Nonparametric methods for heavy tailed vector data: A survey with applications from finance and hydrology

Mark M. Meerschaert[a*] and Hans-Peter Scheffler[b]

[a]Department of Mathematics, University of Nevada, Reno NV 89557 USA

[b]Fachbereich Mathematik, University of Dortmund, 44221 Dortmund, Germany

Many real problems in finance and hydrology involve data sets with power law tails. For multivariable data, the power law tail index usually varies with the coordinate, and the coordinates may be dependent. This paper surveys nonparametric methods for modeling such data sets. These models are based on a generalized central limit theorem. The limit laws in the generalized central limit theorem are operator stable, a class that contains the multivariate Gaussian as well as marginally stable random vectors with different tail behavior in each coordinate. Modeling this kind of data requires choosing the right coordinates, estimating the tail index for those coordinates, and characterizing dependence between the coordinates. We illustrate the practical application of these methods with several example data sets from finance and hydrology.

1. Introduction

Heavy tailed random variables with power law tails $P(|X| > x) \approx Cx^{-\alpha}$ are observed in many real world applications. Estimation of the tail parameter α is important, because it determines which moments exist. If $\alpha < 2$ then the variance is infinite, and if $\alpha < 1$ the mean is also undefined. For a heavy tailed random vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$ the tail index α_i for the ith component may vary with i. Choosing the wrong coordinates can mask variations in tail index, since the heaviest tail will dominate.

Modeling dependence is more complicated when $\alpha_i < 2$ since the covariance matrix is undefined. In order to model the tail behavior and dependence structure of heavy tailed vector data, a generalized central limit theorem [10,11,27] can be used. A nonparametric characterization of the heavy tailed vector data identifies the operator stable limit. Since the data distribution belongs to the generalized domain of attraction of that operator stable limit, it inherits the tail behavior, natural coordinate system, and tail dependence structure of that limit.

This paper surveys some nonparametric methods for modeling heavy tailed vector data. These methods begin by estimating the appropriate coordinate system, to unmask variations in tail behavior. Then the tail index is estimated for each coordinate, and finally the dependence structure can be characterized. In the parlance of operator stable laws, the first two steps estimate the exponent of the operator stable law, and the last step

*Partially supported by NSF grants DES-9980484 and DMS-0139927.

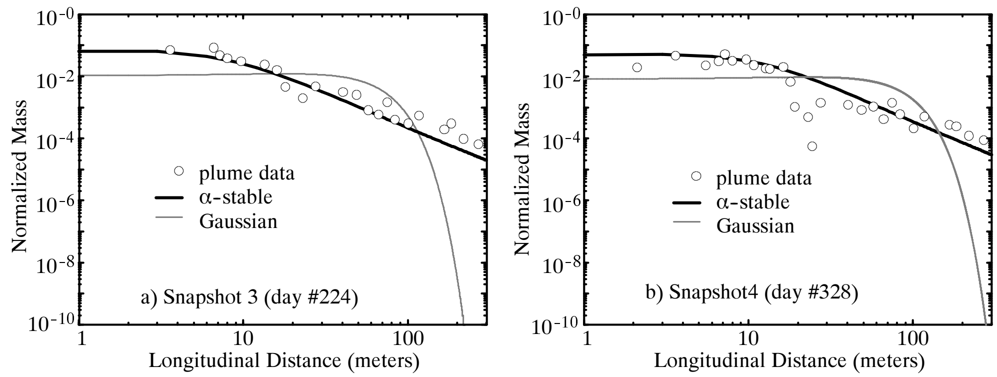

Figure 1. Heavy tailed α-stable densities provide a superior fit to relative concentration of a tracer released in an underground aquifer (from [6]).

estimates its spectral measure. All of the methods presented here apply to data whose distribution is attracted to an operator stable limit, not just to operator stable data. Therefore these methods are extremely robust.

2. Examples of data sets with heavy tails

Applications of heavy tailed random variables and random vectors occur in many areas, including hydrology and finance. Anderson and Meerschaert [4] find heavy tails in a river flow with $\alpha \approx 3$, so that the variance is finite but the fourth moment is infinite. Tessier, et al. [39] find heavy tails with $2 < \alpha < 4$ for a variety of river flows and rainfall accumulations. Hosking and Wallis [14] find evidence of heavy tails with $\alpha \approx 5$ for annual flood levels of a river in England. Benson, et al. [5,6] model concentration profiles for tracer plumes in groundwater using stochastic models whose heavy tails have $1 < \alpha < 2$, so that the mean is finite but the variance is infinite. Figure 1 shows the best-fitting Gaussian and α-stable densities plotted against relative concentration of a passive tracer. The straight lines on these log-log graphs indicate power law tails for the α-stable densities. Heavy tail distributions with $1 < \alpha < 2$ are used in physics to model anomalous diffusion, where a cloud of particles spreads faster than classical Brownian motion predicts [7,18,38]. More applications to physics with $0 < \alpha < 2$ are cataloged in Uchaikin and Zolotarev [40]. Resnick and Stărică [34] examine the quiet periods between transmissions for a networked computer terminal, and find heavy tails with $0 < \alpha < 1$, so that the mean and variance are both infinite. Several additional applications to computer science, finance, and signal processing appear in Adler, Feldman, and Taqqu [3]. More applications to signal processing can be found in Nikias and Shao [30].

Mandelbrot [22] and Fama [9] pioneered the use of heavy tail distributions in finance. Mandelbrot [22] presents graphical evidence that historical daily price changes in cotton have heavy tails with $\alpha \approx 1.7$, so that the mean exists but the variance is infinite. Jansen and de Vries [15] argue that daily returns for many stocks and stock indices have heavy tails with $3 < \alpha < 5$, and discuss the possibility that the October 1987 stock market plunge might be just a heavy tailed random fluctuation. Loretan and Phillips [21] use similar methods to estimate heavy tails with $2 < \alpha < 4$ for returns from numerous stock market indices and exchange rates. This indicates that the variance is finite but the fourth moment is infinite. Both daily and monthly returns show heavy tails with similar values of α in this study. Rachev and Mittnik [33] use different methods to find heavy tails with $1 < \alpha < 2$ for a variety of stocks, stock indices, and exchange rates. McCulloch [23] uses similar methods to re-analyze the data in [15,21], and obtains estimates of $1.5 < \alpha < 2$. This is important because the variance of price returns is finite if $\alpha > 2$ and infinite if $\alpha < 2$. While there is disagreement about the true value of α, depending on which model is employed, all of these studies agree that financial data is typically heavy tailed, and that the tail parameter α varies between different assets.

