DISCUSSION PAPERS IN STATISTICS
AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 2/07

Tyler's M-Estimator, Random Matrix Theory, and Generalized Elliptical Distributions with Applications to Finance

by

Gabriel Frahm
Uwe Jaekel

2nd version
October 21, 2008

DISKUSSIONSBEITRÄGE ZUR
STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

This page intentionally left blank.

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 2/07

Tyler's M-Estimator, Random Matrix Theory, and Generalized Elliptical Distributions with Applications to Finance

by

Gabriel Frahm[1]
Uwe Jaekel[2]

2nd version
October 21, 2008

Abstract

In recent publications standard methods of random matrix theory have been applied to principal components analysis of high-dimensional financial data. We discuss the fundamental results and potential shortcomings of random matrix theory in the light of the stylized facts of empirical finance. It is shown that the Marčenko-Pastur law generally fails when analyzing the empirical distribution function of the eigenvalues given by the sample covariance matrix of generalized elliptically distributed data. As an alternative we derive a random matrix referred to as the *spectral estimator* which is distribution-free within the class of generalized elliptical distributions. We show that the spectral estimator corresponds to Tyler's M-estimator. Substituting the sample covariance matrix by the spectral estimator resolves the problems which are due to the stylized facts and the Marčenko-Pastur law remains valid. This holds even if the data are not generalized elliptically distributed but mutually independent.

AMS Subject Classification: Primary 15A52, Secondary 62H25.

[1]University of Cologne, Statistics & Econometrics, Department of Economic and Social Statistics, Albertus-Magnus-Platz, 50923 Cologne, Germany. Phone: +49 221 470-4267, email: frahm@statistik.uni-koeln.de.

[2]University of Applied Sciences Koblenz, RheinAhrCampus, Department of Mathematics and Technology, Südallee 2, 53424 Remagen, Germany. Phone: +49 2642 932-334, email: jaekel@rheinahrcampus.de.

This page intentionally left blank.

Tyler's M-Estimator, Random Matrix Theory, and Generalized Elliptical Distributions with Applications to Finance*

Gabriel Frahm†
University of Cologne, Chair of Statistics and Econometrics
Meister-Ekkehart-Straße 9, 50937 Cologne, Germany

Uwe Jaekel‡
University of Applied Sciences Koblenz, RheinAhrCampus
Department of Mathematics and Technology
Südallee 2, 53424 Remagen, Germany

October 21, 2008

Abstract

In recent publications standard methods of random matrix theory have been applied to principal components analysis of high-dimensional financial data. We discuss the fundamental results and potential shortcomings of random matrix theory in the light of the stylized facts of empirical finance. It is shown that the Marčenko-Pastur law generally fails when analyzing the empirical distribution function of the eigenvalues given by the sample covariance matrix of generalized elliptically distributed data. As an alternative we derive a random matrix referred to as the *spectral estimator* which is distribution-free within the class of generalized elliptical distributions. We show that the spectral estimator corresponds to Tyler's M-estimator. Substituting the sample covariance matrix by the spectral estimator resolves the problems which are due to the stylized facts and the Marčenko-Pastur law remains valid. This holds even if the data are not generalized elliptically distributed but mutually independent.

AMS 2000 MSC: 15A52; 62H25; 62P05.

Key words: eigenspectrum, eigenvalue, financial data, generalized elliptical distribution, heavy tail, Marčenko-Pastur law, principal components analysis, random matrix theory, spectral density, stylized facts, tail dependence, tail index, Tyler's M-estimator.

*A previous version of the manuscript has been presented under the title *Applicability of Random Matrix Theory Given the Stylized Facts of Empirical Finance* at the Fourth World Congress of the Bachelier Finance Society in Tokyo, 2006.

†Email: frahm@statistik.uni-koeln.de, phone: +49 221 470-4267.

‡Email: jaekel@rheinahrcampus.de, phone: +49 2642 932-334.

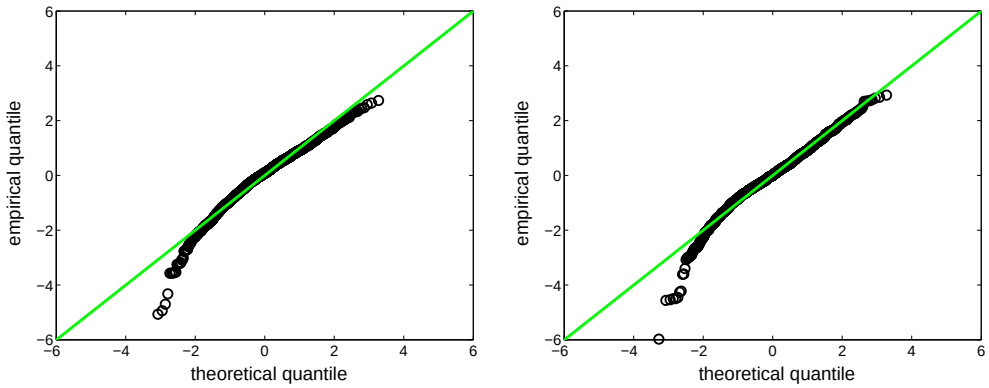

Figure 1: Normal QQ-plots of GARCH(1,1) residuals of daily log-returns of NASDAQ (left hand) and S&P 500 (right hand) from 1993-01-01 to 2000-06-30 ($n = 1892$).

1 Stylized Facts of Empirical Finance

1.1 Motivation

Distributions of short-term financial data usually exhibit some *stylized facts*, e.g. heavy tails or at least leptokurtosis, extremal or tail dependence, skewness and other kinds of asymmetries, volatility clusters or even long-memory, and so on. This holds especially if log-price changes (so-called *log-returns*) of stocks, stock indices, and foreign exchange rates are considered. Furthermore, high-frequency data generally are non-stationary, have jumps, and are strongly dependent. Indeed, there is a vast and still growing literature on that topic, e.g. Bouchaud et al. (1997), Breymann et al. (2003), Ding et al. (1993), Eberlein and Keller (1995), Embrechts et al. (1997, Chapter 6), Engle (1982), Fama (1965), Junker and May (2005), Mandelbrot (1963), McNeil et al. (2005, Section 4.1.1), and Mikosch (2003, Chapter 1).

Figure 1 shows normal QQ-plots of GARCH$(1, 1)$ residuals given by daily log-returns of the NASDAQ and S&P 500 stock indices from 1993-01-01 to 2000-06-30. Here the particular choice of the indices is rather arbitrary and the phenomena discussed later on can be observed for many stocks or stock indices. The QQ-plots clearly indicate that the normal distribution hypothesis is not appropriate for the left tails of the distributions whereas the Gaussian law seems to be acceptable for the right tails. Hence the probability of extreme losses is higher than suggested by the normal distribution assumption.

Figure 2 shows the joint distribution of the GARCH residuals considered above. Essentially there are four effects which can be observed by the scatter plot:

1. The main part of the distribution seems to be *elliptically contoured*.
2. However, we can observe a few outliers or *extreme values*, and
3. almost all extremes occur *simultaneously* whereas
4. the outliers are *not symmetrically* distributed.

