Maximum likelihood estimation and diagnostics for stable distributions

John P. Nolan

ABSTRACT A program for maximum likelihood estimation of general stable parameters is described. The Fisher information matrix is computed, making large sample estimation of stable parameters a practical tool. In addition, diagnostics are developed for assessing the stability of a data set. Applications to simulated data, stock price data, foreign exchange rate data, radar data and ocean wave energy are presented.

1 Introduction

Stable distributions are a rich class of distributions that include the Gaussian and Cauchy distributions in a family that allows skewness and heavy tails. The class was characterized by Paul Lévy (1924) in his study of normalized sums of i.i.d. terms. The general stable distribution is described by four parameters: an index of stability $\alpha \in (0, 2]$, a skewness parameter β, a scale parameter γ and a location parameter δ. There are multiple parameterizations for stable laws and much confusion has been caused by these different parameterizations. The lack of closed formulas for densities and distribution functions for all but a few stable distributions (Gaussian, Cauchy and Lévy) has been a major drawback to the use of stable distributions by practitioners. This paper shows that the computational problems have now been resolved and it is feasible to fit stable models to data and to use diagnostics to assess the goodness of fit.

Stable distributions have been proposed as a model for many types of physical and economic systems. There are several reasons for using a stable distribution to describe a system. The first is where there are solid theoretical reasons for expecting a non-Gaussian stable model, e.g. reflection off a rotating mirror yielding a Cauchy distribution, hitting times for a Brownian motion yielding a Lévy distribution, the gravitational field of stars yielding the Holtsmark distribution; see Feller (1971) for these and other examples. The second reason is the Generalized Central Limit Theorem which states that the only possible non-trivial limit of normalized sums of i.i.d. terms is stable. It has been argued that many observed quantities are the sum of many small terms - the price of a stock, the noise in a communication system, etc. and hence a stable model should be used to describe such systems. The third argument for modeling with stable distributions is empirical: many large data sets exhibit heavy tails and skewness. The strong empirical evidence for these features combined with the Generalized Central Limit Theorem is used by many to justify the use of stable models. Examples in finance and economics are given in Mandelbrot (1963), Fama (1965), Embrechts, Klüppelberg, and Mikosch (1997), Cheng and Rachev (1995), McCulloch (1996); in telecommunication systems by Stuck and Kleiner (1974), Zolotarev (1986), Willinger, Taqqu, Sherman and Wilson (1995), and Nikias and Shao (1995). Such data sets are poorly described by a Gaussian model, but possibly can be described by a stable distribution. Several recent monographs focus on stable models: Zolotarev (1986), Christoph and Wolf (1993), Samorodnitsky and Taqqu (1994), Janicki and Weron (1994), and Nikias and Shao (1995). The related topic of modeling with heavy tailed distributions is discussed in the books by Embrechts, Klüppelberg and Mikosch (1997) and Adler, Feldman and Taqqu (1998).

Skeptics of stable models recoil from the implicit assumption of infinite variance in the non-Gaussian stable model and have proposed other models for observed heavy tailed and skewed data sets, e.g. mixture models, time varying variances, etc. Such models can have very heavy tails, see §8.4 of Embrechts, Klüppelberg and Mikosch for a discussion of the heavy tailed behavior of ARCH and GARCH models with normal innovations. The same people who argue that the population is inherently bounded and therefore must have a finite variance, routinely use the normal distribution - with unbounded support - as a model for this same population. The variance is but one measure of spread for a distribution, and it is not appropriate for all problems. From an applied point of view, what we generally care about is capturing the shape of a distribution.

We propose that the practitioner approaches this dispute as an agnostic. The fact is that until now we have not really been able to compare data sets to a proposed stable model. In this paper we show that maximum likelihood estimation of all four stable parameters is feasible, even for large data sets. And perhaps just as important, it is now feasible to use diagnostics to assess whether a stable model accurately describes the data. In some cases there are solid theoretical reasons for believing that a stable model is appropriate; in other cases we will be pragmatic: if a stable distribution describes the data accurately and parsimoniously with four parameters, then we accept it as a model for the observed data.

This paper is organized in the following way. The remainder of this section describes some parameterizations for stable distributions and some basic properties, then we discuss previous work on methods of estimating stable parameters. Section 2 describes our program to do maximum likelihood (ML) estimation for all four stable parameters. In addition, the Fisher information matrix is computed for a grid of parameter values, so large sample confidence interval estimates for the parameters can be made. Section 3 discusses diagnostics for assessing whether a data set is stable or not. The program STABLE will perform the estimation and diagnostics and is available on the Web at `http://www.cas.american.edu/~jpnolan`, and clicking on the link to stable distributions. Examples of stable ML estimation for several data sets are given in Section 4. Finally, we give a discussion of our results in Section 5.

1.1 Parameterizations and basic properties

There are at least half a dozen different parameterizations of stable distributions. All involve different specifications of the characteristic function and are useful for various technical reasons. The parameterization most often used now, e.g. Samorodnitsky and Taqqu (1994), is the following: $X \sim S(\alpha, \beta, \gamma, \delta_1; 1)$ if the characteristic function of X is given by

$$E \exp(itX) = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left[1 - i\beta(\tan\frac{\pi\alpha}{2})(\text{sign } t)\right] + i\delta_1 t\right) & \alpha \neq 1 \\ \exp\left(-\gamma|t| \left[1 + i\beta\frac{2}{\pi}(\text{sign } t)\ln|t|\right] + i\delta_1 t\right) & \alpha = 1. \end{cases}$$

The range of parameters are $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, scale $\gamma > 0$, and location $\delta_1 \in R$. (We prefer not to use σ for the scale parameter, since variances do not exist unless $\alpha = 2$, and even when $\alpha = 2$, the standard stable scale parameter is not the standard deviation. Likewise, we prefer not to use μ for the location parameter, because means do not always exist and even when they do, the location parameter and the mean differ in some parameterizations.)

A more useful parameterization in appplications is a variation of the (M) parameterization of Zolotarev: we will say $X \sim S(\alpha, \beta, \gamma, \delta_0; 0)$ if the characteristic function of X is given by

$$E \exp(itX) =$$

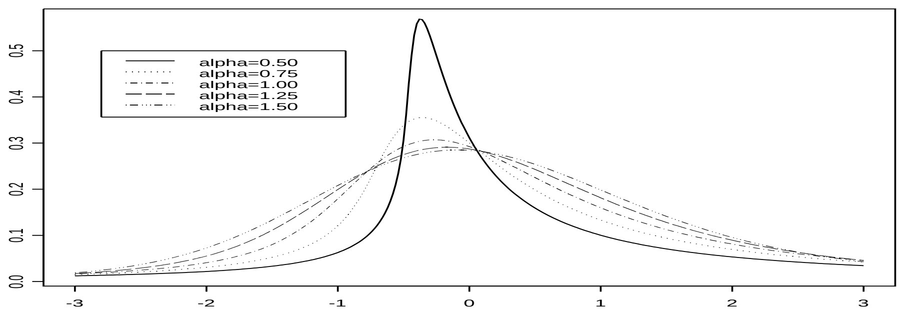

FIGURE 1. Stable densities in the $S(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization. $\beta = 0.5$, $\gamma = 1$, $\delta_0 = 0$, and α as indicated.

$$\begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left[1 + i\beta(\tan\frac{\pi\alpha}{2})(\text{sign } t)((\gamma|t|)^{1-\alpha} - 1)\right] + i\delta_0 t\right) & \alpha \neq 1 \\ \exp\left(-\gamma|t| \left[1 + i\beta\frac{2}{\pi}(\text{sign } t)(\ln|t| + \ln\gamma)\right] + i\delta_0 t\right) & \alpha = 1. \end{cases}$$

The value of this representation is that the characteristic functions (and hence the corresponding densities and d.f.) are jointly continuous in all four parameters. Accurate numerical calculations of the corresponding densities show that in this representation α and β have a much clearer meaning as measures of the heaviness of the tails and skewness parameters, see Figure 1. In contrast, in the standard parameterization, the mode of $X \sim S(\alpha, \beta, \gamma, \delta_1; 1)$ with $\beta \neq 0$ tends to $(\text{sign } \beta)\infty$ as $\alpha \uparrow 1$, is near δ_1 when $\alpha = 1$, and tends to $-(\text{sign } \beta)\infty$ as $\alpha \downarrow 1$.

The parameters α, β and γ have the same meaning for the two parameterizations, while the location parameters of the two representations are related by $\delta_1 = \delta_0 - \beta(\tan\frac{\pi\alpha}{2})\gamma$ if $\alpha \neq 1$; $\delta_1 = \delta_0 - \beta\frac{2}{\pi}\gamma\ln\gamma$ when $\alpha = 1$. The particular form of the characteristic function was chosen to make the $S(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization a location and scale family: if $Y \sim S(\alpha, \beta, \gamma, \delta_0; 0)$, then for any $a \neq 0$, b, $aY + b \sim S(\alpha, (\text{sign } a)\beta, |a|\gamma, a\delta_0 + b; 0)$. We will base the likelihood calculations below on the $S(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization because it is the simplest scale-location parameterization which is jointly continuous in all four parameters. Some authors sidestep the discontinuity at $\alpha = 1$ by saying that the probability that $\alpha = 1$ is zero therefore you can ignore it; Buckle (1995) assumes that you know beforehand that either $\alpha < 1$ or $\alpha > 1$ and restricts his prior for α to the appropriate interval. The shape of the data is what we really care about, and that is similar when α is near or at 1; the standard parameterizations simply masks that with a shift. Clearly it is preferable to let the data determine what α is and not make assumptions about the parameters, even if

α is not near 1. Finally, the use of the $S(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization has the technical advantage of reducing the correlation between the parameter estimates, especially when α is near 1. More information on parameterizations, modes of stable densities and generalizations to multivariate stable laws can be found in Nolan (1998).