3. Generalized central limit theorem

For heavy tailed random vectors, a generalized central limit theorem applies. If $\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots$ are IID random vectors on $\mathbb{R}^d$ we say that $\boldsymbol{X}$ belongs to the *generalized domain of attraction* of some full dimensional random vector $\boldsymbol{Y}$ on $\mathbb{R}^d$, and we write $\boldsymbol{X} \in \mathrm{GDOA}(\boldsymbol{Y})$, if

$$A_n(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n - \boldsymbol{b}_n) \Rightarrow \boldsymbol{Y} \tag{3.1}$$

for some $d \times d$ matrices A_n and vectors $\boldsymbol{b}_n \in \mathbb{R}^d$. The limits in (3.1) are called *operator stable* [17,27,37]. If $E(\|\boldsymbol{X}\|^2)$ exists then the classical central limit theorem shows that $\boldsymbol{Y}$ is multivariable normal, a special case of operator stable. In this case, we can take $A_n = n^{-1/2}I$ and $\boldsymbol{b}_n = nE(\boldsymbol{X})$. If (3.1) holds with $A_n = n^{-1/\alpha}I$ for any $\alpha \in (0, 2]$, then $\boldsymbol{Y}$ is multivariable stable with index α. In this case we say that $\boldsymbol{X}$ belongs to the generalized domain of normal[2] attraction of $\boldsymbol{Y}$. See [27,35] for more information.

Matrix powers provide a natural extension of the stable index α, allowing the tail index to vary with the coordinate. Let $\exp(A) = I + A + A^2/2! + A^3/3! + \cdots$ be the usual exponential operator for $d \times d$ matrices, and let $E = \mathrm{diag}(1/\alpha_1, \ldots, 1/\alpha_d)$ be a diagonal matrix. If (3.1) holds with $A_n = n^{-E} = \exp(-E \ln n)$, then A_n is diagonal with entries n^{-1/α_i} for $i = 1, \ldots, d$. Write $\boldsymbol{X}_t = (X_1(t), \ldots, X_d(t))'$, $\boldsymbol{b}_t = (b_1(t), \ldots, b_d(t))'$, $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$, and project (3.1) onto its ith coordinate to see that

$$\frac{X_i(1) + \cdots + X_i(n) - b_i(n)}{n^{1/\alpha_i}} \Rightarrow Y_i \quad \text{for each } i = 1, \ldots, d. \tag{3.2}$$

Each coordinate $X_i(t)$ is in the domain of attraction of a stable law Y_i with index α_i, and the matrix E specifies every tail index. The limit $\boldsymbol{Y}$ is called *marginally stable*, a special case of operator stable. The matrix E, called an *exponent* of the operator stable random vector $\boldsymbol{Y}$, plays the role of the stable index α for stable laws. The matrix E need not be diagonal. Diagonalizable exponents involve a change of coordinates, degenerate eigenvalues thicken probability tails by a logarithmic factor, and complex eigenvalues introduce rotational scaling, see Meerschaert [24].

A proof of the generalized central limit theorem for matrix scaling can be found in Meerschaert and Scheffler [27]. Since $\boldsymbol{Y}$ is infinitely divisible, the Lévy representation

[2] This terminology refers to the special form of the norming, not the limit!

(Theorem 3.1.11 in [27]) shows that the characteristic function $E[e^{i\boldsymbol{k}\cdot\boldsymbol{Y}}]$ is of the form $e^{\psi(\boldsymbol{k})}$ where

$$\psi(\boldsymbol{k}) = i\boldsymbol{b}\cdot\boldsymbol{k} - \frac{1}{2}\boldsymbol{k}\cdot\Sigma\boldsymbol{k} + \int_{\boldsymbol{x}\neq 0}\left(e^{i\boldsymbol{k}\cdot\boldsymbol{x}} - 1 - \frac{i\boldsymbol{k}\cdot\boldsymbol{x}}{1+\|\boldsymbol{x}\|^2}\right)\phi(d\boldsymbol{x}) \tag{3.3}$$

for some $\boldsymbol{b}\in\mathbb{R}^d$, some nonnegative definite symmetric $d\times d$ matrix Σ and some Lévy measure ϕ. The Lévy measure satisfies $\phi\{\boldsymbol{x}:\|\boldsymbol{x}\|>1\}<\infty$ and

$$\int_{0<\|\boldsymbol{x}\|<1}\|\boldsymbol{x}\|^2\phi(d\boldsymbol{x})<\infty.$$

For a multivariable stable law,

$$\phi\{\boldsymbol{x}:\|\boldsymbol{x}\|>r, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\in B\} = Cr^{-\alpha}M(B)$$

where M is a probability measure on the unit sphere that is not supported on any $d-1$ dimensional subspace of $\mathbb{R}^d$. We call M the *spectral measure*[3]. If $\phi=0$ then $\boldsymbol{Y}$ is normal with mean $\boldsymbol{b}$ and covariance matrix Σ. If $\Sigma=0$ then a necessary and sufficient condition for (3.1) to hold is that

$$nP(A_n\boldsymbol{X}\in B)\to\phi(B) \quad \text{as } n\to\infty \tag{3.4}$$

for Borel subsets B of $\mathbb{R}^d\setminus\{0\}$ whose boundary have ϕ-measure zero, where ϕ is the Lévy measure of the limit $\boldsymbol{Y}$. Proposition 6.1.10 in [27] shows that the convergence (3.4) is equivalent to *regular variation* of the probability distribution $\mu(B)=P(\boldsymbol{X}\in B)$, an analytic condition that extends the idea of power law tails. If (3.4) holds then Proposition 6.1.2 in [27] shows that the Lévy measure satisfies

