Modeling financial data with stable distributions

John P. Nolan
Department of Mathematics and Statistics
American University

Revised 23 May 2002, Corrections 23 March 2005, Printed April 26, 2005

Abstract

Stable distributions are a class of probability distributions that allow heavy tails and skewness. In addition to theoretical reasons for using stable laws, they are a rich family that can accurately model different kinds of financial data. We review the basic facts, describe programs that make it practical to use stable distributions, and give examples of these distributions in finance. A non-technical introduction to multivariate stable laws is also given.

1 Basic facts about stable distributions

Stable distributions are a class of probability laws that have intriguing theoretical and practical properties. Their applications to financial modeling comes from the fact that they generalize the normal (Gaussian) distribution and allow heavy tails and skewness, which are frequently seen in financial data. In this chapter, we focus on the basic definition and properties of stable laws, and show how they can be used in practice. We give no proofs; interested readers can find these in Zolotarev (1983), Samorodnitsky and Taqqu (1994), Janicki and Weron (1994), Uchaikin and Zolotarev (1999), Rachev and Mittnik (2000) and Nolan (2003).

The defining characteristic, and reason for the term *stable*, is that they retain their shape (up to scale and shift) under addition: if $X, X_1, X_2, \ldots, X_n$ are independent, identically distributed stable random variables, then for every n

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} c_n X + d_n \tag{1}$$

for some constants $c_n > 0$ and d_n. The symbol $\stackrel{d}{=}$ means equality in distribution, i.e. the right and left hand sides have the same distribution. The law is called *strictly stable* if $d_n = 0$ for all n. Some authors use the term *sum stable* to emphasize the stability under addition and to distinguish it from other concepts, e.g. max-stable, min-stable, etc. The normal distributions satisfy this property: the sum of normals is normal. Likewise the Cauchy laws and the Lévy laws (see below) satisfy this property. The class of all laws that satisfy (1) is described by four parameters, which we call $(\alpha, \beta, \gamma, \delta)$, see Figure 1 for some density graphs. In general, there are no closed form formulas for stable densities f and cumulative distribution functions F, but there are now reliable computer programs for working with these laws.

The parameter α is called the *index* of the law or the *index of stability* or *characteristic exponent* and must be in the range $0 < \alpha \leq 2$. The constant c_n in (1) must be of the form $n^{1/\alpha}$. The parameter β is called the *skewness* of the law, and must be in the range $-1 \leq \beta \leq 1$. If $\beta = 0$, the distribution is symmetric, if $\beta > 0$ it is skewed toward the right, if $\beta < 0$, it is skewed toward the left. The parameters α and β determine the shape of the distribution. The parameter γ is a scale parameter, it can be any positive number. The parameter δ is a location parameter, it shifts the distribution right if $\delta > 0$, and left if $\delta < 0$.

A confusing issue with stable parameters is that there are multiple definitions of what the parameters mean. There are at least 10 different definitions of stable parameters, see Nolan (2003). The reader should be careful in reading the literature and verify what parameterization is being used. We will describe two different parameterizations, which we denote by $\mathbf{S}(\alpha, \beta, \gamma, \delta_0; 0)$ and $\mathbf{S}(\alpha, \beta, \gamma, \delta_1; 1)$. The first is what we will use in all our applications, because it has better numerical behavior and intuitive meaning. The second parameterization is more commonly used in the literature, so it is important to understand it. The parameters α, β and γ have the same meaning in the two parameterizations, only the location parameter is different. To distinguish between the two, we will sometimes use a subscript to indicate which parameterization is being used: δ_0 for the location parameter in the $\mathbf{S}(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization and δ_1 for the location parameter in the $\mathbf{S}(\alpha, \beta, \gamma, \delta_1; 1)$ parameterization.

Definition 1 A random variable X is $\mathbf{S}(\alpha, \beta, \gamma, \delta_0; 0)$ if it has characteristic function

$$\begin{aligned} &E \exp(iuX) \qquad\qquad (2) \\ &= \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left[1 + i\beta (\tan \frac{\pi\alpha}{2})(\operatorname{sign} u)(|\gamma u|^{1-\alpha} - 1)\right] + i\delta u\right) & \alpha \neq 1 \\ \exp\left(-\gamma |u| \left[1 + i\beta \frac{2}{\pi} (\operatorname{sign} u) \ln(\gamma |u|)\right] + i\delta u\right) & \alpha = 1. \end{cases} \end{aligned}$$

Definition 2 A random variable X is $\mathbf{S}(\alpha, \beta, \gamma, \delta_1; 1)$ if it has characteristic function

$$E \exp(iuX) = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left[1 - i\beta (\tan \frac{\pi\alpha}{2})(\operatorname{sign} u)\right] + i\delta u\right) & \alpha \neq 1 \\ \exp\left(-\gamma |u| \left[1 + i\beta \frac{2}{\pi} (\operatorname{sign} u) \ln |u|\right] + i\delta u\right) & \alpha = 1. \end{cases} \qquad (3)$$

The location parameters are related by

$$\delta_0 = \begin{cases} \delta_1 + \beta\gamma \tan \frac{\pi\alpha}{2} & \alpha \neq 1 \\ \delta_1 + \beta \frac{2}{\pi} \gamma \ln \gamma & \alpha = 1 \end{cases}, \qquad \delta_1 = \begin{cases} \delta_0 - \beta\gamma \tan \frac{\pi\alpha}{2} & \alpha \neq 1 \\ \delta_0 - \beta \frac{2}{\pi} \gamma \ln \gamma & \alpha = 1 \end{cases} \qquad (4)$$

Note that if $\beta = 0$, the parameterizations coincide. When $\alpha \neq 1$ and $\beta \neq 0$, the parameterizations differ by a shift $\beta\gamma \tan \frac{\pi\alpha}{2}$, which gets infinitely large as

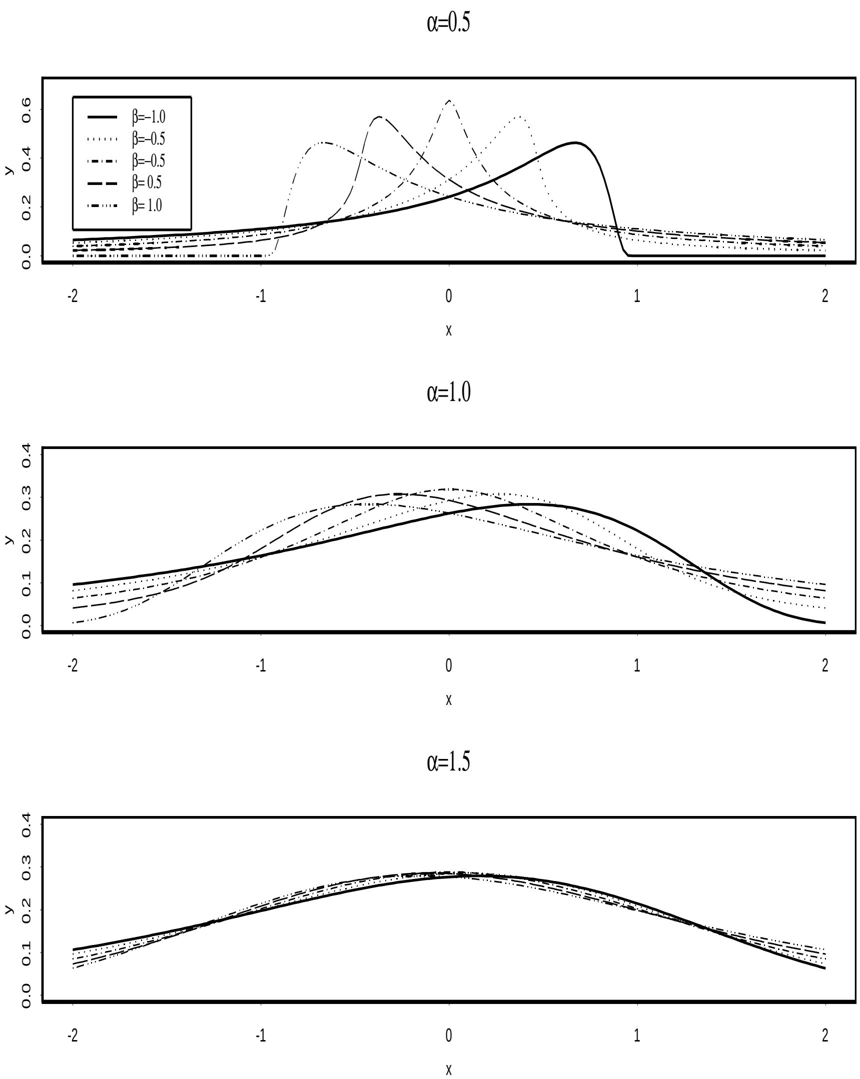

Figure 1: Standardized stable densities for different α and β. The top graph includes a Lévy(1,-1)=$\mathbf{S}(1/2,1,1,0;0) = \mathbf{S}(1/2,1,1,-1;1)$ graph and the middle graph includes a Cauchy(1,0)=$\mathbf{S}(1,0,1,0;0) = \mathbf{S}(1,0,1,0;1)$ graph.