Basic properties of stable distributions can be found in Samorodnitsky and Taqqu (1994). Some of the prominent properties are: heavy tails that are asymptotically Pareto (see equation (1) below), possible skewness of the distributions, and smooth unimodal densities with no closed formula. Let $f(x|\alpha, \beta, \gamma, \delta_0)$ be the density of a $S(\alpha, \beta, \gamma, \delta_0; 0)$ distribution. Known facts about stable densities in the standard parameterization show that $f(x|\alpha, -\beta, \gamma, \delta_0) = f(-x|\alpha, \beta, \gamma, -\delta_0)$ and

$$\text{support}\, f(x|\alpha, \beta, \gamma, \delta_0) = \begin{cases} [\delta_0 - \gamma \tan \frac{\pi\alpha}{2}, \infty) & \alpha < 1 \text{ and } \beta = 1 \\ (-\infty, \delta_0 + \gamma \tan \frac{\pi\alpha}{2}] & \alpha < 1 \text{ and } \beta = -1 \\ (-\infty, +\infty) & \text{otherwise.} \end{cases}$$

Note that for a totally skewed ($\beta = \pm 1$) distribution when $\alpha < 1$, the finite endpoint of the support goes to (sign β)∞ as $\alpha \uparrow 1$. It can be shown that the mode and most of the distribution stays concentrated near δ_0, so that only a very small probability is far out on that tail. In fact, the light tail in the totally skewed cases decays faster than Pareto by (1).

2 Other estimators of stable parameters

Several methods have been proposed for estimating stable parameters. For the index of stability α, the earliest approach is just to plot the empirical distribution function of observed data on a log-log scale. It is well known, e.g. Samorodnitsky and Taqqu (1994), pg. 16, that the asymptotic tail behavior of stable laws is Pareto when $\alpha < 2$, i.e.

$$\lim_{x \to \infty} x^\alpha P(X > x) = c_\alpha (1 + \beta)\gamma^\alpha. \tag{1}$$

(For this reason, the phrase stable Paretian is sometimes used for non-Gaussian stable distributions). Thus the tail of the empirical distribution function on a log-log scale should approach a straight line with slope $-\alpha$ if the data is stable. While simple and direct in principle, this method is unreliable in practice. The main problem is that it has not been known when the Pareto tail behavior actually occurs. McCulloch (1997) shows that using the generalized Pareto model suggested by DuMouchel (1983) or the Hill estimator on stable data when $1 < \alpha < 2$ leads to overestimates of

α. McCulloch points out that several researchers have used such misleading tail estimators of α to conclude that various data sets were not stable. In Fofack and Nolan (1999), it is shown that where the Pareto behavior starts to occur depends heavily on the parameterization, and that even when we shift so that the mode is near or at zero, the place where the power decay starts to be accurate is a complicated function of α and β. In particular, when α is close to 2, one must get extremely far out on the tail before the power decay is accurate.

A second approach to estimating stable parameters is based on quantiles of stable distributions. Fama and Roll (1968) noticed certain patterns in tabulated quantiles of stable distributions and proposed estimates of α, scale γ and location δ for symmetric ($\beta = 0$) stable distributions. McCulloch (1986) extended these ideas to the general (nonsymmetric) case, eliminated the bias and obtained consistent estimators for all four stable parameters in terms of five sample quantiles (the 5th, 25th, 50th, 75th and 95th percentiles). If the data set is stable and if the sample set is large, this last method gives reliable estimates of stable parameters.

Since closed forms are known for the characteristic functions of stable laws, several researchers have based estimates on the empirical characteristic functions. Press (1972) seems to have been the first to do this. Several modifications have been made to this approach, see Paulson, Holcomb and Leitch (1975), Feuerverger and McDunnough (1981a) and (1981b), Koutrouvelis (1980) and (1981), Kogon and Williams (1998).

For symmetric stable distributions ($\beta = \delta = 0$), Nikias and Shao (1995) estimate parameters using fractional and negative moments. They also describe a method based on $\log|X|$.

Maximum likelihood estimation has been done in certain cases. While not easily accessible, DuMouchel (1971) gives a wealth of information on estimating stable parameters at a remarkably early date. An approximate maximum likelihood method was developed based on grouping the data set into bins, and using a combination of means to compute the density (the fast Fourier transform for central values of x and series expansions for tails) to compute an approximate log likelihood function. This function was then numerically maximized. See also DuMouchel (1973a), (1973b), (1975) and (1983). For the special case of ML estimation for symmetric stable distributions, see Brorsen and Yang (1990) and McCulloch (1998). Finally, Brant (1984) proposes a method for approximating the likelihood using the characteristic function directly.

A Bayesian approach using Monte Carlo Markov chain methods was proposed by Buckle (1995). The papers by Akgiray and Lamoureux (1989) and Kogon and Williams (1997) give simulation based comparison of several of the methods described above.

3 Maximum Likelihood Estimation

As stated above, we will use the $S(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization in what follows. To simplify notation in this section, we denote the parameter vector by $\vec{\theta} = (\alpha, \beta, \gamma, \delta_0)$ and the density by $f(x|\vec{\theta})$. The parameter space is $\Theta = (0, 2] \times [-1, 1] \times (0, \infty) \times (-\infty, \infty)$. The log likelihood function for an i.i.d. stable sample $X_1, \ldots, X_n$ is given by

$$\ell(\vec{\theta}) = \sum_{i=1}^{n} \log f(X_i|\vec{\theta}).$$

Of course the difficulty in evaluating this is that there are no known closed formulas for general stable densities. Zolotarev (1986) details the much of what is known about stable densities. When α is a special kind of rational number, there are expressions for densities in terms of special functions (Fresnel integrals, MacDonald functions, Whittaker functions). Putting aside the computational difficulties of evaluating the required special functions, knowing densities at isolated values of the parameter space is not helpful when one is trying to maximize a likelihood. Hoffman-Jørgensen (1994) expressed the general density in terms of what he called "incomplete hypergeometric" functions and Zolotarev (1995) expressed the general density in terms of Meijer G-functions. Unfortunately, neither of these representations are practical for actually evaluating stable densities.

The program STABLE described in Nolan (1997) gives reliable computations of stable densities for values of $\alpha > 0.1$ and any β, γ and δ_0. That program was improved to give more accurate density calculations on the tails, which we found to be necessary for accurate likelihood calculations. The program now includes a fast, pre-computed spline approximation to stable densities for $\alpha \geq 0.4$, routines for maximum likelihood estimation of stable parameters, and diagnostics for assessing the stablity of a data set.

The time required to fit a data set of $n = 10,000$ points with a stable model is approximately three seconds on a 333 MHz Pentium II computer using an approximate gradient based search to maximize the likelihood. The quantile estimator of McCulloch (1985) is used as an initial approximation to the parameters and then a constrained (by the parameter space) quasi-Newton method is used to maximize.

3.1 Asymptotic normality and Fisher information matrix

DuMouchel (1971) and (1973a) show that when $\vec{\theta}_0$ is on the interior of the parameter space Θ, the ML estimator follows the standard theory, so it is consistent and asymptotically normal with mean $\vec{\theta}_0$ and covariance matrix given by $n^{-1}B$, where $B = (b_{ij})$ is the inverse of the 4×4 Fisher information matrix I. The entries of I are given by

$$I_{ij} = \int_{-\infty}^{\infty} \frac{\partial f}{\partial \theta_i} \frac{\partial f}{\partial \theta_j} \frac{1}{f} \, dx.$$

We have written a program to numerically compute these integrals. It computes the density f to high accuracy, then numerically computes the partials. The resulting values for the integrands are then numerically integrated.

When $\vec{\theta}$ is near the boundary of the parameter space the finite sample behavior of the estimators is not precisely known. Intuitively, the distribution of the estimator gets skewed away from the boundary. When $\vec{\theta}$ is on the boundary of the parameter space, i.e. $\alpha = 2$ or $\beta = \pm 1$, the asymptotic normal distribution for the estimators tends to a degenerate distribution at the boundary point and the ML estimators are super-efficient. See DuMouchel (1971) for more information on these cases.

The general theory shows that away from the boundary of Θ, large sample confidence intervals for each of the parameters are given by

$$\hat{\theta}_i \pm z_{\alpha/2} \frac{\sigma_{\hat{\theta}_i}}{\sqrt{n}},$$

where $\sigma_{\hat{\theta}_1}, \ldots, \sigma_{\hat{\theta}_4}$ are the square roots of the diagonal entries of B. The values of $\sigma_{\hat{\theta}_i}, i = 1, \ldots, 4$ have been computed and are plotted in Figure 2 when $\gamma = 1$ and $\delta_0 = 0$. The correlation coefficients $\rho_{ij} = b_{ij}/\sqrt{b_{ii}b_{jj}}$, have also been computed and are plotted in Figure 3. These values are tabulated in the Appendix on a grid of α, β values. When $\beta < 0$, the standard deviations are the same as for $|\beta|$ and the correlation coefficients are expressed in terms of the $\beta > 0$ case as $(-1)^{i+j}\rho_{ij}$. For a general scale γ and location δ_0, $\sigma_{\hat{\alpha}}$, $\sigma_{\hat{\beta}}$ and ρ_{ij} are unchanged, but $\sigma_{\hat{\gamma}}$ and $\sigma_{\hat{\delta}_0}$ are γ times the tabulated value.

The grid values were chosen to give a spread over the parameter space and show behavior near the boundary of the parameter space: $\alpha = 2$ and $|\beta| = 1$. (Because of computation difficulties, these values have not been tabulated for $\alpha < 1/2$). We note that when $\beta = 0$, stable densities are symmetric and all the correlation coefficients involving β are 0. When $\beta = 1$, $\sigma_{\hat{\beta}} = 0$ and all the correlation coefficients involving β are undefined.