$$t\phi(d\boldsymbol{x}) = \phi(t^{-E}d\boldsymbol{x}) \quad \text{for all } t>0 \tag{3.5}$$

for some $d\times d$ matrix E. Then it follows from the characteristic function formula that $\boldsymbol{Y}$ is operator stable with exponent E, and that for $\boldsymbol{Y}_n$ IID with $\boldsymbol{Y}$ we have

$$n^{-E}(\boldsymbol{Y}_1+\cdots+\boldsymbol{Y}_n-\boldsymbol{b}_n)\stackrel{d}{=}\boldsymbol{Y} \tag{3.6}$$

for some $\boldsymbol{b}_n$, see Theorem 7.2.1 in [27]. Hence operator stable laws belong to their own GDOA, so that the probability distribution of $\boldsymbol{Y}$ also varies regularly, and sums of IID operator stable random vectors are again operator stable with the same exponent E. If $E=aI$ then $\boldsymbol{Y}$ is multivariable stable with index $\alpha=1/a$, and (3.4) is equivalent to the balanced tails condition

$$\frac{P(\|\boldsymbol{X}\|>r, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}\in B)}{P(\|\boldsymbol{X}\|>r)}\to M(B) \quad \text{as } r\to\infty \tag{3.7}$$

for all Borel subsets B of the unit sphere $S=\{\boldsymbol{\theta}\in\mathbb{R}^d:\|\boldsymbol{\theta}\|=1\}$ whose boundary has M-measure zero.

[3]Some authors call CM the spectral measure.

4. Spectral decomposition theorem

For any $d \times d$ matrix E there is a unique *spectral decomposition* based on the real parts of the eigenvalues, see for example Theorem 2.1.14 in [27]. Write the minimal polynomial of E as $f_1(\boldsymbol{x}) \cdots f_p(\boldsymbol{x})$ where every root of f_j has real part a_j and $a_1 < \cdots < a_p$. Define $V_j = \ker f_j(E)$ and let $d_j = \dim V_j$. Then we may write $\mathbb{R}^d = \mathbb{V}_1 \oplus \cdots \oplus V_p$, and in any basis that respects this direct sum decomposition we have

$$E = \begin{pmatrix} E_1 & 0 & \cdots & 0 \\ 0 & E_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & E_p \end{pmatrix} \tag{4.1}$$

where E_i is a $d_i \times d_i$ matrix, every eigenvalue of E_i has real part equal to a_i, and $d_1 + \cdots + d_p = d$. This is called the spectral decomposition with respect to E. Given a nonzero vector $\boldsymbol{\theta} \in \mathbb{R}^d$, write $\boldsymbol{\theta} = \boldsymbol{\theta}_1 + \cdots + \boldsymbol{\theta}_p$ with $\boldsymbol{\theta}_i \in V_i$ for each $i = 1, \ldots, p$ and define

$$\alpha(\boldsymbol{\theta}) = \min\{1/a_i : \boldsymbol{\theta}_i \neq 0\}. \tag{4.2}$$

If $\boldsymbol{Y}$ is operator stable with exponent E, then the probability distribution of $\boldsymbol{Y}$ varies regularly with exponent E, and Theorem 6.4.15 in [27] shows that for any small $\delta > 0$ we have

$$r^{-\alpha(\boldsymbol{\theta})-\delta} < P(|\boldsymbol{Y} \cdot \boldsymbol{\theta}| > r) < r^{-\alpha(\boldsymbol{\theta})+\delta}$$

for all $r > 0$ sufficiently large. In other words, the tail behavior of $\boldsymbol{Y}$ is dominated by the component with the heaviest tail. This also means that $E(|\boldsymbol{Y} \cdot \boldsymbol{\theta}|^\beta)$ exists for $0 < \beta < \alpha(\boldsymbol{\theta})$ and diverges for $\beta > \alpha(\boldsymbol{\theta})$. Theorem 7.2.1 in [27] shows that every $a_i \geq 1/2$, so that $0 < \alpha(\boldsymbol{\theta}) \leq 2$. If we write $\boldsymbol{Y} = \boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_p$ with $\boldsymbol{Y}_i \in V_i$ for each $i = 1, \ldots, p$, then projecting (3.6) onto V_i shows that $\boldsymbol{Y}_i$ is an operator stable random vector on V_i with some exponent E_i. We call this the spectral decomposition of $\boldsymbol{Y}$ with respect to E. Since every eigenvalue of E_i has the same real part a_i we say that $\boldsymbol{Y}_i$ is spectrally simple, with index $\alpha_i = 1/a_i$. Although $\boldsymbol{Y}_i$ might not be multivariable stable, it has similar tail behavior. For any small $\delta > 0$ we have

$$r^{-\alpha_i-\delta} < P(\|\boldsymbol{Y}_i\| > r) < r^{-\alpha_i+\delta}$$

for all $r > 0$ sufficiently large, so $E(\|\boldsymbol{Y}_i\|^\beta)$ exists for $0 < \beta < \alpha_i$ and diverges for $\beta > \alpha_i$.

If $\boldsymbol{X} \in \mathrm{GDOA}(\boldsymbol{Y})$ then Theorem 8.3.24 in [27] shows that the limit $\boldsymbol{Y}$ and norming matrices A_n in (3.1) can be chosen so that every V_i in the spectral decomposition of $\mathbb{R}^d$ with respect to the exponent E of $\boldsymbol{Y}$ is A_n-invariant for every n, and $V_1, \ldots, V_p$ are mutually perpendicular. Then the probability distribution of $\boldsymbol{X}$ is regularly varying with exponent E and $\boldsymbol{X}$ has the same tail behavior as $\boldsymbol{Y}$. In particular, for any small $\delta > 0$ we have

$$r^{-\alpha(\boldsymbol{\theta})-\delta} < P(|\boldsymbol{X} \cdot \boldsymbol{\theta}| > r) < r^{-\alpha(\boldsymbol{\theta})+\delta} \tag{4.3}$$

for all $r > 0$ sufficiently large. In this case, we say that $\boldsymbol{Y}$ is spectrally compatible with $\boldsymbol{X}$, and we write $\boldsymbol{X} \in \mathrm{GDOA}_c(\boldsymbol{Y})$.

