PORTFOLIO MODELING WITH HEAVY TAILED RANDOM VECTORS

MARK M. MEERSCHAERT AND HANS-PETER SCHEFFLER

Abstract. Since the work of Mandelbrot in the 1960's there has accumulated a great deal of empirical evidence for heavy tailed models in finance. In these models, the probability of a large fluctuation falls off like a power law. The generalized central limit theorem shows that these heavy-tailed fluctuations accumulate to a stable probability distribution. If the tails are not too heavy then the variance is finite and we find the familiar normal limit, a special case of stable distributions. Otherwise the limit is a nonnormal stable distribution, whose bell-shaped density may be skewed, and whose probability tails fall off like a power law. The most important model parameter for such distributions is the tail thickness α, which governs the rate at which the probability of large fluctuations diminishes. A smaller value of α means that the probability tails are fatter, implying more volatility. In fact, when $\alpha < 2$ the theoretical variance is infinite. A portfolio can be modeled using random vectors, where each entry of the vector represents a different asset. The tail parameter α usually depends on the coordinate. The wrong coordinate system can mask variations in α, since the heaviest tail tends to dominate. A judicious choice of coordinate system is given by the eigenvectors of the sample covariance matrix. This isolates the heaviest tails, associated with the largest eigenvalues, and allows a more faithful representation of the dependence between assets.

1. Introduction

In order to construct a useful probability model for an investment portfolio, we must consider the dependence between assets. If we accept the premise that price changes are heavy tailed, then we are lead to consider random vectors with heavy tails. In this paper, we survey those portions of the theory of heavy tailed random vectors that seem relevant to portfolio analysis. The most flexible models recognize the possibility that the thickness of probability tails varies in different directions, implying the need for matrix scaling. A judicious change of coordinates often simplifies the model, and may uncover features masked by the original coordinates. The original coordinates are the price changes (or returns) for each asset. The new coordinates can be interpreted

Date: 12 December 2000.

Key words and phrases. multivariable regular variation, moment estimates, moving averages, generalized domains of semistable attraction, R-O varying measures.

as market indices, chosen to capture certain features of the market. In some popular heavy-tailed finance models, the tails are so heavy that the theoretical variance of price changes is undefined. For these models, the theoretical covariance matrix is also undefined. Of course the sample variance and the sample covariance matrix can always be computed for any data set, but these statistics are not estimating the usual model parameters. One of the most interesting discoveries in heavy tailed modeling is that, in the infinite variance case, the sample covariance matrix actually contains quite a bit of important information about the underlying distribution. In fact, the eigenvectors of this matrix provide a very useful coordinate system. We illustrate the application of this principle, and we also include a previously unpublished proof, extending the method to more general heavy tailed vector models with time dependence.

2. Heavy tails

A probability distribution has heavy tails if some of its moments fail to exist. Suppose that X is a random variable with density $f(x)$ so that

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

The kth moment of the random variable X is defined by an improper integral

$$\mu_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx.$$

The mean $\mu = \mu_1$, variance $\sigma^2 = \mu_2 - \mu_1^2$, skewness and kurtosis depend on these moments. Because μ_k is an improper integral, it may not exist. If $f(x)$ is a normal density, a lognormal density, or any other density whose tails fall off exponentially then all of the moments μ_k exist. But if $f(x)$ has heavy tails that fall off like a power law, then some of the moments μ_k will not exist. The simplest example of a heavy tailed distribution is a Pareto, invented to model the distribution of incomes. A Pareto random variable satisfies $P(X > x) = Cx^{-\alpha}$ so that the probability of large outcomes falls off like a power law. The Pareto density is defined by

$$f(x) = \begin{cases} C\alpha x^{-\alpha-1} & \text{for } x > C^{1/\alpha} \\ 0 & \text{otherwise} \end{cases}$$

so that

$$\mu_k = \int_{C^{1/\alpha}}^{\infty} C\alpha x^{k-\alpha-1}dx = \alpha C^{k/\alpha} \int_1^{\infty} y^{k-\alpha-1}dy = \alpha C^{k/\alpha} \left[\frac{y^{k-\alpha}}{k-\alpha}\right]_{y=1}^{\infty}$$

using the substitution $x = C^{1/\alpha}y$. If $k < \alpha$ then the limit at infinity is zero and $\mu_k = \alpha C^{k/\alpha}/(\alpha - k)$, but if $k \geq \alpha$ then this improper integral diverges, so that the kth moment does not exist.

Pareto distributions are closely related to some other familiar distributions. If U has a uniform distribution on $(0,1)$, then $X = U^{-1/\alpha}$ has a Pareto distribution with tail parameter α. To check this, write

$$P(X > x) = P(U^{-1/\alpha} > x) = P(U < x^{-\alpha}) = x^{-\alpha}.$$

If X is Pareto with $P(X > x) = x^{-\alpha}$, then $Y = \ln X$ has an exponential distribution with rate α. To see this, note that

$$P(Y > y) = P(\ln X > y) = P(X > e^y) = (e^y)^{-\alpha} = e^{-\alpha y}.$$

Some other familiar distributions have Pareto-like power law tails, causing some moments to diverge. If Y has a Student-t distribution with ν degrees of freedom, then $P(|Y| > y) \sim Cy^{-\alpha}$ where $\alpha = \nu$.[1] Then $E(Y^k)$ exists only for $k < \nu$. If Y has a Gamma distribution with density proportional to $y^{p-1}e^{-qy}$ then the log-Gamma random variable X defined by $Y = \ln X$ satisfies $P(X > x) \sim Cx^{-\alpha}$ for x large, where $\alpha = q$. Some other distributions with Pareto-like tails are the stable and operator stable distributions, which will be discussed later in this paper.

Heavy tailed random variables with $P(|X| > x) \sim Cx^{-\alpha}$ are observed in many real world applications. Estimation of the tail parameter α is important, because it determines which moments exist. Anderson and Meerschaert [5] 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. [74] find heavy tails with $2 < \alpha < 4$ for a variety of river flows and rainfall accumulations. Hosking and Wallis [28] find evidence of heavy tails with $\alpha \approx 5$ for annual flood levels of a river in England. Benson, et al. [9, 10] 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. 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 [11, 32, 73]. More applications to physics with $0 < \alpha < 2$ are cataloged in Uchaikin and Zolotarev [75]. Resnick and Stărică [66] 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 [2]. More applications to signal processing can be found in Nikias and Shao [54].

Mandelbrot [38] and Fama [18] pioneered the use of heavy tail distributions in finance. Mandelbrot [38] 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 [30] argue that daily returns for many stocks and stock indices have heavy tails with $3 < \alpha < 5$, and

[1] Here $f(x) \sim g(x)$ means that $f(x)/g(x) \to 1$ as $x \to \infty$.

discuss the possibility that the October 1987 stock market plunge might be just a heavy tailed random fluctuation. Loretan and Phillips [37] 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 [62] use different methods to find heavy tails with $1 < \alpha < 2$ for a variety of stocks, stock indices, and exchange rates. McCulloch [40] uses similar methods to re-analyze the data in [30, 37], 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.

Portfolio analysis involves the joint probability distribution of several prices or returns $X_1, \ldots, X_d$, where d is the number of assets in the portfolio. It is natural to model this set of numbers as a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$. We say that $\boldsymbol{X}$ has heavy tails if $E(\|\boldsymbol{X}\|^k)$ is undefined for some $k = 1, 2, 3, \ldots$. Let us consider the practical problem of portfolio modeling. We choose d assets and research historical performance to obtain data of the form $X_i(t)$ where $i = 1, \ldots, d$ is the asset and $t = 0, \ldots, n$ is the time variable. Typically the distribution of values $X_i(0), \ldots, X_i(n)$ has a heavy tail whose parameter α_i can be estimated from this data. The research of Jansen and de Vries [30], Loretan and Phillips [37], and Rachev and Mittnik [62] indicates, not surprisingly, that α_i will vary depending on the asset. Then the random vectors $\boldsymbol{X}_t = (X_1(t), \ldots, X_d(t))'$ will have heavier tails in some directions than in others. Despite this well known fact, most existing research on heavy tailed portfolio modeling has assumed that the probability tails are the same in every direction. Nolan, Panorska and McCulloch [58] consider such a model, based on the multivariable stable distribution, for a vector of two exchange rates. They argue that α is the same for both.[2] Rachev and Mittnik [62] use a multivariable stable model for portfolio analysis, so that α is the same for every asset. The same approach was also applied to portfolio analysis by Bawa, Elton and Gruber [7], Belkacem, Véhel and Walter [8], Chamberlain, Cheung and Kwan [14], Fama [19], Gamba [22], Press [60], Rachev and Han [63], and Ziemba [77]. If this modeling approach can be enhanced to allow α_i to vary with the asset, a more realistic and flexible representation of financial portfolios can be achieved. The goal of this paper is to show how this can be accomplished, using modern central limit theory.

[2] Example 8.1 gives an alternative operator stable model for the same data set.

3. Central limit theorems

Normal and log-normal models are popular in finance because of their simplicity and familiarity. Their use can also be justified by the central limit theorem. If $X, X_1, X_2, X_3, \ldots$ are independent and identically distributed (IID) random variables with mean $m = E(X)$ and finite variance $\sigma^2 = E[(X-m)^2]$ then the central limit theorem says that

$$\frac{(X_1 + \cdots + X_n) - nm}{n^{1/2}} \Rightarrow Y \tag{3.1}$$

where Y is a normal random variable with mean zero and variance σ^2, and $\Rightarrow$ means convergence of probability distributions. Essentially, (3.1) means that $X_1 + \cdots + X_n$ is approximately normal (with mean nm and variance $n\sigma^2$) for n large. If the summands X_i represent independent price shocks, then their sum is the price change over a period of time. If price changes are accumulations of many IID shocks, then they should be normally distributed. If price changes accumulate multiplicatively, taking logs changes the product into a sum, leading to a log-normal model.

For portfolio analysis, we need to consider a vector of prices. Suppose that $\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots$ are IID random vectors on a d-dimensional Euclidean space $\mathbb{R}^d$. If $\boldsymbol{X} = (X_1, \ldots, X_d)'$ then the mean $\boldsymbol{m} = E(\boldsymbol{X})$ is a vector with ith entry $m_i = E(X_i)$, the covariance matrix C is a $d \times d$ matrix with ij entry $c_{ij} = \mathrm{Cov}(X_i, X_j) = E[(X_i - m_i)(X_j - m_j)]$, and the central limit theorem says that

$$\frac{(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n) - n\boldsymbol{m}}{\sqrt{n}} \Rightarrow \boldsymbol{Y} \tag{3.2}$$

where $\boldsymbol{Y}$ is a normal random vector with mean zero and covariance matrix $C = E[\boldsymbol{Y}\boldsymbol{Y}']$. In this case, it simplifies the analysis to change coordinates. If the matrix P defines the change of coordinates then it follows from (3.2) that

$$\frac{(P\boldsymbol{X}_1 + \cdots + P\boldsymbol{X}_n) - nP\boldsymbol{m}}{\sqrt{n}} \Rightarrow P\boldsymbol{Y} \tag{3.3}$$

where $P\boldsymbol{Y}$ is multivariate normal with mean zero and covariance matrix $PCP' = E[(P\boldsymbol{Y})(P\boldsymbol{Y})']$. If we take the new coordinate system defined by the eigenvectors of the covariance matrix C, then the limit $P\boldsymbol{Y}$ has independent normal marginals. The eigenvalues of C determine the variance of each marginal, so their square roots measure volatility. The corresponding marginals of $P\boldsymbol{X}$ are all linear combinations of the original assets, chosen to be asymptotically independent. This coordinate system is one of the cornerstones of Markowitz's theory of optimal portfolios, see for example Elton and Gruber [16].

For heavy tailed random variables, the central limit theorem may not hold, because the second moment might not exist. An extended central limit theorem applies in this case. If $X, X_1, X_2, X_3, \ldots$ are IID random variables we say that X belongs to the domain of attraction of some random variable Y, and we write $X \in \mathrm{DOA}(Y)$, if

$$\frac{(X_1 + \cdots + X_n) - b_n}{a_n} \Rightarrow Y. \tag{3.4}$$

For mathematical reasons we exclude the degenerate case where $Y = c$ with probability one. The limits in (3.4) are called stable. If $E(X^2)$ exists then the classical central limit theorem shows that Y is normal, a special case of stable. In this case, we can take $a_n = n^{1/2}$ and $b_n = nE(X)$. If X has heavy tails with $P(|X| > r) \sim Cr^{-\alpha}$ then the situation depends on the tail thickness α. If $\alpha > 2$ then $E(X^2)$ exists and sums are asymptotically normal. But if $0 < \alpha \leq 2$ then $E(X^2) = \infty$ and (3.4) holds with $a_n = n^{1/\alpha}$ as long as a tail balancing condition holds:

$$\frac{P(X > r)}{P(|X| > r)} \to p \quad \text{and} \quad \frac{P(X < -r)}{P(|X| > r)} \to q \quad \text{as } r \to \infty \tag{3.5}$$

for some $0 \leq p, q \leq 1$ with $p + q = 1$.

A proof of the extended central limit theorem can be found in Gnedenko and Kolmogorov [23], see also Feller [20] and Meerschaert and Scheffler [48]. The condition for $X \in \mathrm{DOA}(Y)$ is stated in terms of *regular variation*. A function $f(r)$ varies regularly if

$$\lim_{r \to \infty} \frac{f(\lambda r)}{f(r)} = \lambda^{\rho} \quad \text{for all } \lambda > 0. \tag{3.6}$$

For Y stable with index $0 < \alpha < 2$, so that Y is not normal, a necessary and sufficient condition for $X \in \mathrm{DOA}(Y)$ is that $P(|X| > r)$ varies regularly with index $-\alpha$ and (3.5) holds for some $0 \leq p, q \leq 1$ with $p + q = 1$. If we have $P(|X| > r) \sim Cr^{-\alpha}$ then it is easy to see that $P(|X| > r)$ varies regularly with index $-\alpha$, but the definition also allows a slightly more general tail behavior. For example, if $P(|X| > r) \sim Cr^{-\alpha} \log r$ then $P(|X| > r)$ still varies regularly with index $-\alpha$. The norming constants a_n in (3.4) can always be chosen according to the formula $nP(|X| > a_n) \to C$. If we have $P(|X| > r) \sim Cr^{-\alpha}$ this leads to $a_n = n^{1/\alpha}$. In practical applications, it is common to assume that $P(|X| > r) \sim Cr^{-\alpha}$ because a practical procedure exists for estimating the parameters C, α for a given heavy tailed data set.[3]

[3]See Section 8.