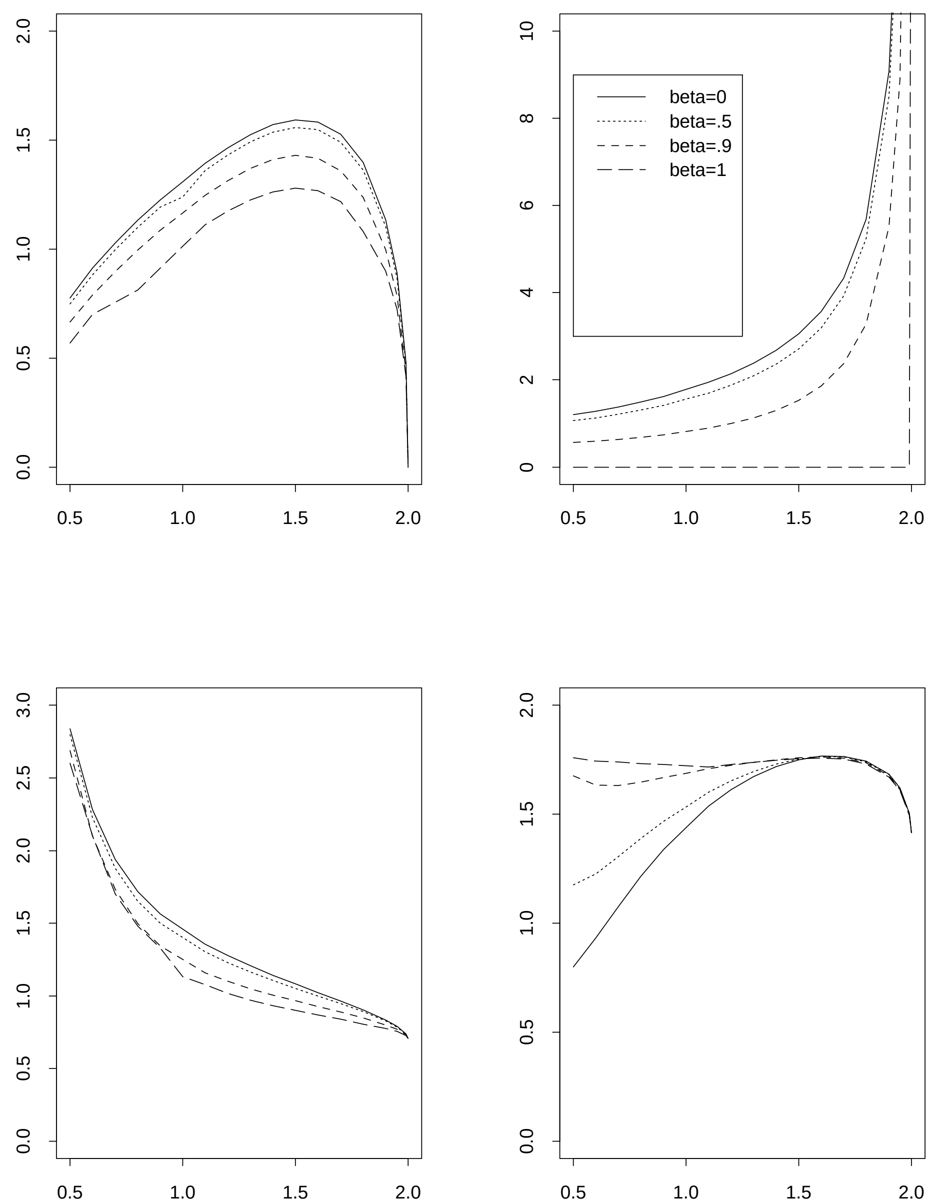

FIGURE 2. Graphs of the standard deviations of the estimators σ_{θ_j} as a function of α, β indicated by line type, $\gamma = 1$ and $\delta_0 = 0$. Upper left is $\sigma_{\hat{\alpha}}$, upper right is $\sigma_{\hat{\beta}}$, lower left is $\sigma_{\hat{\gamma}}$, lower right is $\sigma_{\hat{\delta}_0}$.

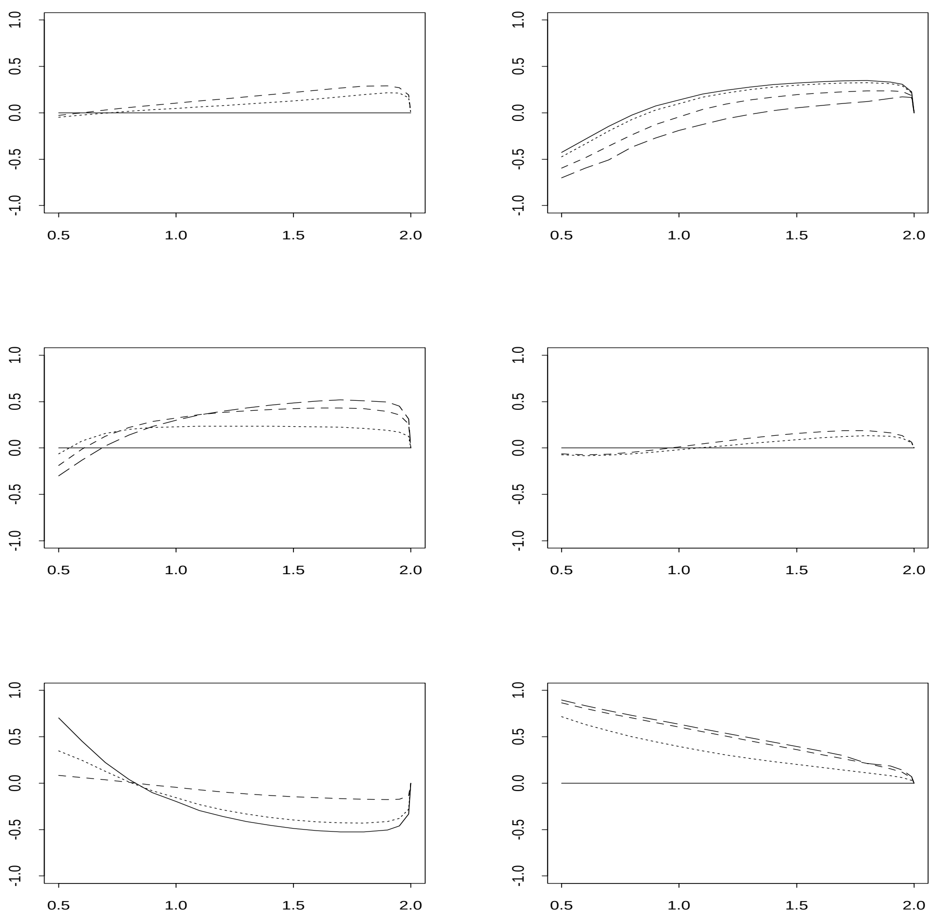

FIGURE 3. Graphs of the correlation of the estimators ρ_{ij} as a function of α, β indicated by line type (as in previous figure), $\gamma = 1$ and $\delta_0 = 0$. Upper left plot is $\rho_{\alpha,\beta}$, upper right plot is $\rho_{\alpha,\gamma}$, middle left plot is ρ_{α,δ_0}, middle right plot is $\rho_{\beta,\gamma}$, lower left plot is ρ_{β,δ_0}, lower right plot is ρ_{γ,δ_0}.

The correlation coefficients directly tell how the parameter estimates are related, and they are useful in estimating the information matrix for a subset of the parameters: express the full covariance matrix in terms of the σ_{θ_i}'s and ρ_{ij}'s, invert to get the full information matrix, delete the rows and columns corresponding to the known parameters, and reinvert to get the covariance matrix for the remaining parameters.

A few of these values have been given in DuMouchel (1971, pg. 93). For α not near 1, most of the values given there agree with our results. That author uses the $S(\alpha, \beta, \gamma, \delta_1; 1)$ parameterization, so near $\alpha = 1$, one would expect different values and numerical problems.

Some general observations about accuracy of parameter estimates can now be made. The parameter of most interest is usually α. Twice the standard error of $\hat{\alpha}$, $2\,SE(\hat{\alpha}) = 2\sigma_{\hat{\alpha}}/\sqrt{n}$, is plotted in Figure 4 for $0.5 \leq \alpha \leq 2$, n=100, 1000 and 10000 and $\beta = 0$, $\beta = 1$. (The graphs for $0 < |\beta| < 1$ are between the given ones.) Unless α is close to 2, it is clear that a large data set will be necessary to get small confidence intervals, e.g. when $\alpha = 1.5$ and $\beta = 0$, sample sizes of 100, 1000 and 10000 yields $S.E.(\hat{\alpha})$'s of 0.318, 0.100 and 0.0318 respectively. Since no other estimation method is asymptotically more efficient than ML, any other method of estimating α will likely yield larger confidence intervals. In contrast, when $\alpha \uparrow 2$, $S.E.(\hat{\alpha})$ approaches 0. Similar calculations of standard errors for $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}_0$ also show that large samples will generally be necessary for small confidence intervals. As an extreme, as $\alpha \uparrow 2$, $S.E.(\hat{\beta}) \to \infty$. In practice, this is of little import because β means little as $\alpha \uparrow 2$.

4 Diagnostics for assessing stability

In principle, it is not surprising that one can fit a data set better with the 4 parameter stable model than with the 2 parameter normal model. The relevant question is whether or not the stable fit actually describes the data well. Any procedure for estimating stable parameters will find a "best fit" by its criteria: the maximum likelihood approach maximizes $\ell(\vec{\theta})$ directly, the quantile methods try to match certain data quantiles with those of stable distributions, the characteristic function based methods fit the empirical characteristic function. All will give some values for parameter estimates, even if the shape of the observed distribution is not similar to the fitted distribution, e.g. the data is multi-modal, has gaps in its support, etc. It is necessary to have some means of assessing whether the resulting fit is reasonable.

