Journal of Multivariate Analysis **71**, 145–159 (1999)
Article ID jmva.1999.1835, available online at http://www.idealibrary.com on IDEAL®

Moment Estimator for Random Vectors with Heavy Tails*

Mark M. Meerschaert

University of Nevada
E-mail: mcubed@unr.edu

and

Hans-Peter Scheffler

University of Dortmund, Dortmund, Germany

Received November 5, 1997

If a set of independent, identically distributed random vectors has heavy tails, so that the covariance matrix does not exist, there is no reason to expect that the sample covariance matrix conveys useful information. On the contrary, this paper shows that the eigenvalues and eigenvectors of the sample covariance matrix contain detailed information about the probability tails of the data. The eigenvectors indicate a set of marginals which completely determine the moment behavior of the data, and the eigenvalues can be used to estimate the tail thickness of each marginal. The paper includes an example application to a data set from finance.

AMS 1991 subject classifications: 62F12, 62H11, 60E07, 90A09.

Key words and phrases: operator stable laws; generalized domains of attraction; regular variation; heavy tails.

1. INTRODUCTION

A probability distribution has heavy tails if some of its moments fail to exists. Heavy tail probability distributions are important in applications to electrical engineering, geology, hydrology, and physics; see, for example, Brockwell and Davis [1], Feller [3], Hosking and Wallis [6], Janicki and Weron [8], Leadbetter *et al.* [11], Nikias and Shao [20], Resnick and Stărică [24], and Samorodnitsky and Taqqu [25]. Mandelbrot [13] and

* This research was supported by a research scholarship from the Volkswagen Stiftung Research in Pairs program at Oberwolfach, Germany.

0047-259X/99 $30.00

Fama [2] pioneered the use of heavy tail distributions in finance. Jansen and de Vries [9], Loretan and Phillips [12], and McCulloch [14] present extensive empirical evidence of heavy tail price fluctuations in stock markets, futures markets, and currency exchange rates. Mittnik and Rachev [19] and Nolan *et al.* [21] discuss multivariable heavy tail models in finance. These models are used for portfolio analysis involving several different stock issues or mutual funds.

Moment estimation for heavy tail random vectors is complicated by the fact that the component with the heaviest tail tends to dominate. Suppose that $X=(U, V)$ where U is normal and V is Cauchy, so that EV does not exists. The first marginal U has light tails and hence a finite variance. The projection of X onto any other radial direction has heavy tails since it has both a normal and a Cauchy component, so that its mean does not exist. Now rotate X so that the normal component lies along the diagonal. Then the projections of X onto the coordinate axes (i.e., the marginal distributions) have heavy tails. If we only consider these marginal distributions, we will only detect the Cauchy-like tails. In order to discover the variations in tail behavior, we must consider every radial direction.

In this paper we show how to simultaneously estimate the thickness of heavy tails in every radial direction. Since tail behavior is dominated by the component with the heaviest tail, the tail thickness will be the same in almost every direction. It is surprising that the lighter tails are even detectable, since these directions lie on a set of measure zero on the unit sphere. However we will show that it is possible to detect the full range of tail behavior for a very general class of distributions. Our procedure yields a coordinate system in which the marginals determine the complete tail behavior, as well as a tail thickness estimate for each marginal. Then the tail behavior in any direction is determined by the heaviest tail marginal which has a nonvanishing component in this direction. The coordinate vectors are simply the eigenvectors of the sample covariance matrix, and the tail thickness in each eigenvector direction depends on the corresponding eigenvalue. Several different one dimensional tail estimators already exist, and these can also be used to verify our estimates by applying them to the marginals along the eigenvector directions.

At the end of this paper we present a practical application of our estimation procedure. Nolan *et al.* [21] fit a multivariable stable model to a portfolio of two currency exchange rates. This model assume a uniform tail thickness in every radial direction. We apply our multivariable moment estimator to the same data, and find that the tail thickness varies significantly with direction. This result is verified by applying standard one variables tail estimation methods in the directions indicated by our procedure. We conclude that a model which allows the tail thickness to vary with radial direction is useful to obtain a more accurate picture of the tail behavior.

2. RESULTS

In this section we present a very robust moment estimation method for random vectors with heavy tails. We assume only that the data belong to the generalized domain of attraction of some operator stable law, i.e. that the generalized central limit theorem applies. Suppose that X, X_1, X_2, X_3, ... are i.i.d. random vectors on $\mathbb{R}^d$ and that Y is full dimensional. We say that X belongs to the generalized domain of attraction of Y if there exist linear operators A_n on $\mathbb{R}^d$ and nonrandom vectors $a_n \in R^d$ such that

$$A_n(X_1 + \cdots + X_n) + a_n \Rightarrow Y. \tag{2.1}$$

The limit Y is operator stable with some exponent B. This means that, given Y_1, Y_2, Y_3, ... i.i.d. with Y, for every n there exists $b_n \in R^d$ such that $n^{-B}(Y_1 + \cdots + Y_n) - b_n$ has the same distribution as Y. Here $n^{-B} = \exp(-B \log n)$ where $\exp(A) = I + A + A^2/2! + A^3/3! + \cdots$ is the usual exponential operator. If $E\,\|X\|^2 < \infty$ then we have the classical central limit theorem with Y multivariate normal, $A_n = n^{-1/2} I$, $a_n = -nA_n EX$, $B = (-1/2)\, I$, and $b_n = 0$. In this paper we will focus on the case where Y is nonnormal, so that X has heavy tails.

Our first result is based on the spectral decomposition of Meerschaert [16]. Given an orthonormal basis $x_1, ..., x_d$ for $\mathbb{R}^d$ we write $\theta^{(i)} = \langle \theta, x_i \rangle$ for the ith coordinate of a vector $\theta \in \mathbb{R}^d$, and $X^{(i)} = \langle X, x_i \rangle$ for the ith marginal of X in this coordinate system.