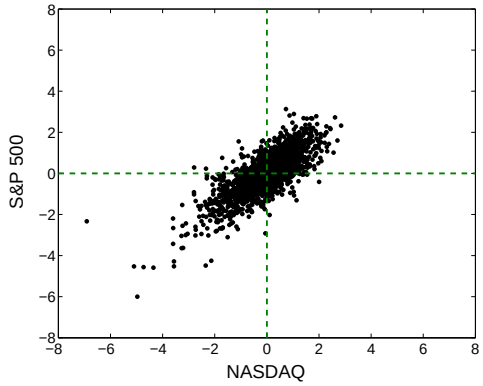

Figure 2: NASDAQ vs. S&P 500 GARCH(1,1) residuals from 1993-01-01 to 2000-06-30 ($n = 1892$).

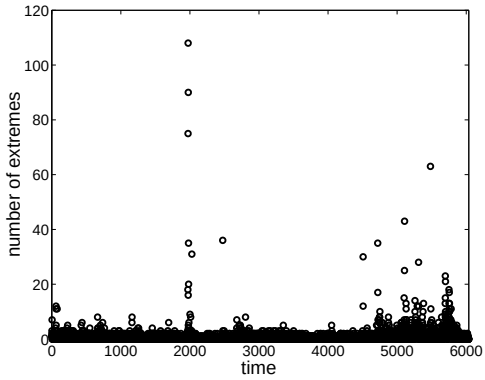

Figure 3: Number of extremes in the S&P 500 during 1980-01-02 to 2003-11-26.

The effect of *simultaneous* extremes can be observed more precisely in Figure 3. It shows the total numbers of S&P 500 stocks whose absolute values of daily log-returns exceeded 10% for each trading day during 1980-01-02 to 2003-11-26. On the 19th October 1987 (the so-called *Black Monday*) there occurred 239 extremes. This number is suppressed for the sake of transparency. This figure points out the concomitance of extremes. If the extremes of each stock would occur independently then the number of extremal events (no matter if losses or profits) should be small and more or less constant over time. Obviously, this is not the case. In contrast one can see the October crash of 1987 and several extremes which occur permanently since the beginning of the bear market in 2000. Hence there is an increasing tendency of simultaneous losses. The phenomenon of simultaneous extremes is often denoted by *asymptotic dependence*, *extremal* or *tail dependence* and is part of copula theory as well as multivariate extreme value theory. We will avoid a formal definition of copulas or tail dependence. A profound treatment of copula theory can be found, e.g., in Joe (1997) and Nelsen (2006) whereas Mikosch (2003, Chapter 4) gives a nice overview on extreme value theory. Our arguments are based on the fact that financial data exhibit tail dependence. Indeed, this is indicated by many empirical studies (see, e.g., Breymann et al., 2003, Junker and May, 2005).

1.2 Generalized Elliptical Distributions

1.2.1 Elliptically Symmetric Distributions

It is well-known that the multivariate normal distribution does neither allow for heavy tails nor for tail dependence. To overcome that problem members of the traditional class of elliptically symmetric distributions (Cambanis et al., 1981, Fang et al., 1990, Kelker, 1970) are often proposed for the modeling of financial data (cf., e.g., Bingham and Kiesel, 2002, Eberlein and Keller, 1995).

In the following definition the term *unit hypersphere* refers to the manifold

$$\mathcal{S}^{k-1} := \left\{ u \in \mathbb{R}^k : \|u\| = 1 \right\}$$

and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^k$.

Definition 1 (Elliptical Distribution). *A d-dimensional random vector X is said to be elliptically distributed if and only if there exist*

1. *a k-dimensional random vector U, uniformly distributed on the unit hypersphere,*
2. *a nonnegative random variable $\mathcal{R}$ being stochastically independent of U,*
3. *a vector $\mu \in \mathbb{R}^d$, and a matrix $\Lambda \in \mathbb{R}^{d\times k}$ such that*

$$X \stackrel{\mathrm{d}}{=} \mu + \Lambda \mathcal{R} U \,.$$

In the following discussion we will call $\Sigma := \Lambda\Lambda^{\mathsf{T}}$ the *dispersion matrix* of X and $\mathcal{R}$ its *generating variate*. Many well-known multivariate distributions belong to the class of elliptically contoured distributions. For instance, the multivariate Gaussian distribution is elliptical since it can be represented by

$$X \stackrel{\mathrm{d}}{=} \mu + \Lambda \sqrt{\chi_k^2}\, U.$$

Further, the multivariate symmetric α-stable or, synonymously, the *sub-Gaussian distribution* is given by

$$X \stackrel{\mathrm{d}}{=} \mu + \Lambda \sqrt{S_{\alpha/2}\chi_k^2}\, U \,,$$

where $0 < \alpha < 2$ and $\mathcal{S}_{\alpha/2}$ is a positive $\alpha/2$-stable distributed random variable with skewness parameter $\beta = 1$. Further, $\mathcal{S}_{\alpha/2}$ and χ_k^2 are stochastically independent. For $\alpha = 1$ we obtain the multivariate symmetric Cauchy distribution, i.e. the multivariate t-distribution with one degree of freedom (see below).

The multivariate t-distribution with $\nu > 0$ degrees of freedom is given by

$$X \stackrel{\mathrm{d}}{=} \mu + \Lambda \sqrt{\nu\chi_k^2/\chi_\nu^2}\, U \,,$$

where χ_k^2 and χ_ν^2 are stochastically independent. Note that ν also corresponds to the *tail index* or *regular variation index* of X (Mikosch, 2003). For $\nu \to \infty$ we obtain

the multivariate normal distribution as a special case of the multivariate t-distibution. Moreover, the multivariate symmetric generalized hyperbolic distribution is given by

$$X \stackrel{\mathrm{d}}{=} \mu + \Lambda\sqrt{\zeta\chi_k^2}U\,,$$

where ζ is generalized inverse Gaussian distributed. Again, ζ and χ_k^2 are stochastically independent. For $\zeta = \nu/\chi_\nu^2$ (i.e. ζ is inverse gamma distributed) once again the multivariate t-distribution with ν degrees of freedom occurs. The generalized hyperbolic distribution also contains the hyperbolic, the normal-inverse Gaussian, and the generalized Laplace distribution (or, synonymously, the variance-gamma distribution). For a nice overview of the generalized hyperbolic distribution and its statistical properties see e.g. McNeil et al. (2005, Section 3.2.3) and Prause (1999, Chapter 1).

The main fact that we would like to point out for the further discussion is that elliptical distributions possess two sorts of dependencies, viz

1. *linear* dependencies, which can be expressed by the dispersion matrix Σ and
2. *nonlinear* dependencies imposed by the generating variate $\mathcal{R}$.

Hence, the generating variate $\mathcal{R}$ not only defines the particular elliptical distribution family and – provided $\mathcal{R}$ is regularly varying – the heaviness of the tails but also that part of the dependence structure which cannot be reduced to a basis transformation caused by the matrix Λ. Particularly, that means that the components of an elliptically distributed random vector can be highly dependent even if they are uncorrelated!

For instance, consider the 2-dimensional random vector $U = (U_1, U_2)$ uniformly distributed on the unit circle. Obviously, the components of U are uncorrelated. Nevertheless, each component of U heavily depends on the other component by the relation

$$U_2 = \pm\sqrt{1 - U_1^2}\,.$$

Of course, that sort of nonlinear dependence has nothing to do with tail dependence. But if U is multiplied by a *regularly varying* or say *heavy tailed* generating variate $\mathcal{R}$ then the tail index of $\mathcal{R}$ carries over to the spherical random vector $\mathcal{R}U$ and thus to the elliptical random vector $X =_{\mathrm{d}} \mu + \Lambda\mathcal{R}U$ (Hult and Lindskog, 2002, Schmidt, 2002). Particularly, the *tail-dependence coefficients* and also the *extremal dependence coefficient* of X are essentially determined by the tail index of $\mathcal{R}$ (Frahm, 2006, Frahm et al., 2003, Hult and Lindskog, 2002, Schmidt, 2002).