Stable distributions are typically specified in terms of their characteristic functions (Fourier transforms). If Y is stable with density $f(y)$ its characteristic function

$$E[e^{ikY}] = \int_{-\infty}^{\infty} e^{iky} f(y)dy$$

is of the form $e^{\psi(k)}$ where

$$\psi(k) = \begin{cases} ibk - \sigma^\alpha |k|^\alpha \left(1 - i\beta \operatorname{sign}(k) \tan\left(\frac{\pi\alpha}{2}\right)\right) & \text{for } \alpha \neq 1, \\ ibk - \sigma^\alpha |k|^\alpha \left(1 + i\beta \left(\frac{2}{\pi}\right) \operatorname{sign}(k) \ln|k|\right) & \text{for } \alpha = 1. \end{cases} \tag{3.7}$$

The entire class of nondegenerate stable laws on $\mathbb{R}^1$ is given by these formulas with *index* $\alpha \in (0, 2]$, *scale* $\sigma \in (0, \infty)$, *skewness* $\beta \in [-1, +1]$, and *center* $b \in (-\infty, \infty)$. The stable distribution with these parameters will be written as $S_\alpha(\sigma, \beta, b)$ using the notation of Samorodnitsky and Taqqu [68]. The skewness $\beta = p - q$ governs the deviations of the distribution from symmetry, so that $f(y)$ is symmetric if $\beta = 0$. The scale σ and the center b have the usual meaning that if Y has a $S_\alpha(1, \beta, 0)$ distribution then $\sigma Y + b$ has a $S_\alpha(\sigma, \beta, b)$ distribution, except that for $\alpha = 1$ and $\beta \neq 0$ multiplication by σ introduces a nonlinear change in the shift. The stable index α governs the tails of Y, and in fact $P(|Y| > r) \sim Cr^{-\alpha}$ where

$$\sigma^\alpha = \begin{cases} C \cdot \frac{\Gamma(2-\alpha)}{1-\alpha} \cdot \cos\left(\frac{\pi\alpha}{2}\right) & \text{for } \alpha \neq 1, \\ C \cdot \frac{\pi}{2} & \text{for } \alpha = 1. \end{cases} \tag{3.8}$$

in the nonnormal case $0 < \alpha < 2$. The tails are balanced so that

$$\frac{P(Y > r)}{P(|Y| > r)} \to p \quad \text{and} \quad \frac{P(Y < -r)}{P(|Y| > r)} \to q \quad \text{as } r \to \infty \tag{3.9}$$

Stable laws belong to their own domain of attraction, but more is true. In fact, if Y_n are IID with Y then

$$\frac{(Y_1 + \cdots + Y_n) - b_n}{n^{1/\alpha}} \stackrel{d}{=} Y \tag{3.10}$$

for some b_n, where $\stackrel{d}{=}$ indicates that both sides have the same probability distribution. Sums of IID stable laws are again stable with the same α, β. Although there is no closed analytical formula for stable densities, the efficient computational method of Nolan [56, 59] can be used to plot density curves. Nolan [57] uses these methods to compute maximum likelihood estimators for the stable parameters, see also Mittnik, *et al.* [51, 52].

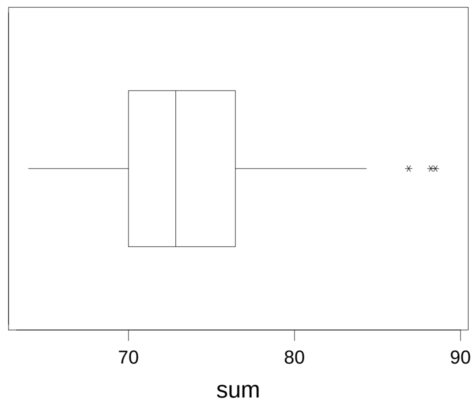

FIGURE 1. Sums of 50 Pareto variables with $\alpha = 3$. Their distribution is skewed to the right with several outliers.

If X_n is the price change on day n then the accumulation of these changes will be approximately stable, assuming that X_n are IID with X and $P(|X| > x) \sim Cx^{-\alpha}$. If $\alpha < 2$, as in the cotton prices considered in Mandelbrot [38], then the price obtained by adding these changes will be approximately stable with a power law tail. The balancing parameters p and q describe the probability that a large change in price will be positive or negative, respectively. The scale σ (or equivalently, the *dispersion* C) depends on the price units (e.g., US dollars). If $2 < \alpha < 4$ then the sum of these price changes will be asymptotically normal. However, the rule of thumb that sums look normal for $n \geq 30$ is no longer reliable. The heavy tails slow the rate of convergence in the central limit theorem. To illustrate the point, we simulated Pareto random variables with $\alpha = 3$, using the fact that if U is uniform on $(0, 1)$ then $U^{-1/\alpha}$ is Pareto with tail parameter α. We summed $n = 50$ of these random variables, and repeated the simulation 100 times to get an idea of the distribution of these sums. The boxplot in Figure 1 indicates that the distribution of the resulting sums is skewed to the right, with some outliers. The normal probability plot in Figure 2 indicates a significant deviation from normality. The moral of this story is that for heavy tailed random variables with $\alpha > 2$, sums eventually converge to a normal limit, but slower than usual.

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

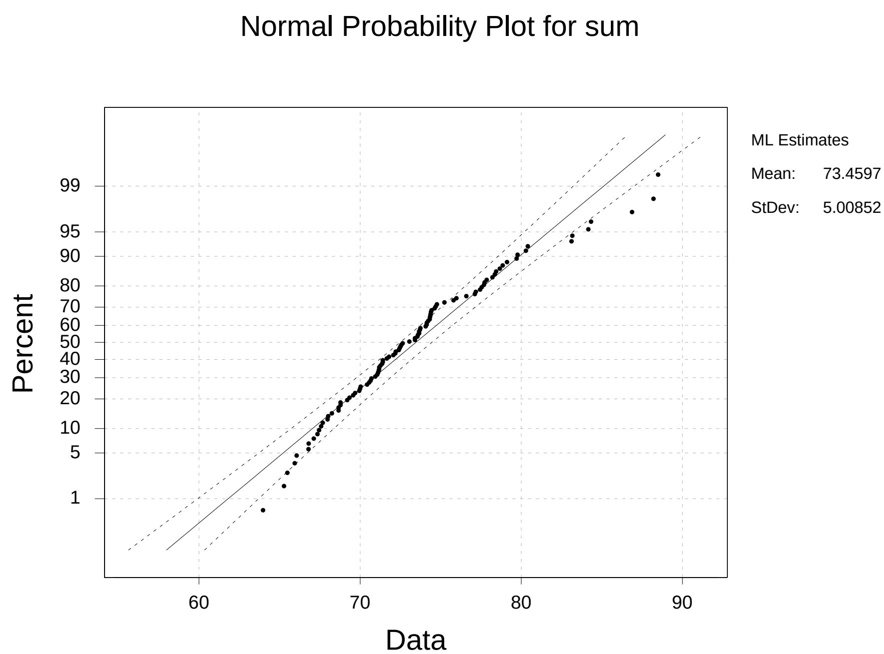

FIGURE 2. Sums of 50 Pareto variables with $\alpha = 3$. Upper tail shows systematic deviation from normal distribution.

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

$$A_n(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n - \boldsymbol{b}_n) \Rightarrow \boldsymbol{Y} \tag{3.11}$$

for some $d \times d$ matrices A_n and vectors $\boldsymbol{b}_n \in \mathbb{R}^d$. The limits in (3.11) are called operator stable [31, 72]. 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 $\boldsymbol{X}$ has heavy tails with $P(\|\boldsymbol{X}\| > r) \sim Cr^{-\alpha}$ then the situation depends on the tail thickness α. If $\alpha > 2$ then $E(\|\boldsymbol{X}\|^2)$ exists and sums are asymptotically normal. But if $0 < \alpha < 2$ then $E(\|\boldsymbol{X}\|^2) = \infty$ and (3.11) holds with $A_n = n^{-1/\alpha} I$ as long as a tail balancing condition holds:

$$\frac{P(\|\boldsymbol{X}\| > r, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in B)}{P(\|\boldsymbol{X}\| > r)} \to M(B) \quad \text{as } r \to \infty \tag{3.12}$$

for all Borel subsets[4] B of the unit sphere $S = \{\theta \in \mathbb{R}^d : \|\theta\| = 1\}$ whose boundary has M-measure zero, where M is a probability measure on the unit sphere which is not supported on any $d-1$ dimensional subspace of $\mathbb{R}^d$. A proof of the generalized central limit theorem can be found in Rvačeva [67] or Meerschaert and Scheffler [48]. In this case, where the tails of $\boldsymbol{X}$ fall off at the same rate in every direction, the limit $\boldsymbol{Y}$ is multivariable stable [68], a special case of operator stable.

If $\boldsymbol{Y}$ is multivariable stable with density $f(\boldsymbol{y})$ its characteristic function

$$E[e^{i\boldsymbol{k}\cdot\boldsymbol{Y}}] = \int e^{i\boldsymbol{k}\cdot\boldsymbol{y}} f(\boldsymbol{y})d\boldsymbol{y}$$

is of the form $e^{\psi(\boldsymbol{k})}$ where

$$\psi(\boldsymbol{k}) = i\boldsymbol{b}\cdot\boldsymbol{k} - \sigma^\alpha \int_{\|\theta\|=1} |\theta\cdot\boldsymbol{k}|^\alpha \left(1 - i\,\mathrm{sign}(\theta\cdot\boldsymbol{k})\tan\left(\frac{\pi\alpha}{2}\right)\right) M(d\theta)$$

for $\alpha \neq 1$ and

$$\psi(\boldsymbol{k}) = i\boldsymbol{b}\cdot\boldsymbol{k} - \sigma^\alpha \int_{\|\theta\|=1} |\theta\cdot\boldsymbol{k}| \left(1 + i\left(\frac{2}{\pi}\right)\mathrm{sign}(\theta\cdot\boldsymbol{k})\ln|\theta\cdot\boldsymbol{k}|\right) M(d\theta)$$

for $\alpha = 1$. The entire class of multivariable stable laws on $\mathbb{R}^d$ is given by these formulas with index $\alpha \in (0, 2]$, scale $\sigma > 0$, mixing measure M and center $\boldsymbol{b} \in \mathbb{R}^d$. We say that $\boldsymbol{Y}$ has distribution $S_\alpha(\sigma, M, \boldsymbol{b})$ in this case. The mixing measure M is a probability distribution on the unit sphere in $\mathbb{R}^d$ that governs the tails of $\boldsymbol{Y}$, so that $f(\boldsymbol{y})$ is symmetric if M is symmetric. The center $\boldsymbol{b}$ and scale σ have the usual meaning that if $\boldsymbol{Y}$ has a $S_\alpha(1, M, 0)$ distribution then $\sigma\boldsymbol{Y} + \boldsymbol{b}$ has a $S_\alpha(\sigma, M, \boldsymbol{b})$ distribution, except when $\alpha = 1$. The stable index α governs the tails of Y in the nonnormal case ($0 < \alpha < 2$). In fact, $P(\|\boldsymbol{Y}\| > r) \sim Cr^{-\alpha}$ where C is given by (3.8). The mixing measure M is a multivariable analogue of the skewness β. If $d = 1$ then $M\{+1\} = p$ and $M\{-1\} = q$, since the unit sphere on $\mathbb{R}^1$ is the two point set $\{-1, +1\}$. In this case, $\boldsymbol{Y}$ is stable with skewness $\beta = p - q$. The tails of a multivariable stable random vector are balanced so that

$$\frac{P(\|\boldsymbol{Y}\| > r, \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|} \in B)}{P(\|\boldsymbol{Y}\| > r)} \to M(B) \quad \text{as } r \to \infty. \tag{3.13}$$

If $d = 1$ this reduces to the tail balancing condition (3.9) for stable random variables. Multivariable stable laws belong to their own domain of attraction, and if $\boldsymbol{Y}_n$ are IID with $\boldsymbol{Y}$ then

$$\frac{(\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_n) - \boldsymbol{b}_n}{n^{1/\alpha}} \stackrel{d}{=} \boldsymbol{Y} \tag{3.14}$$

[4]The class of Borel subsets is the smallest class that include open sets and is closed under complements and countable unions.

for some $\boldsymbol{b}_n$, so that sums of IID multivariable stable laws are again multivariable stable with the same α. When $\boldsymbol{Y}$ is nonnormal multivariable stable with distribution $S_\alpha(\sigma, M, \boldsymbol{b})$ for some $0 < \alpha < 2$, the necessary and sufficient condition for $\boldsymbol{X} \in \mathrm{DOA}(\boldsymbol{Y})$ is that $P(\|\boldsymbol{X}\| > r)$ varies regularly with index $-\alpha$ and the balanced tails condition (3.12) holds.

Example 3.1. The mixing measure governs the radial direction of large price jumps. Take R_i IID Pareto random variables with $P(R > r) = Cr^{-\alpha}$. Take Θ_i to be IID random unit vectors with distribution M, independent of (R_i). Then $\boldsymbol{X}_i = R_i\Theta_i$ are IID random vectors with $P(\|\boldsymbol{X}_i\| > r) = Cr^{-\alpha}$ and

$$\frac{P(\|\boldsymbol{X}_i\| > r, \frac{\boldsymbol{X}_i}{\|\boldsymbol{X}_i\|} \in B)}{P(\|\boldsymbol{X}_i\| > r)} = P(\Theta_i \in B) = M(B)$$

for any Borel subset B of the unit sphere, and so $\boldsymbol{X}_i \in \mathrm{DOA}(\boldsymbol{Y})$ where $\boldsymbol{Y}$ is multivariable stable with distribution $S_\alpha(\sigma, M, \boldsymbol{b})$ for any $\boldsymbol{b} \in \mathbb{R}^d$. We can take $A_n = n^{-1/\alpha}I$ in (3.11), and $\boldsymbol{b}$ depends on the choice of centering $\boldsymbol{b}_n$. We call these heavy tailed random vectors *multivariable Pareto*. If we use a multivariable Pareto model for large jumps in the vector of prices for a portfolio, the parameter α governs the radius and the mixing measure M governs the angle of large jumps. Sums of these IID jumps are asymptotically multivariable stable with the same index α and mixing measure M. The radius $R = \|\boldsymbol{Y}\|$ satisfies $P(R > r) \sim Cr^{-\alpha}$ and the distribution of the radial component $\Theta = \boldsymbol{Y}/\|\boldsymbol{Y}\|$ conditional on $P(\|\boldsymbol{Y}\| > r)$ tends to M as $r \to \infty$ in view of the tail balancing condition (3.13). In other words, multivariable stable random vectors are asymptotically multivariable Pareto on their tails. In a multivariable stable model for price jumps, the mixing measure determines the direction of large jumps. If M is discrete with $M(\theta_i) = p_i$, then it follows from the characteristic function formulas that $\boldsymbol{Y}$ can be represented as the sum of independent stable components laid out along the θ_i directions, and the methods of Nolan [56, 59] can be used to plot multivariable stable densities, see Byczkowski, Nolan and Rajput [13]. The same idea is used by Modarres and Nolan [53] to simulate stable random vectors with discrete mixing measures. For an arbitrary mixing measure, multivariable stable laws can be simulated using sums of independent, identically distributed multivariable Pareto laws. If $0 < \alpha < 1$ then the random vector $n^{-1/\alpha}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$ is approximately $S_\alpha(\sigma, M, 0)$ where C is given by (3.8). If $1 < \alpha < 2$ then $n^{-1/\alpha}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n - nE\boldsymbol{X}_1)$ is approximately $S_\alpha(\sigma, M, 0)$ where C is given by (3.8) and

$$E(\boldsymbol{X}_1) = E(R_1)E(\Theta_1) = \frac{\alpha C^{1/\alpha}}{\alpha - 1}\int_{\|\theta\|=1} \theta M(d\theta).$$

Remark 3.2. Previously a different type of multivariable Pareto distribution was considered by Arnold [6], see also Kotz, *et al.*, [33].