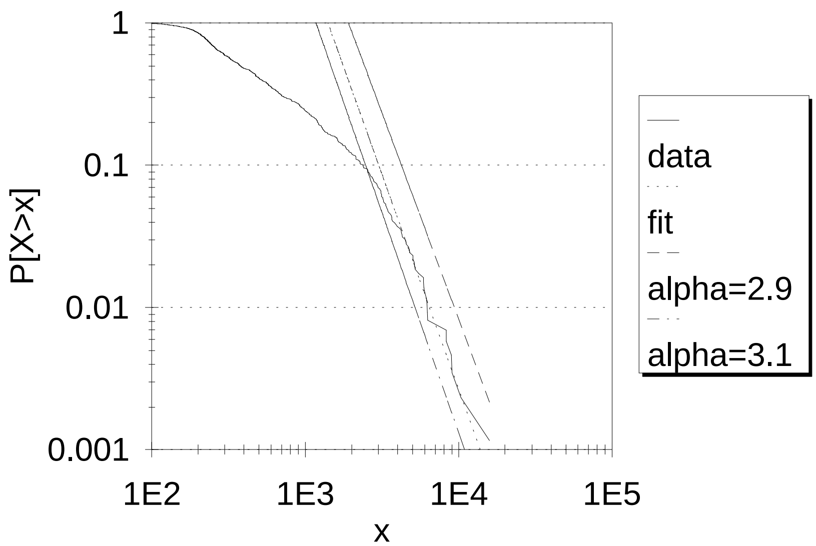

Figure 2. Monthly average river flows (cubic feet per second) for the Salt river near Roosevelt AZ exhibit heavy tails with $\alpha \approx 3$ (from [4]).

5. Nonparametric methods for tail estimation

Mandelbrot [22] pioneered a graphical estimation method for tail estimation. If $y = P(X > r) \approx Cr^{-\alpha}$ then $\log y \approx \log C - \alpha \log r$. Ordering the data so that $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ we should have approximately that $r = X_{(i)}$ when $y = i/n$. Then a plot of $\log X_{(i)}$ versus $\log(i/n)$ should be approximately linear with slope $-\alpha$. If $P(X > r) \approx Cr^{-\alpha}$ for r large, then the upper tail should be approximately linear. We call this a *Mandelbrot plot*. Figure 2 shows a Mandelbrot plot used to estimate the tail index for a river flow time series.

The most popular numerical estimator for α is due to Hill [13], see also Hall [12]. Assuming that $P(X > r) = Cr^{-\alpha}$ for large values of $r > 0$, the maximum likelihood estimates for α and C based on the $m + 1$ largest observations are

$$\begin{aligned} \hat{\alpha} &= \left[\frac{1}{m} \sum_{i=1}^{m} \left(\ln X_{(i)} - \ln X_{(m+1)} \right) \right]^{-1} \\ \hat{C} &= \frac{m}{n} X_{(m+1)}^{\hat{\alpha}} \end{aligned} \tag{5.1}$$

where m is to be taken as large as possible, but small enough so that the tail condition

$P(X > r) = Cr^{-\alpha}$ remains a useful approximation. Finding the best value of m is a practical challenge, and creates a certain amount of controversy [8]. Jansen and de Vries [15] use Hill's estimator with a fixed value of $m = 100$ for several different assets. Loretan and Phillips [21] tabulate several different values of m for each asset. Hill's estimator $\hat{\alpha}$ is consistent and asymptotically normal with variance α^2/m, so confidence intervals are easy to construct. These intervals clearly demonstrate that the tail parameters in Jansen and de Vries [15] and Loretan and Phillips [21] vary depending on the asset. Painter, Cvetkovic, and Selroos [32] apply Hill's estimator to data on fluid flow in fractures, to estimate two parameters of interest. Their α_i estimates also show a significant difference between the two parameters. In all of these studies, an appropriate model for vector data must allow α_i to vary with i.

Aban and Meerschaert [1] develop a more general Hill's estimator to account for a possible shift in the data. If $P(X > r) = C(r - s)^{-\alpha}$ for r large, the maximum likelihood estimates for α and C based on the $m + 1$ largest observations are

$$\begin{aligned} \hat{\alpha} &= \left[\frac{1}{m} \sum_{i=1}^{m} \left(\ln(X_{(i)} - \hat{s}) - \ln(X_{(m+1)} - \hat{s}) \right) \right]^{-1} \\ \hat{C} &= \frac{m}{n} (X_{(m+1)} - \hat{s})^{\hat{\alpha}} \end{aligned} \tag{5.2}$$

where $\hat{s}$ is obtained by numerically solving the equation

$$\hat{\alpha}(X_{(m+1)} - \hat{s})^{-1} = (\hat{\alpha} + 1) \frac{1}{m} \sum_{i=1}^{m} (X_{(i)} - \hat{s})^{-1} \tag{5.3}$$

over $\hat{s} < X_{(m+1)}$. Once the optimal shift is computed, $\hat{\alpha}$ comes from Hill's estimator applied to the shifted data. One practical implication is that, since the Pareto model is not shift-invariant, it is a good idea to try shifting the data to get a linear Mandelbrot plot.

Meerschaert and Scheffler [25] propose a robust estimator

$$\hat{\alpha} = \frac{2 \ln n}{\ln n + \ln \hat{\sigma}^2} \tag{5.4}$$

based on the sample variance $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2$, where as usual $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$ is the sample mean. This tail estimator is consistent for IID data in the domain of attraction of a stable law with index $\alpha < 2$. Like Hill's estimator [34], it is also consistent for moving averages. If X is attracted to a normal limit, then $\hat{\alpha} \to 2$. It is interesting, and even somewhat ironic, that the sample variance can be used to estimate tail behavior, and hence tells us something about the spread of typical values, even in this case $0 < \alpha < 2$ where the variance is undefined.