THEOREM 1. *If X belongs to the generalized domain of attraction of some nonnormal operator stable law on $\mathbb{R}^d$, then there exists an orthonormal basis $x_1, ..., x_d$ and scalars $0 < \rho_d \leqslant \cdots \leqslant \rho_1 < 2$ such that $E\,|X^{(i)}|^\rho < \infty$ for $0 < \rho < \rho_i$ and $E\,|X^{(i)}|^\rho = \infty$ for $\rho > \rho_i$. Furthermore, for any unit vector θ we have $E\,|\langle X, \theta \rangle| < \infty$ for $0 < \rho < \rho(\theta)$ and $E\,|\langle X, \theta \rangle| = \infty$ for $\rho > \rho(\theta)$, where $\rho(\theta) = \min\{\rho_i : \theta^{(i)} \neq 0\}$.*

For purposes of comparison, note that a multivariable stable law with index α satisfies Theorem 1 in any orthonormal coordinate system, with every $\rho_i = \alpha$. Our next result yields the range of the index function $\rho(\theta)$. These numbers determine the moment behavior of X.

THEOREM 2. *Suppose X belongs to the generalized domain of attraction of some nonnormal operator stable law on $\mathbb{R}^d$, and let $\lambda_{n1} \leqslant \cdots \leqslant \lambda_{nd}$ denote the eigenvalues of the matrix $M_n = X_1 X_1' + \cdots + X_n X_n'$. Then $2 \log n / \log \lambda_{ni} \to \rho_i$ in probability for all $i = 1, ..., d$.*

Note that the tail indices $\rho_1, ..., \rho_d$ are not necessarily distinct. Let $\alpha_p < \cdots < \alpha_1$ denote the distinct values, i.e. the range of $\rho(\theta)$. Define

$V_j = \text{span}\{x_i : \rho_i = \alpha_j\}$ and π_j orthogonal projection onto V_j. Then X has moments up to order α_j on the space V_j. The following result shows that the orthonormal basis in Theorem 1 can be approximated by the eigenvectors of the sample covariance matrix, at least when $p \leqslant 3$. We suspect that this results holds for any value of p, but we have not been able to prove this.

THEOREM 3. *Suppose X belongs to the generalized domain of attraction of some nonnormal operator stable law on $\mathbb{R}^d$, and let $\theta_{n1}, ..., \theta_{nd}$ denote unit eigenvectors corresponding to the eigenvalues of the matrix $M_n = X_1 X_1' + \cdots + X_n X_n'$. Define $V_{nj} = \text{span}\{\theta_{ni} : \rho_i = \alpha_j\}$ and π_{nj} orthogonal projection onto V_{nj}. Then $\pi_{n1} \to \pi_1$ and $\pi_{np} \to \pi_p$ in probability. If $p \leqslant 3$ then $\pi_{nj} \to \pi_j$ in probability for all $j = 1, ..., p$.*

Because unit eigenvectors are not unique (we can always replace θ_{ni} by $-\theta_{ni}$), we state Theorem 3 in terms of projections. We can also restate this result directly in terms of the eigenvectors. For example, if $\rho_d < \cdots < \rho_1$ then all of the subspaces V_{nj} are one dimensional, and Theorem 3 implies that $\theta_{ni} \to x_i$ in probability for all $i = 1, ..., d$, assuming that the eigenvectors θ_{ni} are chosen with the appropriate sign. We can also restate Theorem 3 in terms of subspaces. Let $\mathscr{G}_{k,d}$ denote the space of all linear subspaces of R^d having dimension k. Topologize so that $V_n \to V$ in $\mathscr{G}_{k,d}$ if and only if $P_n \to P$ where P_n, P are the orthogonal projection operators onto V_n, V respectively. the space $\mathscr{G}_{k,d}$ is called a Grassman manifold. Theorem 3 shows that $V_{nj} \to V_j$ in probability. We can also estimate the entire index function $\rho(\theta)$ by combining Theorems 2 and 3. Since both the tail index and the direction in which it applies must be estimated, the following result is the best possible.

THEOREM 4. *Define $\hat{\rho}_n(\theta) = 2 \log n / \log \lambda_{ni}$ where i is the largest integer such that $\langle \theta, \theta_{ni} \rangle \neq 0$. If $p \leqslant 3$ then $\hat{\rho}_n \to \rho$ in the following sense: There exists a sequence of random linear operators I_n on $\mathbb{R}^d$ with $I_n \to I$ in probability such that $\hat{\rho}_n(I_n \theta) \to \rho(\theta)$ in probability for every unit vector $\theta \in \mathbb{R}^d$.*

If $E \|X\|^2 < \infty$ then the classical central limit theorem applies. Then $E \|X\|^\rho < \infty$ exists for all $0 < \rho < 2$, and the higher order moments may or may not exist. Although we have not attempted a complete analysis, the following result shows that our estimator remains useful in this case.

THEOREM 5. *Suppose $E \|X\|^2 < \infty$ and let $\lambda_{n1} \leqslant \cdots \leqslant \lambda_{nd}$ denote the eigenvalues of the matrix $M_n = X_1 X_1' + \cdots + X_n X_n'$. Then $2 \log n / \log \lambda_{ni} \to 2$ almost surely for all $i = 1, ..., d$.*

3. PROOFS

Recall that $X, X_1, X_2, X_3, ...$ are independent random vectors on $\mathbb{R}^d$ with common distribution μ, and Y is a random vector on $\mathbb{R}^d$ whose distribution ν is full, i.e. it cannot be supported on any $d-1$ dimensional affine subspace of $\mathbb{R}^d$. We assume that μ belongs to the generalized domain of attraction of ν, so that

$$A_n \mu^n * \delta(a_n) \to \nu, \tag{3.1}$$

where A_n are linear operators on $\mathbb{R}^d$, μ^n is the n-fold convolution product of μ with itself, $*$ denotes convolution, and $\delta(a_n)$ is the unit mass at $a_n \in R^d$. Sharpe [26] calls the class of all possible full limits in (3.1) the operator stable laws. Sharpe shows that a full operator stable law ν on a finite dimensional real vector space V is infinitely divisible, and there exists a (not necessarily unique) linear operator B on V called an exponent such that $\nu^t = t^B \nu * \delta(b_t)$ for all $t > 0$, where ν^t is the t-fold convolution power of ν, $t^B = \exp(B \log t)$, $t^B \nu(dx) = \nu(t^{-B} dx)$ and $b_t \in V$.