1.2.2 Asymmetric Distributions

Elliptical distributions inherit many nice properties from the multivariate Gaussian distribution. For instance, they are closed under affine transformations, the marginal distributions are also elliptical, and even the conditional distributions remain elliptical. Especially, the distributions considered above are *infinitely divisible* which is an appealing property for the modeling of financial data (Bingham and Kiesel, 2002). Further, due to the simple stochastic representation of elliptical distributions they are

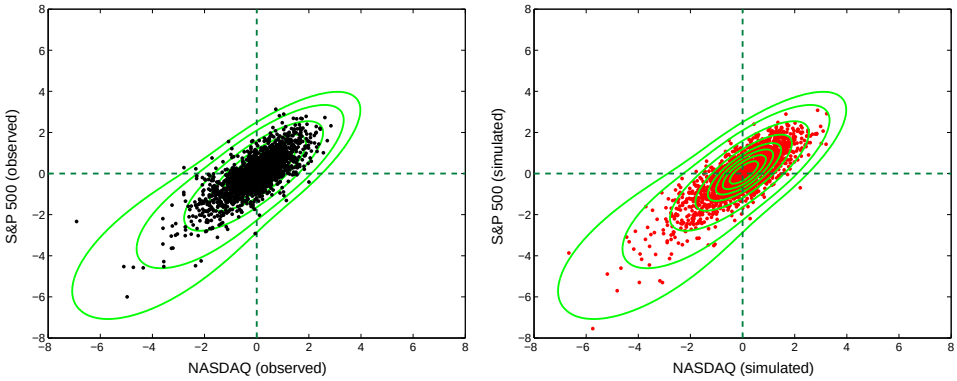

Figure 4: Observed GARCH(1,1) residuals of NASDAQ and S&P 500 (left hand) and simulated generalized elliptically distributed residuals ($n = 1892$) (right hand).

appropriate for modeling of *high-dimensional* financial data. However, elliptical distributions suffer from the property of symmetry. The pictures above show that extremes of financial data are not always symmetrically distributed. For that reason we will bear on the class of *generalized elliptical distributions* (Frahm, 2004, Chapter 3).

Definition 2 (Generalized Elliptical Distribution). *A d-dimensional random vector X is said to be generalized elliptically distributed if and only if there exist*

1. *a k-dimensional random vector U, uniformly distributed on the unit hypersphere,*
2. *a random variable $\mathcal{R}$,*
3. *a vector $\mu \in \mathbb{R}^d$, and a matrix $\Lambda \in \mathbb{R}^{d \times k}$ such that*

$$X \stackrel{\mathrm{d}}{=} \mu + \Lambda \mathcal{R} U \, .$$

Note that all the components of elliptical distributions, i.e. the location vector μ, the dispersion matrix Σ, and the generating variate $\mathcal{R}$ are preserved, but generally $\mathcal{R}$ can be *negative* and even more it may *depend* on U. This fact allows for the modeling of tail dependence and asymmetry. It is worth to point out that the class of generalized elliptical distributions not only includes the traditional class of elliptically symmetric distributions but also a relatively new class known as *skew-elliptical distributions* (Branco and Dey, 2001, Liu and Dey, 2004). This can be obtained by a modeling technique called *hidden truncation* (Arnold and Beaver, 2004, Frahm, 2004, p. 47). However, skew-elliptical distributions have been introduced especially for the modeling of skewness and heavy tails rather than tail dependence (Branco and Dey, 2001).

Figure 4 shows once again the joint distribution of the GARCH residuals of NASDAQ and S&P 500 log-returns from 1993-01-01 to 2000-06-30 which are also given in Figure 2. The right hand side of Figure 4 contains $n = 1892$ simulated GARCH residuals on the basis of a generalized elliptical distribution (where the green curves are the corresponding density contours). More precisely, the generating variate $\mathcal{R}$ corresponds

to $\sqrt{\nu\chi_2^2/\chi_\nu^2}$ but the number of degrees of freedom ν depends on the 2-dimensional random vector $U = (U_1, U_2)$ due to

$$\nu = 4 + 996 \cdot \delta\left(\frac{\Lambda U}{\|\Lambda U\|}, w\right)^3,$$

where $\delta(v, w) := \angle(v, w)/\pi = \arccos(v^\mathsf{T} w)/\pi$ measures the 'distance' between

$$V := \frac{\Lambda U}{\|\Lambda U\|} \tag{1}$$

and a *reference vector*

$$w := (-\cos(\pi/4), -\sin(\pi/4)).$$

Note that GARCH residuals have zero mean and unit variance by definition. Hence, for modeling the linear dependence structure we may concentrate on the correlation coefficient of the observed GARCH residuals of NASDAQ and S&P 500 which corresponds to $\rho = 0.78$. That means for the transformation matrix Λ we may choose the Cholesky root

$$\sqrt{\begin{bmatrix} 1 & 0.78 \\ 0.78 & 1 \end{bmatrix}} = \begin{bmatrix} 1 & 0 \\ 0.78 & 0.63 \end{bmatrix}.$$

Hence, if V is *close* to the reference vector (that means close to the 'perfect loss scenario') then the corresponding random vector is supposed to be t-distributed with almost $\nu = 4$ degrees of freedom (since $\delta \approx 0$). In contrast, a random vector exposed to the *opposite* direction is assumed to be nearly Gaussian distributed (since $\delta \approx 1$ and thus ν is large). Admittedly, this specific parameterization is rather arbitrary. However, by comparing the right hand side of Figure 4 (i.e. the simulated data) with its left hand side (i.e. the observed GARCH residuals) we can see that the chosen model is able to reproduce the stylized facts observed in Figure 2.

1.3 Conclusions

In virtue of the previous findings our conclusions are as follows:

1. The class of generalized elliptical distributions contains many well-known multivariate distributions. Specifically, it includes the class of elliptically symmetric and skew-elliptical distributions.

2. High-dimensional time series reflecting the stylized facts of empirical finance can be readily modeled by means of generalized elliptical distributions.

3. This class of distributions seems to be an appropriate model for financial data to investigate standard methods of random matrix theory.

The problem is that there is a *tremendous* amount of generalized elliptical distribution families which can be considered for the modeling of financial data. Later on we will see that the results given by standard methods of random matrix theory heavily depend on the underlying assumptions concerning the dependence structure of the data and this is essentially determined by the particular generalized elliptical distribution family or, more precisely, by the generating variate $\mathcal{R}$. Thus we aim at finding an alternative approach that is *distribution-free* within the class of generalized elliptical distributions such that standard methods of random matrix theory can be applied given the stylized facts of empirical finance despite that we do not know the 'true' distribution family or generating variate.

2 Random Matrix Theory

2.1 Principal Components Analysis

Recall that – by the Spectral Decomposition Theorem – every positive semi-definite $d \times d$ matrix Σ can be decomposed by

$$\Sigma = \mathcal{O}\mathcal{D}\mathcal{O}^{\mathsf{T}},$$

where $\mathcal{O}$ is an orthonormal matrix of eigenvectors of Σ and $\mathcal{D}$ is a diagonal matrix containing its eigenvalues $\lambda_1, \ldots, \lambda_d \geq 0$. So we obtain

$$\Sigma = \mathcal{O}\sqrt{\mathcal{D}}\left(\mathcal{O}\sqrt{\mathcal{D}}\right)^{\mathsf{T}},$$

where $\sqrt{\mathcal{D}}$ is a diagonal matrix containing the roots of the main diagonal elements of $\mathcal{D}$ and we may define $\Lambda := \mathcal{O}\sqrt{\mathcal{D}}$.