The use of a diagnostic depends on what you are planning to do with a

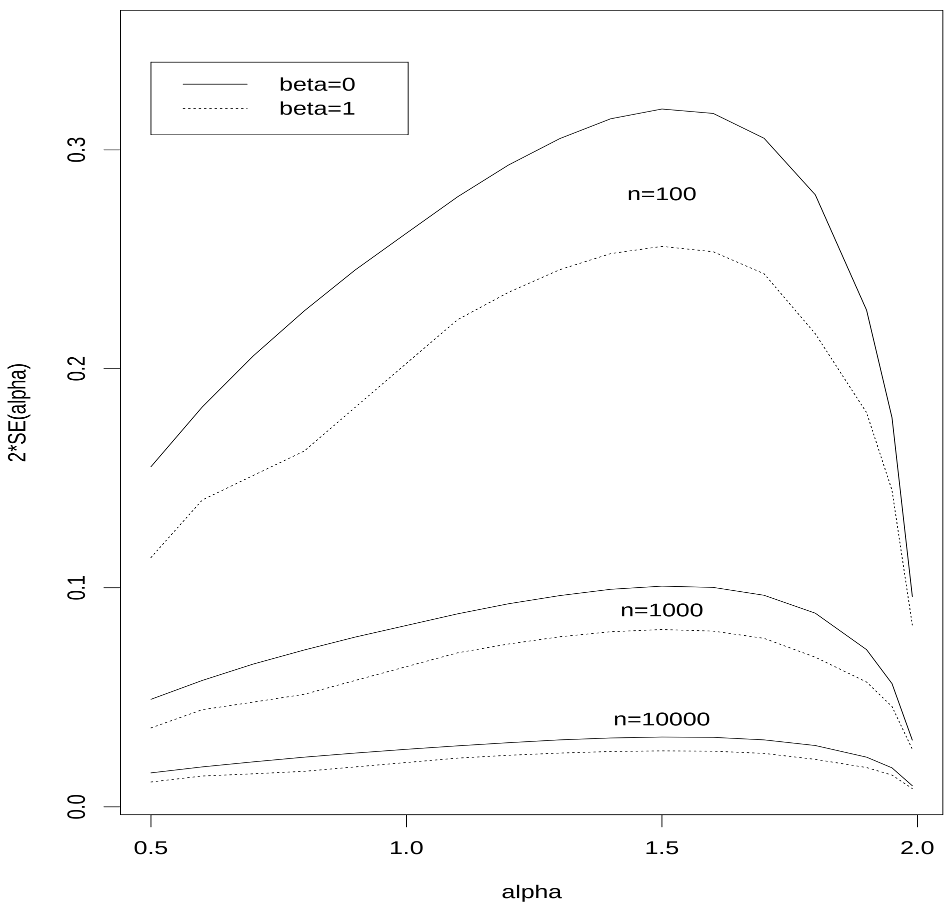

FIGURE 4. Graph of twice the standard error of $\hat{\alpha}$ as a function of α for various sample sizes.

data set. For testing residuals from a regression analysis, departures from normality around the center of the distribution are usually not important; outliers are important because they can affect the validity of normal theory conclusions. In the re-insurance field, one is only concerned with extreme events and there one wants to estimate tails of the claim distribution as accurately as possible. In a model of stock prices or exchange rates, one may be interested in the shape of the whole distribution.

While non-Gaussian stable distributions are heavy-tailed distributions, most heavy-tailed distributions are not stable. One can try to fit a heavy-tailed data set with a stable distribution, but it is inappropriate in many cases. As DuMouchel (1983) points out, making a statement about the tails is quite distinct from making a statement about the entire distribution. We amplify this point by an example similar to one used by DuMouchel. Define for $0 < \alpha < 2$, $x_0 > 0$

$$g(x) = g(x|\alpha, x_0) = \begin{cases} c_1 e^{-x^2/2} & |x| \le x_0 \\ c_2 |x|^{-(1+\alpha)} & |x| > x_0, \end{cases}$$

where c_1 and c_2 are chosen to make g continuous and $\int g(x)dx = 1$: $c_1 = c_1(\alpha, x_0) = [\sqrt{2\pi}(2\Phi(x_0) - 1) + (2/\alpha)x_0 \exp(-x_0^2/2)]^{-1}$, $c_2 = c_1 \exp(-x_0^2/2)$ $x_0^{1+\alpha}$. A random variable X with density g has a normal density in the interval $-x_0 \le x \le x_0$, a Pareto tail, and has fraction $p = P(|X| \le x_0) = c_1\sqrt{2\pi}(2\Phi(x_0) - 1)$ in the normal part of the density and $1 - p$ on the Pareto tails. For any finite x_0, this density has infinite variance and is in the domain of attraction of a symmetric stable distribution with index of stability α. Suppose we observe a sample of size n from such a distribution and try to fit it with a stable distribution. If $(1 - p)n$ is small, we will likely have few observations from the Pareto part of the distribution and we will not be able to detect the heavy tails. Any reasonable estimation scheme would lead to an $\hat{\alpha} \approx 2$. On the other hand, if $(1 - p)n$ is large, then one would get an $\hat{\alpha}$ intermediate between the true α and 2, because the central part of the data is coming from a non-heavy tailed density. An incorrect model is being fit to the data, so it is no surprise that we get the "wrong" α. DuMouchel's argument to let the tails speak for themselves is sound, though his suggestion to use the upper 10% of the sample to fit the tail is generally not appropriate, see McCulloch (1997) and Fofack and Nolan (1999). We mentioned above that the eventual power decay on a stable tail may take a long time to occur; for an arbitrary distribution, there is no general statement that can be made about what fraction of the tail is appropriate. (For a recent summary of work on tail estimation, see Beirlant, Vynckier and Teugels (1996).)

The diagnostics we are about to discuss are an attempt to detect non-stability. As with any other family of distributions, it is not possible to prove that a given data set is or is not stable. We note that even testing for normality is still an active field of research, e.g. Brown and Hettmansperger (1996). The best we can do is determine whether or not the data are consistent with the hypothesis of stability. These tests will fail if the departure from stability is small or occurs in an unobserved part of the range.

The first step we suggest is to do a smoothed density plot of the data. If there are clear multiple modes or gaps in the support, then the data cannot come from a stable distribution. If the smoothed density is plausibly stable, proceed with a stable fit and compare the fitted distribution with the data using EDA (Exploratory Data Analysis) techniques.

We note a practical problem with q-q plots for heavy tailed data. While using q-q plots to compare simulated stable data sets with the exact corresponding cumulative d.f., we routinely had two problems with extreme values: most of the data is visually compressed to a small region and the high tail variability leads one to doubt the stability of the data set. To illustrate this point, we simulated a stable data set with $n = 1,000$ values using the Chambers, Mallows and Stuck (1976) method. The values of the parameters used were $\alpha = 1.3$, $\beta = 0.5$, $\gamma = 5$, and $\delta_0 = 10$. (This same data set is fit using maximum likelihood in the next section.) Figure 5 (left) shows a standard q-q plot of the data vs. the known stable distribution. On the tails there seems to be an unacceptably large amount of fluctuation around the theoretical straight line. For heavy tailed stable distributions, we should expect such fluctuations for the following reason. If $X_{(i)}$ is the i^{th} order statistic from an i.i.d. stable sample of size n, $p = (i-0.5)/n$ and and x_p is the p^{th} percentile, then for n large, the distribution of $X_{(i)}$ is approximately normal with $EX_{(i)} = x_p$ and $\text{Var}(X_{(i)}) = p(1-p)/nf(x_p)^2$. Figure 5 (left) also shows pointwise 95% confidence bounds around the expected value. A heavy tailed distribution should show much larger extremes than a normal sample, e.g. values in the hundreds for this example. Furthermore, the standard errors for the extreme values are also very large. The maximum in this sample was $X_{(1000)} = 618.64$, the corresponding population quantile is $x_{0.9995} = 839.56$, with 95% confidence interval (-955,2634). In sum, q-q plots will generally appear non-linear on the tails, even when the data set is stable.

One technique we tried to lessen this effect was to use a “thinned” q-q plot on large data sets. We illustrate the idea in Figure 5 (right), where only 100 values (of the 1,000 in the sample) are plotted. This gives a point at every 1% of the data, eliminating the most extreme values. While this method eliminates the worst behavior on the tails, information is lost and

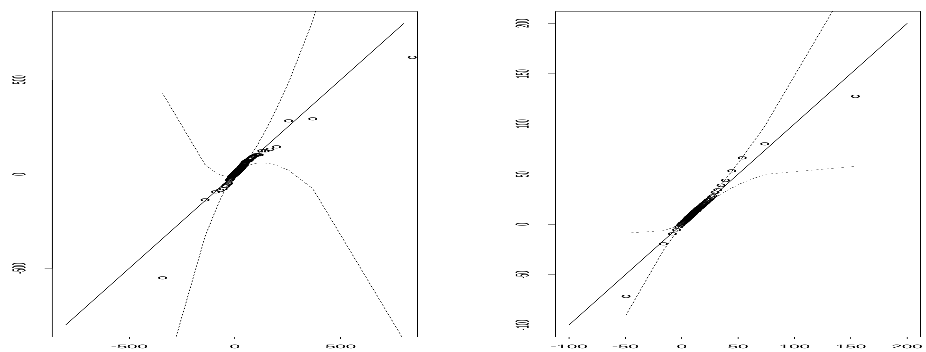

FIGURE 5. Diagnostics for simulated stable data set with $n = 1,000$ data points and $\alpha = 1.3$, $\beta = 0.5$, $\gamma = 5$ and $\delta_0 = 10$. Left graph is a q-q plot for data vs. exact, right is a thinned q-q plot obtained by using every 10^{th} value of the original.