6. The right coordinate system

Equation 4.3 shows that tail behavior is dominated by the component with the heaviest tail. In order to unmask variations in tail behavior, one has to find a coordinate system

$\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ where $\alpha_i = \alpha(\boldsymbol{\theta}_i)$ varies with i. This is equivalent to estimating the spectral decomposition. A useful estimator is based on the (uncentered) sample covariance matrix

$$M_n = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t'. \tag{6.1}$$

If $\boldsymbol{X}_t$ are IID with $\boldsymbol{X}$, then $\boldsymbol{X}_t\boldsymbol{X}_t'$ are IID random elements of the vector space $\mathcal{M}_s^d$ of symmetric $d \times d$ matrices, and the extended central limit theorem applies (see Section 10.2 in [27] for complete proofs). If the probability distribution of $\boldsymbol{X}$ is regularly varying with exponent E and (3.4) holds with $t\phi\{d\boldsymbol{x}\} = \phi\{t^{-E}d\boldsymbol{x}\}$ for all $t > 0$, then the distribution of $\boldsymbol{X}\boldsymbol{X}'$ is also regularly varying with

$$nP(A_n\boldsymbol{X}\boldsymbol{X}'A_n' \in B) \to \Phi(B) \quad \text{as } n \to \infty \tag{6.2}$$

for Borel subsets B of $\mathcal{M}_s^d$ that are bounded away from zero and whose boundary has Φ-measure zero. The exponent ξ of the limit measure $\Phi\{d(\boldsymbol{x}\boldsymbol{x}')\} = \phi\{d\boldsymbol{x}\}$ is defined by $\xi M = EM + ME'$ for $M \in \mathcal{M}_s^d$. If every eigenvalue of E has real part $a_i > 1/2$, then

$$nA_nM_nA_n' \Rightarrow W \tag{6.3}$$

holds with W operator stable. The centered sample covariance matrix is defined by

$$\Gamma_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)'$$

where $\bar{\boldsymbol{X}}_n = n^{-1}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$ is the sample mean. For heavy tailed data, Theorem 10.6.15 in [27] shows that Γ_n and M_n have the same asymptotics. In practice, it is common to mean-center the data, so it does not matter which form we choose.

Since M_n is symmetric and nonnegative definite, there exists an orthonormal basis of eigenvectors for M_n with nonnegative eigenvalues. Sort the eigenvalues

$$\lambda_1 \le \cdots \le \lambda_d$$

and the associated unit eigenvectors

$$\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$$

so that $M_n\boldsymbol{\theta}_j = \lambda_j\boldsymbol{\theta}_j$ for each $j = 1, \ldots, d$. In the spectral decomposition (4.1) each block E_i is a $d_i \times d_i$ matrix, and every eigenvalue of E_i has the same real part a_i for some $1/2 \le a_1 < \cdots < a_p$. Let $D_0 = 0$ and $D_i = d_1 + \cdots + d_i$ for $1 \le i \le p$. Now Theorem 10.4.5 in [27] shows that

$$\frac{2\log n}{\log n + \log \lambda_j} \xrightarrow{P} \alpha_i \quad \text{as } n \to \infty$$

for any $D_{i-1} < j \le D_i$, where $\alpha_i = 1/a_i$ is the tail index. This is a multivariable version of the one variable tail estimator (5.4). Furthermore, Theorem 10.4.8 in [27] shows that the eigenvectors $\boldsymbol{\theta}_j$ converge in probability to V_1 when $j \le D_1$, and to V_p

when $j > D_{p-1}$. This shows that the eigenvectors estimate the coordinate vectors in the spectral decomposition, at least for the lightest and heaviest tails. Meerschaert and Scheffler [29] show that the same results hold for moving averages. If $p \leq 3$, this gives a practical method for determining the right coordinate system for modeling heavy tail data: Simply use the eigenvalues of the sample covariance matrix as the coordinate vectors.

Example 6.1. Meerschaert and Scheffler [26] consider $n = 2853$ daily exchange rate log-returns $X_1(t)$ for the German Deutsch Mark and $X_2(t)$ for the Japanese Yen, both taken against the US Dollar. Divide each entry by .004, which is the approximate median for both $|X_1(t)|$ and $|X_2(t)|$. This has no effect on the eigenvectors but helps to obtain good estimates of the tail thickness. Then compute

$$M_n = \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} X_1(t)^2 & X_1(t)X_2(t) \\ X_1(t)X_2(t) & X_2(t)^2 \end{pmatrix} = \begin{pmatrix} 3.204 & 2.100 \\ 2.100 & 3.011 \end{pmatrix}$$

which has eigenvalues $\lambda_1 = 1.006$, $\lambda_2 = 5.209$ and associated unit eigenvectors $\boldsymbol{\theta}_1 = (0.69, -0.72)'$, $\boldsymbol{\theta}_2 = (0.72, 0.69)'$. Next compute

$$\begin{aligned} \hat{\alpha}_1 &= \frac{2\ln 2853}{\ln 2853 + \ln 1.006} = 1.998 \\ \hat{\alpha}_2 &= \frac{2\ln 2853}{\ln 2853 + \ln 5.209} = 1.656 \end{aligned} \tag{6.4}$$

indicating that one component fits a finite variance model but the other fits a heavy tailed model with $\alpha = 1.656$.