Hence, any generalized elliptically distributed random vector X with dispersion matrix Σ can be represented by

$$X \stackrel{\mathrm{d}}{=} \mu + \mathcal{O}\sqrt{\mathcal{D}}\,Y\,,$$

where $Y := \mathcal{R}U$ is a d-dimensional random vector of latent variables called the *factors* or *principal components* of X. We may assume for convenience that the main diagonal elements of $\mathcal{D}$ are given in decreasing order. That means X is mainly driven by the first principal component Y_1 and its impact on X can be measured by the largest eigenvalue of Σ. The next eigenvalue of Σ quantifies the influence of the second principal component Y_2 and so on. Hence the eigenspectrum of Σ contains useful information about the *linear* dependence structure of X.

2.2 The Marčenko-Pastur Law

Since Σ is an unknown parameter the dispersion matrix must be estimated. Of course, this estimator will be a random matrix. Random matrix theory (RMT) has its origin both in mathematical statistics by the results of John Wishart and in statistical physics dealing with the distribution of eigenvalues of *high-dimensional* randomly generated matrices. RMT found its first application in nuclear physics when trying to model the

energy levels of complex nuclei. It was mainly developed by Arnold (1967, 1971), Grenander (1963), Marčenko and Pastur (1967), Pastur (1972, 1973), and Wigner (1955, 1958). A review of the state of the art of RMT can be found in Bai (1999) and Mehta (1991).

Since we are interested in the eigenspectrum of covariance or dispersion matrices we will only consider symmetric random matrices. Thus the corresponding eigenvalues are always real. The empirical distribution function of the eigenvalues of a random matrix is defined as follows.

Definition 3 (Empirical Distribution Function of Eigenvalues). *Let M be a $d \times d$ symmetric random matrix with eigenvalues $\lambda_1, \ldots, \lambda_d$. Then the function*

$$F_d : \lambda \longmapsto \frac{1}{d} \cdot \sum_{i=1}^{d} \mathbb{1}_{\lambda_i \leq \lambda}$$

is called the empirical distribution function of the eigenvalues of M.

However, note that each eigenvalue of a random matrix is not a *random variable* in the formal sense since there is no single-valued mapping $M \mapsto \lambda_i$ ($i \in \{1, \ldots, d\}$) but rather $M \mapsto \lambda(M)$ where $\lambda(M)$ denotes the set of all eigenvalues of M. This can be simply fixed by assuming that the eigenvalues $\lambda_1, \ldots, \lambda_d$ are sorted either in an increasing or decreasing order.

First of all consider a sample of n independent copies of a d-dimensional random vector U which is uniformly distributed on the unit hypersphere $\mathcal{S}^{d-1}$ and thus possesses the covariance matrix I_d/d (see, e.g., Fang et al., 1990, p. 34). Multiplying the sample covariance matrix by the number of dimensions d should give an appropriate estimator for the true and normalized eigenspectrum $\lambda_1 = \ldots = \lambda_d = 1$. Indeed, if n grows to infinity such that $n/d \to \infty$ then the empirical distribution of the eigenvalues given by the normalized sample covariance matrix converges to a Dirac mass at point 1. This holds especially if d remains fixed. In contrast, for $n \to \infty$ and $n/d \to q < \infty$ the number of eigenvalues grows to infinity with the same rate as the sample size. In that case the empirical distribution of the eigenvalues does *not* converge to the Dirac mass and there can be large fluctuations of eigenvalues around 1 even if n is large. This can be seen as a curse of dimensionality problem which prevents a direct application of principal components analysis to *high-dimensional* data. This holds even if the distributional assumption concerning the data (e.g. the Gaussian distribution hypothesis) is fulfilled.

Theorem 1 (Marčenko and Pastur (1967)). *Let $U_1^{(d)}, \ldots, U_n^{(d)}$ be a sample of mutually independent d-dimensional random vectors uniformly distributed on the unit hypersphere ($n, d = 1, 2, \ldots$). Consider the random matrix*

$$Q := \frac{d}{n} \cdot \sum_{t=1}^{n} U_t^{(d)} U_t^{(d)\mathsf{T}}$$

and the empirical distribution function F_d of its eigenvalues. Suppose that $n, d \to \infty$, but $n/d \to q$ where $0 < q < \infty$. Then there exists a distribution function F_{MP} such that

$$F_d \xrightarrow{\mathrm{p}} F_{\mathrm{MP}}(\cdot\,; q),$$

at all points where F_{MP} is continuous. Moreover, it holds

$$F_{\mathrm{MP}}(\lambda\,;q) = F_{\mathrm{MP}}^{\mathrm{Dir}}(\lambda\,;q) + F_{\mathrm{MP}}^{\mathrm{Leb}}(\lambda\,;q)\,, \qquad \forall\,\lambda \in \mathbb{R}\,,$$

where the Dirac part is given by

$$F_{\mathrm{MP}}^{\mathrm{Dir}}(\cdot\,;q):\ \lambda \longmapsto \begin{cases} 1-q, & \lambda \geq 0,\ 0<q<1, \\ 0, & \textit{else}, \end{cases}$$

and the Lebesgue part by $F_{\mathrm{MP}}^{\mathrm{Leb}}(\cdot\,;q):\ \lambda \mapsto \int_0^{\lambda^+} f_{\mathrm{MP}}^{\mathrm{Leb}}(x\,;q)\,dx$ with

$$f_{\mathrm{MP}}^{\mathrm{Leb}}(\cdot\,;q):\ \lambda \longmapsto \frac{q}{2\pi}\cdot\frac{\sqrt{(\lambda_{\max}-\lambda)^+(\lambda-\lambda_{\min})^+}}{\lambda}$$

for all $\lambda > 0$ where

$$\lambda_{\min,\max} := \left(1 \pm \frac{1}{\sqrt{q}}\right)^2.$$

Proof. Marčenko and Pastur (1967). □

In the following Q will be called *Marčenko-Pastur operator*. The next theorem implies that the Marčenko-Pastur law (MPL) F_{MP} does not only hold for the Marčenko-Pastur operator but also for Pearson's correlation or, synonymously, cross correlation matrix if the data are mutually independent.

Theorem 2 (Yin (1986)). *Let $X_1^{(d)}, \ldots, X_n^{(d)}$ be a sample of mutually independent d-dimensional random vectors whose components are also mutually independent, have zero mean and unit variance. Then the empirical distribution function of the eigenvalues of*

$$S := \frac{1}{n}\cdot\sum_{t=1}^{n} X_t^{(d)} X_t^{(d)\mathsf{T}}$$

converges almost surely to the distribution function F_{MP} given by Theorem 1 as $n, d \to \infty$, but $n/d \to q$ where $0 < q < \infty$.

Proof. Yin (1986). □

Now the superscript '(d)' in '$U_t^{(d)}$' and '$X_t^{(d)}$' will be dropped for notational convenience. However, we should bear in mind that U_t and X_t are d-dimensional random vectors and the dimension grows with $n \to \infty$ such that $n/d \to q$.