4. Matrix scaling

The multivariable stable model is the basis for the work of Nolan, Panorska and McCulloch [58] on exchange rates, and the portfolio models in Rachev and Mittnik [62]. Under the assumptions of this model, the probability tail of the random vector $\boldsymbol{X}_t$ is assumed to fall off at the same power law rate in every radial direction. Suppose that $\boldsymbol{X}_t = (X_1(t), \ldots, X_d(t))'$ where $X_i(t)$ is the price change of the ith asset on day t. If $\boldsymbol{X}_t$ belongs to the domain of attraction of some multivariable stable random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ with index α, and that (3.11) holds with $A_n = n^{-1/\alpha} I$. Projecting onto the ith coordinate axis shows that

$$\frac{X_i(1) + \cdots + X_i(n) - b_i(n)}{n^{1/\alpha}} \Rightarrow Y_i \tag{4.1}$$

where $\boldsymbol{b}_n = (b_1(n), \ldots, b_d(n))'$, so that Y_i is stable with index α and $X_i(t)$ belongs to the domain of attraction of Y_i. According to Jansen and de Vries [30], Loretan and Phillips [37], and Rachev and Mittnik [62], the stable index α_i should vary depending on the asset. Then (4.1) is replaced by

$$\frac{X_i(1) + \cdots + X_i(n) - b_i(n)}{n^{1/\alpha_i}} \Rightarrow Y_i \quad \text{for each } i = 1, \ldots, d \tag{4.2}$$

so that Y_i is stable with index α_i. Mittnik and Rachev [50] seem to have been the first to apply such models to a problem in finance, see also Section 8.6 in Rachev and Mittnik [62]. Assuming the joint convergence

$$A_n \left[\begin{pmatrix} X_1(1) \\ X_2(1) \\ \vdots \\ X_d(1) \end{pmatrix} + \cdots + \begin{pmatrix} X_1(n) \\ X_2(n) \\ \vdots \\ X_d(n) \end{pmatrix} - \begin{pmatrix} b_1(n) \\ b_2(n) \\ \vdots \\ b_d(n) \end{pmatrix} \right] \Rightarrow \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_d \end{pmatrix} \tag{4.3}$$

and changing to vector-matrix notation we get (3.11) with diagonal norming matrices

$$A_n = \begin{pmatrix} n^{-1/\alpha_1} & 0 & \cdots & 0 \\ 0 & n^{-1/\alpha_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & n^{-1/\alpha_d} \end{pmatrix} \tag{4.4}$$

which we will also write as $A_n = \operatorname{diag}(n^{-1/\alpha_1}, \ldots, n^{-1/\alpha_d})$. The matrix scaling is natural since we are dealing with random vectors, and it allows a more realistic portfolio model. The ith marginal Y_i of the operator stable limit vector $\boldsymbol{Y}$ is stable with index α_i, so the tail behavior of $\boldsymbol{Y}$ varies with angle. The convergence (3.11) with A_n diagonal was first considered in Resnick and Greenwood [65], see also Meerschaert [43].

Matrix notation also leads to a natural analogue of the stable index α. Let $\exp(A) = I + A + A^2/2! + A^3/3! + \cdots$ be the usual exponential operator for $d \times d$ matrices. This operator occurs, for example, in the theory of linear differential equations. If $A = \operatorname{diag}(a_1, \ldots, a_d)$ then an easy matrix computation using the Taylor series formula $e^x = 1 + x + x^2/2! + x^3/3! + \cdots$ shows that $\exp(A) = \operatorname{diag}(e^{a_1}, \ldots, e^{a_d})$. See Hirsch and Smale [27] or Section 2.2 of [48] for details and additional information. Now define $E = \operatorname{diag}(1/\alpha_1, \ldots, 1/\alpha_d)$. Then the norming matrices A_n in (4.4) can also be written in the more compact form $A_n = n^{-E} = \exp(-E \ln n)$, since $-E \ln n = \operatorname{diag}(-(1/\alpha_1)\ln n, \ldots, -(1/\alpha_d)\ln n)$ and $e^{-(1/\alpha_i)\ln n} = n^{-1/\alpha_i}$. The matrix E, called an *exponent* of the operator stable random vector $\boldsymbol{Y}$, plays the role of the stable index α. This 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 [42]. The case of a diagonalizable exponent plays an important role in Example 8.1.

The generalized central limit theorem for matrix scaling can be found in Meerschaert and Scheffler [48]. Matrix scaling allows for a limit with both normal and nonnormal components. Since $\boldsymbol{Y}$ is infinitely divisible, the Lévy representation (Theorem 3.1.11 in [48]) shows that the characteristic function $E[e^{i\boldsymbol{k}\cdot\boldsymbol{Y}}]$ is of the form $e^{\psi(\boldsymbol{k})}$ where

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

for some $\boldsymbol{b} \in \mathbb{R}^d$, some nonnegative definite symmetric $d \times d$ matrix C 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,

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

and the characteristic function formulas for multivariable stable laws follow by a lengthy computation, see Section 7.3 in Meerschaert and Scheffler [48] for complete details. If $\phi = 0$ then $\boldsymbol{Y}$ is normal with mean $\boldsymbol{b}$ and covariance matrix C. If $C = 0$ then a necessary and sufficient condition for (3.11) to hold is that

$$nP(A_n\boldsymbol{X} \in B) \to \phi(B) \quad \text{as } n \to \infty \tag{4.5}$$

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 [48] shows that the convergence (4.5) is equivalent to regular variation of the probability distribution $\mu(B) = P(\boldsymbol{X} \in B)$. If (4.5) holds then Proposition 6.1.2 in [48] shows that the Lévy measure satisfies

$$t\phi(dx) = \phi(t^{-E}dx) \quad \text{for all } t > 0 \tag{4.6}$$

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

$$n^{-E}(\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_n - \boldsymbol{b}_n) \stackrel{d}{=} \boldsymbol{Y} \tag{4.7}$$

for some $\boldsymbol{b}_n$, see Theorem 7.2.1 in [48]. 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 (4.5) is equivalent to the balanced tails condition (3.12).

Example 4.1. Multivariable Pareto random vectors with matrix scaling extend the model in Example 3.1. Suppose $\boldsymbol{Y}$ is operator stable with exponent E and Lévy measure ϕ. Define

$$F_{r,B} = \{s^E\theta : s > r, \theta \in B\}$$

and let $\lambda(B) = \phi(F_{1,B})$ for any Borel subset B of the unit sphere S whose boundary has λ-measure zero.[5] Let $C = \lambda(S)$ and define the probability measure $M(B) = \lambda(B)/C$. Take R_i IID standard Pareto random variables with $P(R > r) = Cr^{-1}$, Θ_i IID random unit vectors with distribution M and independent of (R_i), and finally let $\boldsymbol{X}_i = R_i^E\Theta_i$. Since $t^E F_{1,B} = F_{t,B}$ we have $\phi(F_{t,B}) = \phi(t^E F_{1,B}) = t^{-1}\phi(F_{1,B}) = Ct^{-1}M(B)$ in view of (4.6). Then

$$\begin{aligned} nP(n^{-E}\boldsymbol{X}_i \in F_{t,B}) &= nP(R_i^E\Theta_i \in F_{nt,B}) \\ &= nP(R_i > nt, \Theta_i \in B) \\ &= nC(nt)^{-1}M(B) = \phi(F_{t,B}) \end{aligned}$$

for $n > 1/t$, so that (4.5) holds for the sets $F_{t,B}$ with $A_n = n^{-E}$. Then $\boldsymbol{X}_i \in \text{GDOA}(\boldsymbol{Y})$. Operator stable laws can be simulated using sums of these IID random vectors. If every eigenvalue of E has real part greater than one, then $n^{-E}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$ is approximately operator stable with exponent E and Lévy measure ϕ. If every eigenvalue of E has real part less than one, then $n^{-E}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n - n\boldsymbol{m})$ is approximately operator stable with exponent E and Lévy measure ϕ where

$$\boldsymbol{m} = C\int_{\|\theta\|=1}\int_C^\infty r^E\theta\frac{dr}{r^2}M(d\theta)$$

[5]The measure λ is called the spectral measure of $\boldsymbol{Y}$.

is the mean of $\boldsymbol{X}_1$.

5. The spectral decomposition

The tail behavior of an operator stable random vector $\boldsymbol{Y}$ is determined by the eigenvalues of its exponent E. If $E = (1/\alpha)I$ then $\boldsymbol{Y}$ is multivariable stable and $P(|\boldsymbol{Y}\cdot\theta| > r) \sim C_\theta r^{-\alpha}$ for any $\theta \neq 0$. If $E = \text{diag}(a_1, \ldots, a_d)$ then $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ where Y_i is a stable random variable with index $\alpha_i = 1/a_i$. This requires $0 < \alpha_i \leq 2$ so that $a_i \geq 1/2$. 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 [48]. This decomposition allows us to write $E = PBP^{-1}$ where P is a change of coordinates matrix and B is block-diagonal with

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & B_p \end{pmatrix} \tag{5.1}$$

where B_i is a $d_i \times d_i$ matrix, every eigenvalue of B_i has real part equal to a_i, $a_1 < \cdots < a_p$, and $d_1 + \cdots + d_p = d$. Let $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$, $\boldsymbol{e}_2 = (0, 1, 0, \ldots, 0)'$, …, $\boldsymbol{e}_d = (0, \ldots, 0, 1)'$ be the standard coordinates for $\mathbb{R}^d$ and define $\boldsymbol{p}_{ik} = P\boldsymbol{e}_j$ when $j = d_1 + \cdots + d_{i-1} + k$ for some $k = 1, \ldots, d_i$. Then

$$V_i = \text{span}\{\boldsymbol{p}_{i1}, \ldots, \boldsymbol{p}_{id_i}\} = \left\{ \sum_{k=1}^{d_i} t_k \boldsymbol{p}_{ik} : t_1, \ldots, t_{d_i} \text{ real} \right\}$$

is a d_i-dimensional subspace of $\mathbb{R}^d$. Any vector $\boldsymbol{y} \in \mathbb{R}^d$ can be written uniquely in the form $\boldsymbol{y} = \boldsymbol{y}_1 + \cdots + \boldsymbol{y}_p$ with $\boldsymbol{y}_i \in V_i$ for each $i = 1, \ldots, p$. This is called the spectral decomposition of $\mathbb{R}^d$ with respect to E. Since B is block-diagonal and $E = PBP^{-1}$, every $E\boldsymbol{p}_{ik}$ is a linear combination of $\boldsymbol{p}_{i1}, \ldots, \boldsymbol{p}_{id_i}$ and therefore $E\boldsymbol{y}_i \in V_i$ for every $\boldsymbol{y}_i \in V_i$. This means that V_i is an E-invariant subspace of $\mathbb{R}^d$. Given a nonzero vector $\theta \in \mathbb{R}^d$, write $\theta = \theta_1 + \cdots + \theta_p$ with $\theta_i \in V_i$ for each $i = 1, \ldots, p$ and define

$$\alpha(\theta) = \max\{1/a_i : \theta_i \neq 0\}. \tag{5.2}$$

Since the probability distribution of $\boldsymbol{Y}$ varies regularly with exponent E, Theorem 6.4.15 in [48] shows that for any small $\delta > 0$ we have

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

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 $E(|\boldsymbol{Y}\cdot\theta|^\beta)$ exists for $0 < \beta < \alpha(\theta)$ and diverges for $\beta > \alpha(\theta)$. 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 (4.7) 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 [48] shows that the limit $\boldsymbol{Y}$ and norming matrices A_n in (3.11) 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

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

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

Example 5.1. If $\boldsymbol{Y}$ is operator stable with exponent $E = aI$ then (4.7) shows that $\boldsymbol{Y}$ is multivariable stable with index $\alpha = 1/a$. Then $p = 1$, $P = I$, and $B = E$. There is only one spectral component, since the tail behavior is the same in every radial direction. If asset price change vectors are IID with $\boldsymbol{X} = (X_1, \ldots, X_d)' \in \mathrm{GDOA}(\boldsymbol{Y})$, then every asset has the same tail behavior. If θ_j measures the amount of the jth asset in a portfolio, price changes for this portfolio are IID with the random variable $\boldsymbol{X} \cdot \theta = X_1\theta_1 + \cdots + X_d\theta_d$. Since the probability tails of $\boldsymbol{X}$ are uniform in every direction, the probability of a large jump in price falls off like $r^{-\alpha}$ for any portfolio.

Example 5.2. If $\boldsymbol{Y}$ is operator stable with exponent $E = \mathrm{diag}(a_1, \ldots, a_d)$ where $a_1 < \cdots < a_d$ then $p = d$, $P = I$, $B = E$, $B_i = a_i$ and V_i is the ith coordinate axis. The spectral decomposition of $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ with respect to E is $\boldsymbol{Y} = \boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_d$ with $\boldsymbol{Y}_i = Y_i\boldsymbol{e}_i$, the ith marginal laid out along the ith coordinate axis. Projecting (4.7) onto the ith coordinate axis shows that Y_i is stable with index $\alpha_i = 1/a_i$, so that $P(|Y_i| > r) \sim C_i r^{-\alpha_i}$. If $\theta \neq 0$ then $P(|\boldsymbol{Y} \cdot \theta| > r)$ falls off like $r^{-\alpha(\theta)}$ where $\alpha(\theta) = \max\{\alpha_i : \theta_i \neq 0\}$. In other words, the heaviest tail dominates. If asset price change vectors are IID with $\boldsymbol{X} \in \mathrm{GDOA}_c(\boldsymbol{Y})$, then the assets are arranged in order of increasing tail thickness. If θ_i measures the amount of the ith asset in a portfolio, the probability of a large jump in price falls off like $r^{-\alpha(\theta)}$.

Example 5.3. If $\boldsymbol{Y}$ is operator stable with exponent $E = \mathrm{diag}(\beta_1, \ldots, \beta_d)$ then $B_i = a_i I$ for some $a_i \geq 1/2$ and d_i counts the number of diagonal entries β_j for which $\beta_j = 1/\alpha_i$. The matrix P sorts $\beta_1, \ldots, \beta_d$ in increasing order, and the vectors $\boldsymbol{p}_{ik}$ are the coordinates $\boldsymbol{e}_j$ for which $\beta_j = a_i$. The vectors $\boldsymbol{Y}_i$ are multivariable stable with index $\alpha_i = 1/a_i$, so that $P(\|\boldsymbol{Y}_i\| > r) \sim C_i r^{-\alpha_i}$. For nonzero vectors $\theta \in V_i$ we have $P(|\boldsymbol{Y} \cdot \theta| > r) = P(|\boldsymbol{Y}_i \cdot \theta| > r) \sim C_\theta r^{-\alpha_i}$ by the balanced tails condition for multivariable stable laws. For any other nonzero vector θ, $P(|\boldsymbol{Y} \cdot \theta| > r) \sim C_\theta r^{-\alpha(\theta)}$ where $\alpha(\theta) = \max\{1/\beta_j : \theta_j \neq 0\}$. Again, the heaviest tail dominates. If asset price change vectors are IID with $\boldsymbol{X} \in \mathrm{GDOA}_c(\boldsymbol{Y})$, then $\boldsymbol{X}$ has essentially the same tail behavior as $\boldsymbol{Y}$, and P sorts the assets in order of increasing tail thickness.

Example 5.4. Take $B = \mathrm{diag}(a_1, \ldots, a_d)$ where $a_1 < \cdots < a_d$ and P orthogonal, so that $P^{-1} = P'$. If $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ is operator stable with exponent $E = PBP^{-1}$ then $p = d$, $B_i = a_i$ and $V_1, \ldots, V_d$ are the coordinate axes in the new coordinate system defined by the vectors $\boldsymbol{p}_i = P\boldsymbol{e}_i$ for $i = 1, \ldots, d$. The spectral component $\boldsymbol{Y}_i$ is the stable random variable $\boldsymbol{Y} \cdot \boldsymbol{p}_i$ with index $\alpha_i = 1/a_i$, laid out along the V_i axis. Since $Y_j = \boldsymbol{Y} \cdot \boldsymbol{e}_j$ is a linear combination of stable laws of different indices, it is not stable. The change of coordinates P rotates the coordinate axes to make the marginals stable. Since $n^{-PBP^{-1}} = Pn^{-B}P^{-1}$ it follows from (4.7) that

$$\begin{aligned} Pn^{-B}P^{-1}(\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_n - \boldsymbol{b}_n) &\stackrel{d}{=} \boldsymbol{Y} \\ n^{-B}(P^{-1}\boldsymbol{Y}_1 + \cdots + P^{-1}\boldsymbol{Y}_n - P^{-1}\boldsymbol{b}_n) &\stackrel{d}{=} P^{-1}\boldsymbol{Y} \end{aligned}$$

so that $\boldsymbol{Y}_0 = P^{-1}\boldsymbol{Y}$ is operator stable with exponent B. Then the tail behavior of $\boldsymbol{Y} = P\boldsymbol{Y}_0$ follows from Example 5.2 and the change of coordinates. If we write $\theta = \theta_1\boldsymbol{p}_1 + \cdots + \theta_d\boldsymbol{p}_d$ in these coordinates then $P(|\boldsymbol{Y} \cdot \theta| > r) \sim C_\theta r^{-\alpha(\theta)}$ where $\alpha(\theta) = \max\{\alpha_i : \theta_i \neq 0\}$. If asset price change vectors are IID with $\boldsymbol{X} \in \mathrm{GDOA}_c(\boldsymbol{Y})$, then the tail behavior of $\boldsymbol{X}$ is essentially the same as $\boldsymbol{Y}$. In particular, taking $\theta = \boldsymbol{p}_1$ gives a portfolio with the lightest probability tails.