Now the eigenvectors can be used to find a new coordinate system that unmasks the variations in tail behavior. Let

$$P = \begin{pmatrix} 0.69 & -0.72 \\ 0.72 & 0.69 \end{pmatrix}$$

be the change of coordinates matrix whose ith row is the eigenvector $\boldsymbol{\theta}_i$, and let $\boldsymbol{Z}_t = P\boldsymbol{X}_t$ be the same data in the new, rotated coordinate system. The random vectors $\boldsymbol{Z}_t \in \mathrm{GDOA}_c(\boldsymbol{Y})$ where $\boldsymbol{Y} = (Y_1, Y_2)'$ is operator stable with exponent

$$E = \begin{pmatrix} 0.50 & 0 \\ 0 & 0.60 \end{pmatrix}$$

since $0.50 = 1/1.998$ and $0.60 = 1/1.656$. Hence the random variables $Z_1(t)$ can be modeled as belonging to the domain of attraction of a normal limit Y_1, while the random variables $Z_2(t)$ are attracted to a stable limit Y_2 with index $\alpha = 1.656$. Inverting $\boldsymbol{Z}_t = P\boldsymbol{X}_t$ we obtain

$$\begin{aligned} X_1(t) &= 0.69Z_1(t) + 0.72Z_2(t) \\ X_2(t) &= -0.72Z_1(t) + 0.69Z_2(t). \end{aligned} \tag{6.5}$$

Both exchange rates have a common heavy-tailed term $Z_2(t)$, so both have heavy tails with the same tail index $\alpha = 1.656$. It is tempting to interpret $Z_2(t)$ as the common

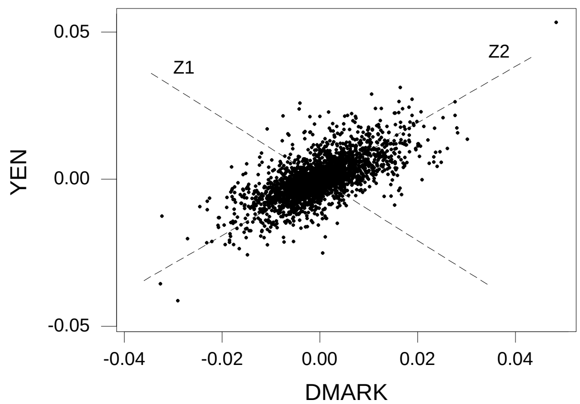

Figure 3. Exchange rates against the US dollar. The new coordinates uncover variations in the tail parameter α.

influence of fluctuations in the US dollar, and the remaining light-tailed factor $Z_1(t)$ as the accumulation of other price shocks independent of the US dollar.

Once the new coordinate system is identified, any one variable tail estimator can be used to approximate the α_i. Applying Hill's estimator to the $Z_i(t)$ data with $m \approx 500$ yields estimates similar to those obtained here, providing another justification for a model with $\alpha_1 \neq \alpha_2$. This exchange rate data was also analyzed by Nolan, Panorska and McCulloch [31] using a multivariable stable model with the same tail index $\alpha \approx 1.6$ for both exchange rates. Rotating to a new coordinate system unmasks variations in the tail index that are not apparent in the original coordinates. Kozubowski, et al. [19] compare the fit of both stable and geometric stable laws to the rotated data $Z_2(t)$, see Figure 4. Since operator geometric stable laws (where the number n of summands in (3.1) is replaced by a geometric random variable) have the same tail behavior as operator stable laws, and even the same domains of attraction (see Theorem 3.1 in [20]), the spectral decomposition and its eigenvector estimator are the same for both models.

7. Modeling dependence

For heavy tailed random vector data with tail index $\alpha_i < 2$, the covariance matrix is undefined. In this case, dependence can be modeled using the spectral measure. Suppose that $\boldsymbol{X}_t$ are IID in the generalized domain of attraction of an operator stable law $\boldsymbol{Y}$ with no normal component, so that $\Sigma = 0$ in (3.3). In this case, the log-characteristic function

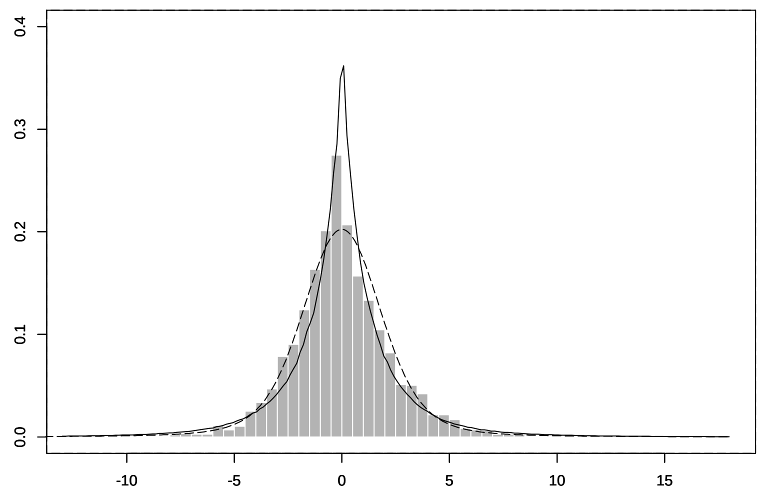

Figure 4. Two models for the rotated exchange rate data Z_2: stable (dotted line) and geometric stable (solid line).

can be written in the form

$$\psi(\boldsymbol{k}) = i\boldsymbol{b}\cdot\boldsymbol{k} + C\int_0^{2\pi}\int_0^\infty \left(e^{i\boldsymbol{k}\cdot r^E\boldsymbol{\theta}} - 1 - \frac{i\boldsymbol{k}\cdot r^E\boldsymbol{\theta}}{1+r^2}\right)\frac{dr}{r^2}M(d\boldsymbol{\theta}) \tag{7.1}$$

where $C > 0$, and M is a probability measure on the unit circle $S = \{x \in \mathbb{R}^d : \|x\| = 1\}$ called the *spectral measure* of $\boldsymbol{X}$. Equation (7.1) comes from applying a disintegration formula (Theorem 7.2.5 in [27]) to the Lévy measure in (3.3). The spectral measure M determines the dependence between the components of $\boldsymbol{X}$. For example (cf. [35] for the multivariate stable case and Meerschaert and Scheffler [28] for the general case), the components of $\boldsymbol{X}$ are independent if and only if M is supported on the coordinate axes.