2.3 Application to Principal Components Analysis

Note that in the preceding theorem it is assumed that the data are already standardized. Thus, for applying the MPL we can calculate the cross correlation matrix

$$\widehat{R} := \left[\frac{1}{n} \cdot \sum_{t=1}^{n} \left(\frac{X_{it} - \hat{\mu}_i}{\hat{\sigma}_i} \right) \left(\frac{X_{jt} - \hat{\mu}_j}{\hat{\sigma}_j} \right)^{\mathsf{T}} \right],$$

where $\hat{\mu}_k$ denotes the sample mean and $\hat{\sigma}_k$ the standard deviation of $X_{k1}, \ldots, X_{kn}$ ($k = 1, \ldots, d$). Indeed, the eigenspectrum of the cross correlation matrix converges to the MPL provided the data are mutually independent. However, it should be noted that by standardizing the data not the dispersion matrix Σ but rather the so-called *pseudo-correlation matrix* ρ is investigated. This matrix is defined by the equation $\Sigma = \sigma\rho\sigma$, where σ is a diagonal matrix containing the roots of the main diagonal elements of Σ and ρ is a square matrix whose main diagonal elements are equal to 1.

Under the null hypothesis $\rho = I_d$ the eigenspectrum of a high-dimensional cross-correlation matrix should be consistent with the MPL. More precisely, empirical eigenvalue distributions close to the MPL indicate that the components of the considered random vector are *uncorrelated* and all apparent correlations are due to 'random noise'. In contrast, the more ρ diverges from the identity matrix, the more eigenvalues are expected to exceed the *Marčenko-Pastur upper bound* $\lambda_{\max}$ given in Theorem 1 and vice versa. The exceeding eigenvalues are considered as 'signals' or 'information'. This argument is used by many authors for rejecting the null hypothesis and quantifying the number of principal components which are essentially responsible for the total variation of financial data, say the 'driving risk factors' of financial markets (see, e.g., Bouchaud and Potters, 2000, Laloux et al., 1999, Plerou et al., 1999, 2002). Indeed, since Theorem 2 does not require any specific distribution (only the second moments must be finite) and the data even do not have to be identically distributed but only stochastically independent, it seems to be a perfect justification for applying the MPL to heavy tailed financial data. We will see that this is a fallacy in the context of elliptically contoured and even more for generalized elliptically distributed data.

Note that for principal components analysis we are usually interested in analyzing the dispersion matrix Σ rather than the pseudo-correlation matrix ρ and the null hypothesis corresponds to $\Sigma = \sigma^2 I_d$ where σ^2 is a positive number. Hence, the MPL can be applied to the empirical distribution function of the eigenvalues of

$$\widehat{S} := \frac{1}{n} \cdot \sum_{t=1}^{n} \left(\frac{X_t - \hat{\mu}}{\hat{\sigma}} \right) \left(\frac{X_t - \hat{\mu}}{\hat{\sigma}} \right)^{\mathsf{T}} = \widehat{\Sigma}/\hat{\sigma}^2,$$

where $\widehat{\Sigma}$ denotes the sample covariance matrix and

$$\hat{\sigma}^2 := \frac{\mathrm{tr}(\widehat{\Sigma})}{d} = \frac{1}{d} \cdot \sum_{i=1}^{d} \lambda_i =: \bar{\lambda},$$

and $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\widehat{\Sigma}$. Note that $\mathrm{tr}(\widehat{S}) = d$ and the MPL can be simply applied to the empirical distribution function of the *normalized eigenvalues* $\lambda_1^*, \ldots, \lambda_d^*$ where $\lambda_i^* := \lambda_i/\bar{\lambda}$ ($i = 1, \ldots, d$). We will only consider this kind of normalization in the subsequent discussion.

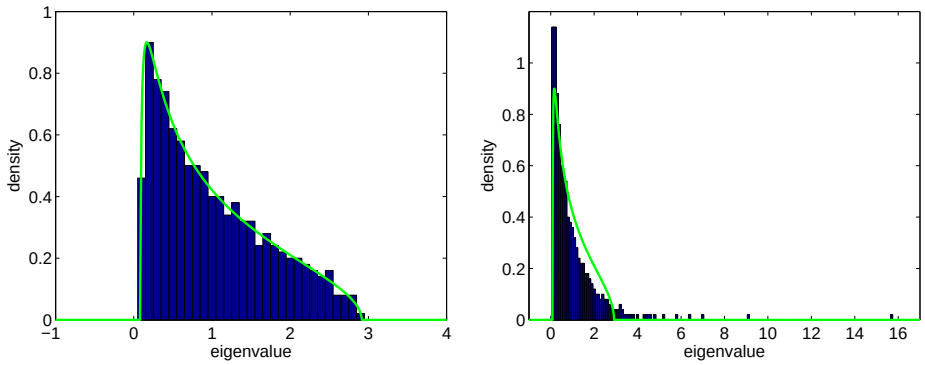

Figure 5: Eigenspectra obtained by the sample covariance matrix of univariate (left) and multivariate (right) uncorrelated t-distributed data ($n = 1000, d = 500$) with five degrees of freedom.

3 Pitfall and Alternative

3.1 Sample Covariance Matrix

Consider a sample ($n = 1000$) of 500-dimensional random vectors where the vector components are standardized t-distributed with $\nu = 5$ degrees of freedom and mutually independent. Indeed, several empirical studies show that daily log-returns of stocks typically possess between three and seven degrees of freedom after fitting a multivariate t-distribution (see, e.g., McNeil et al., 2005, p. 85). On the left hand side of Figure 5 we see that the eigenspectrum obtained by the sample covariance matrix is consistent with the MPL. In contrast, let the data be *jointly* t-distributed possessing the same parameters and each vector component being uncorrelated. Now – as indicated by the right hand side of Figure 5 – the eigenspectrum obtained by the sample covariance matrix does *not* correspond to the MPL. Actually, there are 26 *spurious* eigenvalues exceeding the Marčenko-Pastur upper bound $\lambda_{\max} = (1+1/\sqrt{2}\,)^2 = 2.91$ and the largest eigenvalue even corresponds to 15.69. That means over 5% of the eigenvalues are erroneously considered as signals or information! Moreover, the sum of the eigenvalues larger than $\lambda_{\max}$ divided by the number of dimensions, i.e. the contribution of the large eigenvalues to the total variation of the simulated data considered in Figure 5 corresponds to 24%.

Note that in the former case (i.e. mutually independence of *all* data) the considered random vector is *not* generalized elliptically distributed and although the vector components are heavy tailed they are not tail-dependent. It is well-known that the multivariate Gaussian distribution is the only elliptical distribution where uncorrelatedness and stochastical independence are equivalent. That means that the components of a random vector possessing any other elliptical distribution cannot be stochastically independent even if they are uncorrelated. Hence, in the latter case (where the random vectors are multivariate t-distributed) tail dependence is present and the principal assumption of Theorem 2 is violated. This is the reason why the MPL generally does not work for generalized elliptically distributed random vectors.

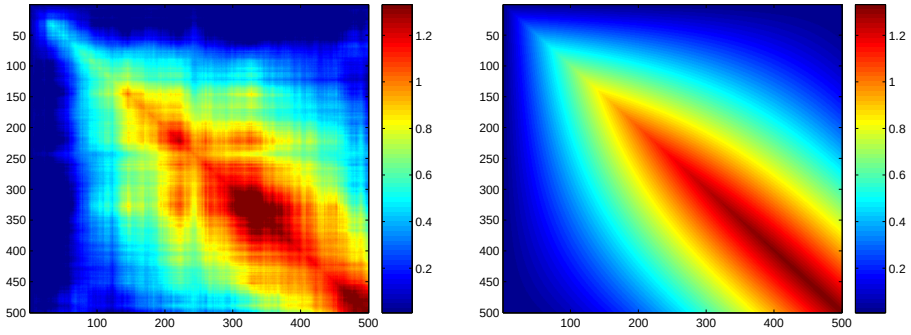

Figure 6: Normalized sample covariance matrix for multivariate t-distributed data ($n = 1000, d = 500$) with three degrees of freedom (left) and the corresponding true dispersion matrix (right).