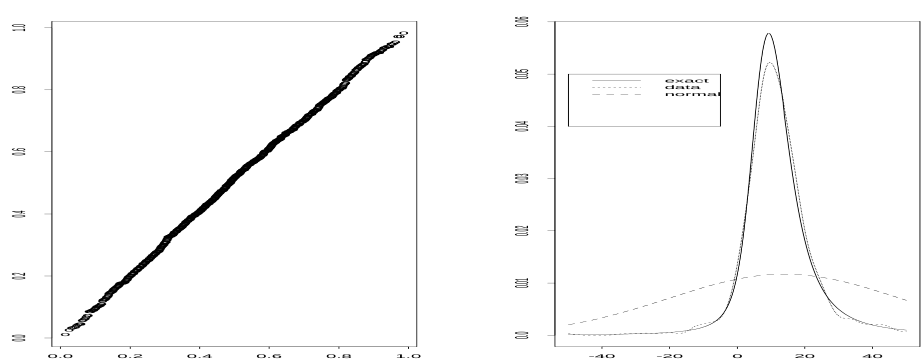

FIGURE 6. Simulated stable data set with n=1000. Left graph is a stabilized p-p plot of data vs. exact $S(1.3, 0.5, 5, 10)$ distribution, right graph shows the smoothed data density (dots) and fitted density (solid) and normal fit (dashes).

the confidence intervals for sample quantiles can still be huge, especially when α gets small.

Our proposed solution is to use a modified p-p plot. Standard p-p plots tend to emphasize behavior around the mode of the distribution, where they have more variation, and necessarily pinch the curve near the tails. In Michael (1983), a "stabilized" p-p plot was defined that eliminates this non-uniformity by using a transformation. (The word stabilized refers to making the variance in the p-p plot uniform, and has nothing to do with stable distributions.) The result is that the acceptance regions for a p-p plot become straight lines spaced a uniform distance above and below the diagonal. A stabilized p-p plot for the simulated data is shown in Figure 6 (left).

For density plots, we smoothed the data with a Gaussian kernel with standard deviation given by a "width" parameter. We found that the commonly suggested width of 2(inter-quartile range)$n^{-1/3}$ works reasonably when the tails of the data are not too heavy, say $\alpha > 1.5$, but works poorly for heavier tailed data. For such cases, we used trial and error to find a width parameter that was a small as possible without showing oscillations from individual data points. The density plots give a good sense of whether the fit matches the data near the mode of the distribution, but generally is uninformative on the tails where both the fitted density and the smoothed data density are small. Figure 6 (right) shows the smoothed data density from the simulated data, the exact population density and a normal fit (using the sample mean and sample standard deviation). We note that both the skewness and the leptokurtosis (a higher thin peak and heavy tails) in the data set are poorly described by a normal fit.

Finally, we tried comparing distribution functions, but did not find it very helpful. Because of the curvature in the distribution functions, it is hard to compare the fitted and empirical d.f. visually, especially on the tails.

5 Applications

5.1 Simulated stable data set

A stable data set with $\alpha = 1.3$, $\beta = 0.5$, $\gamma = 5$ and $\delta_0 = 10$ and $n = 1,000$ values was generated using the method of Chambers, Mallows and Stuck (1976). The quantile estimators of the parameters are: $\hat{\alpha} = 1.204$, $\hat{\beta} = .496$, $\hat{\gamma} = 4.840$, and $\hat{\delta}_0 = 10.426$. The maximum likelihood estimates with 95% confidence intervals are $\hat{\alpha} = 1.284 \pm .092$, $\hat{\beta} = .466 \pm 0.130$, $\hat{\gamma} = 5.111 \pm .369$, $\hat{\delta}_0 = 10.338 \pm .537$. The stabilized p-p plot and smoothed density are

country	α	β	γ	δ_0
Australia	1.479 ± 0.047	0.033 ± 0.080	0.00413 ± 0.00013	-0.00015 ± 0.00022
Austria	1.559 ± 0.047	-0.119 ± 0.092	0.00285 ± 0.00009	0.00014 ± 0.00015
Belgium	1.473 ± 0.047	-0.061 ± 0.080	0.00306 ± 0.00010	0.00009 ± 0.00016
Canada	1.574 ± 0.047	-0.051 ± 0.093	0.00379 ± 0.00012	0.00004 ± 0.00020
Denmark	1.545 ± 0.047	-0.119 ± 0.090	0.00272 ± 0.00008	0.00022 ± 0.00014
France	1.438 ± 0.047	-0.146 ± 0.078	0.00245 ± 0.00008	0.00028 ± 0.00013
Germany	1.495 ± 0.047	-0.182 ± 0.085	0.00244 ± 0.00008	0.00019 ± 0.00013
Italy	1.441 ± 0.046	-0.043 ± 0.076	0.00266 ± 0.00009	0.00017 ± 0.00014
Japan	1.511 ± 0.047	-0.148 ± 0.086	0.00368 ± 0.00012	0.00013 ± 0.00019
Netherlands	1.467 ± 0.047	-0.167 ± 0.081	0.00244 ± 0.00008	0.00016 ± 0.00013
Norway	1.533 ± 0.047	-0.070 ± 0.088	0.00253 ± 0.00008	0.00005 ± 0.00013
Spain	1.512 ± 0.047	-0.007 ± 0.083	0.00268 ± 0.00008	0.00012 ± 0.00014
Sweden	1.517 ± 0.047	-0.081 ± 0.085	0.00256 ± 0.00008	0.00006 ± 0.00013
Switzerland	1.599 ± 0.047	-0.179 ± 0.100	0.00295 ± 0.00009	0.00014 ± 0.00016
United States	1.530 ± 0.047	-0.088 ± 0.088	0.00376 ± 0.00012	0.00009 ± 0.00020

TABLE 1. Exchange rate analysis. Parameter estimates and 95% confidence intervals with sample size of $n = 4274$.

visually indistinguishable from the ones in Figure 6, so new diagnostics are not shown.

5.2 Exchange rate data

Daily exchange rate data for 15 different currencies were recorded (in U.K. Pounds) over a 16 year period (2 January 1980 to 21 May 1996). The data was transformed by $y_t = \ln(x_{t+1}/x_t)$, giving $n = 4,274$ data values. The transformed data was fit with a stable distribution; results are shown in Table 1. The data are likely non-stationary over such a time period and there are questions about the dependence in the values, nevertheless we will do a naive fit here to illustrate the method.

Figure 7 shows a stabilized p-p plot and smoothed density for the German Mark data set. The data sets are clearly not normal: the heavy tails in the data causes the sample variance to be large, and the normal fit poorly describes both the center and the tails of the distribution. The granularity at the center of the graph is from the days where the exchange rate was unchanged on successive days. As another measure of non-normality, the ratio of the stable fit log likelihood to the normal log likelihood was computed for each currency. The ratio of the log likelihoods for the ML stable fit to the normal fit were computed and the values ranged from 113 to 1041.

Plots for the other currencies were similar, showing that the stable fit does a good job of describing the exchange rate data. We note in passing that the currency with the heaviest tails ($\hat{\alpha} = 1.441$) was the Italian Lire, while the one with the lightest tails ($\hat{\alpha} = 1.530$) was the Swiss Franc.

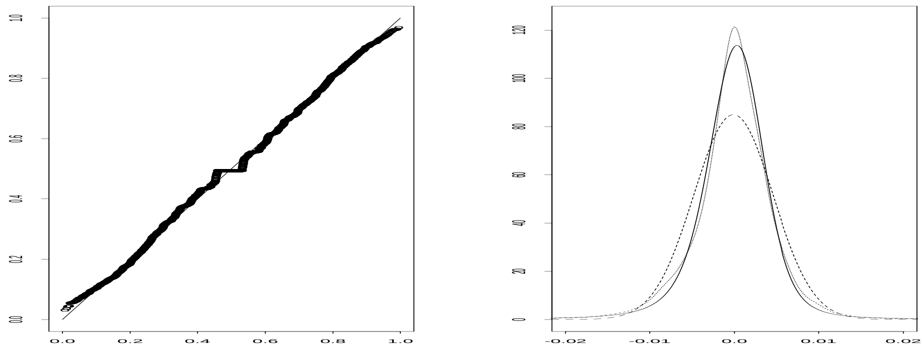

FIGURE 7. Stabilized p-p and density plots for the German mark exchange rate data. n=4274.

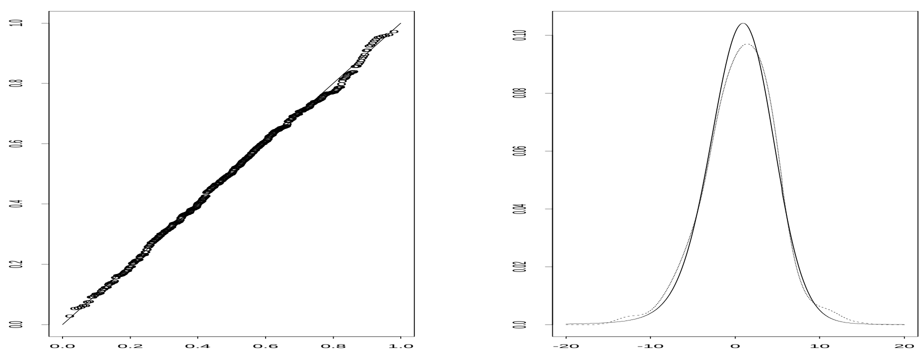

FIGURE 8. Stabilized p-p plot and densities for the CRSP stock price data, n=480.