Example 5.5. Suppose that $\boldsymbol{Y}$ is operator stable with exponent $E = PBP^{-1}$ where P is orthogonal and B is given by (5.1), with $d_i \times d_i$ blocks $B_i = a_i I$ for some $1/2 \leq a_1 < \cdots < a_p$. Let $D_0 = 0$ and $D_i = d_1 + \cdots + d_i$ for $1 \leq i \leq p$. Then $\boldsymbol{p}_{ik} = P\boldsymbol{e}_j$ when $j = D_{i-1} + k$ for some $k = 1, \ldots, d_i$ and $V_i = \mathrm{span}\{\boldsymbol{p}_{ik} : k = 1, \ldots, d_i\}$. To avoid double subscripts we will also write $\boldsymbol{q}_j = P\boldsymbol{e}_j$, so that $\boldsymbol{q}_j = \boldsymbol{p}_{ik}$ when $j = D_{i-1} + k$ for some $k = 1, \ldots, d_i$. The jth column of the matrix P is the vector $\boldsymbol{q}_j$, and

$$E\boldsymbol{q}_j = PBP^{-1}\boldsymbol{q}_j = PB\boldsymbol{e}_j = Pa_i\boldsymbol{e}_j = a_iP\boldsymbol{e}_j = a_i\boldsymbol{q}_j$$

when $\boldsymbol{q}_j \in V_i$, so that $\boldsymbol{q}_j$ is a unit eigenvector of the matrix E with corresponding eigenvalue a_i. The spectral component

$$\boldsymbol{Y}_i = \sum_{k=1}^{d_i} (\boldsymbol{Y} \cdot \boldsymbol{p}_{ik})\boldsymbol{p}_{ik}$$

is the orthogonal projection of $\boldsymbol{Y}$ onto the d_i-dimensional subspace V_i. The random vector $\boldsymbol{Y}_i$ is multivariable stable with index $\alpha_i = 1/a_i$, so that $P(\|\boldsymbol{Y}_i\| > r) \sim C_i r^{-\alpha_i}$, and every marginal $Y_{ik} = \boldsymbol{Y} \cdot \boldsymbol{p}_{ik}$ is stable with the same index α_i. The change of coordinates P rotates the coordinate axes to find a set of orthogonal unit eigenvectors for E, so that the marginals of $\boldsymbol{Y}$ in the new coordinate system are all stable random variables. The matrix P also sorts the corresponding eigenvalues in increasing order. For any nonzero vector $\theta \in \mathbb{R}^d$, $P(|\boldsymbol{Y} \cdot \theta| > r) \sim C_\theta r^{-\alpha(\theta)}$ where $\alpha(\theta) = \alpha_i$ for the largest i such that the orthogonal projection of θ onto the subspace V_i is not equal to zero. If asset price change vectors are IID with $\boldsymbol{X} \in \mathrm{GDOA}_c(\boldsymbol{Y})$, then the tail behavior of $\boldsymbol{X}$ is essentially the same as $\boldsymbol{Y}$. If $\theta = \theta_1 \boldsymbol{e}_1 + \cdots \theta_d \boldsymbol{e}_d$ so that θ_i measures the amount of the ith asset in a portfolio, price changes for this portfolio are IID with $\boldsymbol{X} \cdot \theta = X_1\theta_1 + \cdots + X_d\theta_d$. In particular, any $\theta \in V_1$ gives a portfolio with the lightest probability tails.

6. Sample covariance matrix

Given a data set of price changes (or log returns) $X_1, X_2, \ldots, X_n$ for a given asset, the kth sample moment

$$\hat{\mu}_k = \frac{1}{n} \sum_{t=1}^{n} X_t^k$$

estimates the kth moment $\mu_k = E(X^k)$. These sample moments are used to estimate the mean, variance, skewness and kurtosis of the data. If X_t are IID with $P(|X_t| > r) \sim C r^{-\alpha}$, then X_t^k are also IID and heavy tailed with

$$P(|X_t^k| > r) = P(|X_t| > r^{1/k}) \sim C r^{-\alpha/k}$$

so the extended central limit theorem applies. Recall from Section 2 that μ_k exists for $k < \alpha$ and diverges for $k \geq \alpha$. If $\alpha > 4$ then $\mathrm{Var}(X_t^2) = \mu_4 - \mu_2^2$ exists and the central limit theorem (3.1) implies that

$$n^{1/2}(\hat{\mu}_2 - \mu_2) = n^{-1/2} \sum_{t=1}^{n} (X_t^2 - \mu_2) \Rightarrow Y \tag{6.1}$$

where Y is normal. When $2 < \alpha < 4$, the mean $\mu_2 = E(X_t^2)$ of these summands exists but $\mathrm{Var}(X_t^2)$ is infinite, and the extended central limit theorem (3.4) implies that

$$n^{1-2/\alpha}(\hat{\mu}_2 - \mu_2) = n^{-2/\alpha} \sum_{t=1}^{n} (X_t^2 - \mu_2) \Rightarrow Y$$

where Y is stable with index $\alpha/2$. When $0 < \alpha < 2$ the mean $\mu_2 = E(X_t^2)$ of the squared price change diverges, and the extended central limit theorem implies that

$$n^{1-2/\alpha}\hat{\mu}_2 = n^{-2/\alpha}\sum_{t=1}^{n} X_t^2 \Rightarrow Y$$

where again Y is stable with index $\alpha/2$. In this case, the sample second moment $\hat{\mu}_2$ exists but the second moment μ_2 does not. When $0 < \alpha < 2$, or when $2 < \alpha < 4$ and $\mu_1 = 0$, the sample variance

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{\mu}_1)^2 = \hat{\mu}_2 - \hat{\mu}_1^2 \tag{6.2}$$

is asymptotically equivalent to the sample second moment, see for example Anderson and Meerschaert [4]. Since we can always center to zero expectation when $2 < \alpha < 4$, both have the same asymptotics. If $\alpha > 4$ the sample variance is asymptotically normal, and when $0 < \alpha < 4$ the sample variance is asymptotically stable. Since the variance is a measure of price volatility, the sample variance estimates volatility. Confidence intervals for the variance are based on normal asymptotics when $\alpha > 4$ and stable asymptotics when $2 < \alpha < 4$. When $\alpha < 2$ the variance is undefined, but the sample variance still captures some important features of the data, see Section 8.

Suppose that $\boldsymbol{X}_t = (X_1(t), \ldots, X_d(t))'$ where $X_i(t)$ is the price change of the ith asset on day t. The covariance matrix characterizes dependence between price changes of different assets over the same day, and the sample covariance matrix estimates the covariance matrix. As before, it is simpler to begin with the uncentered estimate

$$M_n = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{X}_t\boldsymbol{X}_t' \tag{6.3}$$

where $\boldsymbol{X}'$ denotes the transpose of the vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$ and hence

$$\boldsymbol{X}\boldsymbol{X}' = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}(X_1, \ldots, X_d) = \begin{pmatrix} X_1X_1 & \cdots & X_1X_d \\ X_2X_1 & \cdots & X_2X_d \\ \vdots & & \vdots \\ \vdots & & \\ X_dX_1 & \cdots & X_dX_d \end{pmatrix}$$

is an element of the vector space $\mathcal{M}_s^d$ of symmetric $d \times d$ matrices. The ij entry of M_n is

$$M_n(i,j) = \frac{1}{n}\sum_{t=1}^{n} X_i(t)X_j(t)$$

which estimates $E(X_iX_j)$. If $\boldsymbol{X}_t$ are IID with $\boldsymbol{X}$, then $\boldsymbol{X}_t\boldsymbol{X}_t'$ are IID random matrices and we can apply the central limit theorems from Section 3 (see Section 10.2 in [48] for complete proofs). If the probability distribution of $\boldsymbol{X}$

is regularly varying with exponent E and (4.5) 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

$$nP(A_n\boldsymbol{X}\boldsymbol{X}'A_n' \in B) \to \Phi(B) \quad \text{as } n \to \infty \tag{6.4}$$

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 $M \in \mathcal{M}_s^d$. Using the matrix norm

$$\|M\| = \left(\sum_{i=1}^{d}\sum_{j=1}^{d} M(i,j)^2\right)^{1/2}$$

we get

$$\|\boldsymbol{X}\boldsymbol{X}'\|^2 = \sum_{i=1}^{d}\sum_{j=1}^{d}(X_iX_j)^2 = \left(\sum_{i=1}^{d} X_i^2\right)\left(\sum_{j=1}^{d} X_j^2\right) = \|\boldsymbol{X}\|^4$$

so that $\|\boldsymbol{X}\boldsymbol{X}'\| = \|\boldsymbol{X}\|^2$. If every eigenvalue of E has real part $a_i < 1/4$, then $E(\|\boldsymbol{X}\boldsymbol{X}'\|^2) = E(\|\boldsymbol{X}\|^4) < \infty$ and the multivariable central limit theorem (3.2) shows that

$$n^{1/2}(M_n - C) = n^{-1/2}\sum_{t=1}^{n}(\boldsymbol{X}_t\boldsymbol{X}_t' - C) \Rightarrow W \tag{6.5}$$

where W is a Gaussian random matrix and C is the (uncentered) covariance matrix $C = E(\boldsymbol{X}\boldsymbol{X}')$. The estimates of Jansen and de Vries [30] and Loretan and Phillips [37] indicate tail estimates in the range $2 < \alpha < 4$. In this case, every eigenvalue of E has real part $1/4 < a_i < 1/2$. Then $E(\|\boldsymbol{X}\boldsymbol{X}'\|^2) = E(\|\boldsymbol{X}\|^4) = \infty$, but $E(\|\boldsymbol{X}\boldsymbol{X}'\|) = E(\|\boldsymbol{X}\|^2) < \infty$ so the covariance matrix $C = E(\boldsymbol{X}\boldsymbol{X}')$ exists. Now the generalized central limit theorem (3.11) gives

$$nA_n(M_n - C)A_n' = A_n\left(\sum_{t=1}^{n}(\boldsymbol{X}_t\boldsymbol{X}_t' - C)\right)A_n' \Rightarrow W \tag{6.6}$$

where the limit W is a nonnormal operator stable random matrix. The estimates in Rachev and Mittnik [62] give tail estimates in the range $1 < \alpha < 2$, so that every eigenvalue of E has real part $a_i > 1/2$. Then $E(\|\boldsymbol{X}\boldsymbol{X}'\|) = E(\|\boldsymbol{X}\|^2) = \infty$ and the covariance matrix $C = E(\boldsymbol{X}\boldsymbol{X}')$ diverges. In this case,

$$nA_nM_nA_n' \Rightarrow W \tag{6.7}$$

holds with W operator stable. Since the covariance matrix is undefined, there is no reason to believe that the sample covariance matrix contains useful information. However, we will see in Section 8 that even in this case the sample covariance matrix characterizes the most important distributional features of the random vector $\boldsymbol{X}$.

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. In the heavy tailed case $a_i > 1/4$, Theorem 10.6.15 in [48] shows that Γ_n and M_n have the same asymptotics, similar to the one dimensional case. In practice, it is common to mean-center the data, so it does not matter which form we choose.

7. Dependent random vectors

Suppose that $\boldsymbol{X}_t = (X_1(t), \ldots, X_d(t))'$ where $X_i(t)$ represents the price change (or log return) of the ith asset on day t. A model where $\boldsymbol{X}_t$ are IID with $\boldsymbol{X} \in \mathrm{GDOA}(\boldsymbol{Y})$ allows dependence between the price changes $X_i(t)$ and $X_j(t)$ on the same day t, which is commonly observed in practice. If we also want to model dependence between days, we need to relax the IID assumption. A wide variety of time series models can be mathematically reduced to a linear moving average. This reduction may involve integer or fractional differencing, detrending and deseasoning, and nonlinear mappings. Asymptotics for the underlying moving average are established in Section 10.6 of [48]. Assume that $\boldsymbol{Z}, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3, \ldots$ are IID random vectors on $\mathbb{R}^d$ whose probability distribution is regularly varying with exponent E, so that

$$nP(A_n\boldsymbol{Z} \in B) \to \phi(B) \quad \text{as } n \to \infty \tag{7.1}$$

for Borel subsets B of $\mathbb{R}^d \setminus \{0\}$ whose boundary have ϕ-measure zero, and $t\phi(d\boldsymbol{x}) = \phi(t^{-E}d\boldsymbol{x})$ for all $t > 0$. If every eigenvalue of E has real part $a_i > 1/2$ then $\boldsymbol{Z} \in \mathrm{GDOA}(\boldsymbol{Y})$ and

$$A_n(\boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_n - n\boldsymbol{b}_n) \Rightarrow \boldsymbol{Y} \tag{7.2}$$

where $\boldsymbol{Y}$ is operator stable with exponent E and Lévy measure ϕ. Define the moving average process

$$\boldsymbol{X}_t = \sum_{j=0}^{\infty} C_j \boldsymbol{Z}_{t-j} \tag{7.3}$$

where C_j are $d \times d$ real matrices. If every eigenvalue of E has real part $a_i < a_p$ then the moving average (7.3) is well defined as long as

$$\sum_{j=0}^{\infty} \|C_j\|^{\delta} < \infty \tag{7.4}$$

for some $\delta < 1/a_p$ with $\delta \leq 1$. If every eigenvalue of E has real part $a_i < 1/2$, then $E(\|X_t\|^2)$ exists and the asymptotics are normal, see Brockwell and Davis [12]. If every eigenvalue of E has real part $a_i > 1/2$, and if for each j either

$C_j = 0$, or else C_j^{-1} exists and $A_n C_j = C_j A_n$ for all n, then Theorem 10.6.2 in [48] shows that

$$A_n \left(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n - n \sum_{j=0}^{\infty} C_j \boldsymbol{b}_n \right) \Rightarrow \sum_{j=0}^{\infty} C_j \boldsymbol{Y}. \tag{7.5}$$

The limit in (7.5) is operator stable with no normal component and Lévy measure $\sum_j C_j \phi$, where $C_j \phi = 0$ if $C_j = 0$ and otherwise $C_j \phi(d\boldsymbol{x}) = \phi(C_j^{-1} d\boldsymbol{x})$.