Proof of Theorem 1. Write the minimal polynomial of B as $f_1(x) \cdots f_p(x)$ where every root of f_j has real part a_j and $a_1 < \cdots < a_p$. Then $a_1, ..., a_p$ is the real spectrum of B and so $a_1 > 1/2$ since ν has no normal component; see, for example, Jurek and Mason [10, Theorem 4.6.5]. Define $V_j = \ker f_j(B)$ and let $d_j = \dim V_j$. Then $B = B_1 \oplus \cdots \oplus B_p$ where B_j is the restriction of B to V_j, and every eigenvalue of the $d_j \times d_j$ matrix B_j has real part equal to a_j. Theorem 4.2 of Meerschaert [16] shows that we can always choose ν spectrally compatible with μ, so that $A_n = A_{n1} \oplus \cdots \oplus A_{np}$ for all n, where A_{nj} is the restriction of A_n to V_j. The remark following the proof of that theorem shows that we may also assume without loss of generality that the subspaces V_j are mutually orthogonal. (Note that for ν spectrally compatible with μ the subspaces $V_j = W_j$ in that remark.) Let $\{x_i = 1 \leqslant i \leqslant b_1\}$ be an arbitrary orthonormal basis for V_1, and for $j = 2, ..., p$ take $\{x_i : b_{j-1} < i \leqslant b_j\}$ an arbitrary orthonormal basis for V_j, where $b_j = d_1 + \cdots + d_j$. Then $x_1, ..., x_d$ is an orthonormal basis for $\mathbb{R}^d$. Define $\alpha_j = 1/a_j$ for $j = 1, ..., p$ and let $\rho_i = \alpha_j$ for all $b_{j-1} < i \leqslant b_j$. Theorem 3.3 of Meerschaert [16] shows that for all $\theta \neq 0$ in $\mathbb{R}^d$

$$\int |\langle x, \theta \rangle|^\rho \, \mu(dx)$$

exists for $0 < \rho < \rho(\theta)$ and diverges for $\rho > \rho(\theta)$, where $\rho(\theta) = \min\{\rho_i : \theta^{(i)} \neq 0\}$, and $\theta^{(i)} = \langle \theta, x_i \rangle$ is the ith coordinate of the vector $\theta \in \mathbb{R}^d$ in this coordinate system. If $\theta = x_i$ then this integral is $E\,|X^{(i)}|^\rho$, and more generally it represents $E\,|\langle X, \theta \rangle|^\rho = E\,|\sum_i \theta^{(i)} X^{(i)}|^\rho$. Then $X^{(i)}$ has absolute moments up to order ρ_i, and linear combinations of the $X^{(i)}$ have moments up to the order of the component with the heaviest tail. This concludes the proof of Theorem 1.

Let $\mathscr{M}_s^d$ denote the vector space of $d \times d$ symmetric matrices with real entries. Define the outer product mapping $T: \mathbb{R}^d \to \mathscr{M}_s^d$ by $Tx = xx'$. Using the usual inner product on $\mathscr{M}_s^d$ we have $\langle M, N \rangle = \sum_{i,j} M_{ij} N_{ij}$ where M_{ij} denotes the ij entry of the matrix M. Then it is easy to check that $\langle Tx, Ty \rangle = \langle x, y \rangle^2$ so that $\|Tx\| = \|x\|^2$ in the associated Euclidean norms. Let $M_n = X_1 X_1' + \cdots + X_n X_n' = TX_1 + \cdots + TX_n$. Meerschaert and Scheffler [17] show that

$$A_n M_n A_n^* \Rightarrow W, \tag{3.2}$$

where W is a random element of the vector space $\mathscr{M}_s^d$. The limit W is full and operator stable on a subspace of $\mathscr{M}_s^d$ and W is invertible with probability one. Note that M_n is symmetric and nonnegative definite so that there exists an orthonormal basis of eigenvectors $\theta_{n1}, ..., \theta_{nd}$ corresponding to the nonnegative eigenvalues $\lambda_{n1} \leqslant \cdots \leqslant \lambda_{nd}$ of M_n.

Lemma 1. *For all $\theta \neq 0$ in $\mathbb{R}^d$ the random variable $\langle W, T\theta \rangle$ has a density with respect to Lebesgue measure.*

Proof. Let ϕ denote the Lévy spectral measure of v; see, for example, Jurek and Mason [10, p. 33]. Since v is full and has no normal component, ϕ cannot be concentrated on any $d-1$ dimensional subspace of $\mathbb{R}^d$. Define $L = \operatorname{span}\{\operatorname{supp}(T\phi)\}$ so that L is a linear subspace of $\mathscr{M}_s^d$ and $T\phi$ is full on L. Meerschaert and Scheffler [17] show that W is full and operator stable on L, and then Theorem 4.10.2 of Jurek and Mason [10] shows that W has a density on L. Since ϕ is full on $\mathbb{R}^d$ we can choose $y, ..., y_d$ linearly independent in $\mathbb{R}^d$ with all $y_i \in \operatorname{supp}(\phi)$. Otherwise $\operatorname{span}\{\operatorname{supp}(\phi)\}$ has dimension less than d and ϕ would not be full. Then, by the lemma to the proof of Theorem 2 in Meerschaert and Scheffler [17], $Ty_i \in L$ for all $i = 1, ..., d$. If $T\theta$ is perpendicular to L, then $\langle Ty_i, T\theta \rangle = \langle y_i, \theta \rangle^2 = 0$ for all $i = 1, ..., d$, which is a contradiction. Then $\langle W, T\theta \rangle$ is a one dimensional marginal of W on L and so it has a density by the Fubini theorem.

