilon > 0$ *sufficiently small but independent of n, the probability that* $(\partial/\partial\alpha)\bar{g} < 0$ *for all* $2 - \varepsilon < \alpha < 2$ *approaches* 1 *as* $n \to \infty$.

PROOF. (The proof of Proposition 3 is due to M. Woodroofe, 1973). This result is not surprising, since the result that $i(2) = \infty$ implies that $E((\partial/\partial\alpha)\bar{g}_\alpha) \to -\infty$ as $\alpha \to 2$. The details of the proof are omitted here.

Since the consistency of $\hat{\alpha}$ implies $P(\hat{\alpha} < 2 - \varepsilon) \to 0$ the event $\bar{g}_2 < 0$ is asymptotically equivalent to $\hat{\alpha} < 2$. Thus, as $n \to \infty$ $P(\hat{\alpha} < 2) - S_{1,-1}(-nb_n) \to 0$.

The final result, that $P(\alpha < 2) \to 0$, follows from

PROPOSITION 4. *The quantity* $nb_n \to \infty$ *as* $n \to \infty$, *so that* $S_{1,-1}(-nb_n) \to 0$.

The proof is tedious and will be omitted. It uses the facts that

$$\int_0^\infty g_2(x)s_2(x)\, dx = 0, \quad \text{that} \quad g_2(x)s_2(x) \sim -x^{-3}, \quad \text{and that} \quad |R - \sin R| < R^3.$$

Details are available from the author.

This establishes the result that $P(\hat{\alpha} < 2) \to 0$ when there is no nuisance parameter.

The same result also holds if the scale parameter c is estimated by maximum likelihood. The proof, which is omitted for the sake of brevity, involves showing that $n^{-1}\sum_{i=1}^{n} g_2(x_i/c)$ and $n^{-1}\sum_{i=1}^{n} g_2(x_i/\hat{c})$ have the same limiting distribution.

6. Acknowledgments. I would like to thank Professor Michael Woodroofe for the benefit of discussions on the topics of Section 2 and for his contribution of Propositions 2 and 3 in the Appendix. Thanks also to professor Richard Roll for supplying the Treasury Bill data used in the examples, to Professors Joseph Gastwirth and Peter Hall for their comments on earlier drafts of this paper, and to the editor and the referees for very helpful comments on matters of exposition.

REFERENCES

ANSCOMBE, F. (1961). Estimating a mixed-exponential response law. *J. Amer. Statist. Assoc.* **56** 493–502.

BERGSTRÖM, H. (1952). On some expansions of stable distributions. *Arkiv. fur Matik II* 375–378.

BLATTBERG, R. C. and GONEDES, N. J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *J. Bus. Univ. Chicago* **47** 244–280.

BREIMAN, L., STONE, C. J. and GINS, J. D. (1979). New methods for estimating tail probabilities and extreme value distributions. Tech. Report #TSC-PD-A266-1, Technology Service Corp. Santa Monica, California 90405.

BREIMAN, L., STONE, C. J., AND GINS, J. D. (1981). Futher development of new methods for estimating tail probabilities and extreme value distributions. Tech. Report #TSD-PD-A243-1, Technology Service Corp. Santa Monica, California 90405.

BROTHERS, K. M., DUMOUCHEL, W. H., and PAULSON, A. S. (1981). Fractiles of stable laws, School of Management, Technical Report, Rensselaer Polytechnic Institute, Troy, New York 12181.

CHERNOFF, H. and LEHMANN, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25** 579–586.

DUMOUCHEL, W. (1971). Stable distributions in statistical inference. Ph.D. thesis, Yale Univ. (Available from Univ. Microfilms, Ann Arbor, Michigan).

DUMOUCHEL, W. (1973a). Stable distributions in statistical inference: 1. Symmetric stable distribution compared to other symmetric long-tailed distributions. *J. Amer. Statist. Assoc.* **68** 469–477.

DUMOUCHEL, W. (1973b). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Ann. Statist.* **3** 948–957.

DUMOUCHEL, W. (1975). Stable distributions in statistical inference: 2. Information from stably distributed samples. *J. Amer. Statist. Assoc.* **70** 386–393.

DUMOUCHEL, W. and OLSHEN, R. A. (1975). On the distributions of claim costs. In *Credibility: Theory and Applications* (ed. P. M. Kahn). Academic, New York, 23–46.

FAMA, E. (1965). The behavior of stock market prices. *J. Bus. Univ. Chicago* **38** 34–105.

FAMA, E. and ROLL, R. (1968). Some properties of symmetric stable distributions. *J. Amer. Statist. Assoc.* **63** 817–836.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications, Vol.* 2. Wiley, New York.

FEUERVERGER, A. and MCDUNNOUGH, P. (1981). On some Fourier methods for inference. *J. Amer. Statist. Assoc.* **76** 379–387.

FISK, P. R. (1961). Estimation of location and scale parameters in a truncated grouped sech square distribution. *J. Amer. Statist. Assoc.* **56** 692–702.

HALL, PETER (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568.

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

HOAGLIN, D. and PETERS, S. (1979). Software for exploring distribution shape. Unpublished manuscript #P-909, Laboratory for Information and Decision Systems, M.I.T.

HOLT, D. R. and CROW, E. L. (1973). Tables and graphs of the stable probability density functions. *J. Research of National Bureau of Standards—B. Math. Sci.* **77B** 143–198.

KOUTROUVELIS, I. A. (1980). Regression-type estimation of the parameters of stable laws. *J. Amer. Statist. Assoc.* **75** 918–928.

MANDELBROT, B. (1960). The Pareto-Levy law and the distribution of income. *Internat. Econ. Review* **1** 79–106.

MANDELBROT, B. (1963). The variation of certain speculative prices. *J. Bus. Univ. Chicago* **26** 394–419.

MANDELBROT, B. (1969). Long-run linearity, locally Gaussian processes, H-spectra, and infinite variances. *Internat. Econ. Review* **10** 82–111.

MCCULLOUGH, H. (1978). Continuous time processes with stable increments. *J. Bus. Univ. Chicago* **51** 601–619.

PAULSON, A. S., RUTHERFORD, D. C. and ULSETH, G. W. (1981). Some applications of stable laws to collective risk theory. School of Management, Technical Report, Rensselaer Polytechnic Institute, Troy, New York 12181.

PRESS, S. J. (1975). Stable distributions: probability, inference, and applications in finance—a survey and a review of recent results. In *Statistical Distributions in Scientific Work* **1** 87–102. G. P. Patel et al. (Eds.). Reidel, Dordrecht, Holland.

ROLL, R. (1968). The efficient market model applied to U.S. treasury bill rates. Doctoral dissertation, Graduate School of Business, University of Chicago.

SAMUELSON, P. A. (1976). Limited liability, shortselling, bounded utility, and infinite-variance stable distributions. *J. Financial Quant. Anal.* **11** 485–503.

WEISSMAN, I. (1978). Estimation of parameters and quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73** 812–815.

WOODROOFE, M. (1973). Personal communication.

STATISTICS CENTER
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 MASSACHUSETTS AVENUE
ROOM E40-111
CAMBRIDGE, MASSACHUSETTS 02139