If every eigenvalue of E has real part $a_i < 1/2$, then both the mean $\boldsymbol{m} = E(\boldsymbol{X}_t)$ and the lag h covariance matrix

$$\Gamma(h) = E[(\boldsymbol{X}_t - \boldsymbol{m})(\boldsymbol{X}_{t+h} - \boldsymbol{m})']$$

exist. The matrix $\Gamma(h)$ tells us when price changes on day t are correlated with price changes (of the same asset or some other asset) h days later. These correlations are useful to identify leading indicators, and they are the basic tools of time series modeling. The sample covariance matrix at lag $h \geq 0$ for the moving average $\boldsymbol{X}_t$ is defined by

$$\hat{\Gamma}_n(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (\boldsymbol{X}_t - \bar{\boldsymbol{X}})(\boldsymbol{X}_{t+h} - \bar{\boldsymbol{X}})' \tag{7.6}$$

where $\bar{\boldsymbol{X}} = (\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)/n$. If every eigenvalue of E has real part $a_i < 1/4$, then $E(\|\boldsymbol{X}_t\|^4) < \infty$ and $\hat{\Gamma}_n(h)$ is asymptotically normal, see Brockwell and Davis [12]. If every eigenvalue of E has real part $1/4 < a_i < 1/2$, the estimates of Jansen and de Vries [30] and Loretan and Phillips [37], then

$$A_n \left(\sum_{t=1}^{n} \boldsymbol{Z}_t \boldsymbol{Z}_t' - D \right) A_n' \Rightarrow U \tag{7.7}$$

as in Section 6, where U is a nonnormal operator stable random matrix and $D = E(\boldsymbol{Z}\boldsymbol{Z}')$. Then Theorem 10.6.15 in [48] shows that

$$n A_n \left(\hat{\Gamma}_n(h) - \Gamma(h) \right) A_n' \Rightarrow \sum_{j=0}^{\infty} C_j U C_{j+h}' \tag{7.8}$$

for any $h \geq 0$, The asymptotics (7.8) determine which elements of the sample covariance matrix $\hat{\Gamma}_n(h)$ are statistically significantly different from zero.

If every eigenvalue of E has real part $a_i > 1/2$, as in the estimates of Rachev and Mittnik [62], then

$$A_n \left(\sum_{t=1}^{n} \boldsymbol{Z}_t \boldsymbol{Z}_t' \right) A_n' \Rightarrow U \tag{7.9}$$

and Theorem 10.6.15 in [48] shows that

$$nA_n\hat{\Gamma}_n(h)A_n' \Rightarrow \sum_{j=0}^{\infty} C_j U C_{j+h}' \tag{7.10}$$

for any $h \geq 0$. In this case the covariance matrix $\Gamma(h)$ does not exist, but the sample covariance matrix $\hat{\Gamma}_n(h)$ still contains useful information about the time series $\boldsymbol{X}_t$ of price changes. In the next section, we will explain this apparent paradox.

8. Tail estimation

Given a set of price changes (or log-returns) $X_1, \ldots, X_n$ for some asset, it is important to estimate the tail behavior. If the price changes X_t are identically distributed[6] with X and $P(X > r) \sim Cr^{-\alpha}$, then the dispersion C and the tail index α determine the central limit behavior, as well as the extreme value behavior, of the price change distribution. Mandelbrot [38] pioneered a graphical estimation method for C and α. 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 $-\alpha$ and $\log C$ can be estimated from the vertical axis intercept. If $P(X > r) \approx Cr^{-\alpha}$ for r large, then the upper tail should be approximately linear. We call this a *Mandelbrot plot*. Several Mandelbrot plots for stock market and exchange rate returns appear in Loretan and Phillips [37] as evidence of heavy tails with $2.5 < \alpha < 3$. Replacing X by $-X$ gives information about the left tail. Least squares estimators for α based on the Mandelbrot plot were proposed by Schultze and Steinebach [71], see also Csörgo and Viharos [15].

The most popular numerical estimator for C and α is due to Hill [26], see also Hall [25]. Sort the data in decreasing order to obtain the *order statistics* $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$. 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

$$\begin{aligned} \hat{\alpha} &= \left[\frac{1}{m} \sum_{i=1}^{m} \left(\ln X_{(i)} - \ln X_{(m+1)} \right) \right]^{-1} \\ \hat{C} &= \frac{m}{n} X_{(m+1)}^{\hat{\alpha}} \end{aligned} \tag{8.1}$$

where m is to be taken as large as possible, but small enough so that the tail condition $P(X > r) = Cr^{-\alpha}$ remains valid. Replacing X by $-X$ gives estimates for the left tail. Replacing X by $|X|$ gives estimates for the combined

[6] Note that we are not assuming IID here.

tail. This is often advantageous, because it allows us to combine the data from both tails, and increase the number m of order statistics used. Finding the best value of m is a challenge, and creates a certain amount of controversy. Jansen and de Vries [30] use Hill's estimator with a fixed value of $m = 100$ for several different assets. Loretan and Phillips [37] 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 [30] and Loretan and Phillips [37] vary depending on the asset.

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

$$\begin{aligned}\hat{\alpha} &= \left[\frac{1}{m}\sum_{i=1}^{m}\bigl(\ln(X_{(i)} - \hat{s}) - \ln(X_{(m+1)} - \hat{s})\bigr)\right]^{-1} \\ \hat{C} &= \frac{m}{n}(X_{(m+1)} - \hat{s})^{\hat{\alpha}}\end{aligned} \tag{8.2}$$

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

$$\hat{\alpha}(X_{(m+1)} - \hat{s})^{-1} = (\hat{\alpha} + 1)\frac{1}{m}\sum_{i=1}^{m}(X_{(i)} - \hat{s})^{-1} \tag{8.3}$$

over $\hat{s} < X_{(m+1)}$. Once the optimal shift is computed, $\hat{\alpha}$ and $\hat{C}$ come 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.

If X_t is the sum of many IID price shocks, then it can be argued that the distribution of X_t must be (at least approximately) stable with distribution $S_\alpha(\sigma, \beta, b)$. Maximum likelihood estimation for the stable parameters is now practical, using the efficient method of Nolan [56] for computing stable densities, see also Mittnik, *et al.* [51, 52]. Since the stable index $0 < \alpha \leq 2$, the stable MLE for α cannot possibly agree with the estimates found in Jansen and de Vries [30] and Loretan and Phillips [37]. Rachev and Mittnik [62] use a stable model for price changes, and their estimates yield $1 < \alpha < 2$ for a variety of assets. McCulloch [41] argues that the $\alpha > 2$ estimates found in Jansen and de Vries [30] and Loretan and Phillips [37] are inflated due to a distributional misspecification. The Pareto tail of a stable random variable X disappears as $\alpha \to 2$, so that it may be impossible to take m large enough for a reliable estimate, see Fofack and Nolan [21] for a more detailed discussion. The estimator in [1] corrects for the fact that Hill's $\hat{\alpha}$ is not shift-invariant, and may go some distance towards correcting the problem identified by McCulloch [41].

Maximum likelihood estimation is quite sensitive to deviations from the proscribed distribution, and it is no surprise that the MLE computations of Jansen and de Vries [30] and Loretan and Phillips [37], based on the Pareto model, differ significantly from the estimates of Rachev and Mittnik [62], based on a stable model. Part of the controversy stems from the fact that the range of α is limited to $(0, 2]$ for the stable model. Akgiray and Booth [3] interpret the results of Hill's estimator for stock returns as evidence against the stable model. Actual finance data does not exactly fit either the stable or Pareto-tail models, and in our opinion, parameter estimates are only valid with respect to the model used to obtain them, so that Pareto-based estimates of $\alpha > 2$ in no way invalidate the stable model.

Meerschaert and Scheffler [44] propose a robust estimator

$$\hat{\alpha} = \frac{2 \ln n}{\ln n + \ln \hat{\sigma}^2} \tag{8.4}$$

based on the sample variance (6.2). This estimator can be applied whenever $X \in \mathrm{DOA}(Y)$ and Y is stable with index $0 < \alpha < 2$. Then X can be stable or Pareto, or any distribution with balanced power-law tails. The estimator is also applicable to dependent data, since it also applies when $X_t = \sum_j c_j Z_{t-j}$, Z_t is IID with $Z \in \mathrm{DOA}(Y)$, and Y is stable with index $0 < \alpha < 2$. The estimator is based on the simple idea that

$$\begin{aligned} n^{1-2/\alpha} \hat{\sigma}^2 &\Rightarrow Y \\ \ln(n\hat{\sigma}^2) - \frac{2}{\alpha} \ln n &\Rightarrow \ln Y \\ 2 \ln n \left(\frac{\ln(n\hat{\sigma}^2)}{2 \ln n} - \frac{1}{\alpha} \right) &\Rightarrow \ln Y \end{aligned}$$

so that $\ln(n\hat{\sigma}^2)/(2 \ln n)$ estimates $1/\alpha$. If X has heavy tails with $\alpha \geq 2$ then $\hat{\alpha} \to 2$. In this case, we can apply the estimator to X^k, which also has heavy tails with tail parameter α/k. 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.

Portfolio modeling requires a vector model to incorporate dependence between price changes for different assets. In these vector models, the sample variance is replaced by the sample covariance matrix. For heavy tailed price changes with infinite variance, the covariance matrix does not exist. Even so, we will see that the sample covariance matrix is a very useful tool for portfolio modeling. Suppose that $\boldsymbol{X}_t = (X_1(t), \ldots, X_d(t))'$ where $X_i(t)$ is the price change of the ith asset on day t. If $\boldsymbol{X}_t$ are identically distributed with $\boldsymbol{X}$

and if $\boldsymbol{X}$ has heavy tails with $P(\|\boldsymbol{X}\| > r) \sim Cr^{-\alpha}$ then the vector norms $\|\boldsymbol{X}_1\|, \ldots, \|\boldsymbol{X}_n\|$ can be used to estimate the tail parameter α. Alternatively, we can apply one variable tail estimators to the ith marginal to get an estimate $\hat{\alpha}_i$ of the tail parameter. If the probability tails of $\boldsymbol{X}$ fall off at the same rate $r^{-\alpha}$ in every radial direction, then these estimates should all be reasonably close. In that case, we might assume that $\boldsymbol{X}$ is multivariable stable with distribution $S_\alpha(\sigma, M, \boldsymbol{b})$. The mean $\boldsymbol{b}$ can be estimated using the sample mean in the usual case $1 < \alpha < 2$. Several estimators now exist for the scale σ and the mixing measure M, or equivalently, for the spectral measure $\lambda(d\theta) = \sigma^\alpha M(d\theta)$. Those estimators are surveyed in another paper in this volume [35], so we will not dwell on them here. If $\alpha > 2$, one might consider the multivariable Pareto laws introduced in Example 3.1. If $P(\|\boldsymbol{X}\| > r) \sim Cr^{-\alpha}$ and the balanced tails condition (3.13) holds for some mixing measure M, then the tail behavior of $\boldsymbol{X}$ is multivariable Pareto. Multivariable stable random vectors have this property with $0 < \alpha < 2$. If $\alpha > 2$ then multivariable Pareto could offer a reasonable alternative, which to our knowledge has not been pursued in the finance literature.

While experts disagree on the range of α for typical assets, there seems to be general agreement that the tail index depends on the asset. Then it is appropriate to assume that the probability distribution of $\boldsymbol{X}$ varies regularly with some exponent E. For IID random vectors, a method for estimating the exponent E can be found in Section 10.4 of [48]. In Section 9 we show that the same methods also apply to dependent random vectors which are identically distributed. The method is applicable when the eigenvalues of E all have real part $a_i > 1/2$, the infinite variance case. To be concrete, we adopt the model of Example 5.5, which is the simplest model flexible enough for realism. This model assumes that E has a set of d mutually orthogonal unit eigenvectors. Note that if the eigenvalues of E are all distinct then these unit eigenvectors are unique up to a factor of ± 1. On the other hand, if $E = aI$ for some $a > 1/2$ then any set of d mutually orthogonal unit vectors can be used.

Recall the spectral decomposition $E = PBP^{-1}$ from Example 5.5, where P is orthogonal and B is given by (5.1), with $d_i \times d_i$ blocks $B_i = a_i I$ for some $1/2 \le a_1 < \cdots < a_p$. Let $D_0 = 0$ and $D_i = d_1 + \cdots + d_i$ for $1 \le i \le p$. Then $\boldsymbol{q}_j = P\boldsymbol{e}_j$ is a unit eigenvector of the matrix E and the d_i dimensional subspace $V_i = \operatorname{span}\{\boldsymbol{q}_j : D_{i-1} < j \le D_i\}$ contains every eigenvector of E with associated eigenvalue a_i. Our estimator for E is based on the sample covariance matrix M_n defined in (6.3). Since M_n is symmetric and nonnegative definite, there exists an orthonormal basis of eigenvectors for M_n with nonnegative eigenvalues. Eigenvalues and eigenvectors of M_n are easily computed using standard numerical routines, see for example Press *et al.* [61]. Sort the eigenvalues

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

and the associated unit eigenvectors

$$\theta_1, \ldots, \theta_d$$

so that $M_n \theta_j = \lambda_j \theta_j$ for each $j = 1, \ldots, d$. Now Theorem 10.4.5 in [48] shows that

$$\frac{\log n + \log \lambda_j}{2 \log n} \to a_i \quad \text{as } n \to \infty$$

in probability for any $D_{i-1} < j \leq D_i$. This is a multivariable analogue for the one variable tail estimator (8.4). Furthermore, Theorem 10.4.8 in [48] shows that the eigenvectors θ_j converge in probability to V_1 when $j \leq D_1$, 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.

Now we illustrate the practical application of the multivariable tail estimator. Recall that $\boldsymbol{X}_t = (X_1(t), \ldots, X_d(t))'$ where $X_i(t)$ is the price change of the ith asset on day t. Compute the (uncentered) sample covariance matrix M_n using the formula (6.3) and then compute the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_d$ and the associated eigenvectors

$$\begin{aligned} \theta_1 &= (\theta_1(1), \ldots, \theta_d(1))' \\ &\vdots \\ \theta_d &= (\theta_1(d), \ldots, \theta_d(d))' \end{aligned} \tag{8.5}$$

of the matrix M_n. A change of coordinates is essential to the method. Write

$$Z_j(t) = \boldsymbol{X}_t \cdot \theta_j = X_1(t)\theta_1(j) + \cdots + X_d(t)\theta_d(j)$$

for each $j = 1, \ldots, d$. Our portfolio model is based on these new coordinates. Let

$$\hat{\alpha}_j = \frac{2 \log n}{\log n + \log \lambda_j}$$

for each $j = 1, \ldots, n$. Since the eigenvalues are sorted in increasing order we will have $\hat{\alpha}_1 \geq \cdots \geq \hat{\alpha}_d$. Our model assumes that $Z_j(t)$ are identically distributed with Z_j, and the tail parameter $\hat{\alpha}_j$ governs the jth coordinate Z_j. If $\hat{\alpha}_j < 2$ then $P(|Z_j| > r)$ falls off like $r^{-\hat{\alpha}_j}$ and if $\hat{\alpha}_j \geq 2$ then a finite variance model for Z_j is adequate. We can also use any other one variable tail estimator to get α_j for each of the new coordinates $Z_j(t)$. The new coordinates unmask variations in α that would go undetected in the original coordinates.

Example 8.1. We look at a data set of $n = 2853$ 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. We 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 we 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 $\theta_1 = (0.69, -0.72)'$, $\theta_2 = (0.72, 0.69)'$. Next we compute

$$\begin{aligned} \hat\alpha_1 &= \frac{2\ln 2853}{\ln 2853 + \ln 1.006} = 1.998 \\ \hat\alpha_2 &= \frac{2\ln 2853}{\ln 2853 + \ln 5.209} = 1.656 \end{aligned} \tag{8.6}$$

indicating that $Z_1(t) = 0.69X_1(t) - 0.72X_2(t)$ fits a finite variance model but $Z_2(t) = 0.72X_1(t) + 0.69X_2(t)$ fits a heavy tailed model with $\alpha = 1.656$. Then we can model $\boldsymbol{Z}_t = (Z_1(t), Z_2(t))'$ as being identically distributed with the random vector $\boldsymbol{Z} = (Z_1, Z_2)'$ where $P(|Z_2| > r) \approx C_1 r^{-1.656}$ and $\mathrm{Var}(Z_1) < \infty$. The simplest model with these properties is to take $Z_1(t)$ normal and $Z_2(t)$ stable with index $\alpha = 1.656$ and independent of $Z_1(t)$.