$\alpha \to 1$. In particular, the mode of a $\mathbf{S}(\alpha, \beta, \gamma, \delta_1; 1)$ density tends toward ∞ (if $\operatorname{sign}(\alpha - 1)\beta > 0$) or $-\infty$ (otherwise) as $\alpha \to 1$. When α is near 1, computing stable densities and cumulatives in this range is numerically difficult and estimating parameters is unreliable. When $\alpha = 1$, the 0-parameterization is a simple scale family, but the 1-parameterization is not. From the applied point of view, it is preferable to use the $\mathbf{S}(\alpha, \beta, \gamma, \delta_0; 0)$ parameterization, which is jointly continuous in all four parameters. The arguments for using the $\mathbf{S}(\alpha, \beta, \gamma, \delta_1; 1)$ parameterization are historical and algebraic simplicity. It seems unavoidable that both parameterizations will be used, so users of stable distributions should know both and state clearly which they are using.

There are three cases where one can write down closed form expressions for the density and verify directly that they are stable - normal, Cauchy and Lévy distributions.

Example 1 Normal or Gaussian distributions. $X \sim N(\mu, \sigma^2)$ if it has a density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

Gaussian laws are stable with $\alpha = 2$ and $\beta = 0$; more precisely $N(\mu, \sigma^2) = \mathbf{S}(2, 0, \sigma/\sqrt{2}, 0; 0) = \mathbf{S}(2, 0, \sigma/\sqrt{2}, 0; 1)$.

Example 2 Cauchy distributions. $X \sim \text{Cauchy}(\gamma, \delta)$ if it has density

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\delta)^2} \quad -\infty < x < \infty.$$

Cauchy laws are stable with $\alpha = 1$ and $\beta = 0$; more precisely, $\text{Cauchy}(\gamma, \delta) = \mathbf{S}(1, 0, \gamma, \delta; 0) = \mathbf{S}(1, 0, \gamma, \delta; 1)$.

Example 3 Lévy distributions. $X \sim \text{Lévy}(\gamma, \delta)$ if it has density

$$f(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x-\delta)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\delta)}\right), \quad \delta < x < \infty.$$

These are stable with $\alpha = 1/2$, $\beta = 1$; $\text{Lévy}(\gamma, \delta) = \mathbf{S}(1/2, 1, \gamma, \gamma + \delta; 0) = \mathbf{S}(1/2, 1, \gamma, \delta; 1)$.

The graphs in Figure 1 show several qualitative features of stable laws. First, stable distributions have densities and are unimodal. These facts are not obvious: since there is no general formula for stable densities, indirect arguments must be used and it is quite involved to prove unimodality. Second, the $-\beta$ curve is a reflection of the β curve. Third, when α is small, the skewness is significant, when α is near 2, the skewness parameter matters less and less. The support of a stable density is either all of $(-\infty, \infty)$ or a half-line. The latter case occurs if and only if ($0 < \alpha < 1$ and $\beta = \pm 1$). More precisely, the support of density $f(x|\alpha,\beta,\gamma,\delta;k)$ for a $\mathbf{S}(\alpha,\beta,\delta,\gamma;k)$ law is

$$\begin{cases} [\delta-\gamma\tan\frac{\pi\alpha}{2},\infty) & \alpha<1,\beta=+1,k=0 \\ (-\infty,\delta+\gamma\tan\frac{\pi\alpha}{2}] & \alpha<1,\beta=-1,k=0 \\ [\delta,\infty) & \alpha<1,\beta=+1,k=1 \\ (-\infty,\delta] & \alpha<1,\beta=-1,k=1 \\ (-\infty,\infty) & \text{otherwise.} \end{cases}$$

In particular, to model a positive distribution, a $\mathbf{S}(\alpha,1,\delta,0;1)$ distribution with $\alpha<1$ is used.

When $\alpha=2$, the normal law has light tails and all moments exist. Except for the normal law, all stable laws have heavy tails with an asymptotic power law (Pareto) decay. The term *stable Paretian* distributions is used to distinguish the $\alpha<2$ cases from the normal case. For $X\sim\mathbf{S}(\alpha,\beta,1,0;0))$ with $0<\alpha<2$ and $-1<\beta\le 1$, then as $x\to\infty$,

$$\begin{aligned} P(X>x) &\sim c_\alpha(1+\beta)x^{-\alpha} \\ f(x|\alpha,\beta;0) &\sim \alpha c_\alpha(1+\beta)x^{-(\alpha+1)} \end{aligned}$$

where $c_\alpha=\Gamma(\alpha)\left(\sin\frac{\pi\alpha}{2}\right)/\pi$. When $\beta=-1$, the right tail decays faster than any power. The left tail behavior is similar by the reflection property mentioned above.

One consequence of these heavy tails is that only certain moments exist. This is not a property restricted to stable laws: any distribution with power law decay will not have certain moments. When $\alpha<2$, it can be shown that the variance does not exist and that when $\alpha\le 1$, the mean does not exist. If we use fractional moments, then the p^{th} absolute moment $E|X|^p=\int|x|^p f(x)dx$ exists if and only if $p<\alpha$. We stress that this is a population moment, and by definition it is finite when the integral just above converges. If the tails are too heavy, the integral will diverge. In contrast, the sample moments of all orders will exist: one can always compute the variance of a sample. The problem is that it does not tell you much about stable laws because the sample variance does not converge to a well-defined population moment (unless $\alpha=2$).

If X,X_1,X_2 are i.i.d. stable, then for any $a,b>0$,

$$aX_1+bX_2\overset{d}{=}cX+d,$$

for some $c>0,-\infty<d<\infty$. This condition is equivalent to (1) and can be taken as a definition of stability. More generally, linear combinations of independent stable laws with the same α are stable: if $X_j\sim\mathbf{S}(\alpha,\beta_j,\gamma_j,\delta_j;k)$ for $j=1,\ldots,n$, then

$$a_1X_1+a_2X_2+\cdots+a_nX_n\sim\mathbf{S}(\alpha,\beta,\gamma,\delta;k) \tag{5}$$

where $\beta=\left(\sum_{j=1}^n\beta_j(\operatorname{sign}a_j)|a_j\gamma_j|^\alpha\right)/\sum_{j=1}^n|a_j\gamma_j|^\alpha$, $\gamma^\alpha=\sum_{j=1}^n|a_j\gamma_j|^\alpha$, and

$$\delta=\begin{cases}\sum\delta_j+\gamma\beta\tan\frac{\pi\alpha}{2} & k=0,\alpha\ne 1 \\ \sum\delta_j+\beta\frac{2}{\pi}\gamma\ln\gamma & k=0,\alpha=1 \\ \sum\delta_j & k=1.\end{cases}$$

This is a generalization of (1): it allows different skewness, scales and locations in the terms. It is essential that all the αs are the same: adding two stable random variables with different αs does not give a stable law.

2 Appropriateness of stable models

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) and Uchaikin and Zolotarev (1999) for these and other examples. The second reason is the Generalized Central Limit Theorem, see below, which states that the only possible non-trivial limit of normalized sums of independent identically distributed terms is stable. It is argued that some observed quantities are the sum of many small terms, e.g. the price of a stock, 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 to justify the use of stable models. Examples in finance and economics are given in Mandelbrot (1963), Fama (1965), Roll (1970), Embrechts, Klüppelberg, and Mikosch (1997), and Rachev and Mittnik (2000). Such data sets are poorly described by a Gaussian model, some can be well described by a stable distribution.

The classical Central Limit Theorem says that the normalized sum of independent, identical terms with a finite variance converges to a normal distribution. The *Generalized Central Limit Theorem* shows that if the finite variance assumption is dropped, the only possible resulting limits are stable. Let X_1, $X_2, X_3, \ldots$ be a sequence of independent, identically distributed random variables. There exists constants $a_n > 0$, b_n and a non-degenerate random variable Z with

$$a_n(X_1 + \cdots X_n) - b_n \xrightarrow{d} Z \tag{6}$$

if and only if Z is stable. A random variable X is in the *domain of attraction* of Z if there exists constants $a_n > 0$, b_n such that (6) holds when $X_1, X_2, X_3, \ldots$ are independent identically distributed copies of X.

The Generalized Central Limit Theorem says that the only possible distributions with a domain of attraction are stable. Characterizations of distributions in the domain of attraction of a stable law are in terms of tail probabilities. The simplest is: if X is a random variable with $x^\alpha P(X > x) \to c^+ \geq 0$ and $x^\alpha P(X < -x) \to c^- \geq 0$, with $c^+ + c^- > 0$ for some $0 < \alpha < 2$ as $x \to \infty$, then X is in the domain of attraction of an α-stable law. In this case, the scaling constants must be of the form $a_n = an^{-1/\alpha}$.

Even if we accept that large data sets have heavy tails, is it ever reasonable to use a stable model? One of the arguments against using stable models is that they have infinite variance, which is inappropriate for real data that have bounded range. However, bounded data are routinely modeled by normal distributions which have infinite support. The only justification for this is that the normal distribution gives a usable description of the shape of the distribution, even though it is clearly inappropriate on the tails for any problem with naturally bounded data. The same justification can be used for stable models: does a stable fit gives an accurate description of the shape of the distribution? The variance is one measure of spread; the scale γ in a stable model is another. Perhaps practioners are so used to using the variance as *the* measure of spread, that they automatically retreat from models without a variance. The parameters δ and γ can play the role of the scale and location usually played by the mean and variance. For the normal distribution, the first and second moment completely specify the distribution; for most distributions they do not.

We propose that the practitioner approach this dispute as an agnostic. The fact is that until recently we have not really been able to compare data sets to a proposed stable model. The next Section shows that estimation of all four stable parameters is feasible and that there are methods 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.