Lemma 2. $(\log \lambda_{nd}/2 \log n) \to a_p$ *in probability.*

Proof. For $\delta > 0$ arbitrary write

$$\begin{aligned} P&\left[\left|\frac{\log \lambda_{nd}}{2 \log n} - a_p\right| > \delta\right] \\ &\leqslant P[\log \lambda_{nd} > 2(a_p + \delta) \log n] + P[\log \lambda_{nd} < 2(a_p - \delta) \log n] \\ &= P[\max_{\|\theta\| = 1} \langle M_n \theta, \theta \rangle > n^{2(a_p + \delta)}] + P[\max_{\|\theta\| = 1} \langle M_n \theta, \theta \rangle < n^{2(a_p - \delta)}]. \end{aligned}$$

Now choose ρ so that $2a_p < \rho^{-1} < 2(a_p + \delta)$ and note that $E\,\|X\|^{2\rho} < \infty$ by Theorem 3 of Hudson *et al.* [7] and hence $E\,\|TX\|^{\rho} = E\,\|X\|^{2\rho} < \infty$. Furthermore since $a_p > 1/2$ we have $\rho < 1$ so that $|x+y|^{\rho} \leqslant |x|^{\rho} + |y|^{\rho}$. Then

$$\begin{aligned} P[\max_{\|\theta\|=1} \langle M_n\theta, \theta\rangle > n^{2(a_p+\delta)}] &= P[\,\|M_n\| > n^{2(a_p+\delta)}] \\ &\leqslant P\left[\sum_{i=1}^{n} \|TX_i\| > n^{2(a_p+\delta)}\right] \\ &\leqslant \left(\frac{1}{n^{2(a_p+\delta)}}\right)^{\rho} E\left|\sum_{i=1}^{n} \|TX_i\|\right|^{\rho} \\ &\leqslant n^{1-2\rho(a_p+\delta)} E\,\|TX\|^{\rho} \end{aligned}$$

which tends to zero as $n \to \infty$ by choice of ρ. Let $\bar{L}_j = V_1 \oplus \cdots \oplus V_j$. Theorem 3.2 of Meerschaert [16] shows that for any $x \in \bar{L}_j \backslash \bar{L}_{j-1}$ we have $n^{a_j - \varepsilon} < \|(A_n^*)^{-1}\, x\| < n^{a_j + \varepsilon}$ for all large n. (Note that since the subspaces V_i are mutually orthogonal we have $V_i = V_i^*$, thus for v spectrally compatible with μ we have $L_i^* = V_1 \oplus \cdots \oplus V_i$ in that result.) Choose $\theta_0 \in \bar{L}_p \backslash \bar{L}_{p-1}$ and write $(A_n^*)^{-1}\,\theta_0 = r_n\theta_n$ where $r_n > 0$ and $\|\theta_n\| = 1$. Choose $\varepsilon < \delta$ so that $n^{a_p - \varepsilon} < r_n < n^{a_p + \varepsilon}$ for all large n. Then

$$\begin{aligned} P[\max_{\|\theta\|=1} \langle M_n\theta, \theta\rangle < n^{2(a_p-\delta)}] &\leqslant P[\langle M_n\theta_0, \theta_0\rangle < n^{2(a_p-\delta)}] \\ &= P[\langle A_n M_n A_n^*\,\theta_n, \theta_n\rangle < r_n^{-2} n^{2(a_p-\delta)}] \\ &\leqslant P[\langle A_n M_n A_n^*\,\theta_n, \theta_n\rangle < n^{2(\varepsilon-\delta)}]. \end{aligned}$$

Since $\|\theta_n\| = 1$ for all n, given any subsequence there exists a further subsequence n' along which $\theta_{n'} \to \theta$, where $\|\theta\| = 1$. Then $\langle A_{n'} M_{n'} A_{n'}^* \theta_{n'}, \theta_{n'}\rangle = \langle A_{n'} M_{n'} A_{n'}^*, T\theta_{n'}\rangle \Rightarrow \langle W, T\theta\rangle$ by continuous mapping. For any $\varepsilon_1 > 0$ there exists $\rho > 0$ such that $P[\langle W, T\theta\rangle < \rho] < \varepsilon_1/2$, since by Lemma 1 $\langle W, T\theta\rangle$ has a Lebesgue density. Now choose n_0 such that $(n')^{2(\varepsilon-\delta)} < \rho$ and $|P[\langle A_{n'} M_{n'} A_{n'}^*, T\theta_{n'}\rangle < \rho] - P[\langle W, T\theta\rangle < \rho]| < \varepsilon_1/2$ for all $n' \geqslant n_0$. Then

$$\begin{aligned} P[\langle A_{n'} M_{n'} A_{n'}^*, T\theta_{n'}\rangle < (n')^{2(\varepsilon-\delta)}] &\leqslant P[\langle A_{n'} M_{n'} A_{n'}^*, T\theta_{n'}\rangle < \rho] \\ &\leqslant P[\langle W, T\theta\rangle < \rho] + \varepsilon_1/2 < \varepsilon_1 \end{aligned}$$

for $n' \geqslant n_0$. Since for any subsequence there is a further subsequence along which $P[\langle A_n M_n A_n^*\,\theta_n, \theta_n\rangle < n^{2(\varepsilon-\delta)}] \to 0$, this convergence holds along the entire sequence, which concludes the proof.

LEMMA 3. *$(\log \lambda_{n1}/2 \log n) \to a_1$ in probability.*

Proof. The proof is similar to Lemma 2. Without loss of generality M_n is invertible, and (3.2) along with continuous mapping implies that $(A_n M_n A_n^*)^{-1} \Rightarrow W^{-1}$. Write $P[|\log \lambda_{n1}/2 \log n - a_1| > \delta] \leqslant P[\lambda_{n1} > n^{2(a_1+\delta)}] + P[\lambda_{n1} < n^{2(a_1-\delta)}]$, use $1/\lambda_{n1} = \max\{\langle M_n^{-1}\theta, \theta\rangle : \|\theta\| = 1\}$ for the first term and $\lambda_{n1} = \min\{\langle M_n\theta, \theta\rangle : \|\theta\| = 1\}$ for the second term.