Next we explain the operator stable model based on these estimates. The random vectors $\boldsymbol{Z}_t$ are operator stable with exponent

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

since $0.50 = 1/1.998$ and $0.60 = 1/1.656$. The change of coordinates matrix

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

so that

$$\boldsymbol{Z}_t = \begin{pmatrix} Z_1(t) \\ Z_2(t) \end{pmatrix} = \begin{pmatrix} 0.69 & -0.72 \\ 0.72 & 0.69 \end{pmatrix}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = P\boldsymbol{X}_t.$$

Since

$$P^{-1} = \begin{pmatrix} 0.69 & 0.72 \\ -0.72 & 0.69 \end{pmatrix}$$

(rounded off to two decimal places) we also have

$$\boldsymbol{X}_t = P^{-1}\boldsymbol{Z}_t = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} 0.69 & 0.72 \\ -0.72 & 0.69 \end{pmatrix}\begin{pmatrix} Z_1(t) \\ Z_2(t) \end{pmatrix}$$

so that

$$\begin{aligned} X_1(t) &= 0.69Z_1(t) + 0.72Z_2(t) \\ X_2(t) &= -0.72Z_1(t) + 0.69Z_2(t). \end{aligned} \tag{8.7}$$

Both exchange rates have a common heavy-tailed stable factor $Z_2(t)$ and so both exchange rates have heavy tails with the same tail index $\alpha = 1.656$. It

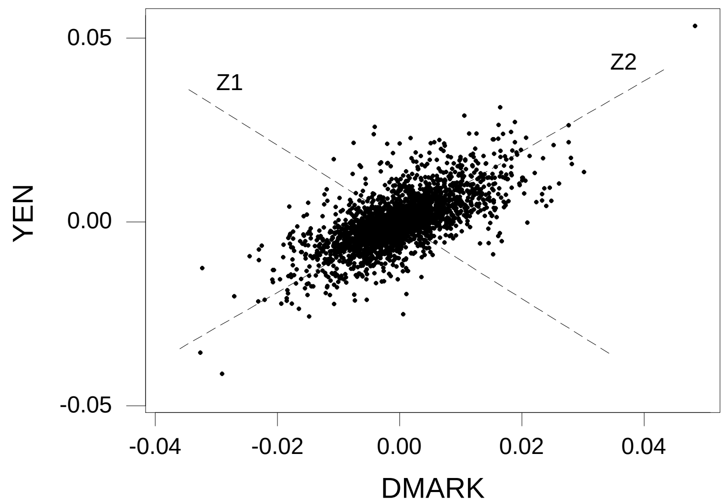

FIGURE 3. Exchange rates against the US dollar. The new coordinates uncover variations in the tail parameter α.

is tempting to interpret $Z_2(t)$ as the common 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.

We also take the opportunity to fill in the details of Example 5.4 in this simple case. The original data $\boldsymbol{X}_t = P^{-1}\boldsymbol{Z}_t$ is modeled as operator stable with exponent

$$E = PBP^{-1} = \begin{pmatrix} 0.55 & 0.05 \\ 0.05 & 0.55 \end{pmatrix}.$$

In this case, $Z_1(t)$ and $Z_2(t)$ are independent so the density of $\boldsymbol{Z}_t$ is the product of the two marginal densities, and then the density of $\boldsymbol{X}_t$ can be obtained by a simple change of variables. The columns of the change of variables matrix P are the eigenvectors θ_j of the sample covariance matrix, which estimate the theoretical coordinate system vectors $\boldsymbol{p}_j$ in the spectral decomposition.

Remark 8.2. This exchange rate data in Example 8.1 was also analyzed by Nolan, Panorska and McCulloch [58] using a multivariable stable model. Since both marginals $X_1(t)$ and $X_2(t)$ have heavy tails with the same α, there is no obvious reason to employ a more complicated model. However, the change of coordinates in Example 8.1 uncovers variations in the tail parameter α, an important modeling insight.

Remark 8.3. Kotz, Kozubowski and Podgórski [34] employ a very different model for the data in Example 8.1, based on the Laplace distribution. This distribution, and its multivariable analogues, assume exponential probability tails for the data. These models have heavier tails than the Gaussian, but they have moments of all orders.

Remark 8.4. The simplistic model in Example 8.1 assumes that the two factors Z_1 and Z_2 are independent. If we assume that $\boldsymbol{Z}$ is operator stable with Z_1 normal and Z_2 stable then these components must be independent, in view of the general characteristic function formula for operator stable laws. Another alternative is to assume that Z_1 is stable with index $\alpha = 1.998$, very close to a normal distribution. In this case, the two components can be dependent. The dependence is captured by the mixing measure or spectral measure, see Example 4.1. Scheffler [69] provides a method for estimating the spectral measure from data for an operator stable random vector with a known exponent. This provides a more flexible model including dependence between the two factors.

9. Tail estimator proof for dependent random vectors

In this section, we provide a proof that the multivariable tail estimator of Section 8 is still valid for certain sequences of dependent heavy tailed random vectors. We say that a sequence (B_n) of invertible linear operators is regularly varying with index $-E$ if for any $\lambda > 0$ we have

$$B_{[\lambda n]} B_n^{-1} \to \lambda^{-E} \quad \text{as } n \to \infty.$$

For further information about regular variation of linear operators see [48], Chapter 4.

In view of Theorem 2.1.14 of [48] we can write $\mathbb{R}^d = V_1 \oplus \cdots \oplus V_p$ and $E = E_1 \oplus \cdots \oplus E_p$ for some $1 \le p \le d$ where each V_i is E invariant, $E_i : V_i \to V_i$ and $\mathrm{Re}(\lambda) = a_i$ for all real parts of the eigenvalues of E_i and some $a_1 < \cdots < a_p$. By Definition 2.1.15 of [48] this is called the spectral decomposition of $\mathbb{R}^d$ with respect to E. By Definition 4.3.13 of [48] we say that (B_n) is spectrally compatible with $-E$ if every V_i is B_n-invariant for all n. Note that in this case we can write $B_n = B_{1n} \oplus \cdots \oplus B_{pn}$ and each $B_{in} : V_i \to V_i$ is regularly varying with index $-E_i$. (See Proposition 4.3.14 of [48].) For the proofs in this section we will always assume that the subspaces V_i in the spectral decomposition of $\mathbb{R}^d$ with respect to E are mutually orthogonal. We will also assume that (B_n) is spectrally compatible with $-E$. Let π_i denote the orthogonal projection operator onto V_i. If we let $P_i = \pi_i + \cdots + \pi_p$ and $L_i = V_i \oplus \cdots \oplus V_p$ then

$P_i : \mathbb{R}^d \to L_i$ is a orthogonal projection. Furthermore, $\bar{P}_i = \pi_1 + \cdots + \pi_i$ is the orthogonal projection onto $\bar{L}_i = V_1 \oplus \cdots \oplus V_i$.

Now assume $0 < a_1 < \cdots < a_p$. Since (B_n) is spectrally compatible with $-E$, Proposition 4.3.14 of [48] shows that the conclusions of Theorem 4.3.1 of [48] hold with $L_i = V_i \oplus \cdots \oplus V_p$ for each $i = 1, \ldots, p$. Then for any $\varepsilon > 0$ and any $x \in L_i \setminus L_{i+1}$ we have

$$n^{-a_i - \varepsilon} \le \|B_n x\| \le n^{-a_i + \varepsilon} \tag{9.1}$$

for all large n. Then

$$\frac{\log \|B_n x\|}{\log n} \to -a_i \quad \text{as } n \to \infty \tag{9.2}$$

and since this convergence is uniform on compact subsets of $L_i \setminus L_{i+1}$ we also have

$$\frac{\log \|\pi_i B_n\|}{\log n} \to -a_i \quad \text{as } n \to \infty. \tag{9.3}$$

It follows that

$$\frac{\log \|B_n\|}{\log n} \to -a_1 \quad \text{as } n \to \infty. \tag{9.4}$$

Since $(B_n')^{-1}$ is regularly varying with index E', a similar argument shows that for any $x \in \bar{L}_i \setminus \bar{L}_{i-1}$ we have

$$n^{a_i - \varepsilon} \le \|(B_n')^{-1} x\| \le n^{a_i + \varepsilon} \tag{9.5}$$

for all large n. Then

$$\frac{\log \|(B_n')^{-1} x\|}{\log n} \to a_i \quad \text{as } n \to \infty \tag{9.6}$$

and since this convergence is uniform on compact subsets of $\bar{L}_i \setminus \bar{L}_{i-1}$ we also have

$$\frac{\log \|\pi_i (B_n')^{-1}\|}{\log n} \to a_i \quad \text{as } n \to \infty. \tag{9.7}$$

Hence

$$\frac{\log \|(B_n')^{-1}\|}{\log n} \to a_p \quad \text{as } n \to \infty. \tag{9.8}$$

Suppose that $\boldsymbol{X}_t$, $t = 1, 2, \ldots$ are $\mathbb{R}^d$-valued random vectors and let M_n be the sample covariance matrix of $(\boldsymbol{X}_t)$ defined by (6.3). Note that M_n is symmetric and positive semidefinite. Let $0 \le \lambda_{1n} \le \cdots \le \lambda_{dn}$ denote the eigenvalues of M_n and let $\theta_{1n}, \ldots, \theta_{dn}$ be the corresponding orthonormal basis of eigenvectors.

Basic Assumptions: Assume that for some exponent E with real spectrum $1/2 < a_1 < \cdots < a_p$ the subspaces V_i in the spectral decomposition of $\mathbb{R}^d$

with respect to E are mutually orthogonal, and there exists a sequence (B_n) regularly varying with index $-E$ and spectrally compatible with $-E$ such that:

(A1) The set $\{n(B_n M_n B_n') : n \geq 1\}$ is weakly relatively compact.

(A2) For any limit point M of this set we have:
   (a) M is almost surely positive definite.
   (b) For all unit vectors θ the random variable $\theta' M \theta$ has no atom at zero.

Now let $\mathbb{R}^d = V_1 \oplus \cdots \oplus V_p$ be the spectral decomposition of $\mathbb{R}^d$ with respect to E. Put $d_i = \dim V_i$ and for $i = 1, \dots, p$ let $b_i = d_i + \cdots + d_p$ and $\bar{b}_i = d_1 + \cdots + d_i$. Our goal is now to estimate the real spectrum $a_1 < \cdots < a_p$ of E as well as the spectral decomposition $V_1, \dots, V_p$. In various situation, these quantities completely describe the moment behavior of the X_t.

Theorem 9.1. *Under our basic assumptions, for $i = 1, \dots, p$ and $\bar{b}_{i-1} < j \leq \bar{b}_i$ we have*

$$\frac{\log(n\lambda_{jn})}{2\log n} \to a_i \quad \textit{in probability as } n \to \infty.$$

The proof of Theorem 9.1 is in parts quite similar to the Theorem 2 in [46]. See also Section 10.4 in [48], and [70]. We include it here for sake of completeness.

Proposition 9.2. *Under our basic assumptions we have*

$$\frac{\log(n\lambda_{dn})}{2\log n} \to a_p \quad \textit{in probability.}$$

Proof. For $\delta > 0$ arbitrary we have

$$P\Big\{\Big|\frac{\log(n\lambda_{dn})}{2\log n} - a_p\Big| > \delta\Big\} \leq P\{\lambda_{dn} > n^{2(a_p+\delta)-1}\} + P\{\lambda_{dn} < n^{2(a_p-\delta)-1}\}.$$

Now choose $0 < \varepsilon < \delta$ and note that by (9.8) we have $\|(B_n')^{-1}\| \leq n^{a_p+\varepsilon}$ for all large n. Using assumption (A1) we obtain for all large n

$$\begin{aligned} P\{\lambda_{dn} > n^{2(a_p+\delta)-1}\} &= P\{\|M_n\| > n^{2(a_p+\delta)-1}\} \\ &\leq P\{\|(B_n')^{-1}\|^2 \|nB_n M_n B_n'\| > n^{2(a_p+\delta)}\} \\ &\leq P\{\|nB_n M_n B_n'\| > n^{2(\delta-\varepsilon)}\} \end{aligned}$$

and the last probability tends to zero as $n \to \infty$.

Now fix any $\theta_0 \in \bar{L}_p \setminus \bar{L}_{p-1}$ and write $(B_n')^{-1}\theta_0 = r_n \theta_n$ for some unit vector θ_n and $r_n > 0$. Theorem 4.3.14 of [48] shows that every limit point of (θ_n) lies in the unit sphere in V_p. Then since (9.5) holds uniformly on compact sets we have for any $0 < \varepsilon < \delta$ that $n^{a_p-\varepsilon} \leq r_n \leq n^{a_p+\varepsilon}$ for all large n. Then for all large n we get

$$\begin{aligned} P\{\lambda_{dn} < n^{2(a_p-\delta)-1}\} &= P\Big\{\max_{\|\theta\|=1} M_n\theta\cdot\theta < n^{2(a_p-\delta)-1}\Big\} \\ &\le P\{M_n\theta_0\cdot\theta_0 < n^{2(a_p-\delta)-1}\} \\ &= P\{nB_nM_nB_n'\theta_n\cdot\theta_n < r_n^{-2}n^{2(a_p-\delta)-1}\} \\ &\le P\{nB_nM_nB_n'\theta_n\cdot\theta_n < n^{2(\varepsilon-\delta)}\}. \end{aligned}$$

Given any subsequence (n') there exists a further subsequence $(n'')\subset(n')$ along which $\theta_n\to\theta$. Furthermore, by assumption (A1) there exists another subsequence $(n''')\subset(n'')$ such that $nB_nM_nB_n'\Rightarrow M$ along (n'''). Hence by continuous mapping (see Theorem 1.2.8 in [48]) we have

$$nB_nM_nB_n'\theta_n\cdot\theta_n \Rightarrow M\theta\cdot\theta \quad \text{along } (n''').$$

Now, given any $\varepsilon_1>0$ by assumption (A2)(b) there exists a $\rho>0$ such that $P\{M\theta\cdot\theta<\rho\}<\varepsilon_1/2$. Hence for all large $n=n'''$ we have

$$\begin{aligned} P\{nB_nM_nB_n'\theta_n\cdot\theta_n < n^{2(\varepsilon-\delta)}\} &\le P\{nB_nM_nB_n'\theta_n\cdot\theta_n<\rho\} \\ &\le P\{M\theta\cdot\theta<\rho\}+\frac{\varepsilon_1}{2} \\ &< \varepsilon_1 \end{aligned}$$

Since for any subsequence there exists a further subsequence along which $P\{nB_nM_nB_n'\theta_n\cdot\theta_n < n^{2(\varepsilon-\delta)}\}\to 0$, this convergence holds along the entire sequence which concludes the proof. □

Proposition 9.3. *Under the basic assumptions we have*

$$\frac{\log(n\lambda_{1n})}{2\log n}\to a_1 \quad \textit{in probability.}$$

Proof. Since the set $\mathrm{GL}(\mathbb{R}^{\mathrm{d}})$ of invertible matrices is an open subset of the vector space of $d\times d$ real matrices, it follows from (A1) and (A2)(a) together with the Portmanteau Theorem (c.f., Theorem 1.2.2 in [48]) that $\lim_{n\to\infty}P\{M_n\in\mathrm{GL}(\mathbb{R}^{\mathrm{d}})\}=1$ holds. Hence we can assume without loss of generality that M_n is invertible for all large n.