3 Computation, simulation, estimation and diagnostics

Until recently, it was difficult to use stable laws in practical problems because of computational difficulties. Most of these difficulties have been resolved by the program STABLE[1], which can compute stable densities, cumulative distribution functions and quantiles. The basic method used in the program are described in Nolan (1997). Later improvements to the program include incorporating the Chambers, Mallows and Stuck (1976) method of simulating stable random variables, improved accuracy in the calculations, and estimation of stable parameters from data sets. Except for α close to 0, it is now possible to quickly and accurately work with stable distributions. We will not discuss details of these programs here, but will focus on the practical problems of estimation and assessing goodness of fit.

The basic estimation problem for stable laws is to estimate the four parameters $(\alpha, \beta, \gamma, \delta)$ from an i.i.d. sample $X_1, X_2, \ldots, X_n$. Because of numerical problems with the 1-parameterization, we will always use the 0-parameterization in estimation. If desired, the parameter δ_1 can be estimated by using (4). There are several methods available for this basic estimation problem: a quan-

[1] The program STABLE is available at `www.mathstat.american.edu` and following the "Faculty" link to the author's homepage.

tile method of McCulloch (1986), a fractional moment method of Ma and Nikias (1995), sample characteristic function (SCF) method of Kogon-Williams (1998) based on ideas of Koutrouvelis, and maximum likelihood (ML) estimation of DuMouchel (1971) and Nolan (2001). These methods have been compared in a large simulation study, Ojeda (2001), who found that the ML estimates are almost always more accurate, with the SCF estimates next best, followed by the quantile method, and finally the moment method. The ML method has the added advantage that one can give large sample confidence intervals for the parameters, based on numerical computations of the Fisher information matrix.

Perhaps just as important as methods of estimation, are diagnostics for assessing the fit. While a Kolmogorov-Smirnov goodness-of-fit test statistic can be computed, giving a correct significance level to such a test when comparing a data set to a fitted distribution is an involved problem. However, one can adapt standard exploratory data analysis graphical techniques to informally evaluate the closeness of a stable fit. We have found that comparing smoothed data density plots to a proposed fit gives a good sense of how good the fit is near the center of the data. P-P plots allow a comparison over the range of the data. For technical reasons we recommend the "variance stabilized" P-P plot of Michael (1983). We found Q-Q plots not as satisfactory for comparing heavy tailed data to proposed fit. One reason for this is visual - by definition a heavy tailed data set will have many more extreme values than a typical sample from finite variance population. This forces a Q-Q plot to be visually compressed, with a few extreme values dominating the plot. Also, the heavy tails imply that the extreme order statistics will have a lot of variability, and hence deviations from an ideal straight line Q-Q plot are hard to assess. The next section shows some examples of these techniques on financial data, more examples can be found in Nolan (1999) and (2001).

There are methods for more complicated estimation problems involving stable laws. For example, regression models with stable residuals have been described by McCulloch (1998) for the symmetric stable case and Ojeda (2001) for the general case. The problem analyzing time series with stable noise is discussed in Section II of Adler, Feldman and Taqqu (1998), in Nikias and Shao (1995), and in Rachev and Mittnik (2000). McCulloch (1996) and Rachev and Mittnik (2000) give methods of pricing options under stable models.

4 Applications to financial data

The first example we consider is the British Pound vs. German Mark exchange rate. The data set has daily exchange rates for the 16 year period from 2 January 1980 to 21 May 1996. The log of the successive exchange rates was computed as $y_t = \ln(x_{t+1}/x_t)$, yielding 4,274 y_t values. The ML parameter estimates with 95% confidence intervals are 1.495 ± 0.047 for α, -0.182 ± 0.085 for β, 0.00244 ± 0.00008 for γ and 0.00019 ± 0.00013 for δ_0. Figure 2 shows a P-P plot and density for the data vs. the stable fit. The third curve in the density plot is the normal/Gaussian fit to the data.

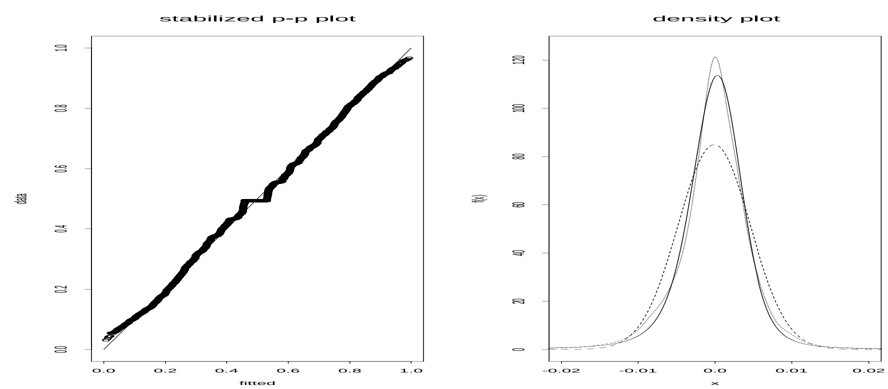

Figure 2: P-P plot and density plot for Pound vs. Mark exchange rate data. On the density plot, the dotted curve is the smoothed data, the solid curve is the stable fit, the dashed curve is a normal fit.

The next example is another exchange rate one, this time from a developing country. This data set consists of monthly exchange rates between the US Dollar and the Tanzanian Shilling, from January 1975 to September 1997. The log of the successive exchange rates were computed as above for this monthly data, giving a data set with $n = 213$ points. The ML parameter estimates with 95% confidence intervals are 1.088 ± 0.185 for α, 0.112 ± 0.251 for β, 0.0300 ± 0.0055 for γ and 0.00501 ± 0.00621 for δ_0. The more extreme fluctuations of the Tanzanian Shilling exchange rate show up in the smaller estimate of α and in the larger estimate of γ. Figure 3 shows the diagnostics, with the third curve again showing a normal/Gaussian fit.

The third example is from the stock market. McCulloch (1997) analyzed 40 years of monthly stock price data from the Center for Research in Security Prices (CRSP). The data set is 480 values of the CRSP value-weighted stock index, including dividends and adjusted for inflation. The ML estimates with 95% confidence intervals are 1.855 ± 0.110 for α, -0.558 ± 0.615 for β, 2.711 ± 0.213 for γ, and 0.871 ± 0.424 for δ_0. Figure 4 shows the goodness of fit.

Stable distributions may be a useful tool in Value at Risk (VaR) calculations. The goal of VaR calculations is to assess the risk in an asset by estimating population quantiles. Stable distributions have two advantages over normal distributions: they can explicitly model both the heavier tails and asymmetry that are frequently found in financial data. Sometimes the normal distribution can give reasonable VaR estimates, because the sample variance is inflated by the extreme values in the sample. If one is lucky, the poor fitting normal distribution may approximate certain quantiles well, at the cost of poorly approximating other quantiles. Additionally, some practioners compensate for the heavy tail behavior by "adjusting" a normal quantile estimate by some empirical factor. If a stable distribution gives a more accurate fit to the sample, then it is more

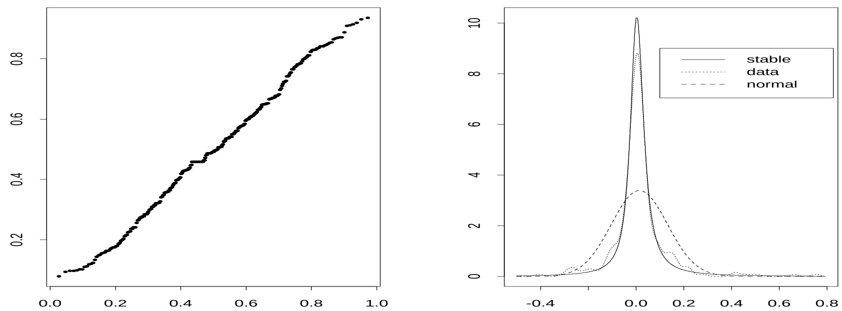

Figure 3: P-P plot and density plot for the Tanzanian Shilling/US Dollar exchange rate.

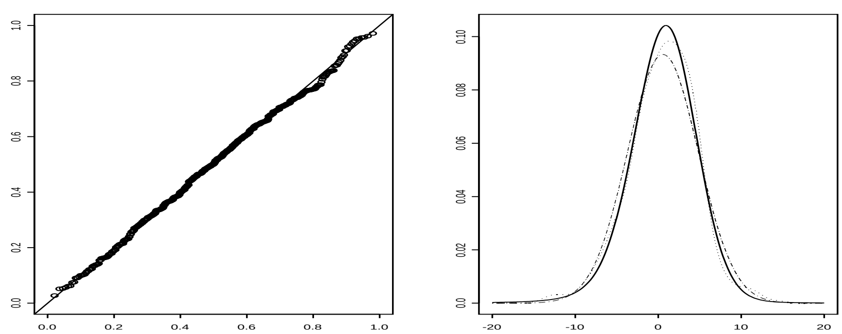

Figure 4: P-P plot and densities for the CRSP stock price data.