5.3 CRSP stock prices

McCulloch (1997) analyzed forty years (January 1953 - December 1992) of monthly stock price data from the Center for Research in Security Prices (CRSP). The data set consists of 480 values of the CRSP value-weighted stock index, including dividends, and adjusted for inflation. The quantile estimates were $\hat{\alpha} = 1.965$, $\hat{\beta} = -1$, $\hat{\gamma} = 2.755$ and $\hat{\delta}_0 = 0.896$. McCulloch (unpublished) used ML with an approximation for symmetric stable distributions to fit this data and obtained $\hat{\alpha} = 1.845$, $\hat{\beta} = 0$, $\hat{\gamma} = 2.712$ and $\hat{\delta}_0 = 0.673$. Our ML estimates with naive 95% confidence intervals are $\hat{\alpha} = 1.855 \pm 0.110$, $\hat{\beta} = -0.558 \pm 0.615$, $\hat{\gamma} = 2.711 \pm 0.213$ and $\hat{\delta}_0 = 0.871 \pm 0.424$. The diagnostics in Figure 8 show a close fit.

We note that the confidence interval for $\hat{\alpha}$ is close to the upper bound of 2 for α and the one for $\hat{\beta}$ is large and extends beyond the lower bound of -1, so the naive confidence intervals cannot be strictly believed.

method	α	β	γ	δ_0
quantile	1.996	1.000	.008579	-.003445
MCMC	1.650	.768	.007900	-.003187
ML	$1.518 \pm .422$	$.743 \pm .651$	$.006828 \pm .001931$	$-.003064 \pm .003359$

TABLE 2. Abbey National share price parameter estimates, $n = 49$.

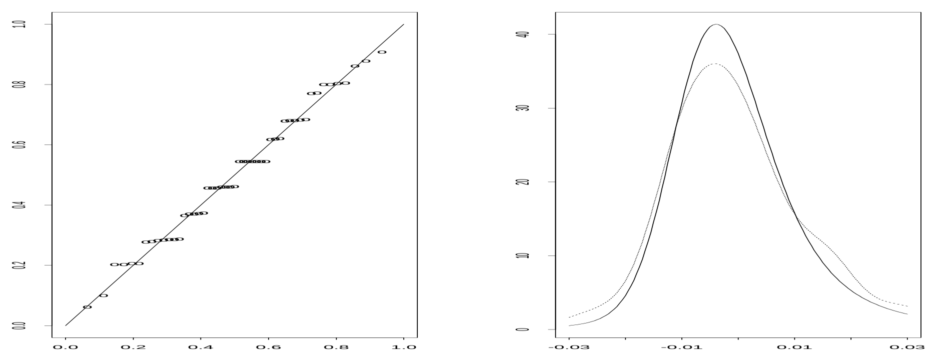

FIGURE 9. Stabilized p-p plot and densities for Abbey share price data, n=49.

5.4 Abbey National share price

Buckle (1995) listed a small data set of stock price data. The price for Abbey National shares was recorded for the period 31 July 1991 through 8 October 1991. The return was defined as $(x_{t+1}/x_t) - 1$, yielding $n = 49$ data points, which were fit with a stable distribution. In the Monte Carlo Markov chain (MCMC) approach used by Buckle, the means of the posterior distributions were given. Table 2 lists these MCMC parameter estimates (transformed to the $S(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization), the quantile estimates, and the ML estimates with naive 95% confidence intervals.

The quantile method fit is essentially a normal distribution with $\alpha = 1.996$, yet highly skewed. This is likely caused by the small sample size: with $n = 49$, the 5^{th} percentile is found by interpolating between the second and third data point. It is hard to detect heavy tails when there is virtually no tail. The MCMC method and ML method reach similar estimates. We tried the diagnostics on this data set and got mixed results, see Figure 9. The data are concentrated on a subset of values and it is not clear how good a stable model is for this small data set. In particular, healthy skepticism is called for when making statements about tail probabilities unless a large data set is available to verify stable behavior.

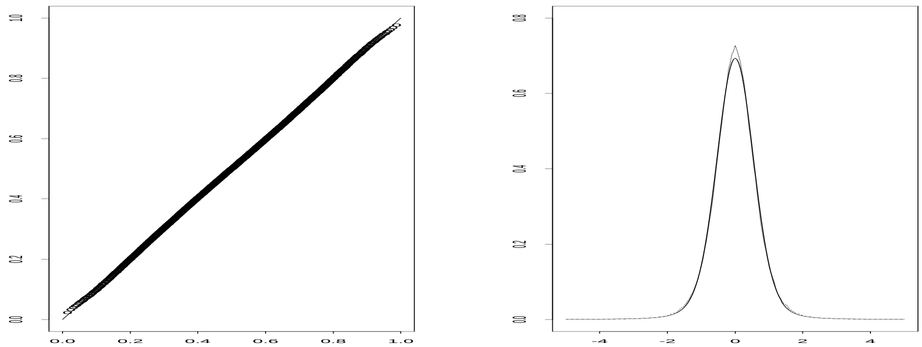

FIGURE 10. Stabilized p-p plot and densities for the in-phase component of sea clutter radar noise, n=320000.

5.5 Radar noise

This is a very large data set with $n = 320,000$ pairs of data points. The two values correspond to the in-phase and quadrature components of sea clutter radar noise. We focus on the in-phase component only in this paper. (Unpublished work shows that the other component is similar, and that the bivariate data set is radially symmetric.) The parameter estimates are $\hat{\alpha} = 1.7966 \pm .0048$, $\hat{\beta} = .0054 \pm .0173$, $\hat{\gamma} = .4402 \pm .0013$ and $\hat{\delta}_0 = -.00060 \pm .00247$. (The quantile based estimators are $\hat{\alpha} = 1.7042$, $\hat{\beta} = .0058$, $\hat{\gamma} = .3981$ and $\hat{\delta}_0 = -.00040$.) With this large sample size, the confidence intervals for the ML parameter estimates are very small. Again the correct question is not how tight the parameter estimates are, but whether or not the fit accurately describes the data. The p-p plot and density plots in Figure 10 show a close stable fit. Because 320,000 data points add little to the p-p plot, we actually show a thinned p-p plot with 1,000 values.

5.6 Ocean wave power

Pierce (1997) proposed using positive α-stable distributions to model inherently positive quantities such as energy or power. One example he uses is the power in ocean waves, which is proportional to the square of the wave height. Pierce used a National Oceanographic and Atmospheric Administration (NOAA) data set with hourly measurements of sea wave. We used the same data set, edited out invalid numbers (99.00) and had 8084 values for the wave height variable WVHT. Pierce compared the data with an α=0.75, $\beta = 1$ distribution (it is not indicated how these values are obtained). Our analysis gave quantile estimates of $\hat{\alpha} = 1.139255$, $\hat{\beta} = 1$,

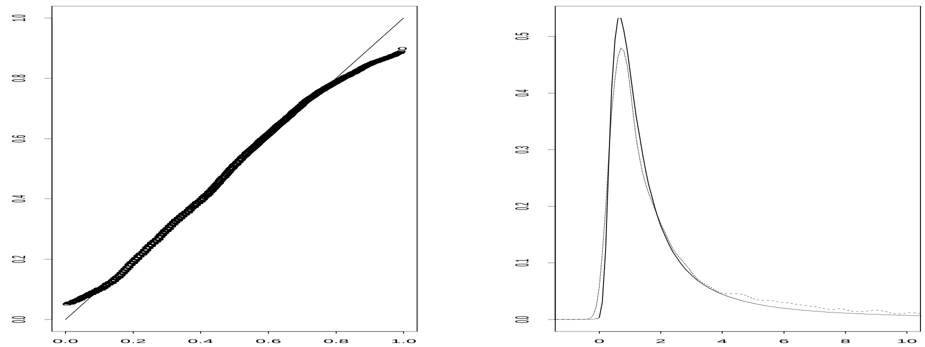

FIGURE 11. Stabilized p-p plot and densities for wave height squared (proportional to power), n=8084.

$\hat{\gamma} = 0.813324$ and $\hat{\delta}_0 = 0.841235$; the ML estimates with naive 95% confidence intervals are $\hat{\alpha}$=0.800 $\pm$ 0.0177, $\hat{\beta}$=1 $\pm$ 0, $\hat{\gamma}$=0.566 $\pm$ 0.018 and $\hat{\delta}_0$=0.965 $\pm$ 0.021. The fact that we get very different estimates of α is an indication that the data set is not stable. The diagnostics in Figure 11 support this idea. The stabilized pp-plot and the density plot show a reasonable fit around the mode, but a poor fit on both tails. As in any problem, it is possible that the energy in waves is stably distributed, but that measurement of extremes (both high and low) of wave height are unreliable, leading to the discrepancies we see on the tails.

The referee kindly pointed out that there is recent work on the related topic of wave heights and wind speed in de Haan and de Ronde (1998).