Given any $\delta>0$ write

$$P\Big\{\Big|\frac{\log(n\lambda_{1n})}{2\log n}-a_1\Big|>\delta\Big\}\le P\{\lambda_{1n}>n^{2(a_1+\delta)-1}\}+P\{\lambda_{1n}<n^{2(a_1-\delta)-1}\}.$$

To estimate the first probability on the right hand side of the inequality above choose a unit vector $\theta_0\in\bar{L}_1$ and write $(B_n')^{-1}\theta_0=r_n\theta_n$ as above. Then, since (9.5) holds uniformly on the unit sphere in $\bar{L}_1=V_1$, for $0<\varepsilon<\delta$ we have

$n^{a_1-\varepsilon} \le r_n \le n^{a_1+\varepsilon}$ for all large n. Therefore for all large n

$$\begin{aligned} P\{\lambda_{1n} > n^{2(a_1+\delta)-1}\} &\le P\Big\{\min_{\|\theta\|=1} M_n\theta\cdot\theta > n^{2(a_1+\delta)-1}\Big\} \\ &\le P\{M_n\theta_0\cdot\theta_0 > n^{2(a_1+\delta)-1}\} \\ &\le P\{nB_nM_nB_n'\theta_n\cdot\theta_n > n^{2(\delta-\varepsilon)}\}. \end{aligned}$$

It follows from assumption (A1) together with the compactness of the unit sphere in $\mathbb{R}^d$ and continuous mapping that the sequence $(nB_nM_nB_n'\theta_n\cdot\theta_n)$ is weakly relatively compact and hence by Prohorov's Theorem this sequence is uniformly tight. Since $\delta > \varepsilon$ it follows that $P\{\lambda_{1n} > n^{2(a_1+\delta)-1}\} \to 0$ as $n \to \infty$.

Since the smallest eigenvalue of M_n is the reciprocal of the largest eigenvalue of M_n^{-1} we have

$$\begin{aligned} P\{\lambda_{1n} < n^{2(a_1-\delta)-1}\} &= P\{\frac{1}{\lambda_{1n}} > n^{2(\delta-a_1)+1}\} \\ &= P\Big\{\max_{\|\theta\|=1} M_n^{-1}\theta\cdot\theta > n^{2(\delta-a_1)+1}\Big\} \\ &= P\{\|M_n^{-1}\| > n^{2(\delta-a_1)+1}\} \\ &\le P\{\|\frac{1}{n}(B_n')^{-1}M_n^{-1}B_n^{-1}\| > \|B_n\|^{-2}n^{2(\delta-a_1)}\} \end{aligned}$$

It follows from (9.4) that for any $0 < \varepsilon < \delta$ there exists a constant $C > 0$ such that $\|B_n\| \le Cn^{-a_1+\varepsilon}$ for all n and hence for some constant $K > 0$ we get $\|B_n\|^{-2} \ge Kn^{2(a_1-\varepsilon)}$ for all n. Note that by assumptions (A1) and (A2)(a) together with continuous mapping the sequence

$$\big(\frac{1}{n}(B_n')^{-1}M_n^{-1}B_n^{-1}\big)$$

is weakly relatively compact and hence by Prohorov's theorem this sequence is uniformly tight. Hence

$$\begin{aligned} &P\{\|\frac{1}{n}(B_n')^{-1}M_n^{-1}B_n^{-1}\| > \|B_n\|^{-2}n^{2(\delta-a_1)}\} \\ &\qquad\qquad \le P\{\|\frac{1}{n}(B_n')^{-1}M_n^{-1}B_n^{-1}\| > Kn^{2(\delta-\varepsilon)}\} \to 0 \end{aligned}$$

as $n \to \infty$. This concludes the proof. □

Proof of Theorem 9.1: Let $\mathcal{C}_j$ denote the collection of all orthogonal projections onto subspaces of $\mathbb{R}^d$ with dimension j. The Courant-Fischer Max-Min

Theorem (see [64],p.51) implies that

$$\lambda_{jn} = \min_{P \in \mathcal{C}_j} \max_{\|\theta\|=1} PM_nP\theta \cdot \theta$$
$$(9.9) \qquad = \max_{P \in \mathcal{C}_{d-j+1}} \min_{\|\theta\|=1} PM_nP\theta \cdot \theta.$$

Note that $P_i^2 = P_i$ and that B_n and P_i commute for all n, i. Furthermore (P_iB_n) is regularly varying with index $E_i \oplus \cdots \oplus E_p$. Since

$$n(P_iB_n)P_iM_nP_i(B_nP_i)' = nP_i(B_nM_nB_n')P_i$$

it follows by projection from our basic assumptions that the sample covariance matrix formed from the L_i valued random variables $P_i\boldsymbol{X}_t$ satisfies again those basic assumptions with $E = E_i \oplus \cdots \oplus E_p$ on L_i. Hence if λ_n denotes the smallest eigenvalue of the matrix $P_iM_nP_i$ it follows from Proposition 9.3 that

$$\frac{\log(n\lambda_n)}{2\log n} \to a_i \quad \text{in probability.}$$

Similarly, the sample covariance matrix formed in terms of the $\bar{L}_i$-valued random vectors $\bar{L}_i\boldsymbol{X}_t$ again satisfies the basic assumptions with $E = E_1 \oplus \cdots \oplus E_i$ as above. Then, if $\bar{\lambda}_n$ denotes the largest eigenvalue of the matrix $\bar{P}_iM_n\bar{P}_i$ it follows from Proposition 9.2 above that

$$\frac{\log(n\bar{\lambda}_n)}{2\log n} \to a_i \quad \text{in probability.}$$

Now apply (9.9) to see that

$$\lambda_n \le \lambda_{jn} \le \bar{\lambda}_n$$

whenever $\bar{b}_{i-1} < j \le \bar{b}_i$. The result now follows easily. □

After dealing with the asymptotics of the eigenvalues of the sample covariance in Theorem 9.1 above we now investigate the convergence of the unit eigenvectors of M_n. Recall that $\pi_i : \mathbb{R}^d \to V_i$ denotes the orthogonal projection onto V_i for $i = 1, \ldots, p$. Define the random projection

$$\pi_{in}(\boldsymbol{x}) = \sum_{j=\bar{b}_{i-1}+1}^{\bar{b}_i} (\boldsymbol{x} \cdot \theta_{jn})\theta_{jn}.$$

Theorem 9.4. *Under the basic assumptions we have $\pi_{1n} \to \pi_1$ and $\pi_{pn} \to \pi_p$ in probability as $n \to \infty$.*

Again the proof is quite similar to the proof of Theorem 3 in [46] and Theorem 10.4.8 in [48]. See also [70]. We include here a sketch of the arguments.

Proposition 9.5. *Under our basic assumptions we have: If $j > \bar{b}_{p-1}$ and $r < p$ then*

$$\pi_r\theta_{jn} \to 0 \quad \textit{in probability.}$$

Proof. Since $\pi_r\theta_{jn} = (\pi_r M_n/\lambda_{jn})\theta_{jn}$ we get

$$\begin{aligned}\|\pi_r\theta_{jn}\| &\le \|\pi_r M_n/\lambda_{jn}\|\\ &\le \frac{\|\pi_r B_n^{-1}\|\|nB_nM_nB_n'\|\|(B_n')^{-1}\|}{n\lambda_{jn}}.\end{aligned}$$

By assumption (A1) together with continuous mapping it follows from Prohorov's theorem that $(n\|B_nM_nB_n'\|)$ is uniformly tight. Also, by (9.7), (9.8) and Theorem 9.1 we get

$$\begin{aligned}&\frac{\log(\|\pi_r B_n^{-1}\|\|nB_nM_nB_n'\|\|(B_n')^{-1}\|)/(n\lambda_{jn})}{\log n}\\ =&\frac{\log\|\pi_r B_n^{-1}\|}{\log n} + \frac{\log\|(B_n')^{-1}\|}{\log n} - \frac{\log(n\lambda_{jn})}{\log n}\\ &\to a_r + a_p - 2a_p < 0 \quad \text{in probability.}\end{aligned}$$

Hence the assertion follows. □

Proposition 9.6. *Under our basic assumptions we have: If $j \le \bar{b}_1$ and $r > 1$ then*

$$\pi_r\theta_{jn} \to 0 \quad \textit{in probability.}$$

Proof. Since $\pi_r\theta_{jn} = (\pi_r M_n^{-1}\lambda_{jn})\theta_{jn}$ we get

$$\begin{aligned}\|\pi_r\theta_{jn}\| &\le \|\pi_r M_n^{-1}\lambda_{jn}\|\\ &\le \|\pi_r B_n'\|\|\frac{1}{n}(B_n')^{-1}M_n^{-1}B_n^{-1}\|\|B_n\|(n\lambda_{jn})\end{aligned}$$

As in the proof of Proposition 9.3 the sequence $(\frac{1}{n}(B_n')^{-1}M_n^{-1}B_n^{-1}\|)$ is uniformly tight and now the assertion follows as in the proof of Proposition 9.5. □

Proof of Theorem 9.4. The proof is almost identical to the proof of Theorem 3 in [46] or Theorem 10.4.8 in [48] and therefore omitted. □

Corollary 9.7. *Under our basic assumptions, if $p \le 3$ then $\pi_{in} \to \pi_i$ in probability for $i = 1, \ldots, p$.*

Proof. Obvious. □

Example 9.8. Suppose that $\boldsymbol{Z}, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ is a sequence of independent and identically distributed (IID) random vectors with common distribution μ. We assume that μ is regularly varying with exponent E. That means that there exists a regularly varying sequence (A_n) of linear operators with index $-E$ such that

$$n(A_n\mu) \to \phi \quad \text{as } n \to \infty. \tag{9.10}$$

For more information on regularly varying measures see [48], Chapter 6.

Regularly varying measures are closely related to the generalized central limit theorem discussed in Section 3. Recall that if

$$A_n(\boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_n - n\boldsymbol{b}_n) \Rightarrow \boldsymbol{Y} \quad \text{as } n \to \infty \tag{9.11}$$

for some nonrandom $\boldsymbol{b}_n \in \mathbb{R}^d$, we say that $\boldsymbol{Z}$ belongs to the generalized domain of attraction of $\boldsymbol{Y}$ and we write $\boldsymbol{Z} \in \text{GDOA}(\boldsymbol{Y})$. Corollary 8.2.12 in [48] shows that $\boldsymbol{Z} \in \text{GDOA}(\boldsymbol{Y})$ and (9.11) holds if and only if μ varies regularly with exponent E and (9.10) holds, where the real parts of the eigenvalues of E are greater than 1/2. In this case, $\boldsymbol{Y}$ has an operator stable distribution and the measure ϕ in (9.10) is the Lévy measure of the distribution of $\boldsymbol{Y}$. Operator stable distributions and Lévy measures were discussed in Section 4, where (9.10) is written in the equivalent form $nP(A_n\boldsymbol{Z} \in dx) \to \phi(dx)$. The spectral decomposition was discussed in Section 5. Theorem 8.3.24 in [48] shows that we can always choose norming operators A_n and limit $\boldsymbol{Y}$ in (9.11) so that $\boldsymbol{Y}$ is spectrally compatible with $\boldsymbol{Z}$, meaning that A_n varies regularly with some exponent $-E$, the subspaces V_i in the spectral decomposition of $\mathbb{R}^d$ with respect to E are mutually orthogonal, and these subspaces are also A_n-invariant for every n. In this case, we write $\boldsymbol{Z} \in \text{GDOA}_c(\boldsymbol{Y})$.

Recall from Section 6 that, since the real parts of the eigenvalues of E are greater than 1/2,

$$nA_nM_nA_n' \Rightarrow W \quad \text{as } n \to \infty \tag{9.12}$$

where M_n is the uncentered sample covariance matrix

$$M_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i\boldsymbol{Z}_i'$$

and W is a random $d \times d$ matrix whose distribution is operator stable. Theorem 10.2.9 in [48] shows that W is invertible with probability one, and Theorem 10.4.2 in [48] shows that for all unit vectors $\theta \in \mathbb{R}^d$ the random variable $\theta \cdot W\theta$ has a Lebesgue density. Then the basic assumptions of this section hold, and hence the results of this section apply.

The tail estimator proven in this section approximates the spectral index function $\alpha(\boldsymbol{x})$ defined in (5.2). This index function provides sharp bounds on the tails and radial projection moments of $\boldsymbol{Z}$. Given a d-dimensional data set $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ with uncentered covariance matrix M_n, let $0 \le \lambda_{1n} \le \cdots \le \lambda_{dn}$ denote the eigenvalues of M_n and $\theta_{1n}, \ldots, \theta_{dn}$ the corresponding orthonormal basis of eigenvectors. Writing $x_j = \boldsymbol{x} \cdot \theta_j$ we can estimate the spectral index $\alpha(\boldsymbol{x})$ by

$$\hat{\alpha}(\boldsymbol{x}) = \max\{\hat{\alpha}_j : x_j \neq 0\}$$

where

$$\hat{\alpha}_j = \frac{2\log n}{\log(n\lambda_{jn})}$$

using the results of this section. Hence the eigenvalues are used to approximate the tail behavior, and the eigenvectors determine the coordinate system to which these estimates pertain. A practical application of this tail estimator appears in Example 8.1.

Example 9.9. The same tail estimation methods used in the previous example also apply to the moving averages considered in Section 7. This result is apparently new. Given a sequence of IID random vectors $\boldsymbol{Z}, \boldsymbol{Z}_j$ whose common distribution μ varies regularly with exponent E, so that (9.10) holds, we define the moving average process

$$\boldsymbol{X}_t = \sum_{j=-\infty}^{\infty} C_j \boldsymbol{Z}_{t-j} \tag{9.13}$$

where we assume that the $d \times d$ matrices C_j fulfill for each j either $C_j = 0$ or C_j is invertible and $A_n C_j = C_j A_n$ for all n. Moreover if a_p denotes the largest real part of the eigenvalues of E we assume further

$$\sum_{j=-\infty}^{\infty} \|C_j\|^\delta < \infty \tag{9.14}$$

for some $\delta < 1/a_p$ with $\delta \leq 1$. Recall from Section 7 that under those conditions $\boldsymbol{X}_t$ is almost surely well defined, and that if the real parts of the eigenvalues of E are greater than $1/2$ we have that

$$nA_n \hat{\Gamma}_n(0) A_n' \Rightarrow M = \sum_{j=-\infty}^{\infty} C_j W C_j' \quad \text{as } n \to \infty. \tag{9.15}$$

where the sample covariance matrix $\hat{\Gamma}_n(h)$ is defined by (7.6) and W is a random $d \times d$ matrix whose probability distribution is operator stable. Suppose that the norming operators A_n are chosen so that (9.11) holds and $\boldsymbol{Z} \in \mathrm{GDOA}_c(\boldsymbol{Y})$. Then in view of our basic assumptions (A1) and (A2) it remains to show:

Lemma 9.10. *Under the assumptions of the paragraph above the limiting matrix M in* (9.15) *is a.s. positive definite and for any unit vector θ the random variable $M\theta \cdot \theta$ has no atom at zero.*

Proof. Since W in (9.12) is a.s. positive definite we have for any $\theta \neq 0$ that $C_j W C_j' \theta \cdot \theta = W C_j' \theta \cdot C_j' \theta \geq 0$ for all j and strictly greater that zero for those j with $C_j \neq 0$. Hence

$$M\theta \cdot \theta = \sum_{j=-\infty}^{\infty} C_j W C_j' \theta \cdot \theta > 0$$

for any $\theta \neq 0$ so M is positive definite.

Moreover if for a given unit vector θ we set $\boldsymbol{z}_j = C_j'\theta$ then $\boldsymbol{z}_{j_0} \neq 0$ for at least one j_0. Since W is almost surely positive definite we have

$$P\{M\theta\cdot\theta < t\} = P\Big\{\sum_{j=-\infty}^{\infty} W\boldsymbol{z}_j\cdot\boldsymbol{z}_j < t\Big\} \leq P\{W\boldsymbol{z}_{j_0}\cdot\boldsymbol{z}_{j_0} < t\} \to 0$$

as $t\to 0$ using the fact that $W\boldsymbol{z}_{j_0}\cdot\boldsymbol{z}_{j_0}$ has a Lebesgue density as above. Hence $M\theta\cdot\theta$ has no atom at zero. □

It follows from (9.15) together with Lemma 9.10 that the $\boldsymbol{X}_t$ defined above fulfill the basic assumptions of this section. Hence it follows from Theorem 9.1 and Theorem 9.4 that the tail estimator used in Example 9.8 also applies to time-dependent data that can be modeled as a multivariate moving average. We can also utilize the uncentered sample covariance matrix (6.3), which has the same asymptotics as long as $E\boldsymbol{Z} = 0$ (c.f. Theorem 10.6.7 and Corollary 10.2.6 in [48]). In either case, the eigenvalues can be used to approximate the tail behavior, and the eigenvectors determine the coordinate system in which these estimates apply.