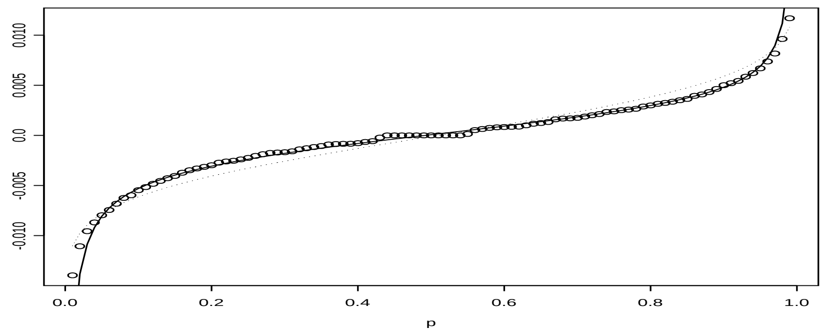

Figure 5: VaR comparison of quantiles for the Deutsch Mark exchange rate data (circles), quantiles predicted by the stable fit (solid line), and quantiles predicted by the normal distribution (dotted line).

likely to accurately predict the VaR values. In order to compare different fits, a plot like Figure 5 can be useful. It uses the Deutsch Mark exchange rate data (log ratios of successive values) described above.

5 Multivariate stable distributions

This section is about d-dimensional stable laws. Such random vectors will be denoted by $\mathbf{X} = (X_1, \ldots, X_d)$. The definition of stability is the same as in (1): for i.i.d. $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots$,

$$\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n \overset{d}{=} a_n \mathbf{X} + \mathbf{b}_n, \tag{7}$$

for some $a_n > 0$, and some vector $\mathbf{b}_n \in \mathbf{R}^d$. As in one dimension, an equivalent definition is that $a\mathbf{X}_1 + b\mathbf{X}_2 \overset{d}{=} c\mathbf{X} + \mathbf{d}$ for all $a, b > 0$.

If $\mathbf{X}$ is a stable random vector, then every one dimensional projection $\mathbf{u} \cdot \mathbf{X} = u_1 X_1 + \cdots + u_d X_d$ is a one dimensional stable random variable with the same index α for every $\mathbf{u}$. The phrase "jointly stable" is sometimes used to stress the fact that the definition forces all the components X_j to be univariate α-stable with one α. Conversely, suppose $\mathbf{X}$ is a random vector with the property that every one-dimensional projection $\mathbf{u} \cdot \mathbf{X}$ is one dimensional stable, e.g. $\mathbf{u} \cdot \mathbf{X} \sim \mathbf{S}(\alpha(\mathbf{u}), \beta(\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u}); 1)$. Then there is one α that is the index of all projections, i.e. $\alpha(\mathbf{u}) = \alpha$ is constant. If $\alpha \geq 1$, then $\mathbf{X}$ is stable. If $\alpha < 1$ and the location parameter function $\delta(\mathbf{u})$ and the vector of location parameters $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_d)$ of the components $X_1, X_2, \ldots, X_d$ (all in the 1 parameterization) are related by

$$\delta(\mathbf{u}) = \mathbf{u} \cdot \boldsymbol{\delta}, \tag{8}$$

then $\mathbf{X}$ is stable. The point here is that we have a way of determining joint stability in terms of univariate stability and, when $\alpha < 1$, equation (8).

We note that (8) holds automatically when $\alpha > 1$, so the condition is only required when $\alpha < 1$. Furthermore, (8) is necessary when $\alpha \neq 1$, so it cannot be dropped. There are examples, e.g. Section 2.2 of Samorodnitsky and Taqqu (1994), where $\alpha < 1$ and all one dimensional projections are stable, but (8) fails and $\mathbf{X}$ is not jointly stable.

One way of parameterizing multivariate stable distributions is to use the above results about one dimensional projections. For any vector $\mathbf{u} \in \mathbf{R}^d$, $\mathbf{u} \cdot \mathbf{X} \sim \mathbf{S}(\alpha, \beta(\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u}); k)$, $k = 0, 1$. Thus we know the (univariate) characteristic function of $\mathbf{u} \cdot \mathbf{X}$ for every $\mathbf{u}$, and hence the joint characteristic function of $\mathbf{X}$. Therefore α and the functions $\beta(\cdot)$, $\gamma(\cdot)$ and $\delta(\cdot)$ completely characterize the joint distribution. In fact, knowing these functions on the sphere $\mathbf{S}^d = \{\mathbf{u} \in \mathbf{R}^d : |\mathbf{u}| = 1\}$ characterizes the distribution.

The functions $\beta(\cdot)$, $\gamma(\cdot)$ and $\delta(\cdot)$ must satisfy certain regularity conditions. The standard way of describing multivariate stable distributions is in terms of a finite measure Λ on $\mathbf{S}^d$, called the spectral measure. Let $\mathbf{X} = (X_1, \ldots, X_d)$ be jointly stable, say

$$\mathbf{u} \cdot \mathbf{X} \sim \mathbf{S}(\alpha, \beta(\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u}); k), \quad k = 0, 1.$$

Then there exists a finite measure Λ on $\mathbf{S}^d$ and a location vector $\boldsymbol{\delta} \in \mathbf{R}^d$ with

$$\begin{aligned}
\gamma(\mathbf{u}) &= \left(\int_{\mathbf{S}^d} |\mathbf{u} \cdot \mathbf{s}|^\alpha \Lambda(d\mathbf{s}) \right)^{1/\alpha} \qquad (9) \\
\beta(\mathbf{u}) &= \frac{\int_{\mathbf{S}^d} |\mathbf{u} \cdot \mathbf{s}|^\alpha \operatorname{sign}(\mathbf{u} \cdot \mathbf{s}) \Lambda(d\mathbf{s})}{\int_{\mathbf{S}^d} |\mathbf{u} \cdot \mathbf{s}|^\alpha \Lambda(d\mathbf{s})} \\
\delta(\mathbf{u}) &= \begin{cases} \boldsymbol{\delta} \cdot \mathbf{u} & k = 1, \alpha \neq 1 \\ \boldsymbol{\delta} \cdot \mathbf{u} - \frac{2}{\pi} \int_{\mathbf{S}^d} (\mathbf{u} \cdot \mathbf{s}) \ln |\mathbf{u} \cdot \mathbf{s}| \Lambda(d\mathbf{s}) & k = 1, \alpha = 1 \\ \boldsymbol{\delta} \cdot \mathbf{u} + (\tan \frac{\pi\alpha}{2}) \beta(\mathbf{u}) \gamma(\mathbf{u}) & k = 0, \alpha \neq 1 \\ \boldsymbol{\delta} \cdot \mathbf{u} - \frac{2}{\pi} \int_{\mathbf{S}^d} (\mathbf{u} \cdot \mathbf{s}) \ln (\mathbf{u} \cdot \mathbf{s}) \Lambda(d\mathbf{s}) & \\ \qquad + \frac{2}{\pi} \beta(\mathbf{u}) \gamma(\mathbf{u}) \ln \gamma(\mathbf{u}) & k = 0, \alpha = 1. \end{cases}
\end{aligned}$$

Thus another way to parameterize is $\mathbf{X} \sim \mathbf{S}(\alpha, \Lambda, \boldsymbol{\delta}; k)$, $k = 0, 1$. If one knows Λ, then the above equations specify the parameter functions $\beta(\cdot)$, $\gamma(\cdot)$ and $\delta(\cdot)$. Going the other direction is more difficult. If one recognizes a certain form for the parameter functions, then one can specify the spectral measure. In the general case, one can numerically invert the map $\Lambda \to (\beta(\cdot), \gamma(\cdot), \delta(\cdot))$ to get a discrete approximation to Λ.

It is possible for $\mathbf{X}$ to be non-degenerate, but singular. For example, $\mathbf{X} = (X_1, 0)$ is formally a two dimensional stable distribution if X_1 is univariate stable, but it is supported on a one dimensional subspace. In what follows, we will always assume that $\mathbf{X}$ is non-singular that is, it has a density on $\mathbf{R}^d$. It can be shown that the following are equivalent: (i) $\mathbf{X}$ is nonsingular, (ii) $\gamma(\mathbf{u}) > 0$ for all non-zero $\mathbf{u} \in \mathbf{R}^d$, and (iii) span support$(\Lambda) = \mathbf{R}^d$.

For $\alpha \geq 1$, the support of non-singular stable $\mathbf{X}$ is all of $\mathbf{R}^d$. When $\alpha < 1$, it can be all of $\mathbf{R}^d$ or a cone, depending on the spectral measure. For A is a subset of $\mathbf{R}^d$, define CCH(A)= closed convex hull of A = closure of $\{\mathbf{x} = \mathbf{a}_1 b_1 + \cdots + \mathbf{a}_n b_n \in \mathbf{R}^d : \mathbf{a}_1, \ldots, \mathbf{a}_n \in A, b_1, \ldots, b_n \geq 0\}$. Note that we only take positive linear combinations of elements of A, so this is not generally the closed span of A. The translate of a cone is denoted by $\mathrm{CCH}(A) + \boldsymbol{\delta} = \{\mathbf{x} + \boldsymbol{\delta} : \mathbf{x} \in \mathrm{CCH}(A)\}$. Then the support of $\mathbf{X} \sim \mathbf{S}(\alpha, \Lambda, \delta; 1)$ is

$$\text{support } \mathbf{X} = \begin{cases} \mathrm{CCH}(\mathrm{support}(\Lambda)) + \boldsymbol{\delta} & \alpha < 1 \\ \mathbf{R}^d & \alpha \geq 1. \end{cases}$$

For example, in the two dimensional case, if the spectral measure is supported in the first quadrant, $\alpha < 1$, and $\boldsymbol{\delta} = 0$, then the support of the corresponding stable distribution is contained in the first quadrant, i. e. both components are positive.