The following method of Scheffler [36] can be used to estimate C, M: Any $\boldsymbol{x} \in \mathbb{R}^d \setminus \{0\}$ can be written uniquely in the form $\boldsymbol{x} = \tau(\boldsymbol{x})^E\boldsymbol{\theta}(\boldsymbol{x})$ for some radius $\tau(\boldsymbol{x}) > 0$ and some direction $\boldsymbol{\theta}(\boldsymbol{x}) \in S$. These are called the Jurek coordinates [16]. Define the order statistics $\boldsymbol{X}_{(i)}$ such that $\tau(\boldsymbol{X}_{(1)}) \geq \cdots \geq \tau(\boldsymbol{X}_{(n)})$ where ties are broken arbitrarily. The estimate $\hat{M}_m$ of the mixing measure based on the m largest order statistics is just the empirical measure based on the points $\boldsymbol{\theta}(\boldsymbol{X}_{(i)})$ on the unit sphere S. In other words, the probability we assign to any sector F of the unit sphere S is equal to the fraction of the points $\{\boldsymbol{\theta}(\boldsymbol{X}_{(i)}) : 1 \leq i \leq m\}$ falling in this sector. The estimator of C is $\hat{C} = (m/n)\tau(\boldsymbol{X}_{(m)})$, which reduces to Hill's estimator of C in the one variable case. This estimator applies when the data belong to the normal[4] generalized domain of attraction of an operator stable law with no normal component.

[4]Here normal means that we can take $A_n = n^{-E}$ in (3.1).

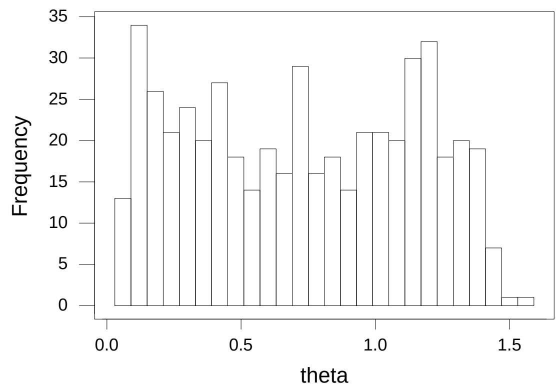

Figure 5. The empirical mixing measure distribution for the fracture flow data is nearly uniform on the interval $[0, \pi/2]$.

Example 7.1. Painter, Cvetkovic, and Selroos [32] performed a detailed simulation of fracture flow networks, and extracted data on fracture aperture q (millimeters) and fluid velocity v (meters/year). Mandelbrot plots of an unpublished data set related to that study indicate that both $X_t(1) = 1/v$ and $X_t(2) = 1/qv$ have heavy tails with $1 < \alpha < 2$. The Hill estimator (5.1) yields $\hat{\alpha}_1 = 1.4$ and $\hat{C}_1 = 0.0065$ for the $1/v$ data, $\hat{\alpha}_2 = 1.05$ and $\hat{C}_2 = 0.028$ for the $1/qv$ data. The $\hat{\alpha}$ estimates were plotted as a function of m to see where they stabilized. Linear regression estimates of the parameters based on a log-log plot are consistent with the results of Hill's estimator. See Aban and Meerschaert [2] for a discussion of linear regression estimates and their relation to Hill's estimator. Since $\alpha_1 \neq \alpha_2$ the original coordinates are appropriate. Also, the eigenvalues of the sample covariance matrix are near the coordinate axes.

We rescale $Y_t(i) = X_t(i)/\hat{C}_i^{a_i}$ so that approximately $P(Y_t(i) > r) = r^{-a_i}$, and then we estimate C, M for the data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$. This rescaling gives a clearer picture of the mixing measure. A histogram of $\{\boldsymbol{\theta}(\boldsymbol{Y}_{(i)}) : 1 \leq i \leq m\}$ for $m = 500$, shown in Figure 5, indicates that these unit vectors are approximately uniformly distributed over the first quadrant of the unit circle. Other values of m in the range $100 < m < 1000$ show similar behavior, and we conclude that the mixing measure M is approximately uniform on the first quadrant of the unit circle, indicating strong dependence between $1/v$ and $1/qv$. The estimator $\hat{C}$ based on the m largest order statistics stabilizes at a value near $\pi/2$ for $100 < m < 1000$, which coincides with the arclength of the first quadrant, so we estimate $CM(d\boldsymbol{\theta}) = d\boldsymbol{\theta}$. Painter, Cvetkovic, and Selroos [32] argue that the $1/v$ and $1/qv$ data are well modeled by stable distributions. In that case, it is reasonable to model the rescaled data $\boldsymbol{Y}$ as operator stable with characteristic function

$$F(\boldsymbol{k}) = \exp\left\{ i\boldsymbol{b}\cdot\boldsymbol{k} + \int_0^{2\pi}\int_0^{\infty}\left(e^{i\boldsymbol{k}\cdot r^E\boldsymbol{\theta}} - 1 - i\boldsymbol{k}\cdot r^E\boldsymbol{\theta}\right)\frac{dr}{r^2}d\boldsymbol{\theta}\right\}$$

where $\boldsymbol{b}$ is the sample mean (this uses an alternative form of the log-characteristic function, see [27] Theorem 3.1.14 and Remark 3.1.15), and $\boldsymbol{k}\cdot r^E\boldsymbol{\theta} = k_1 r^{1/1.4}\cos\boldsymbol{\theta} + k_2 r^{1/1.05}\sin\boldsymbol{\theta}$. The density $f(\boldsymbol{y})$ of this operator stable random vector can then be obtained from the Fourier inversion formula

$$f(\boldsymbol{y}) = (2\pi)^{-1}\int_{\boldsymbol{k}\in\mathbb{R}^2} e^{-i\boldsymbol{k}\cdot\boldsymbol{y}}F(\boldsymbol{k})d\boldsymbol{k}$$

or perhaps more efficiently via inverse fast Fourier transforms.

8. Summary

Vector data sets with heavy tails can be usefully modeled as belonging to the generalized domain of attraction of an operator stable law. This robust model characterizes the tail behavior, which can vary with coordinate, and also the dependence between coordinates. Choosing the right coordinate system is crucial, since variations in tail behavior can otherwise go undetected. A useful coordinate system in this regard is the set of eigenvectors of the sample covariance matrix. Once the right coordinates are chosen, any one variable tail estimator can be used. Then a nonparametric estimator of the spectral measure provides a way to model the dependence between coordinates. These methods have proven useful in a variety of applications to data analysis problems in hydrology and finance.