Since the sample covariance matrix corresponds to the *ML-estimator* for multivariate Gaussian distributed random vectors but correlation and dependence are equivalent in the Gaussian case any sort of *nonlinear* dependence is confounded with linear dependence. Especially, tail dependence may lead to observations where the vector components seem to be highly correlated and the smaller the tail index of the generating variate $\mathcal{R}$, i.e. the heavier the tails of X, the more spurious eigenvalues occur.

3.2 Spectral Estimator

As mentioned before, the true linear dependence structure i.e. the dispersion matrix of a generalized elliptically distributed random vector X in general cannot be estimated efficiently (in the statistical sense) by the sample covariance matrix. If the data stem from a leptokurtic or even regularly varying elliptically distributed random vector both the finite sample and asymptotic (co-)variances of the sample covariance or cross correlation matrix can be very large (see, e.g., Lindskog et al., 2003, Oja, 2003, van Praag and Wesselman, 1989). For example, Figure 6 contains a realization of the (normalized) sample covariance matrix for multivariate t-distributed data ($n = 1000, d = 500$) with three degrees of freedom (left hand side) and the corresponding true dispersion matrix (right hand side).

Large perturbations of the sample covariance matrix can be due to extreme values (e.g. in the case of regular variation) or *contamination* of the data which is typically caused by measurement errors. It is worth to point out that these phenomena can even occur in *low* dimensions. Actually, this is a well-established branch of robust statistics and one can find a large number of robust covariance matrix estimators in the literature such as M-estimators (Maronna, 1976), S-estimators (Lopuhaä, 1989), the MVE- and MCD-estimators proposed by **?**, estimation procedures based on trimming (Gnanadesikan and Kettenring, 1972) and orthogonal projections of the data (Stahel, 1981), etc.

Further, suppose that the data are *generalized* elliptically distributed such that $\mathcal{R}$ and U depend on each other. If the true covariance matrix of X is not a linear function of Σ the sample covariance matrix even will be a *biased* estimator for Σ, generally. In

the following it is shown that there exists a distribution-free alternative to the sample covariance matrix for the class of generalized elliptical distributions. Note that we do not focus on robust but *distribution-free* estimation of Σ and therefore we will not go into the details of robust statistics.

For the following discussion it is assumed that X is a d-dimensional generalized elliptically distributed random vector where μ is supposed to be known, $\Lambda \in \mathbb{R}^{d\times k}$ with $\mathrm{r}(\Lambda) = d$, and $P(\mathcal{R} = 0) = 0$ (that is X has no atom at μ). Further, the random vector V defined by Eq. 1 is referred to as the *unit random vector* generated by Λ.

Due to the stochastic representation of X given by Definition 2 the following relations hold:

$$\frac{X-\mu}{\|X-\mu\|} \overset{\mathrm{d}}{=} \frac{\mathcal{R}\Lambda U}{\|\mathcal{R}\Lambda U\|} \overset{\text{a.s.}}{=} \pm\frac{\Lambda U}{\|\Lambda U\|} = \pm V, \tag{2}$$

where $\pm := \mathrm{sgn}(\mathcal{R})$. Note that the random vector $\pm V$ does not depend on the absolute value of $\mathcal{R}$. Especially, it is completely invariant against extreme outcomes of the generating variate. However, the sign of $\mathcal{R}$ still remains and indeed this may depend on U, anymore.

Suppose for the moment that $\pm$ is known for each realization of $\mathcal{R}$ so that we can easily calculate any realization of V. Then the dispersion matrix of X can be estimated via the method of *maximum likelihood* (ML) but without any distributional assumption concerning $\mathcal{R}$. Even the dependence structure of $\mathcal{R}$ and U is not relevant since the distribution of V depends only on Λ. Hence, the resulting estimator will be distribution-free. That is we have to calculate the density function of V, say $v \mapsto \psi(v)$ and search for $T := \Gamma\Gamma^{\mathsf{T}}$ with

$$\Gamma := \arg\max_{\Lambda} \prod_{t=1}^{n} \psi\left(v_t\,;\Lambda\right).$$

Theorem 3. *Let Λ be a $d \times k$ matrix with $\mathrm{r}(\Lambda) = d$. The density function of the unit random vector generated by Λ corresponds to*

$$v \longmapsto \psi\left(v\right) = \frac{\Gamma(d/2)}{2\pi^{d/2}} \cdot \sqrt{\det(\Sigma^{-1})} \cdot \sqrt{v^{\mathsf{T}}\Sigma^{-1}v}^{\,-d}, \qquad \forall\, v \in \mathcal{S}^{d-1},$$

where $\Sigma = \Lambda\Lambda^{\mathsf{T}}$.

Proof. See, e.g., Frahm (2004, pp. 59–60). □

This distribution is sometimes referred to as the *angular central Gaussian distribution on the sphere* (Tyler, 1987b, Kent and Tyler, 1988) but we will call it simply 'spectral density function'. This is justified by the next two corollaries.

Corollary 1. *Consider Theorem 3. The extremal positions of ψ are given by the space of normalized eigenvectors of Σ, i.e. for any $v \in \mathcal{S}^{d-1}$ satisfying $\Sigma v = \lambda v$ the value $\psi(v)$ is a local extremum of ψ and vice versa.*

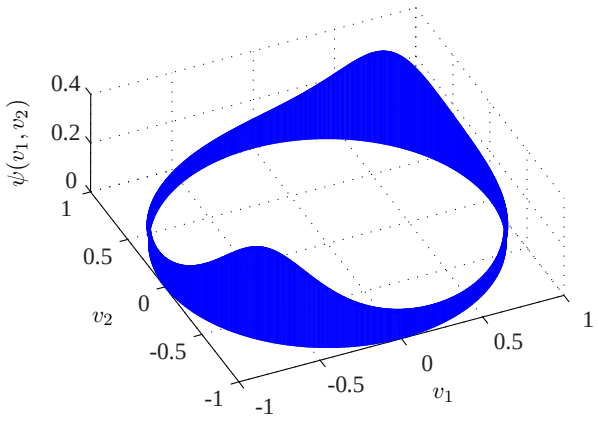

Figure 7: Spectral density of a 2-dimensional unit random vector generated by $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \Sigma_{21} = 0.7$.