Proof of Theorem 2. Let $\mathscr{C}_i$ denote the collection of all orthogonal projection operators onto subspaces of $\mathbb{R}^d$ with dimension i. The Courant–Fischer Max–Min Theorem of linear algebra (see Rao [22]) implies that

$$\begin{aligned}\lambda_{ni} &= \min_{P \in \mathscr{C}_i} \max_{\|\theta\|=1} \langle PM_nP\theta, \theta\rangle \\ &= \max_{P \in \mathscr{C}_{d-i+1}} \min_{\|\theta\|=1} \langle PM_nP\theta, \theta\rangle. \end{aligned} \tag{3.3}$$

Let P_j denote the orthogonal projection operator onto $L_j = V_j \oplus \cdots \oplus V_p$ and $b_j = \dim(L_j) = d_j + \cdots + d_p$, where j is chosen so that $\rho_i = \alpha_j$. Since P_j commutes with both B and A_n, it follows immediately from (3.1) that $P_j\mu$ belongs to the generalized domain of attraction of the operator stable law $P_j\nu$ on L_j whose exponent $B_j \oplus \cdots \oplus B_p$ has real spectrum $a_j, ..., a_p$. Let λ_n denote the smallest eigenvalue of the matrix $P_jM_nP_j$ and apply Lemma 3 above to see that $\log \lambda_n(2 \log n)^{-1} \to a_j$ in probability. Now let $\bar{P}_j$ denote the orthogonal projection onto $\bar{L}_j = V_1 \oplus \cdots \oplus V_j$ and $\bar{b}_j = \dim(\bar{L}_j) = d_1 + \cdots + d_j$. Then $\bar{P}_j\mu$ belongs to the generalized domain of attraction of the operator stable law $\bar{P}_j\nu$ whose exponent $B_1 \oplus \cdots \oplus B_j$ has real spectrum $a_1, ..., a_j$, and we can apply Lemma 2 above to see that $\log \bar{\lambda}_n(2 \log n)^{-1} \to a_j$ in probability, where $\bar{\lambda}_n$ is the largest eigenvalue of the matrix $\bar{P}_jM_n\bar{P}_j$. Now apply (3.3) to see that

$$\lambda_n \leqslant \lambda_{n, \bar{b}_{j-1}+1} \leqslant \lambda_{ni} \leqslant \lambda_{n, \bar{b}_j} \leqslant \bar{\lambda}_n$$

using the fact that $\bar{b}_{j-1} < i \leqslant \bar{b}_j$. The theorem follows easily.

LEMMA 4. *If $i > \bar{b}_{d-1}$ and $j < p$ then $\pi_j\theta_{ni} \to 0$ in probability.*

Proof. Since $M_n\theta_{ni} = \lambda_{ni}\theta_{ni}$ we can write $\pi_j\theta_{ni} = (\pi_jM_n/\lambda_{ni})\,\theta_{ni}$ where

$$\begin{aligned}\|\pi_jM_n/\lambda_{ni}\| &= \|A_n^{-1}A_n\pi_jM_nA_n^*(A_n^*)^{-1}\|/\lambda_{ni} \\ &= \|\pi_jA_n^{-1}A_nM_nA_n^*(A_n^*)^{-1}\|/\lambda_{ni} \\ &\leqslant \|\pi_jA_n^{-1}\|\;\|A_nM_nA_n^*\|\;\|(A_n^*)^{-1}\|/\lambda_{ni},\end{aligned}$$

where $\|A_n M_n A_n^*\| \Rightarrow \|W\|$ by (3.2) and the continuous mapping theorem. Apply Theorem 3.2 of Meerschaert [16] as in the proof of Lemma 2 to see that for any $x \in \bar{L}_j \backslash \bar{L}_{j-1}$ we have $n^{a_j - \varepsilon} < \|(A_n^*)^{-1} x\| < n^{a_j + \varepsilon}$ for all large n, or equivalently $\log \|(A_n^*)^{-1} x\| / \log n \to a_j$. This convergence is uniform on compact subsets of $x \in \bar{L}_j \backslash \bar{L}_{j-1}$ and so we also have $\log \|\pi_j (A_n^*)^{-1}\| / \log n \to a_j$ and $\log \|(A_n^*)^{-1}\| / \log n \to a_p$. Similarly, Theorem 4.1 of Meerschaert [16] shows that for any $\varepsilon > 0$, $x \in L_i \backslash L_{i+1}$ we have $\log \|A_n x\| / \log n \to -a_j$. This convergence is uniform on compact subsets of $x \in L_j \backslash L_{j+1}$ and so we also have $\log \|\pi_j A_n\| / \log n \to -a_j$ and $\log \|A_n\| / \log n \to -a_1$. Then

$$\frac{\log(\|\pi_j A_n^{-1}\| \; \|(A_n^*)^{-1}\| / \lambda_{ni})}{\log n} = \frac{\log \|\pi_j A_n^{-1}\|}{\log n} + \frac{\log \|(A_n^*)^{-1}\|}{\log n} - \frac{\log \lambda_{ni}}{\log n}$$

$$\to a_j + a_p - 2a_p < 0$$

and so $\pi_j M_n / \lambda_{ni} \to 0$ in probability. Since $\|\theta_{ni}\| = 1$ this implies that

$$\|\pi_j \theta_{ni}\| = \|(\pi_j M_n / \lambda_{ni}) \, \theta_{ni}\| \leqslant \|\pi_j M_n / \lambda_{ni}\| \; \|\theta_{ni}\| \to 0$$

in probability, which concludes the proof.

LEMMA 5. *If* $i \leqslant \bar{b}_1$ *and* $j > 1$ *then* $\pi_j \theta_{ni} \to 0$ *in probability.*

Proof. The proof is similar to Lemma 4.