The tail behavior of $\mathbf{X}$ is easiest to describe in terms of the spectral measure. It is best stated in polar form: let $A \subset \mathbf{S}^d$, then

$$\lim_{r \to \infty} \frac{P(\mathbf{X} \in \mathrm{CCH}(A), |\mathbf{X}| > r)}{P(|\mathbf{X}| > r)} = \frac{\Lambda(A)}{\Lambda(\mathbf{S}^d)}.$$

The tail behavior of the densities is more intricate. In the radially symmetric case, $f(\mathbf{x}) \sim c|\mathbf{x}|^{-(d+\alpha)}$ as $|\mathbf{x}| \to \infty$. In other cases, the tail behavior can have very different behavior in different directions. For example, in the bivariate independent case, the joint density factors: $f(x_1, x_2) = f_1(x_1) f_2(x_2)$. The one dimensional results above show $f(x, 0) \sim c_1 x^{-(1+\alpha)}$ along the x-axis, but $f(x, x) \sim c_2 x^{-2(1+\alpha)}$ along the diagonal. The general case is complicated, depending on the nature (discrete, continuous) and spread of the spectral measure.

We now give some examples of bivariate stable densities, see the next section for information on their computation. In all cases, the shift vector $\boldsymbol{\delta} = 0$.

Example 4 The first example uses $\alpha = 1.2$ and a discrete spectral measure with three unit point masses, distributed on the unit circle at angles $\pi/3$, π and $-\pi/3$. A plot of the density surface and level curves are given in Figure 6. The triangular spread of the spectral measure shows up in the triangular shape of the level curves. The contour plot reveals more about the shape of the surface, so the following examples will show only the contour plots.

Example 5 Figure 7 shows the contour plots of the independent components cases when $\alpha = 0.6, 1.6$ and $\beta = 0, 1$. Note that the upper right graph has $\alpha < 1$ and is supported in the first quadrant.

Example 6 Figure 8 shows a mix of different contours, mostly to show the range of possibilities. The upper left plot shows an elliptically contoured stable distribution with $\alpha = 1.5$ and "covariation matrix"

$$R = \begin{pmatrix} 1.0 & 0.7 \\ 0.7 & 1.0 \end{pmatrix}.$$

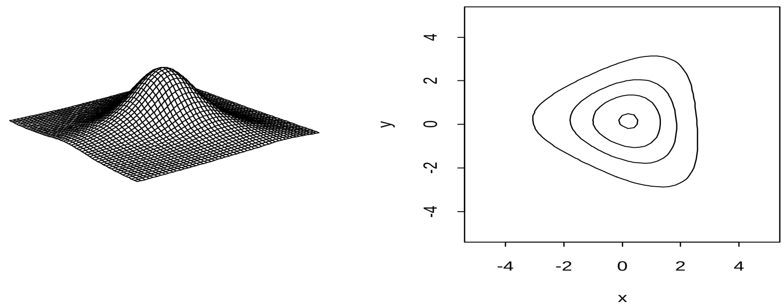

Figure 6: Density surface and level curves for "triangle" example of a bivariate stable law.

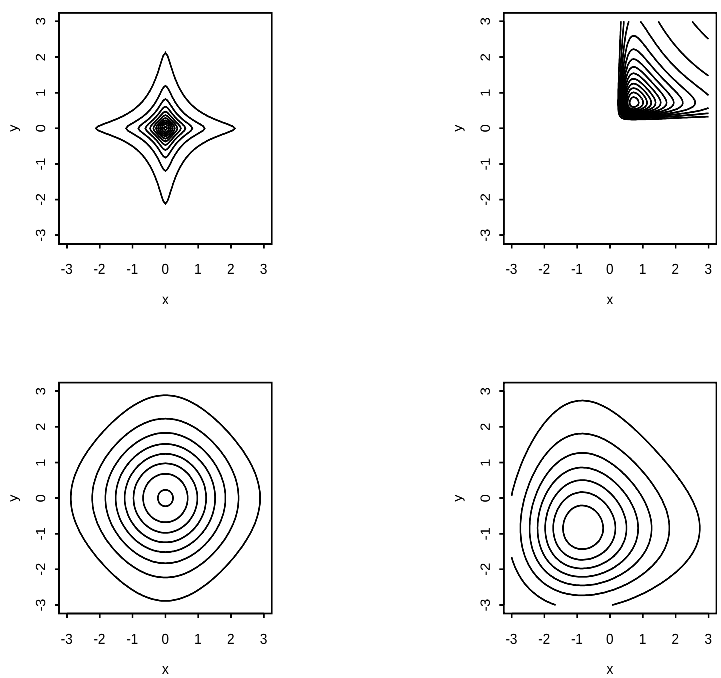

Figure 7: Contour plots for bivariate stable densities with independent $\mathbf{S}(\alpha, \beta, 1, 0; 1)$ components. The plots show $\alpha = 0.6$, $\beta = 0$ in upper left, $\alpha = 0.6$, $\beta = 1$ in upper right, $\alpha = 1.6$, $\beta = 0$ in lower left, and $\alpha = 1.6$, $\beta = 1$ in lower right.

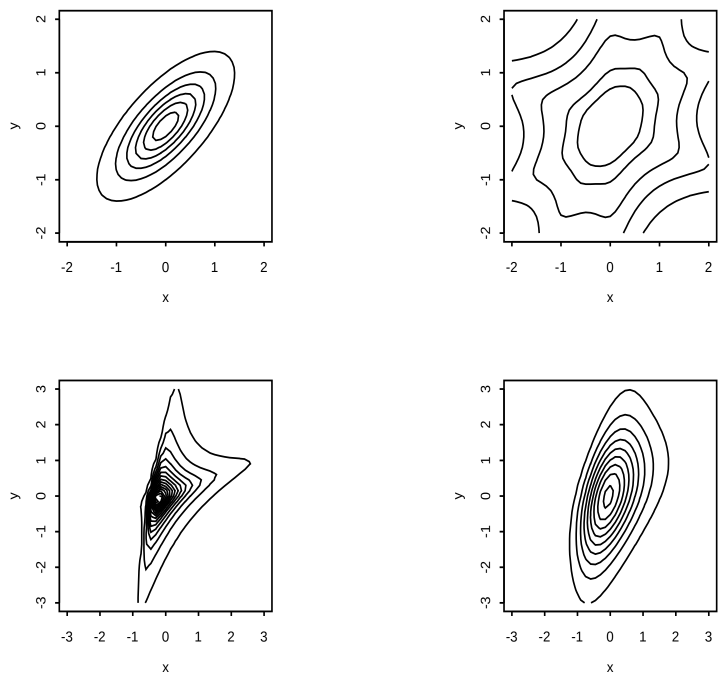

Figure 8: Contours of miscellaneous bivariate stable distributions.

The upper right plot shows a $\alpha = 0.8$ stable distribution with discrete spectral measure having point masses at angles $(-\pi/9, \pi/6, \pi/3, \pi/2)$ and uniform weight $\lambda_j = 0.3$. The lower left plot uses $\alpha = 0.7$ with a discrete spectral measure with point masses at angles $(\pi/9, 4\pi/9, 10\pi/9, 13\pi/9)$ of weight $(.75, 1, .25, 1)$. The lower right plot uses the same discrete spectral measure as the lower left, but with $\alpha = 1.5$.

There are some general statements that can be made about the qualitative behavior of multivariate stable densities. For fixed α, central behavior is determined by overall spread of the spectral measure: if the spectral mass is highly concentrated the density is close to singular, with large values near the center; if the spectral mass is more evenly spread around the sphere, the density is less peaked. On the tails, behavior is determined by the exact distribution of the spectral measure, with the contour lines bulging out in directions where the spectral measure is concentrated. This tail effect is more pronounced for small values of α, where distributions can be highly skewed, and becomes less pronounced as α approaches 2, where contours are all rounded into ellipses.

6 Multivariate computation, simulation, estimation and diagnostics

The computational problems are challenging, and not solved for general multivariate stable distributions. The problems are caused by the both the usual difficulties of working in d dimensions and by the complexity of the possible distributions: spectral measures are an uncountable set of "parameters". The graphs above were computed by the program MVSTABLE (available at the same web-site noted above), which only works in 2 dimensions and has limited accuracy. Density calculations are based on either numerically inverting the characteristic function as described in Nolan and Rajput (1995) or by numerically implementing the symmetric formulas in Abdul-Hamid and Nolan (1998).

One class of accessible models is when the spectral measure is discrete with a finite number of point masses:

$$\Lambda(\cdot) = \sum_{j=1}^{n} \lambda_j 1_{\{\cdot\}}(\mathbf{s}_j). \tag{10}$$

This class is dense in the space of all stable distributions: given an arbitrary spectral measure Λ_1, there is a concrete formula for n and a discrete spectral measure Λ_2 such that the densities of the corresponding stable densities are uniformly close on $\mathbf{R}^d$.

In the case of a discrete spectral measure, the parameter functions $\beta(\cdot)$, $\gamma(\cdot)$ and $\delta(\cdot)$ are computed as finite sums, rather than $(d-1)$-dimensional integrals, which makes all computations easier. It also makes simulation simple in an arbitrary dimension: $\mathbf{X} \sim \mathbf{S}(\alpha, \Lambda, \boldsymbol{\delta}; k)$ can be simulated by the vector sum

$$\mathbf{X} \stackrel{d}{=} \sum_{j=1}^{n} \lambda_j^{1/\alpha} Z_j \mathbf{s}_j + \boldsymbol{\delta},$$

where $Z_1, \ldots, Z_n$ are i.i.d. univariate $\mathbf{S}(\alpha, 1, 1, 0; k)$ random variables.