Proof. Define the Lagrangian

$$\mathcal{L} := -\frac{2}{d} \cdot \log \psi(v) - \ell \left(v^{\mathsf{T}} v - 1 \right).$$

The derivative of $\mathcal{L}$ with respect to v corresponds to

$$\frac{\partial \mathcal{L}}{\partial v} = \frac{2\Sigma^{-1} v}{v^{\mathsf{T}} \Sigma^{-1} v} - 2\ell v \,.$$

Setting $\partial \mathcal{L} / \partial v = 0$ leads to the relation

$$\Sigma^{-1} v = \left(v^{\mathsf{T}} \Sigma^{-1} v \right) \ell v = \ell \lambda^{-1} v \,,$$

where λ^{-1} is an eigenvalue of Σ^{-1} and $\ell = 1$. Note that due to $\mathrm{r}(\Lambda) = d$ the matrix Σ is positive definite, $\lambda > 0$ is an eigenvalue of Σ, and v is the corresponding normalized eigenvector. We have shown that $\psi(v)$ is a stationary point. Since Σ^{-1} is also positive definite both $\varphi : v \mapsto v^{\mathsf{T}} \Sigma^{-1} v$ and $\psi \propto \varphi^{-d/2}$ cannot possess saddle points. That means $\psi(v)$ is a local extremum of ψ. □

Corollary 2. *Consider Theorem 3. The local extrema of ψ can be calculated by*

$$\psi\left(v\right) = \frac{\Gamma(d/2)}{2\pi^{d/2}} \cdot \sqrt{\det(\lambda \Sigma^{-1})} \,,$$

where λ is an eigenvalue of Σ and v is the corresponding normalized eigenvector.

Proof. This is a direct consequence of Corollary 1. □

Figure 7 exemplifies the spectral density of a unit random vector distributed on the unit circle. Note that ψ is symmetric and thus the sign of $\mathcal{R}$ does not matter at all. That means we do not have to know $\pm$ for calculating the ML-estimator based on the spectral density function. Suppose that $X_1, \ldots, X_n$ are n independent copies of X. It

can be shown that the desired ML-estimator, say the 'spectral estimator', is given by the following fixed-point equation (Frahm, 2004, Section 4.2.2):

$$T := \frac{d}{n} \cdot \sum_{t=1}^{n} \frac{V_t V_t^{\mathsf{T}}}{V_t^{\mathsf{T}} T^{-1} V_t}\,, \tag{3}$$

where $V_t := (X_t - \mu)/\|X_t - \mu\|$ for $t = 1, \ldots, n$. Actually, the spectral estimator corresponds to *Tyler's M-estimator* (Tyler, 1983, 1987a), i.e.

$$T = \frac{d}{n} \cdot \sum_{t=1}^{n} \frac{(X_t - \mu)(X_t - \mu)^{\mathsf{T}}}{(X_t - \mu)^{\mathsf{T}} T^{-1} (X_t - \mu)}\,. \tag{4}$$

Note that the solution of that fixed-point equation is only unique up to a scaling constant and thus we will require $\mathrm{tr}(T) = d$ in the subsequent discussion. We also could have considered other constraints like, e.g., $\det(T) = 1$ or fixed the upper left element of T to 1 (Frahm, 2004, p. 64). However, in the context of RMT the first restriction has the advantage that we do not need to normalize the eigenvalues of T in an extra step for applying the MPL.

The spectral estimator possesses several nice properties. For instance, it is strongly consistent, asymptotically normally distributed, and asymptotically efficient among all distribution-free estimators (Tyler, 1987a, Frahm, 2004, Chapter 5). Note that for obtaining these asymptotic properties it is implicitly assumed that the dimension d is *fixed*. We have found only one exception in the literature. Dümbgen (1998) investigated the asymptotic behavior of Tyler's M-estimator for $d \to \infty$ but $n/d \to \infty$. However, this is not the topic of RMT where $n/d \to q < \infty$ is assumed.

Further properties concerning both the existence and convergence of Tyler's M-estimator by applying fixed-point iteration algorithms were derived by Tyler (1987a) as well as Kent and Tyler (1988, 1991). Particularly, Kent and Tyler (1988) proved that for any given sample $x_1, \ldots, x_n$ the fixed-point solution T exists and the sequence (T_i) defined by the simple fixed-point iteration scheme

$$T_{i+1} = \frac{d}{n} \cdot \sum_{t=1}^{n} \frac{(x_t - \mu)(x_t - \mu)^{\mathsf{T}}}{(x_t - \mu)^{\mathsf{T}} T_i^{-1} (x_t - \mu)}\,, \qquad i = 0, 1, \ldots, \tag{5}$$

converges to $\sigma^2 T$ provided the data stem from a continuous distribution in $\mathbb{R}^d$ and $n > d$. Here the initial value T_0 can be any positive definite $d \times d$ matrix and $\sigma^2 > 0$ is a scaling constant depending on the initial value T_0. We can see by Eq. 3 that for the existence of T and convergence of (T_i) it is only required that the distribution of the *projected* data $V_1, \ldots, V_n$ defined by Eq. 1 are continuously distributed and note that

$$V \stackrel{\mathrm{d}}{=} \|\Lambda U\|^{-1} \Lambda U$$

is generalized elliptically distributed with generating variate $\mathcal{R} = \|\Lambda U\|^{-1}$. Hence, Tyler's proof holds for the class of generalized elliptical distributions, too, given the rather weak conditions mentioned at the beginning of this section.

It is worth to point out that the spectral estimator is a robust estimator and its robustness properties (i.e. breakdown point, maximum bias and variance) were already investigated by Adrover (1998), Dümbgen and Tyler (2005), Maronna and Yohai (1990),

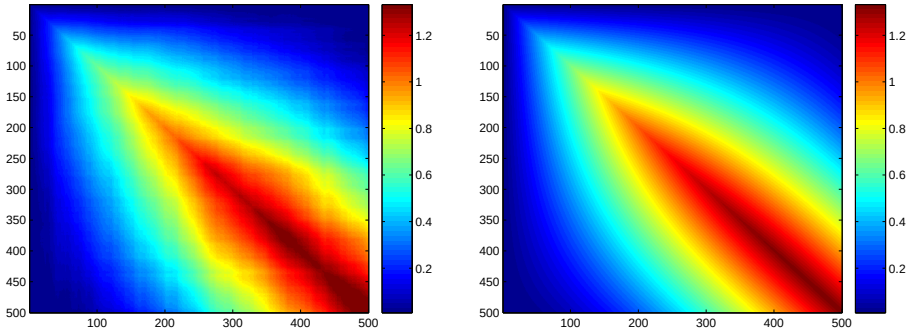

Figure 8: Spectral estimator for multivariate t-distributed data ($n = 1000, d = 500$) with three degrees of freedom (left) and the corresponding true dispersion matrix (right).

and Tyler (1983, 1987a). In particular it has been shown that the Dirac contamination breakdown point of T corresponds to $1/d$ (Maronna and Yohai, 1990) whereas for *any* kind of contamination it is between $1/(d+1)$ and $1/d$ (Adrover, 1998) if the data are elliptically distributed. Due to the arguments given above the same holds for generalized elliptical distributions and the spectral estimator breaks down for $d \to \infty$ if the data are contaminated. Thus when working with financial data it is important to eliminate clusters such as null-returns before applying Tyler's M-estimator in the context of RMT.

The left hand side of Figure 8 contains a realization of the spectral estimator for the multivariate t-distributed data already used for calculating the sample covariance matrix in Figure 6. This can be compared with the corresponding true dispersion matrix on the right hand side of Figure 8 and the sample covariance matrix in Figure 6. Obviously, the spectral estimator provides a robust alternative to the sample covariance matrix. Note that we do not need to investigate the spectral estimator under a 'true' generalized elliptical distribution such as the model proposed at the end of Section 1.2.2. By the relations (2) it was already proved that T depends only on the dispersion matrix Σ and not on the generating variate $\mathcal{R}$ or the relationship between $\mathcal{R}$ and U. That means under generalized elliptically distributed data the spectral estimator would perform as well as e.g. under the multivariate t-distributed data considered in Figure 8 provided the dispersion matrices are equal.