Proof of Theorem 3. We will show that $\pi_{np} \to \pi_p$ in probability. The proof that $\pi_{n1} \to \pi_1$ in probability is similar. Convergence in probability is equivalent to the fact that for any subsequence there exists a further subsequence along which we have almost sure convergence. For any $i > \bar{b}_{p-1}$ and $j < p$ we have $\pi_j \theta_{ni} \to 0$ in probability and so $\|\pi_p \theta_{ni}\| \to 1$ in probability. Then for any subsequence n_1 there exists a further subsequence n_2 such that $\|\pi_p \theta_{n_2, i}(\omega)\| \to 1$ for all $i > \bar{b}_{p-1}$ and all $\omega \in \Omega_0$ where $P[\Omega_0] = 1$. Then $\pi_p \theta_{n_2, i}(\omega)$ is relatively compact so given any subsequence n_3 of n_2 there exists a further subsequence n_4 depending on ω such that $\pi_p \theta_{n_4, i}(\omega) \to \theta_i(\omega)$ for all $i > \bar{b}_{p-1}$. By continuity we have $\|\theta_i(\omega)\| = 1$ and $\langle \theta_i(\omega), \theta_k(\omega) \rangle = 0$ for $i \neq k$, and since $\pi_p = \pi_p^2$ we also have $\pi_p \theta_{n_4, i}(\omega) \to \pi_p \theta_i(\omega)$ so that $\pi_p \theta_i(\omega) = \theta_i(\omega)$, hence $\theta_i(\omega) \in V_p$ and so the vectors $\theta_i(\omega)$ form an orthonormal basis for V_p. Then for any $x \in \mathbb{R}^d$ we have

$$\pi_{n_4, p}(\omega) \, x = \sum_{\bar{b}_{p-1} < i \leqslant \bar{b}_p} \langle x, \theta_{n_4, i}(\omega) \rangle \, \theta_{n_4, i}(\omega)$$

$$\to \sum_{\bar{b}_{p-1} < i \leqslant \bar{b}_p} \langle x, \theta_i(\omega) \rangle \, \theta_i(\omega) = \pi_p x.$$

We have shown that for all $\omega \in \Omega_0$, for any subsequence n_3 of n_2 there exists a further subsequence n_4 such that $\pi_{n_4,\,p}(\omega) \to \pi_p$. This means that for all $\omega \in \Omega_0$ we have $\pi_{n_2,\,p}(\omega) \to \pi_p$. Then for every subsequence n_1 we have shown that there exists a further subsequence n_2 such that $\pi_{n_2,\,p} \to \pi_p$ almost surely, so $\pi_{np} \to \pi_p$ in probability. Finally if $p \leqslant 3$ we need to show that $\pi_{nj} \to \pi_j$ in probability for all $j = 1, ..., p$. If $p < 3$ there is nothing to prove. If $p = 3$ then $\pi_{nj} \to \pi_j$ in probability for $j = 1$ and $j = 3$ and so $\pi_{n2} = I - \pi_{n1} - \pi_{n3} \to I - \pi_1 - \pi_3 = \pi_2$ in probability as well.

Proof of Theorem 4. Given $\theta \neq 0$ choose j so that $\rho(\theta) = 1/a_j = \alpha_j$, and let $\bar{P}_{nj} = \pi_{n1} + \cdots + \pi_{nj}$ denote orthogonal projection onto the space $\bar{L}_j = V_1 \oplus \cdots \oplus V_j$. If $p \leqslant 3$ then Theorem 3 implies that $\bar{P}_{nj} \to \bar{P}_j$ in probability, and so $\theta_n \to \bar{P}_j \theta$ in probability where $\theta_n = \bar{P}_{nj}\theta$. Since $\theta \in \bar{L}_j$ we have $\bar{P}_j \theta = \theta$ so in fact $\theta_n \to \theta$ in probability. Since both $\pi_{nj} \to \pi_j$ and $\theta_n \to \theta$ in probability and since

$$\begin{aligned}\|\pi_{nj}\theta_n - \pi_j\theta\| &\leqslant \|\pi_{nj}\theta_n - \pi_j\theta_n\| + \|\pi_j\theta_n - \pi_j\theta\| \\ &\leqslant \|\pi_{nj} - \pi_j\|\ \|\theta_n\| + \|\pi_j\|\ \|\theta_n - \theta\| \\ &\leqslant \|\pi_{nj} - \pi_j\| + \|\theta_n - \theta\|\end{aligned}$$

we also have $\pi_{nj}\theta_n \to \pi_j\theta$ in probability. Since $\theta \notin \bar{L}_{j-1}$ we have $\pi_j\theta \neq 0$, and so given any $\varepsilon > 0$ any $\delta > 0$ (choose δ smaller than $\|\pi_j\theta\|$) we have $P[\pi_{nj}\theta_n = 0] \leqslant P[\|\pi_{nj}\theta_n - \pi_j\theta\| > \delta] < \varepsilon/2$ for all large n. By construction we always have $\pi_{nr}\theta_n = 0$ when $r > j$. Theorem 2 implies that $\log \lambda_{ni}/(2 \log n) \to a_j$ in probability for all $\bar{b}_{j-1} < i \leqslant \bar{b}_j$, and so we also have

$$P\left[\left|\frac{2 \log n}{\log \lambda_{ni}} - \frac{1}{a_j}\right| > \delta\right] < \frac{\varepsilon}{2d_j}$$

for all $\bar{b}_{j-1} < i \leqslant \bar{b}_j$, for all large n. Then we have

$$\begin{aligned}&P[|\hat{\rho}_n(\theta_n) - \rho(\theta)| > \delta] \\ &\quad\leqslant P[\pi_{nj}\theta_n = 0] + P[\pi_{nr}\theta_n \neq 0\ \exists r > j] \\ &\qquad + P\left[\left|\frac{2 \log n}{\log \lambda_{ni}} - \frac{1}{a_j}\right| > \delta \text{ for some } \bar{b}_{j-1} < i \leqslant \bar{b}_j\right] \\ &\quad\leqslant P[\pi_{nj}\theta_n = 0] + 0 + \sum_{\bar{b}_{j-1} < i \leqslant \bar{b}_j} P\left[\left|\frac{2 \log n}{\log \lambda_{ni}} - \frac{1}{a_j}\right| > \delta\right] \\ &\quad< \frac{\varepsilon}{2} + 0 + d_j \frac{\varepsilon}{2d_j} = \varepsilon\end{aligned}$$

for all large n, so that $\hat{\rho}(\theta_n) \to \tau(\theta)$ in probability. Now let $I_n\theta = \bar{P}_{nj}\theta$ when $\rho(\theta) = 1/a_j$, so that $\theta_n = I_n\theta$. Since $I\theta = \bar{P}_j\theta$ for such θ we see that $\|I_n - I\|$ is bounded above by the maximum of $\|\bar{P}_{nj} - \bar{P}_j\|$ over $j = 1, ..., p$ and hence $I_n \to I$ in probability.