Example 9.11. Suppose now that $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ are IID $\mathbb{R}^d$-valued random vectors with common distribution μ. We assume that μ is $\mathrm{ROV}_\infty(E,c)$, meaning that there exist (A_n) regularly varying with index $-E$, a sequence (k_n) of natural numbers tending to infinity with $k_{n+1}/k_n \to c > 1$ such that

$$k_n(A_{k_n}\mu) \to \phi \quad \text{as } n\to\infty. \tag{9.16}$$

See [48], Section 6.2 for more information on R-O varying measures.

R-O varying measures are closely related to a generalized central limit theorem. In fact, if μ is $\mathrm{ROV}_\infty(E,c)$ and the real parts of the eigenvalues of E are greater than 1/2 then (9.16) is equivalent to

$$A_{k_n}(\boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_{k_n} - k_n\boldsymbol{b}_n) \Rightarrow \boldsymbol{Y} \quad \text{as } n\to\infty,$$

where $\boldsymbol{Y}$ has a so called (c^E, c) operator semistable distribution. See [48], Section 7.1 and Section 8.2 for details. Once again, a judicious choice of norming operators and limits guarantees that $\boldsymbol{Y}$ is spectrally compatible with $\boldsymbol{Z}$, so that A_n varies regularly with some exponent $-E$, the subspaces V_i in the spectral decomposition of $\mathbb{R}^d$ with respect to E are mutually orthogonal, and these subspaces are also A_n-invariant for every n. It follows from Theorem 8.2.5 of [48] that $\boldsymbol{Z}$ has the same moment and tail behavior as for the generalized domain of attraction case considered in Section 5. In particular, there is a spectral index function $\alpha(\boldsymbol{x})$ taking values in the set $\{a_1^{-1},\ldots,a_p^{-1}\}$ where $a_1 < \cdots < a_p$ are the real parts of the eigenvalues of E. Given $\boldsymbol{x}\neq 0$, for any small $\delta > 0$ we have

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

for all $r > 0$ sufficiently large. Then $E(|\boldsymbol{Z} \cdot \boldsymbol{x}|^\beta)$ exists for $0 < \beta < \alpha(\boldsymbol{x})$ and diverges for $\beta > \alpha(\boldsymbol{x})$.

Now let

$$M_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i'$$

denote the sample covariance matrix of $(\boldsymbol{Z}_i)$. Then it follows from Theorem 10.2.3, Corollary 10.2.4, Corollary 10.2.6, Theorem 10.2.9, and Lemma 10.4.2 in [48] that M_n fulfills the basic assumptions (A1) and (A2) of this section. Hence, by Theorem 9.1 and Theorem 9.4 we rediscover Theorem 10.4.5 and Theorem 10.4.8 of [48]. See also [70]. In other words, the approximation $\hat{\alpha}(\boldsymbol{x})$ from Example 9.8 still functions in this more general case, which represents the most general setting in which sums of IID random vectors can approximated in distribution via a central limit theorem.

10. Conclusions

If one believes that asset price changes (or log-returns) have heavy tails, then there is ample reason to seek a model where the tail thickness parameter α varies with the asset. Operator stable random vectors provide such a model, and are justified by a central limit theorem. Matrix-scaled sums of independent, identically distributed random vectors can only converge (in a distributional sense) to an operator stable limit. Such random vectors have regularly varying probability distributions whose tails are governed by a matrix exponent. Time dependent models can be constructed by taking moving averages of these random vectors. If X_i is the price change in the ith asset then the vector of price changes $\boldsymbol{X} = (X_1, \ldots, X_d)'$ can be described by such models. If θ_i measures the amount of the ith asset in a portfolio, price changes for this portfolio are of the form $\boldsymbol{X} \cdot \theta = X_1\theta_1 + \cdots + X_d\theta_d$. The probability of large jumps in price depends on the mix according to a tail index function $\alpha(\theta)$. If $2 < \alpha(\theta) < 4$ we have a finite variance model with infinite fourth moments. Then the sample covariance matrix plays the usual role as a descriptor of dependence between assets, but its asymptotics are operator stable. If $\alpha(\theta) < 2$ indicating heavy tails with infinite variance, the sample covariance matrix still provides some useful information. In particular, the coordinate system that diagonalizes this matrix also identifies the portfolios with the best or worst tail behavior.

References

[1] Aban, I. and M. Meerschaert (2001) Shifted Hill's estimator for heavy tails. *Comm. Statist. Simulation Comput.* **30**, No. 4, to appear.

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

[3] Akgiray, V. and G. Booth (1988) The stable-law model of stock returns. *J. Business Econ. Statist.* **6**, 51–57.

[4] Anderson, P. and M. Meerschaert (1997) Periodic moving averages of random variables with regularly varying tails. *Ann. Statist.* **25**, 771–785.

[5] Anderson, P. and M. Meerschaert (1998) Modeling river flows with heavy tails. *Water Resour. Res.* **34**, 2271–2280.

[6] Arnold, B.C. (1990) A flexible family of multivariate Pareto distributions. *J. Statist. Plann. Inference* **24** 249–258.

[7] Bawa, V., E. Elton and M. Gruber (1979) Simple rules for optimal portfolio selection in stable Paretian markets. *J. Finance* **34**, no. 4, 1041–1047.

[8] Belkacem, L., J. Véhel and C. Walter (2000) CAPM, risk and portfolio selection in α-stable markets. *Fractals* **8**, no. 1, 99–115.

[9] Benson, D., S. Wheatcraft, and M. Meerschaert (2000) Application of a fractional advection-dispersion equation. *Water Resour. Res.* **36**, 1403–1412.

[10] 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.

[11] Blumen, A., G. Zumofen and J. Klafter (1989) Transport aspects in anomalous diffusion: Lévy walks. *Phys Rev. A* **40** 3964–3973.

[12] Brockwell, P. and R. Davis (1991) *Time Series: Theory and Methods.* 2nd Ed., Springer-Verlag, New York.

[13] Byczkowski, T., J.P. Nolan and B. Rajput (1993) Approximation of multidimensional stable densities, *J. Multivariate Anal.* **46**, 13–31.

[14] Chamberlain, T.W., C.S. Cheung and C.C. Kwan (1990) Optimal portfolio selection using the general multi-index model: A stable-Paretian framework. *Decision Sciences* **21**, 563–571.

[15] Csörgo, S. and L. Viharos (1997) Asymptotic normality of least-squares estimators of tail indices. *Bernoulli* **3**, 351–370.

[16] Elton, E. amd M. Gruber (1995) *Modern Portfolio Theory and Investment Analysis.* Wiley, New York.

[17] Embrechts, P., C. Klüppelberg, and T. Mikosch (1997) Modelling extremal events. For insurance and finance. *Applications of Mathematics* **33**, Springer-Verlag, Berlin.

[18] Fama, E. (1965) The behavior of stock market prices. *J. Business* **38**, 34–105.

[19] Fama, E.F. (1965) Portfolio analysis in a stable Paretian market. *Management Science* **11**, 404–419.

[20] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications.* Vol. II, 2nd Ed., Wiley, New York.

[21] Fofack, H. and J. Nolan (1999) Tail Behavior, Modes and Other Characteristics of Stable Distribution. *Extremes* **2**, 39–58.

[22] Gamba, A. (1999) Portfolio analysis with symmetric stable Paretian returns. Current topics in quantitative finance (Venice, 1997), 48–69, *Contrib. Manag. Sci.*, Physica, Heidelberg.

[23] Gnedenko, B. and A. Kolmogorov (1968) *Limit Distributions for Sums of Independent Random Variables.* Translated from the Russian, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition, Addison-Wesley, Reading, Mass.

[24] Hall, P. (1981) A comedy of errors: the canonical form for a stable characteristic function. *Bull. London Math. Soc.* **13**, 23–27.

[25] Hall, P. (1982) On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc. B* **44** 37–42.
[26] Hill, B. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1173.
[27] Hirsch, M. and S. Smale (1974) *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, New York.
[28] Hosking, J. and J. Wallis (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349.
[29] Janicki, A. and A. Weron (1994) *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*. Marcel Dekker, New York.
[30] 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.
[31] Jurek, Z. and J. D. Mason (1993) *Operator-Limit Distributions in Probability Theory.* Wiley, New York.
[32] Klafter, J., A. Blumen and M. Shlesinger (1987) Stochastic pathways to anomalous diffusion. *Phys. Rev. A* **35** 3081–3085.
[33] Kotz, S., N. Balakrishnan and N.L. Johnson (2000) *Continuous Multivariate Distributions.* **1**, Wiley, New York.
[34] Kotz, S., T. Kozubowski and K. Podgórski (2001) *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance.* Birkhäuser, Boston.
[35] Kozubowski, T. J., A. K. Panorska and S. T. Rachev (2001) Statistical issues in modeling multivariate stable portfolios, this volume.
[36] Leadbetter, M., G. Lindgren and H. Rootzén (1980) *Extremes and related properties of random sequences and processes.* Springer-Verlag, New York.
[37] Loretan, M. and P. Phillips (1994) Testing the covariance stationarity of heavy tailed time series. *J. Empirical Finance* **1**, 211–248.
[38] Mandelbrot, B. (1963) The variation of certain speculative prices. *J. Business* **36**, 394–419.
[39] Mandelbrot, B. (1982) *The Fractal Geometry of Nature*. W. H. Freeman, San Francisco.
[40] McCulloch, J. (1996) Financial applications of stable distributions. *Statistical Methods in Finance: Handbook of Statistics* **14**, G. Madfala and C. R. Rao, Eds., Elsevier, Amsterdam, 393–425.
[41] McCulloch, J. (1997) Measuring tail thickness to estimate the stable index α: A critique. *J. Business Econ. Statist.* **15**, 74–81.
[42] Meerschaert, M. (1990) Moments of random vectors which belong to some domain of normal attraction. *Ann. Probab.* **18**, 870–876.
[43] Meerschaert, M. (1991) Regular variation in $\mathbb{R}^k$ and vector-normed domains of attraction. *Stat. Prob. Lett.* **11**, 287–289.
[44] Meerschaert, M. and H. P. Scheffler (1998) A simple robust estimator for the thickness of heavy tails. *J. Statist. Plann. Inf.* **71**, 19–34.
[45] Meerschaert, M. and H. P. Scheffler (1999) Sample covariance matrix for random vectors with heavy tails. *J Theoret. Probab.* **12**, 821–838.
[46] Meerschaert, M. and H. P. Scheffler (1999) Moment estimator for random vectors with heavy tails. *J. Multivariate Anal.* **71**, 145–159.
[47] Meerschaert, M. and H. P. Scheffler (2000) Moving averages of random vectors with regularly varying tails. *J. Time Series Anal.* **21**, 297–328.

[48] Meerschaert, M. and H. P. Scheffler (2001) *Limit Theorems for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*. Wiley, New York.

[49] Meerschaert, M. and H. P. Scheffler (2001) Limit Theorems for continuous time random walks. preprint.

[50] Mittnik, S. and S. Rachev (1993) Reply to comments on Modeling asset returns with alternative stable distributions, and some extensions. *Econometric Rev.* **12**, 347–389.

[51] Mittnik, S., S. Rachev, T. Doganoglu and D. Chenyao (1999) Maximum likelihood estimation of stable paretian models. *Math. Comput. Modelling* **29**, 275–293.

[52] Mittnik, S., T. Doganoglu and D. Chenyao (1999) Computing the probability density function of the stable paretian distribution. *Math. Comput. Modelling* **29**, 235–240.

[53] Modarres, R. and J.P. Nolan (1994) A method for simulating stable random vectors. *Comput. Statist.* **9**, 11–19.

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

[55] Nolan, J. P. and A. Panorska (1997) Data analysis for heavy tailed multivariate samples. Heavy tails and highly volatile phenomena. *Comm. Statist. Stochastic Models* **13**, 687–702.

[56] Nolan, J. P. (1997) Numerical calculation of stable densities and distribution functions. Heavy tails and highly volatile phenomena. *Comm. Statist. Stochastic Models* **13**, 759–774.

[57] Nolan, J. P. (2001) Maximum likelihood estimation of stable parameters. *Levy Processes*. O. Barndorff-Nielsen, T. Mikosch and S. Resnick, Eds., Birkhauser, Boston.

[58] Nolan, J. P., Panorska A. and J. H. McCulloch (2001) Estimation of stable spectral measures. *Math. Comput. Modelling*, to appear.

[59] Nolan, J. P. (2002) *Stable Distributions: Models for Heavy Tailed Data*. Birkhauser, Boston.

[60] Press, S. J. (1982) *Applied Multivariate Analysis*. 2nd ed., Robert E. Krieger, Malabar.

[61] Press, W., B. Flannery, S. Teukolsky and W. Vetterling (1987) *Numerical Recipes*. Cambridge University Press, New York.

[62] Rachev, S. and S. Mittnik (2000) *Stable Paretian Models in Finance*, Wiley, Chichester.

[63] Rachev, S. and S. Han (2000) Portfolio management with stable distributions. *Math. Methods Oper. Res.* **51**, no. 2, 341–352.

[64] Rao, C. R. (1965) *Linear Statistical Inference and Its Applications*. Wiley, New York.

[65] Resnick, S. and P. Greenwood (1979) A bivariate characterization and domains of attraction. *J. Multivariate Anal.* **9**, 206–221.

[66] Resnick, S. and C. Stărică (1995) Consistency of Hill's estimator for dependent data. *J. Appl. Probab.* **32** 139–167.

[67] Rvačeva, E. (1962) On domains of attraction of multidimensional distributions. *Select. Transl. Math. Stat. Prob.* **2**, American Math. Soc., Providence, Rhode Island, 183–205.

[68] Samorodnitsky, G. and M. Taqqu (1994) *Stable non-Gaussian Random Processes*. Chapman & Hall, New York.

[69] Scheffler, H. P. (1999) On estimation of the spectral measure of certain nonnormal operator stable laws. *Statist. Probab. Lett.* **43**, 385–392.

[70] Scheffler, H. P. (1998) *Multivariable R-O Varying Measures and Generalized Domains of Semistable Attraction* Habilitation thesis, University of Dortmund.

[71] Schultze, J. and J. Steinebach (1996) On least squares estimates of an exponential tail coefficient. *Statist. Decisions* **14**, 353–372.

[72] Sharpe, M. (1969) Operator-stable probability distributions on vector groups. *Trans. Amer. Math. Soc.* **136**, 51–65.

[73] Shlesinger, M., G. Zaslavsky, and U. Frisch (1994) Lévy flights and related topics in physics. *Lecture Notes in Physics* **450**, Springer-Verlag, Berlin.

[74] 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.

[75] Uchaikin, V. and V. Zolotarev (1999) *Chance and Stability: Stable Distributions and Their Applications.* VSP, Utrecht.

[76] Weron, R. (1996) On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statist. Probab. Lett.* **28**, no. 2, 165–171.

[77] Ziemba, W. (1974) Choosing investment portfolios when the returns have stable distributions. Mathematical programming in theory and practice (Proc. NATO Adv. Study Inst., Figueira da Foz, 1972), 443–482. North-Holland, Amsterdam.

Mark M. Meerschaert, Department of Mathematics, University of Nevada, Reno, USA

E-mail address: `mcubed@unr.edu`

Hans-Peter Scheffler, Fachbereich Mathematik, University of Dortmund, 44221 Dortmund, Germany

E-mail address: `hps@math.uni-dortmund.de`