Another example where computations are more accessible is the elliptically contoured, or sub-Gaussian, stable distributions described in Section 8. Such densities are easier to compute and simulation is straightforward. Certain sub-stable distributions are also easy to simulate: if $\alpha < \alpha_1$, $\mathbf{X}$ is strictly α_1-stable and A is positive (α/α_1)-stable, then $A^{1/\alpha_1}\mathbf{X}$ is α-stable. Since sums and shifts of multivariate stables are also multivariate stable, one can combine these different classes to simulate a large class of multivariate stable laws.

There are several methods of estimating for multivariate stable distributions. If you know the distribution is isotropic (radially symmetric), then problem 4, pg. 44 of Nikias and Shao (1995) gives a way to estimate α and then the constant scale function/uniform spectral measure from fractional moments. In general one should let the data speak for itself, and see if the spectral measure Λ is constant. The general techniques involve some estimate of α and some estimate of the spectral measure $\hat{\Lambda} = \sum_{k=1}^{m} \lambda_k 1_{\{\cdot\}}(\mathbf{s}_k)$, $\mathbf{s}_k \in \mathbf{S}^d$. Rachev and

Xin (1993) and Cheng and Rachev (1995) use the fact that the directional tail behavior of multivariate stable distributions is Pareto, and base an estimate of Λ on this. Nolan, Panorska and McCulloch (2001) define two other estimates of Λ, one based on the joint empirical/sample ch. f. and one based on the one dimensional projections of the data.

Using the fact that one dimensional projections are univariate stable gives a way of assessing whether a multivariate data set is stable by looking at just one dimensional projections of the data. Fit projections in multiple directions using the univariate techniques described above, and see if they are well described by a univariate stable fit. If so, and if the α's are the same for every direction (and if $\alpha < 1$, the location parameters satisfy (8)), then a multivariate stable model is appropriate. We will illustrate this in examples below.

For the purposes of comparing two multivariate stable distributions, the parameter functions $(\alpha, \beta(\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u}))$ are more useful than Λ itself. This is because the distribution of $\mathbf{X}$ depends more on how Λ distributes mass around the sphere than exactly on the measure. Two spectral measures can be far away in the traditional total variation norm (e.g. one can be discrete and the other continuous), but their corresponding parameter functions and densities can be very close.

The diagnostics suggested for assessing stability of a multivariate data set are:

- Project the data in a variety of directions $\mathbf{u}$ and use the univariate diagnostics described in Section 3 on each of those distributions. Bad fits in any direction indicate that the data is not stable.

- For each direction $\mathbf{u}$, estimate the parameter functions $\alpha(\mathbf{u})$, $\beta(\mathbf{u})$, $\gamma(\mathbf{u})$, $\delta(\mathbf{u})$ by ML estimation. The plot of $\alpha(\mathbf{u})$ should be a constant, significant departures from this indicate that the data has different decay rates in different directions. (Note that $\gamma(\mathbf{t})$ will be a constant iff the distribution is isotropic.)

- Assess the goodness-of-fit by computing a discrete $\hat{\Lambda}$ by one of the methods above. Substitute the discrete $\hat{\Lambda}$ in (9) to compute parameter functions. If it differs from the one obtained above by projection, then either the data is not jointly stable, or not enough points were chosen in the discrete spectral measure approximation.

These techniques are illustrated in the next section.

7 Multivariate application

Here we will examine the joint distribution of the German Mark and the Japanese Yen. The data set is the one described above in the univariate example. We are interested in both assessing whether the joint distribution is bivariate stable and in estimating the fit.

Figure 9 shows a sequence of smoothed density, q-q plot and variance stabilized p-p plot for projections in 8 different directions: $\pi/2$, $\pi/3$, $\pi/4$, $\pi/6$,0,$-\pi/6$, $-\pi/4$, $-\pi/3$. (We restrict to the right half plane because projections in the left half plane are reflections of those in the right half plane). These projections are similar to Figure 2, in fact the fifth row of Figure 9 is exactly the same as Figure 2. Except on the extreme tails, the stable fit does a good job of describing the data.

The projection functions $\alpha(\mathbf{t})$, $\beta(\mathbf{t})$, $\gamma(\mathbf{t})$, and $\delta(\mathbf{t})$ were estimated and used to compute an estimate of the spectral measure using the projection method. The results are shown in Figure 10. It shows a discrete estimate of the spectral measure (with $m = 100$ evenly spaced point masses) in polar form, a cumulative plot of the spectral measure in rectangular form, and then four plots for the parameter estimates $(\alpha(\mathbf{t}), \beta(\mathbf{t}), \gamma(\mathbf{t}), \delta(\mathbf{t}))$. Also on the $\alpha(\mathbf{t})$ plot is a horizontal line showing the average value of all the estimated indices which is taken as the estimate of the common α that should come from a jointly stable distribution. The plots of $\beta(\mathbf{t})$ and $\gamma(\mathbf{t})$ also show the skewness and scale functions computed from the estimated spectral measure substituted into (9). These curves, which are based on a joint estimate of the spectral measure, are indistinguishable from the direct, separate estimates of the directional parameters.

The fitted spectral measure was used to plot the fitted bivariate density shown in Figure 11. The spread of the spectral measure is spiky, and masks a pattern that is more obvious in the density surface: the approximate elliptical contours of the fitted density. This suggests modeling the data by a sub-Gaussian stable distribution, a topic discussed in the next section.

Some comments on these plots. The polar plots of the spectral measure show a unit circle and lines connecting the points (θ_j, r_j), where $\theta_j = 2\pi(j-1)/m$ and $r_j = 1+(\lambda_j/\lambda_{max})$, where $\lambda_{max} = \max \lambda_j$. The polar plots are spiky, because we are estimating a discrete object. What should be looked at is the overall spread of mass, not specific spikes in the plot. In cases where the spectral measure is really smooth, it may be appropriate to smooth these plots out to better show it's true nature. In cases where the measure is discrete, i.e. the independent case, then one wants to emphasize the spikes. So there is no satisfactory general solution and we just plot the raw data.

Finally, most graphing programs will set vertical scale so that the graph fills the graph. This emphasizes minor fluctuations in the data that are not of practical significance. In the graphs below, the vertical scales for the parameter functions $\alpha(\mathbf{t})$, $\beta(\mathbf{t})$, $\gamma(\mathbf{t})$ are respectively [0,2], [-1,1], and [0,1.2 $\times \max \gamma(\mathbf{t})$]. These bounds show how the functions vary over their possible range. For $\delta(\mathbf{t})$, we used the bounds $[-1.2 \times \max|\delta(\mathbf{t})|, 1.2 \times \max|\delta(\mathbf{t})|]$, which visually exaggerates the changes in $\delta(\mathbf{t})$. A scale that depends on $\max \gamma(\mathbf{t})$ may be more appropriate.

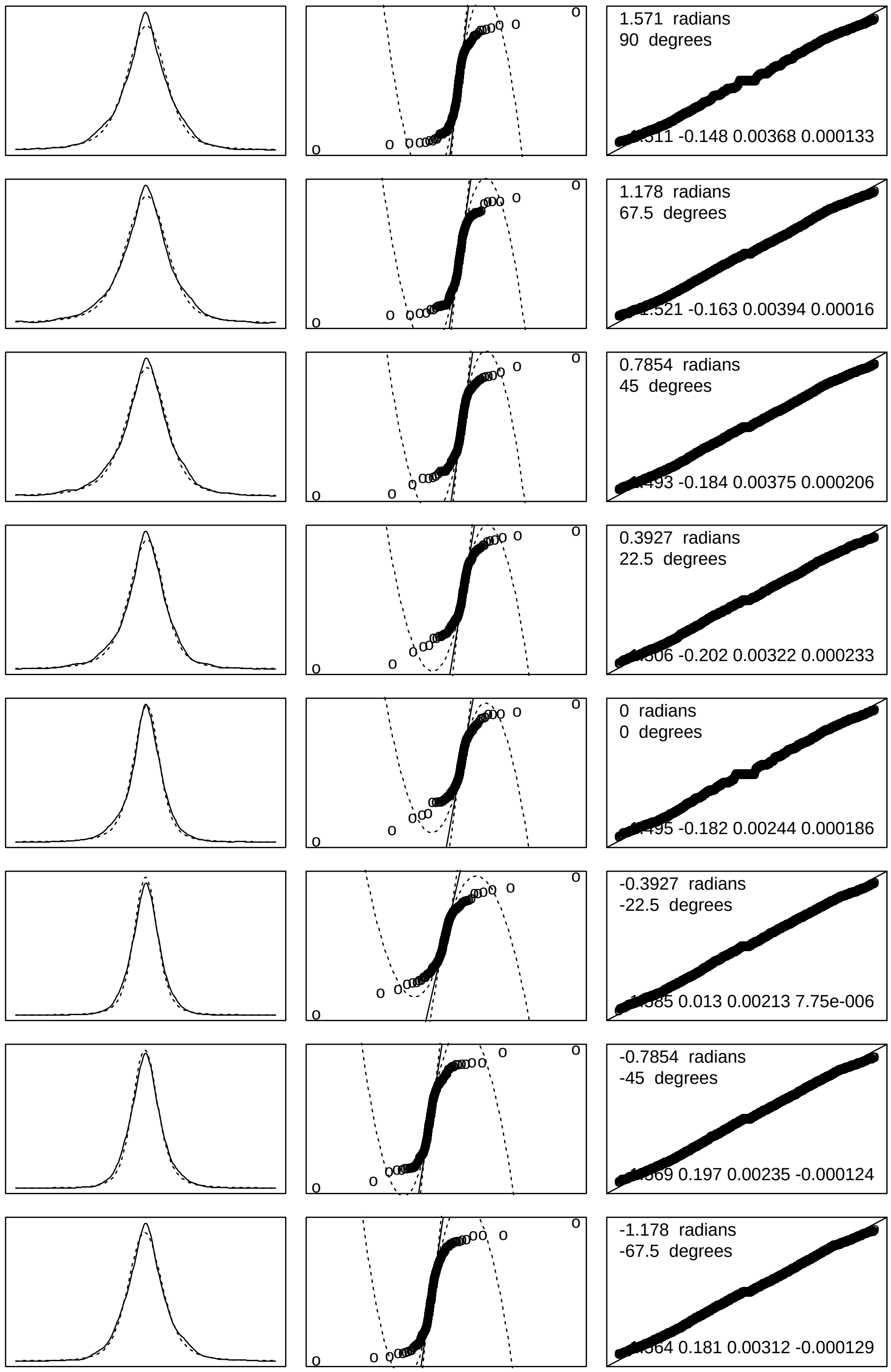

Figure 9: Projection diagnostics for the German Mark and Japanese Yen exchange rates.