5.7 Simulated non-stable data

We simulated several data sets that were not stable and used our diagnostics to assess the fit with a stable model. The first is a data set consisting of a mixture of 9,000 Gaussian random variables with scale 1 and 1,000 Gaussian random variables with scale 10, a "contaminated" normal mixture. The mixture has heavier tails than a pure normal, so one might try to fit it with a stable distribution. However, what we would really like to do is detect that it is not a stably distributed data set. The ML estimates of the parameters are $\alpha = 1.346 \pm .030$ and $\gamma = 1.048 \pm .033$. Here the confidence intervals are small because the sample size of $n = 10,000$, not because we have a good fit. The density plot in Figure 12 shows the smoothed data density and the stable fit. The curves show a systematic difference that indicates departure from a stable distribution. It is interesting to note that in this example, the percentile estimate of α is 1.535, quite different from the ML estimate. This is another indication that the data is not stable: if

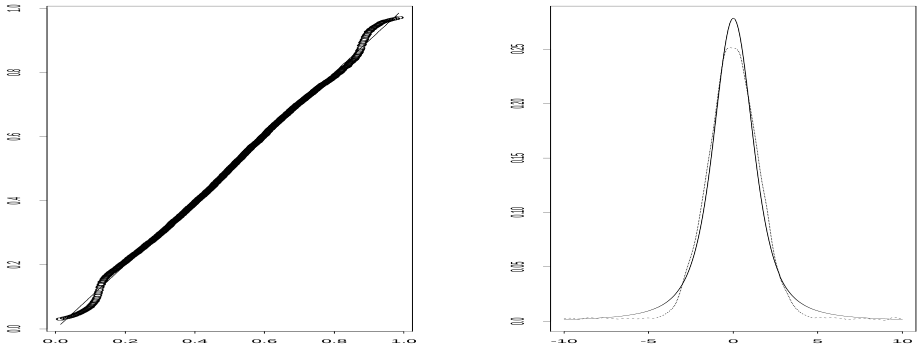

FIGURE 12. Stabilized p-p plot and densities for simulated contaminated normal mixture, n=10000.

the distribution is stable, then all consistent estimators of the parameters should be close when there is a large sample.

The next example is a mixture of two Cauchy distributions with different modes: $\alpha = 1$, $\beta = 0$, $\gamma = 1$, with $\delta_0 = 5$ for 100 data points and then $\delta_0 = -5$ for another 100 data points. Simple diagnostics (not shown) show the bimodality, so a stable model is clearly not appropriate. Still, it is instructive to see what happens if we fit these data with a stable model. The maximum likelihood estimates are $\alpha = 2$, $\gamma = 3.867$, and $\delta_0 = -.395$ (β is meaningless in the normal case). Apparently the likelihood for this data set is dominated by the central terms and is maximized by taking a normal curve with large variance. Even though this is a heavy tailed data set, the use of an inappropriate stable model leads to a light tailed fit!

We briefly mention two other experiments we did. In one experiment, 10000 variables were generated from a Pareto distribution ($F(x) = 1 - x^{-1-\alpha}$; $x > 1$) with $\alpha = 1.5$. The quantile and ML estimates of α were 1.23 and 0.9 respectively, β was essentially 1. This shows that a stable fit to a data set with genuine Pareto tails will give poor estimates of the tail index. In the second experiment 10000 Gamma(2) variates were generated and fit with a stable distribution. The quantile and ML estimates of α were 1.98 and 1.80 respectively, β was essentially 1. This shows that the light tails of the Gamma distribution lead to estimates of α close to the Gaussian case, but the skewed nature of the data showed up in the estimate of β.

6 Discussion

We have shown that ML estimation of general stable parameters is now feasible. The diagnostics show that several large data sets with heavy tails are well described by stable distributions. We also showed that stable models are not a panacea - not all heavy tailed data sets can be well described by stable distributions.

In practice, the decision to use a stable model should be based on the purpose of the model. In cases where a large data set shows close agreement with a stable fit, confident statements can be made about the population. In other cases, one should clearly not use a stable model. In intermediate cases, one could tentatively use a stable model as a descriptive method of summarizing the general shape of the distribution, but not try to make statements about tail probabilities. In such problems, it may be better to use the quantile estimators rather than ML estimators, because the former tries to match the shape of the empirical distribution and ignores the top and bottom 5% of the data.

We have not considered parameters that vary with time, mixture models, etc. While we do not do so here, it is possible to use an information criteria like AIC to compare a stable model to mixture models or GARCH models for a data set. It seems likely that certain problems, e.g. the radar sea clutter problem, have physical explanations that make a stationary model plausible. Other problems, particularly economic time series, may very well have time varying parameters that reflect changes in the underlying conditions for that series. We cannot resolve this issue here. Our main purpose is to make stable models a practical tool that can be used and evaluated by the statistical community.

We note that there are now several methods of estimation for multivariate stable distributions. In the multivariate setting one has to estimate α, a shift vector, and a spectral measure. For references and new work on this problem, see Nolan, Panorska and McCulloch (1996), Nolan and Panorska (1997), and Nolan (1999). One of those methods is based on estimation of one dimensional stable parameters and would be improved with the quick ML algorithm described here.

Acknowledgments: The data sets analyzed above were graciously provided by C. Klüppelberg (exchange rate data), P. Tsakalides (radar data) and J. H. McCulloch (CRSP stock data). R. Jernigan provided discussion and references on EDA techniques.

7 Appendix

Asymptotic standard deviations and correlation coefficients for estimators.

α	β	σ_α	σ_β	σ_γ	σ_{δ_0}	$\rho_{\alpha,\beta}$	$\rho_{\alpha,\gamma}$	ρ_{α,δ_0}	$\rho_{\beta,\gamma}$	ρ_{β,δ_0}	ρ_{γ,δ_0}
.50	.00	.776	1.201	2.840	.799	.000	-.427	.000	.000	.705	.000
	.50	.749	1.065	2.797	1.175	-.048	-.474	-.062	-.075	.349	.716
	.90	.665	.564	2.687	1.676	-.029	-.595	-.190	-.063	.084	.866
	1.00	.569	.000	2.600	1.760	*	-.700	-.300	*	*	.895
.60	.00	.912	1.278	2.283	.933	.000	-.285	.000	.000	.452	.000
	.50	.883	1.129	2.226	1.226	-.025	-.337	.076	-.083	.246	.634
	.90	.789	.594	2.101	1.633	.000	-.485	-.011	-.074	.062	.805
	1.00	.700	.000	2.104	1.743	*	-.595	-.130	*	*	.835
.70	.00	1.029	1.374	1.939	1.076	.000	-.145	.000	.000	.222	.000
	.50	.997	1.210	1.875	1.304	-.004	-.198	.157	-.078	.127	.564
	.90	.897	.633	1.731	1.631	.029	-.360	.126	-.067	.036	.751
	1.00	.756	.000	1.700	1.740	*	-.510	.020	*	*	.781
.80	.00	1.132	1.488	1.717	1.214	.000	-.024	.000	.000	.039	.000
	.50	1.100	1.305	1.652	1.388	.016	-.073	.201	-.063	.014	.502
	.90	.995	.681	1.496	1.647	.056	-.237	.222	-.048	.008	.702
	1.00	.813	.000	1.480	1.733	*	-.370	.140	*	*	.732
.90	.00	1.226	1.619	1.563	1.338	.000	.072	.000	.000	-.101	.000
	.50	1.194	1.418	1.500	1.467	.033	.028	.222	-.043	-.083	.445
	.90	1.086	.739	1.343	1.669	.081	-.127	.286	-.020	-.020	.654
	1.00	.912	.000	1.330	1.728	*	-.270	.230	*	*	.684
1.00	.00	1.309	1.780	1.459	1.438	.000	.137	.000	.000	-.198	.000
	.50	1.240	1.560	1.401	1.534	.048	.098	.228	-.020	-.158	.396
	.90	1.166	.817	1.251	1.688	.104	-.045	.323	.012	-.045	.605
	1.00	1.012	.000	1.133	1.722	*	-.190	.300	*	*	.635
1.10	.00	1.393	1.942	1.355	1.538	.000	.202	.000	.000	-.294	.000
	.50	1.360	1.699	1.302	1.601	.064	.168	.234	.003	-.232	.347
	.90	1.246	.894	1.160	1.708	.127	.037	.361	.044	-.072	.556
	1.00	1.112	.000	1.077	1.717	*	-.125	.357	*	*	.586
1.20	.00	1.465	2.144	1.276	1.613	.000	.244	.000	.000	-.361	.000
	.50	1.432	1.878	1.229	1.653	.078	.214	.235	.026	-.288	.306
	.90	1.315	.998	1.099	1.725	.149	.094	.384	.076	-.094	.507
	1.00	1.175	.000	1.018	1.728	*	-.065	.397	*	*	.539
1.30	.00	1.526	2.383	1.206	1.673	.000	.277	.000	.000	-.413	.000
	.50	1.492	2.094	1.165	1.696	.093	.249	.235	.049	-.333	.268
	.90	1.371	1.129	1.050	1.739	.172	.137	.401	.105	-.114	.458
	1.00	1.226	.000	.971	1.738	*	-.016	.431	*	*	.492
1.40	.00	1.571	2.677	1.142	1.718	.000	.302	.000	.000	-.455	.000
	.50	1.536	2.362	1.107	1.730	.110	.277	.233	.070	-.369	.233
	.90	1.412	1.299	1.006	1.750	.195	.170	.415	.132	-.131	.409
	1.00	1.263	.000	.933	1.748	*	.022	.460	*	*	.444
1.50	.00	1.593	3.054	1.082	1.750	.000	.322	.000	.000	-.487	.000
	.50	1.558	2.712	1.052	1.753	.128	.297	.232	.090	-.397	.201
	.90	1.430	1.529	.966	1.759	.219	.195	.425	.156	-.145	.361
	1.00	1.279	.000	.900	1.755	*	.053	.485	*	*	.396
1.60	.00	1.583	3.565	1.024	1.766	.000	.336	.000	.000	-.510	.000
	.50	1.548	3.194	1.000	1.765	.148	.312	.229	.109	-.417	.171
	.90	1.417	1.857	.928	1.762	.244	.213	.431	.175	-.156	.313
	1.00	1.267	.000	.869	1.758	*	.078	.506	*	*	.348