Proof of Theorem 5. The strong law of large numbers yields that $M_n/n \to M$ almost surely. Then $\|M_n\|/n \to \|M\|$ almost surely and so

$$2 \log n \left(\frac{\log \lambda_{nd}}{2 \log n} - \frac{1}{2} \right) = 2 \log n \left(\frac{\log \|M_n\|}{2 \log n} - \frac{1}{2} \right)$$

$$= \log(\|M_n\|/n) \to \log \|M\|$$

almost surely. It follows easily that $2 \log n / \log \lambda_{nd} \to 2$. Since X is full, M is invertible. In fact, if $\theta \neq 0$ then $\theta' M \theta = E(\theta' X X' \theta) = E \langle X, \theta \rangle^2 > 0$ since X is full, so M is positive definite. Then $\|(M_n/n)^{-1}\| = n \|M_n^{-1}\| \to \|M^{-1}\|$ almost surely and so

$$2 \log n \left(\frac{\log \lambda_{n1}}{2 \log n} - \frac{1}{2} \right) = 2 \log n \left(\frac{-\log \|M_n^{-1}\|}{2 \log n} - \frac{1}{2} \right)$$

$$= -\log(n \|M_n^{-1}\|) \to -\log \|M^{-1}\|$$

almost surely. Then $2 \log n / \log \lambda_{n1} \to 2$, and the theorem follows from the fact that $\lambda_{n1} \leqslant \lambda_{ni} \leqslant \lambda_{nd}$.

4. APPLICATION TO FINANCE

Nolan *et al.* [21] examine $n = 2853$ daily fluctuations in the exchange rates of two foreign currencies versus the US Dollar. They consider the vector data $X_t = (D_t, Y_t)$ where D_t, Y_t denote the exchange rate fluctuations on day t for the Deutsche Mark and Yen, respectively. They fit a model which assumes that $X_1, ..., X_n$ are i.i.d. according to some multivariate α stable distribution, i.e., an operator stable distribution with exponent $B = \alpha^{-1} I$. They pool estimates of α along the coordinate axes and conclude that α is near 1.6. Then they estimate the Lévy spectral measure of the multivariable stable law assuming that the stable index α is known, and find that the preponderance of probability mass lies near the diagonal line with slope $+1$. This reflects that fact that D_t and Y_t are fairly highly correlated.

We apply our estimator to the same exchange rate data assuming only that X_t belongs to the generalized domain of attraction of some operator stable law. We begin by scaling the data. We divide each entry by 0.004 which is the approximate median for both D_t and Y_t. This has no effect on the eigenvectors but helps to obtain good estimates of the tail thickness. Then we compute

$$M_n = \begin{pmatrix} \sum_t D_t^2 & \sum_t D_t Y_t \\ \sum_t D_t Y_t & \sum_t Y_t^2 \end{pmatrix} = \begin{pmatrix} 9142 & 5990 \\ 5990 & 8590 \end{pmatrix}$$

which has eigenvalues $\lambda_{n1} = 2870$, $\lambda_{n2} = 14862$ and associated unit eigenvectors $\theta_{n1} = [0.69, -0.72]'$, $\theta_{n2} = [0.72, 0.69]'$. Now we compute that $2 \ln 2853 / \ln 2870 = 1.998$ and $2 \ln 2853 / \ln 14862 = 1.656$. Then we estimate the tail indices $\alpha_1 \doteq 2.0$ and $\alpha_2 \doteq 1.65$, and the corresponding coordinate directions $x_1 \doteq [1, -1]'$ and $x_2 \doteq [1, 1]'$. Since α_1 is much larger than α_2, the tail behavior of X_t varies significantly with radial direction.

Rotating to the new coordinates we let $N_t = (D_t - Y_t)/\sqrt{2}$ and $S_t = (D_t + Y_t)/\sqrt{2}$. Our estimator predicts that N_t has a finite second moment but S_t only has moments up to order 1.65. This is in complete accord with the findings of Nolan *et al.* Since both marginals $D_t = (S_t + N_t)/\sqrt{2}$ and $Y_t = (S_t - N_t)/\sqrt{2}$ have a nonvanishing S_t component, both have heavy tails with the same tail index $\alpha = 1.65$. Pooling tail estimates from each marginal is misleading in this case, since it fails to detect the lighter tails along the diagonal with slope -1.

We verify our results by applying an alternative one variable tail estimator in the rotated coordinate directions. The most commonly used robust tail estimator, due to Hill [5], estimates the tail index α using the largest order statistics. We apply Hill's estimator to the largest 525 order statistic for each of D_t, Y_t, S_t, and N_t to obtain tail index estimates of 1.6, 1.6, 1.7, and 2.0, respectively. This provides additional evidence that the tails of N_t are lighter than those of S_t, and that the heavier tails determine the tail behavior of both marginals D_t and Y_t. Using asymptotic results of Hall [4] we can also compute individual 90% confidence intervals for each tail estimate, yielding (1.49, 1.72), (1.49, 1.72), (1.60, 1.84), and (1.87, 2.16), respectively. Since the last two intervals do not overlap, we can be reasonably sure that $\alpha_1 \neq \alpha_2$.

5. REMARKS

Our estimator is based on the eigenvalues and eigenvectors of the matrix $M_n = \sum_i X_i X_i'$. The (uncentered) sample covariance matrix $\Gamma_n^* = n^{-1} M_n$ has the same eigenvectors as M_n, so if we substitute Γ_n^* for M_n then all of the results of this paper still holds, except that in Theorem 2 we have $2 \log n/(\log \lambda_{ni} - \log n) \to \rho_i$ where $\lambda_{n1} \leqslant \cdots \leqslant \lambda_{nd}$ are the eigenvalues of Γ_n^*. Theorem 3 of Meerschaert and Scheffler [18] implies that the asymptotic behavior of Γ_n^* is identical to that of the usual sample covariance matrix $\Gamma_n = n^{-1} \sum_i (X_i - \bar{X}_n)(X_i - \bar{X}_n)'$, where $\bar{X}_n = n^{-1} \sum_i X_i$ is the sample mean. Then all of the results of this paper still hold with M_n replaced by Γ_n, except that in Theorem 2 we have $2 \log n/(\log \lambda_{ni} - \log n) \to \rho_i$ where $\lambda_{n1} \leqslant \cdots \leqslant \lambda_{nd}$ are the eigenvalues of Γ_n.