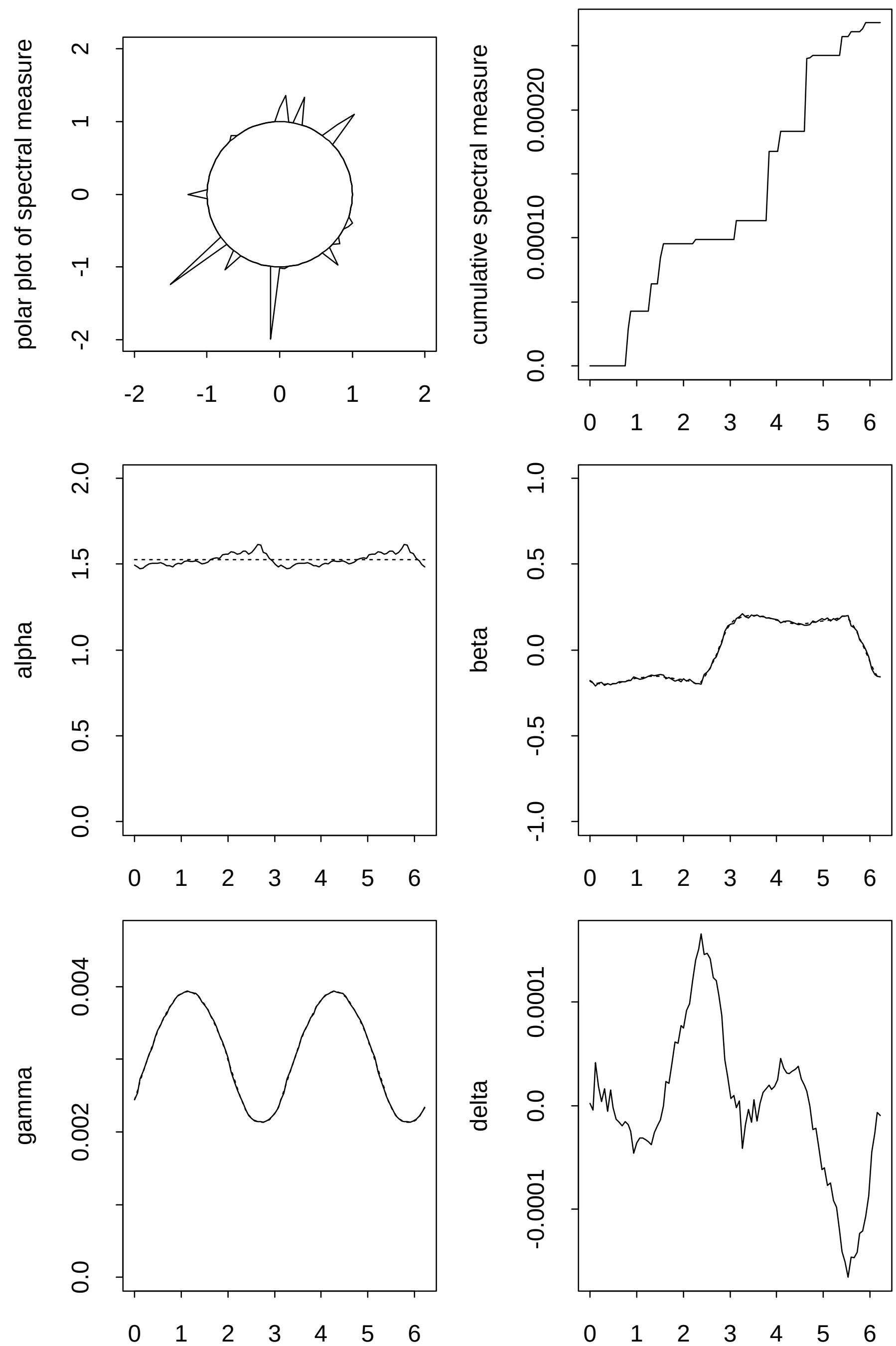

Figure 10: Estimation results for the German Mark and Japanese Yen exchange rates.

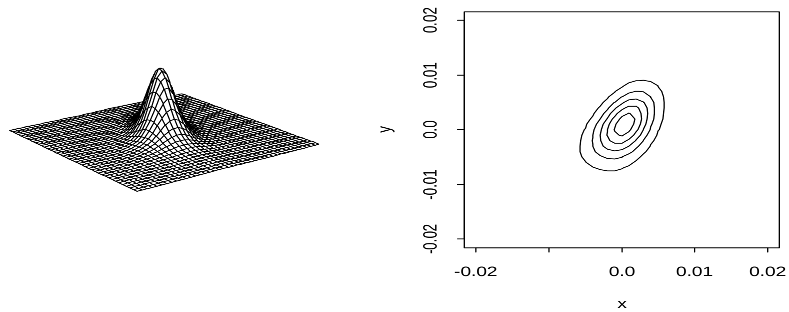

Figure 11: Estimated density surface and level curves for a bivariate stable fit to the German Mark and Japanese Yen exchange rates.

8 Classes of multivariate stable distributions

There may be cases where we believe that a multivariate sample has certain structure. If so, we can fit a stable model that takes this into account. This may give a more parsimonious fit to the model, especially if the data set is high dimensional. Below we fill focus on elliptically contoured distributions and see that it is computationally accessible. The idea here is to estimate an α and a matrix R so that the scale function is closely approximated by $\gamma(\mathbf{u}) = (\mathbf{u}R\mathbf{u})^{\alpha/2}$. The principle can be generalized to other special classes of distributions. Given some parametric model for the scale function $\gamma(\cdot)$, one can fit parameters, or use a nonparametric model (smoothing or loess) for the scale. Or, one can assume a special form of the spectral measure $\Lambda(\cdot)$, which determines the scale function $\gamma(\cdot)$. The methods of estimation described above do this implicitly, by assuming Λ is discrete as in (10). This can be adapted in many ways. If we assume the components of the data are independent, then we can only allow point masses at "poles", i.e. where the coordinate axes intersect the sphere. If we assume the spectral measure is concentrated on some smaller region, then one can allow point masses only in that region.

If we assume the spectral measure is continuous, then one can use some particular model for its density, say as a sum of terms like $\Lambda(d\mathbf{s}) = \sum_{k=1}^{n} \lambda_k(\mathbf{s})d\mathbf{s}$, where the density terms $\lambda_k(\cdot)$ in the sum have some accessible form. If the goal is a computationally accessible model, then an ad hoc approach may be useful. First compute a fit using a discrete spectral measure. If there are clearly defined point masses that are isolated, then include them and try to model the rest as an elliptical model, or using some spectral density.

Since the foreign exchange data seems to be approximately elliptically contoured, there may be interest in categorizing such stable distributions. The main practical advantage to this is that all d-dimensional elliptically contoured stable distributions are parameterized by α and a symmetric, positive definite

$d \times d$ matrix. Since the matrix is symmetric, there are a total of $1 + d(d+1)/2$ parameters. This is quite different from the general stable case, which involves an infinite dimensional spectral measure. Even a discrete approximating measure involves a much larger number of terms: if a "polar grid" is used with each of the angle directions divided up evenly with k subintervals, then there are k^{d-1} point masses to be estimated.

For $\mathbf{X}$ an non-singular symmetric α-stable random vector, the following are equivalent:

- $\mathbf{X}$ is elliptically contoured around the origin.
- $\mathbf{X}$ is sub-Gaussian, i.e. $\mathbf{X} \stackrel{d}{=} A^{1/2}\mathbf{G}$, where $A \sim \mathbf{S}(\alpha, 1, \gamma, 0; 1)$ and $\mathbf{G} \sim N(0, R)$.
- The characteristic function is $E \exp(i\mathbf{u} \cdot \mathbf{X}) = \exp(-(\mathbf{u}R\mathbf{u}^T)^{\alpha/2})$, for some symmetric, positive definite matrix R.

There is a "random volatility" interpretation of sub-Gaussian distributions. Think of $\mathbf{G}$ as an underlying multivariate normal model for the returns on d assets with random scale $A^{1/2}$. In general, A can be any positive random variable, but the product will be α-stable only when A is itself a positive $(\alpha/2)$-stable random variable.