α	β	σ_α	σ_β	σ_γ	σ_{δ_0}	$\rho_{\alpha,\beta}$	$\rho_{\alpha,\gamma}$	ρ_{α,δ_0}	$\rho_{\beta,\gamma}$	ρ_{β,δ_0}	ρ_{γ,δ_0}
1.70	.00	1.526	4.332	.965	1.766	.000	.344	.000	.000	-.524	.000
	.50	1.491	3.928	.947	1.763	.171	.321	.223	.125	-.428	.142
	.90	1.360	2.367	.889	1.757	.267	.226	.433	.187	-.165	.264
	1.00	1.217	.000	.840	1.754	*	.101	.520	*	*	.298
1.80	.00	1.397	5.686	.904	1.744	.000	.346	.000	.000	-.527	.000
	.50	1.363	5.235	.891	1.741	.197	.324	.213	.134	-.430	.112
	.90	1.237	3.281	.848	1.735	.287	.235	.424	.187	-.172	.213
	1.00	1.080	.000	.804	1.730	*	.122	.508	*	*	.214
1.90	.00	1.134	9.082	.835	1.683	.000	.333	.000	.000	-.503	.000
	.50	1.104	8.504	.827	1.681	.217	.313	.191	.128	-.412	.080
	.90	.998	5.517	.800	1.676	.292	.237	.395	.165	-.178	.156
	1.00	.900	.000	.776	1.671	*	.155	.496	*	*	.185
1.95	.00	.889	14.587	.791	1.620	.000	.306	.000	.000	-.461	.000
	.50	.865	13.717	.786	1.618	.214	.290	.170	.107	-.381	.059
	.90	.784	8.911	.769	1.612	.272	.230	.357	.132	-.174	.117
	1.00	.723	.000	.755	1.608	*	.174	.451	*	*	.142
1.99	.00	.480	43.227	.738	1.505	.000	.224	.000	.000	-.332	.000
	.50	.469	39.967	.737	1.503	.162	.216	.125	.054	-.280	.030
	.90	.435	24.597	.731	1.498	.193	.186	.258	.063	-.136	.058
	1.00	.415	.000	.728	1.495	*	.165	.314	*	*	.070
2.00	*	.000	∞	.707	1.415	*	*	*	*	*	.000

8 References

[1] Adler, R., Feldman, R. and Taqqu, M. (eds.) (1998), *A Practical Guide to Heavy Tailed Data*, Birkhäuser, Boston, MA.

[2] Akgiray, V. and Lamoureux, C. G. (1989), Estimation of stable law parameters: a comparative study, *J. Bus. Econ. Stat.* **7**:85-93.

[3] Beirlant, J., Vynckier, P. and Teugels, J. L. (1996), Tail index estimation, Pareto quantile plots and regression diagnostics, *JASA* **91**:1659-1667.

[4] Brant, R. (1984), Approximate likelihood and probability calculations based on transforms, *Ann. Stat.* **12**:989-1005.

[5] Brorsen, B. W. and Yang, S. R. (1990), Maximum likelihood estimates of symmetric stable distribution parameters, *Commun. Stat: Simul.* **19**:1459-1464.

[6] Brown, B. M. and Hettmansperger, T. P. (1996), Normal scores, normal plots and tests for normality, *JASA* **91**:1668-1675.

[7] Buckle, D. J. (1995), Bayesian inference for stable distributions, *JASA* **90**:605-613.

[8] Chambers, J. M., Mallows, C. and Stuck, B. W. (1976), A method for simulating stable random variables, *JASA* **71**: 340-344.

[9] Cheng, B.N., and Rachev, S.T. (1995), Multivariate stable future prices, *Math. Finance* **5**:133-153.

[10] Christoph, G. and Wolf, W. (1993), *Convergence Theorems with a Stable Limit Law*, Akademie Verlag, Berlin.

[11] DuMouchel, W. H. (1971), *Stable Distributions in Statistical Inference*, Ph.D. dissertation, Department of Statistics, Yale University.

[12] DuMouchel, W. H. (1973a), On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution, *Ann. Stat.* **1**:948-957.

[13] DuMouchel, W. H. (1973b), Stable distributions in statistical inference. 1: Symmetric stable distributions compared to other symmetric long-tailed distributions, *JASA* **68**:469-477.

[14] DuMouchel, W. H. (1975), Stable distributions in statistical inference. 2: Information from stably distributed samples, *JASA* **70**:386-393.

[15] DuMouchel, W. H. (1983), Estimating the stable index α in order to measure tail thickness: A critique, *Ann. Stat.* **11**:1019-1031.

[16] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, Heidelberg.

[17] Fama, E. (1965), The behavior of stock prices, *J. of Business* **38**:34-105.

[18] Feller, W. (1975), *An Introduction to Probability Theory and Its Applications, Vol. II*, Second Edition. John Wiley and Sons, New York, New York.

[19] Feuerverger, A. and McDunnough, P. (1981a), On the efficiency of empirical characteristic function procedures. *J. Roy. Stat. Soc, Ser. B* **43**:20-27.

[20] Feuerverger, A. and McDunnough, P. (1981b), On efficient inference in symmetric stable laws and processes. In M. Csörgö, Dawson, D. A., Rao, N. J. K. and Saleh, A. K. (Editors) *Statistics and Related Topics*, pp. 109-122.

[21] Fofack, H. and Nolan, J. P. (1999), Tail behavior, modes and other characteristics of stable distributions, *Extremes* **2**:39-58.

[22] Haan, L. de and Ronde, J. de (1998). Sea and wind: multivariate extremes at work, *Extremes* **1**:7-46.

[23] Hoffmann-Jørgensen, J. (1994), Stable densities, *Theory Prob. Appl.* **38**:350-355.

[24] Janicki, A. and Weron, A. (1994), *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*, Marcel Dekker, New York, New York.

[25] Kogon, S. M. and Williams, D. B. (1998), Characteristic function based estimation of stable parameters. In Adler, R., Feldman, R. and Taqqu, M. (eds.) *A Practical Guide to Heavy Tailed Data*, Birkhäuser, Boston, MA, pp. 311-335.

[26] Koutrouvelis, I. A. (1980), Regression type estimation of the parameters of stable laws, *JASA* **75**: 918-928.

[27] Koutrouvelis, I. A. (1981), An iterative procedure for the estimation of the parameters of stable laws, *Commun. Stat.: Simul.* **10**:17-28.

[28] Lévy, P. (1924), Théorie des erreurs la loi de Gauss et les lois exceptionelles, *Bulletin de la Société de France* **52**:49-85.

[29] Mandelbrot, B. (1963), The variation of certain speculative prices, *J. of Business* **26**:394-419.

[30] McCulloch, J. H. (1986), Simple consistent estimators of stable distribution parameters, *Commun. Statist.: Simul.* **15**: 1109-1136.

[31] McCulloch, J. H. (1996), Financial applications of stable distributions, *Statistical Methods in Finance, Handbook of Statistics, Vol. 14*, Maddala, G. S. and Rao, C. R. (eds.) North-Holland, NY.

[32] McCulloch, J. H. (1997), Measuring Tail Thickness to Estimate the Stable Index Alpha: A Critique, *J. Bus. Econ. Stat.* **15**: 74-81.

[33] McCulloch, J. H. (1998), Linear regression with stable disturbances. In Adler, R., Feldman, R. and Taqqu, M. (eds.) *A Practical Guide to Heavy Tailed Data*, Birkhäuser, Boston, MA. pp. 359-376.

[34] Michael, J. R. (1983), The stabilized probability plot, *Biometrika* **70**:11-17.

[35] Nikias, C. L. and Shao, M. (1995), *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley, New York, New York.

[36] Nolan, J. P. (1997), Numerical computation of stable densities and distribution functions, *Commun. Stat.: Stochastic Models* **13**: 759-774.

[37] Nolan, J. P. (1998), Parameterizations and modes of stable distributions, *Stat. Prob. Letters* **38**: 187-195.

[38] Nolan, J. P., Panorska, A. K. and McCulloch, J. H. (1996), Estimation of stable spectral measures, to appear in *Math. Computer Modelling*.

[39] Nolan, J. P. and Panorska, A. K. (1997), Data analysis for heavy tailed multivariate samples, *Commun. Stat.: Stochastic Models* **13**: 687-702.

[40] Nolan, J. P. (1999), Fitting data and assessing goodness of fit with stable distributions. In *Proceedings of the Conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, American University, Washington, DC, June 3-5, 1999.

[41] Paulson, A. S., Holcomb, E. W., and Leitch, R. (1975), The estimation of the parameters of the stable laws, *Biometrika* **62**: 163-170.

[42] Pierce, R. D. (1997), Application of the positive α stable distribution. In *Signal Processing Workshop on Higher-Order Statistics (SPW-HOS '97)*, IEEE Press, Piscataway, N. J., pp. 420-424.

[43] Press, S. J. (1972), Estimation in univariate and multivariate stable distributions, *JASA* **67**: 842-846.

[44] Samorodnitsky, G. and Taqqu, M. S. (1994), *Stable Non-Gaussian Random Processes*, Chapman and Hall, New York, New York.

[45] Stuck, B. W. and Kleiner, B. (1974), A statistical analysis of telephone noise, *Bell Syst. Tech. J.* **53**:1263-1320.

[46] Willinger, W., Taqqu, M., Sherman, R. and Wilson, D. (1995), Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. In *Proceedings of the ACM/SIGCOMM '95*, pp. 100-113.

[47] Zolotarev, V. M. (1986), *One-Dimensional Stable Distributions, Amer. Math. Soc. Transl. of Math. Monographs, Vol. 65.* Amer. Math. Soc., Providence, RI. (Transl. of the original 1983 Russian)

[48] Zolotarev, V. M. (1995), On representation of densities of stable laws by special functions, *Theory Probab. Appl.* **39**:354-362.

John P. Nolan
Department of Mathematics and Statistics
American University
Washington, DC 20016-8050, USA

E-mail: jpnolan@american.edu