The proof of Theorem 3 also provides information on the rate of convergence of the eigenvectors. For example if $p = 2$ then $n^{a_2 - a_1 - \varepsilon}(\pi_{nj} - \pi_j) \to 0$ in probability for all j, for any $\varepsilon > 0$. In our introductory example (see Section 1) we have $a_1 = 1/2$ for the normal component and $a_2 = 1$ for the Cauchy component, so the rate of convergence is $\sqrt{n}$. Theorem 3 can also be used together with alternative tail estimation methods. Once we change to the coordinates determined by the eigenvectors of the sample covariance matrix, we can apply any one variable tail estimator to determine the tail thickness for each marginal. See McCulloch [15] for a recent survey and critique of several one variable tail estimation methods.

For a multivariable stable law with index α, the tail thickness is the same in every radial direction. One important application of Theorem 3 is to test this hypothesis. We have proven (for p arbitrary) that the eigenvectors corresponding to the smallest and largest eigenvalues of the sample covariance matrix are consistent estimators of the directions in which the tails of X are lightest and heaviest, respectively. Project the data onto these two eigenvector directions, and then apply any of the standard one variable tail estimators to each of these two marginals. If the difference in the tails is statistically signicant, we can reject the hypothesis that the data are multivariable stable.

Models which do not allow the tail thickness parameter α to vary with direction are too restrictive for many practical applications. Resnick and Greenwood [23] were the first to study the asymptotics of a class of models which are now called marginally stable. This means that the marginals of the limit Y in (2.1) are univariate stable. If the norming operators A_n are diagonal this is always so, as may easily be seen by projecting (2.1) onto the coordinate axes. Then the exponent B of the operator stable limit Y is also diagonal with entries a_j where $\alpha_j = 1/a_j$ ranges over the tail index of the stable marginals. The main result of this paper, Theorem 3, shows how to estimate the coordinate system in which a given data set is marginally stable, or in the domain of attraction of a marginally stable law. Further research into the properties of marginally stable laws and their domains of attraction would be useful to provide more robust methods for multivariable data analysis in the presence of heavy tails.

ACKNOWLEDGMENTS

The authors thank John Nolan, Department of Mathematics and Statistics, American University for providing the data on exchange rates, and Ed Keppelmann, Department of Mathematics, University of Nevada for helpful discussions.

REFERENCES

1. P. Brockwell and R. Davis, "Time Series: Theory and Methods," 2nd ed., Springer-Verlag, New York, 1991.
2. E. Fama, The behavior of stock market prices, *J. Business* **38** (1965), 34–105.
3. W. Feller, "An Introduction to Probability Theory and Its Applications," 2nd ed., Vol. II, Wiley, New York, 1971.
4. P. Hall, On some simple estimates of an exponent of regular variation, *J. Roy. Statist. Soc. B* **44** (1982), 37–42.
5. B. Hill, A simple general approach to inference about the tail of a distribution, *Ann. Statist.* **3** (1975), 1163–1173.
6. J. Hosking and J. Wallis, Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics* **29** (1987), 339–349.
7. W. Hudson, J. Veeh, and D. Weiner, Moments of distributions attracted to operator stable laws, *J. Multivariate Anal.* **24** (1988), 1–10.
8. A. Janicki and A. Weron, "Simulation and Chaotic Behavior of α-Stable Stochastic Processes," Dekker, New York, 1994.
9. D. Jansen and C. de Vries, On the frequency of large stock market returns: Putting booms and busts into perspective, *Rev. Econ. Statist.* **23** (1991), 18–24.
10. Z. Jurek and J. D. Mason, "Operator-Limit Distributions in Probability Theory," Wiley, New York, 1993.
11. M. Leadbetter, G. Lindgren, and H. Rootzén, "Extremes and Related Properties of Random Sequences and Processes," Springer-Verlag, New York, 1980.
12. M. Loretan and P. Phillips, Testing the covariance stationarity of heavy tailed time series, *J. Empirical Finance* **1** (1994), 211–248.
13. B. Mandelbrot, The variation of certain speculative prices, *J. Business* **36** (1963), 394–419.
14. J. McCulloch, Financial applications of stable distributions, *in* "Statistical Methods in finance: Handbook of Statistics" (G. S. Madfala and C. R. Rao, Eds.), Vol. 14, pp. 393–425, Elsevier, Amsterdam, 1996.
15. J. McCulloch, Measuring tail thickness to estimate the stable index α: A critique, *J. Business Econ. Statist.* **15** (1997), 74–81.
16. M. Meerschaert, Spectral decomposition for generalized domains of attraction, *Ann. Probab.* **19** (1991), 875–892.
17. M. Meerschaert and H. P. Scheffler, Sample covariance matrix for random vectors with regularly varying tails, *J. Theoret. Probab.*, in press.
18. M. Meerschaert and H. P. Scheffler, Moving averages of random vectors with regularly varying tails, *J. Time Series Anal.*, in press.
19. S. Mittnik and S. Rachev, "Assets and Option Pricing with Alternative Stable Models," Wiley, New York, 1998.
20. C. Nikias and M. Shao, "Signal Processing with Alpha Stable Distributions and Applications," Wiley, New York, 1995.
21. J. Noland, A. Panorska, and J. H. McCulloch, Estimation of stable spectral measures, preprint 1994.

22. C. R. Rao, "Linear Statistical Inference and Its Applications," Wiley, New York, 1965.
23. S. Resnick and P. Greenwood, A bivariate stable characterization and domains of attraction, *J. Multivariate Anal.* **9** (1979), 206–221.
24. S. Resnick and C. Stărică, Consistency of Hill's estimator for dependent data, *J. Appl. Probab.* **32** (1995), 139–167.
25. G. Samorodnitsky and M. Taqqu, "Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance," Chapman & Hall, London, 1994.
26. M. Sharpe, Operator-stable probability distributions on vector groups, *Trans. Amer. Math. Soc.* **136** (1969), 51–65.