Computations with elliptically contoured stable distributions is much simpler than the general stable case. All calculations are essentially reduced to one-dimensional problems: the linear transformation $\mathbf{Y} = R^{-1/2}\mathbf{X}$ gives a radially symmetric distribution. With a radially symmetric density, one only needs to compute it along some one dimensional ray. In symbols, $f(\mathbf{x}) = \det(R)^{-1/2} f(|R^{-1/2}\mathbf{x}^T|, 0, 0, \ldots, 0) = c(R) g(|R^{-1/2}\mathbf{x}'|)$. The univariate function g can be computed for arbitrary dimension d by numerically evaluating the univariate integral

$$g(x) = (2\pi)^{-d/2} \int_0^\infty t^{-d/2} e^{-x^2/(2t)} f(t|\alpha/2, 1, 2(\cos \pi\alpha/4)^{2/\alpha}, 0; 1)\, dt.$$

We next describe ways of assessing a d-dimensional data set to see if it is approximately sub-Gaussian and then estimating the parameters of a sub-Gaussian vector.

First perform a one dimensional stable fit to each coordinate of the data using one of the methods described above, to get estimates $\hat{\boldsymbol{\theta}}_i = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\gamma}_i, \hat{\delta}_i)$. If the α_i's are significantly different, then the data is not jointly α-stable, so it cannot be sub-Gaussian. Likewise, if the β_i's are not all close to 0, then the distribution is not symmetric and it cannot be sub-Gaussian.

If the α_i's are all close, form a pooled estimate of $\alpha = (\sum_{i=1}^d \alpha_i)/d =$ average of the indices of each component. Then shift the data by $\hat{\boldsymbol{\delta}} = (\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_d)$ so the distribution is centered at the origin.

Next, test for sub-Gaussian behavior. This can be accomplished by examining two dimensional projections because of the following result. If $\mathbf{X}$ is a d-dimensional sub-Gaussian α-stable random vector, then every two dimensional projection

$$\mathbf{Y} = (Y_1, Y_2) = (\mathbf{a}_1 \cdot \mathbf{X}, \mathbf{a}_2 \cdot \mathbf{X}) \tag{11}$$

($\mathbf{a}_1, \mathbf{a}_2 \in \mathbf{R}^d$) is a 2-dimensional sub-Gaussian α-stable random vector. Conversely, suppose $\mathbf{X}$ is a d-dimensional α-stable random vector with the property that every two dimensional projection of form (11) is non-singular sub-Gaussian. Then d-dimensional $\mathbf{X}$ is non-singular sub-Gaussian α-stable.

Estimating the $d(d+1)/2$ parameters (upper triangular part) of R can be done in at least two ways. For the first method, set $r_{ii} = \gamma_i^2$, i.e. the square of the scale parameter of the i-th coordinate. Then estimate r_{ij} by analyzing the pair (X_i, X_j) and take $r_{ij} = (\gamma^2(1,1) - r_{ii} - r_{jj})/2$, where $\gamma(1,1)$ is the scale parameter of $(1,1)\cdot(X_i, X_j) = X_i + X_j$. This involves estimating $d + d(d-1)/2 = d(d+1)/2$ one dimensional scale parameters.

For the second method, note that if $\mathbf{X}$ is α-stable sub-Gaussian, then $E\exp(i\mathbf{u}\cdot\mathbf{X}) = \exp(-(\mathbf{u}R\mathbf{u}^T)^{\alpha/2})$, so

$$[-\ln E\exp(i\mathbf{u}\cdot\mathbf{X})]^{2/\alpha} = \mathbf{u}R\mathbf{u}^T = \sum_i u_i^2 r_{ii} + 2\sum_{i<j} u_i u_j r_{ij}.$$

This is a linear function of the r_{ij}'s, so they can be estimated by regression. This method may be more accurate because it uses multiple directions, whereas the first method uses only three directions: (1,0), (0,1) and (1,1). If a two dimensional fit has already been done, then one has already estimated $\gamma(\mathbf{u})$ on a grid. Note that $\mathbf{u}R\mathbf{u}^T = \gamma^2(\mathbf{u})$ is the square of the scale parameter in the direction $\mathbf{u}$. Sample estimates of $\gamma^2(\mathbf{u})$ on a grid of $\mathbf{u}$ points can be used for the middle term above. In both methods, checks should be made to test that the resulting matrix R is positive definite.

The first method was used to estimate the matrix R for the Deutsche Mark-Japanese Yen data set considered above. The estimated matrix $\hat{R}$ was

$$\hat{R} = 10^{-6}\begin{pmatrix} 5.9552 & 4.0783 \\ 4.0783 & 13.9861 \end{pmatrix}.$$

The plot of $\gamma(\mathbf{t})$ shown in the lower left corner of Figure 10 also shows $\sqrt{\mathbf{t}\hat{R}\mathbf{t}^T}$ as a dashed line. It is virtually indistinguishable from the curve of $\gamma(\mathbf{t})$, supporting the idea that a sub-Gaussian stable fit does a good job of fitting the bivariate data.

9 Operator stable distributions

A brief discussion of operator stable laws is given next. The class of operator stable distributions allows different components of $\mathbf{X}$ to be stable with different indices α_j. It is defined by replacing the real scale term a_n in (7) with a matrix scale term A_n, see Jurek and Mason (1993) or Meerschaert and Schefler (2001). This may be of use in analyzing a portfolio, where different assets have different characteristics, e.g. some have Gaussian behavior and some have heavy tailed behavior, possibly with different tail behavior.

One subclass of the operator stable distributions is obtained by building up from independent groups of α-stable laws: suppose $(X_1, \ldots, X_{d_1})$ is a d_1-dimensional α_1-stable distribution, $(X_{d_1+1}, \ldots, X_{d_1+d_2})$ is a d_2-dimensional α_2-stable distribution, ..., $(X_{d_1+d_2+\cdots+d_{k-1}+1}, \ldots, X_{d_1+\cdots+d_k})$ is a d_k-dimensional α_k-stable distribution. If all these groups of distributions are independent, then the vector $\mathbf{X} = (X_1, \ldots, X_d)$, $d = d_1 + \cdots + d_k$, is a d-dimensional operator stable law. Also, for any $d \times d$ matrix A, the vector $\mathbf{Y} = A\mathbf{X}$ is an operator stable law. (One usually requires A to be invertible, otherwise the resulting $\mathbf{Y}$ will not be d-dimensional.)

10 Discussion

We have shown that estimation of general stable parameters is now feasible. The diagnostics show that some financial sets with heavy tails are well described by stable distributions. While they do not give a perfect fit, stable models can give a much better fit than Gaussian models.

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 where there is a poor fit, one should not use a stable model. These models are not a panacea - not all heavy tailed data sets can be well described by stable distributions. 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 actually be better to use the quantile parameter estimates rather than ML estimates, because the former tries to match the shape of the empirical distribution and ignores the top and bottom 5% of the data.

In multivariate problems where the dimension is large, it will be very difficult to model with a stable distribution unless there is some special structure. If some components are independent, then they should be separated out and analyzed alone. If the dependent components are elliptically contoured or have some other special structure, then Section 8 discusses a way to analyze them. In the general stable case, one may try to group the components into smaller dependent groups, estimate within groups, and then try to characterize dependence between groups. We are not aware of work on this topic.

References

[1] Abdul-Hamid, H. and Nolan, J. P. (1998), Multivariate stable densities as functions of their one dimensional projections. *J. Multivariate Analysis* **67**:80-89.

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

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

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

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

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

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

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

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

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

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

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

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

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

[15] Jurek, Z. J. and Mason, J. D. (1993), *Operator-limit distributions is probability theory*, Wiley, New York, New York.

[16] 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.

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

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

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

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

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

[22] 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, New York.

[23] McCulloch, J. H. (1997), Measuring tail thickness to estimate the stable index alpha: a critique, *J. Bus. Econ. Stat.* **15**:74-81.

[24] 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.

[25] Meerschaert, M. M. and Scheffler, H.-P. (2001), *Limit Distributions for Sums of Independent Random Vectors*, Wiley, N.Y., N.Y.

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

[27] Ma, X. and Nikias, C. L. (1995) On blind channel identification for impulsive signal environments, *Proceedings of ICASSP'95*, Detroit, MI.

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

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

[30] 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.

[31] Nolan, J. P. (2001) Maximum likelihood estimation and diagnostics for stable distributions. In *Leẃy Processes*, Ed. Barndorff-Nielson, Mikosch, and Resnick. Birkhäuser, Boston, MA.

[32] Nolan, J. P. (2003) *Stable Distributions: Models for Heavy Tailed Data* Birkhäuser, Boston, MA.

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

[34] Nolan, J. P., Panorska, A. K. and McCulloch, J. H. (2001), Estimation of stable spectral measures, *Mathematical and Computer Modelling*, **34**:1113-1122.

[35] Nolan, J. P. and Rajput, B. (1995) Calculation of multidimensional stable densities. *Communications in Statistics -Simulation* **24**:551-556.

[36] Ojeda, D. (2001) Comparison of stable estimators, Ph.D. Thesis, American University, Department of Mathematics and Statistics.

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

[38] Rachev, S. T. and Mittnik, S. (2000) *Stable Paretian Models in Finance*, Wiley, New York, New York.

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

[40] Uchaikin, V. V. and Zolotarev, V. M. (1999) *Chance and Stability: Stable Distributions and Their Applications*, VSP BV, Zeist.

[41] 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)