REFERENCES

1. I. Aban and M. Meerschaert (2001) Shifted Hill's estimator for heavy tails. *Comm. Statist. Simulation Comput.* **30**, 949–962.
2. Aban, I. and M. Meerschaert (2002) Generalized least squares estimators for the thickness of heavy tails. *J. Statis. Plann. Inf.*, to appear.
3. Adler, R., R. Feldman, and M. Taqqu (1998) *A Practical Guide to Heavy Tails.* Birkhäuser, Boston.
4. Anderson, P. and M. Meerschaert (1998) Modeling river flows with heavy tails. *Water Resour. Res.* **34**, 2271–2280.
5. Benson, D., S. Wheatcraft, and M. Meerschaert (2000) Application of a fractional advection-dispersion equation. *Water Resour. Res.* **36**, 1403–1412.
6. Benson, D., R. Schumer, M. Meerschaert, and S. Wheatcraft (2001) Fractional dispersion, Lévy motions, and the MADE tracer tests. *Transport Porous Med.* **42**, 211–240.
7. Blumen, A., G. Zumofen and J. Klafter (1989) Transport aspects in anomalous diffusion: Lévy walks. *Phys Rev. A* **40** 3964–3973.
8. Drees, H, L. de Haan, and S. Resnick (2000) How to make a Hill plot. *Ann. Statist.* **28**, 254–274.
9. Fama, E. (1965) The behavior of stock market prices. *J. Business* **38**, 34–105.

10. Feller, W. (1971) *An Introduction to Probability Theory and Its Applications.* Vol. II, 2nd Ed., Wiley, New York.
11. Gnedenko, B. and A. Kolmogorov (1968) *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, Mass.
12. Hall, P. (1982) On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc. B* **44** 37–42.
13. Hill, B. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1173.
14. Hosking, J. and J. Wallis (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349.
15. Jansen, D. and C. de Vries (1991) On the frequency of large stock market returns: Putting booms and busts into perspective. *Rev. Econ. Statist.* **23**, 18–24.
16. Jurek, Z.J. (1984) Polar Coordinates in Banach Spaces, *Bull. Ac. Pol. Math.* 32, 61-66.
17. Jurek, Z. and J. D. Mason (1993) *Operator-Limit Distributions in Probability Theory.* Wiley, New York.
18. Klafter, J., A. Blumen and M. Shlesinger (1987) Stochastic pathways to anomalous diffusion. *Phys. Rev. A* **35** 3081–3085.
19. Kozubowski, T. J., M.M. Meerschaert, A. K. Panorska and H. P. Scheffler (2003) Operator geometric stable laws, preprint.
20. Kozubowski, T. J., M.M. Meerschaert and H. P. Scheffler (2003) The operator ν-stable laws, preprint.
21. Loretan, M. and P. Phillips (1994) Testing the covariance stationarity of heavy tailed time series. *J. Empirical Finance* **1**, 211–248.
22. Mandelbrot, B. (1963) The variation of certain speculative prices. *J. Business* **36**, 394–419.
23. McCulloch, J. (1997) Measuring tail thickness to estimate the stable index α: A critique. *J. Business Econ. Statist.* **15**, 74–81.
24. Meerschaert, M. (1990) Moments of random vectors which belong to some domain of normal attraction. *Ann. Probab.* **18**, 870–876.
25. Meerschaert, M. and H. P. Scheffler (1998) A simple robust estimator for the thickness of heavy tails. *J. Statist. Plann. Inf.* **71**, 19–34.
26. Meerschaert, M. and H. P. Scheffler (1999) Moment estimator for random vectors with heavy tails. *J. Multivariate Anal.* **71**, 145–159.
27. Meerschaert, M. and H. P. Scheffler (2001) *Limit Theorems for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice.* Wiley, New York.
28. Meerschaert, M. and H. P. Scheffler (2001) Sample cross-correlations for moving averages with regularly varying tails. *J. Time Series Anal.* **22**, no. 4, 481–492.
29. M.M. Meerschaert and H.P. Scheffler (2003) Portfolio modeling with heavy tailed random vectors. *Handbook of Heavy Tailed Distributions in Finance*, S.T.Rachev (Ed.), Elsevier Science.
30. Nikias, C. and M. Shao (1995) *Signal Processing with Alpha Stable Distributions and Applications.* Wiley, New York.
31. Nolan, J., Panorska A. and J. H. McCulloch (2001) Estimation of stable spectral measures. *Math. Comput. Modelling*, **34**, no. 9-11, 1113–1122.
32. Painter, S., V. Cvetkovic and J. Selroos (2002) Power-law velocity distributions in

fracture networks: Numerical evidence and implications for tracer transport. *Geophysical Research Letters* **29**, no. 14.

33. Rachev, S. and S. Mittnik (2000) *Stable Paretian Models in Finance*, Wiley, Chichester.
34. Resnick, S. and C. Stărică (1995) Consistency of Hill's estimator for dependent data. *J. Appl. Probab.* **32** 139–167.
35. Samorodnitsky, G. and M. Taqqu (1994) *Stable non-Gaussian Random Processes*. Chapman & Hall, New York.
36. Scheffler, H. P. (1999) On estimation of the spectral measure of certain nonnormal operator stable laws. *Statist. Probab. Lett.* **43**, 385–392.
37. Sharpe, M. (1969) Operator-stable probability distributions on vector groups. *Trans. Amer. Math. Soc.* **136**, 51–65.
38. Shlesinger, M., G. Zaslavsky, and U. Frisch (1994) Lévy flights and related topics in physics. *Lecture Notes in Physics* **450**, Springer-Verlag, Berlin.
39. Tessier, Y., S. Lovejoy, P. Hubert and D. Schertzer (1996) Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions. *J. Geophys. Res.*, **101**, 427–440.
40. Uchaikin, V. and V. Zolotarev (1999) *Chance and Stability: Stable Distributions and Their Applications*. VSP, Utrecht.