The following arguments are based on the distribution freeness rather than the robustness of T. Consider once again the Marčenko-Pastur operator Q given in Theorem 1 and note that under the null hypothesis $\Sigma \propto I_d$ the spectral estimator corresponds to

$$T = \frac{d}{n} \cdot \sum_{t=1}^{n} \frac{U_t U_t^{\mathsf{T}}}{U_t^{\mathsf{T}} T^{-1} U_t} \,.$$

Due to the strong consistency of T we know that $U_t^{\mathsf{T}} T^{-1} U_t \to 1$ almost surely (where d is fixed and $n \to \infty$) for every $t = 1, 2, \ldots$. Thus our intuition tells us that $T \sim Q$ for $n \to \infty$ though we have to bear in mind that in the context of RMT also d grows to infinity. Thus it is not clear whether T is strongly consistent for $n, d \to \infty$ but

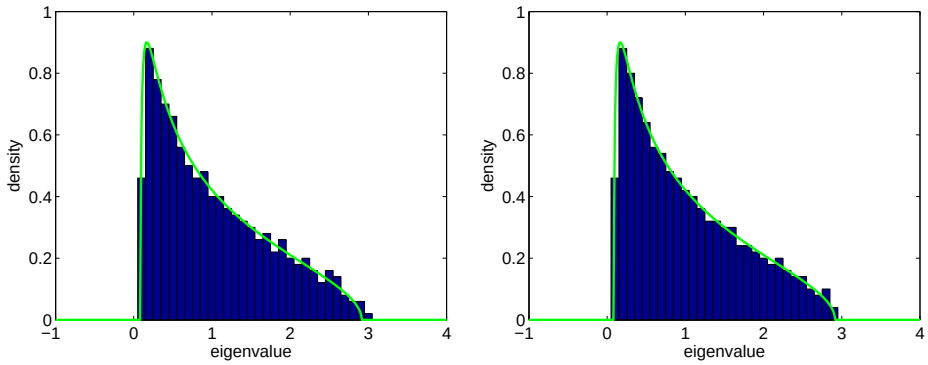

Figure 9: Eigenspectra obtained by the spectral estimator for univariate (left) and multivariate (right) uncorrelated t-distributed data ($n = 1000, d = 500$) with five degrees of freedom.

$n/d \to q < \infty$. However, we expect that the empirical distribution functions of the eigenvalues of T and Q are asymptotically equivalent.

What if the data are not generalized elliptically distributed but – as described in Theorem 2 – standardized and mutually independent? Now consider the random matrix S given by Theorem 2 and Eq. 4 with $\mu = 0$, i.e.

$$T = \frac{1}{n} \cdot \sum_{t=1}^{n} \frac{X_t X_t^{\mathsf{T}}}{X_t^{\mathsf{T}} T^{-1} X_t / d} \,.$$

If T is strongly consistent in the strict sense of RMT we obtain $X_t^{\mathsf{T}} T^{-1} X_t / d \to 1$ almost surely ($n, d \to \infty$, $n/d \to q < \infty$) for every $t = 1, 2, \ldots$ (due to the Strong Law of Large Numbers). Hence, the empirical distribution functions of the eigenvalues of T and S might be asymptotically equivalent, too, provided the conditions of Theorem 2 are fulfilled. Unfortunately, the authors did not resolve the difficulties to prove these two conjectures, yet.

Usually, the true location vector μ is unknown. It can be substituted by a consistent estimator like, e.g., the sample mean or some other robust alternative (Tyler, 1987a). Another possibility is to estimate μ and Σ *simultaneously* as described by Tyler (1987a). However, for applying RMT our simulation studies indicate that the particular choice of the location estimator does not matter at all. For the next simulation study μ was simply substituted by the sample mean $\hat{\mu}$.

Consider once again the sample of 500-dimensional random vectors with sample size $n = 1000$ where the vector components are standardized t-distributed with $\nu = 5$ degrees of freedom and mutually independent. On the left hand side of Figure 9 we can see that the eigenspectrum obtained by the spectral estimator is consistent with the MPL. Indeed, this is also true for the sample covariance matrix (see the left hand side of Figure 5). Now, if the data are jointly t-distributed possessing the same parameters but the vector components are only uncorrelated, the eigenspectrum obtained by the spectral estimator *again* is consistent with the MPL as indicated by the right hand side of Figure 9. Remember that this is not true for the sample covariance matrix (see the right hand side of Figure 5).

Acknowledgements

The authors would like to thank David E. Tyler for his valuable and encouraging comments and Karl Mosler for proof-reading the manuscript and his many important suggestions. We also thank Marc Hallin, Frank Hampel, Ingo Klein, Hans-Rudolf Künsch, Alexander McNeil, Davy Paindaveine, and Bernd Rosenow for fruitful discussions.

References

J.G. Adrover (1998), 'Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution', *Annals of Statistics* **26**, pp. 2301–2320.

B.C. Arnold and R.J. Beaver (2004), 'Elliptical Models Subject to Hidden Truncation or Selective Sampling', in: M.G. Genton, ed., 'Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality', chapter 6, Chapman & Hall.

L. Arnold (1967), 'On the asymptotic distribution of the eigenvalues of random matrices', *Journal of Mathematical Analysis and Applications* **20**, pp. 262–268.

L. Arnold (1971), 'On Wigner's semicircle law for the eigenvalues of random matrices', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **19**, pp. 191–198.

Z.D. Bai (1999), 'Methodologies in spectral analysis of large dimensional random matrices, a review', *Statistica Sinica* **9**, pp. 611–677.

N.H. Bingham and R. Kiesel (2002), 'Semi-parametric modelling in finance: theoretical foundation', *Quantitative Finance* **2**, pp. 241–250.

J.-P. Bouchaud, R. Cont, and M. Potters (1997), 'Scaling in stock market data: stable laws and beyond', in: B. Dubrulle, F. Graner, and D. Sornette, eds., 'Scale Invariance and Beyond. Proceedings of the CNRS Workshop on Scale Invariance, Les Houches, March 1997.', EDP-Springer.

J.-P. Bouchaud and M. Potters (2000), *Theory of Financial Risks*, Cambridge University Press.

M.D. Branco and D.K. Dey (2001), 'A general class of multivariate skew-elliptical distributions', *Journal of Multivariate Analysis* **79**, pp. 99–113.

W. Breymann, A. Dias, and P. Embrechts (2003), 'Dependence structures for multivariate high-frequency data in finance', *Quantitative Finance* **3**, pp. 1–14.

S. Cambanis, S. Huang, and G. Simons (1981), 'On the theory of elliptically contoured distributions', *Journal of Multivariate Analysis* **11**, pp. 368–385.

Z. Ding, C.W.J. Granger, and R.F. Engle (1993), 'A long memory property of stock market returns and a new model', *Journal of Empirical Finance* **1**, pp. 83–106.

L. Dümbgen (1998), 'On Tyler's M-functional of scatter in high dimension', *Annals of the Institute of Statistical Mathematics* **50**, pp. 471–491.

L. Dümbgen and D.E. Tyler (2005), 'On the breakdown properties of some multivariate M-functionals', *Scandinavian Journal of Statistics* **32**, pp. 247–264.

E. Eberlein and U. Keller (1995), 'Hyperbolic distributions in finance', *Bernoulli* **1**, pp. 281–299.

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

R.F. Engle (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation', *Econometrica* **50**, pp. 987–1007.

E.F. Fama (1965), 'The behavior of stock market prices', *Journal of Business* **38**, pp. 34–105.

KT. Fang, S. Kotz, and KW. Ng (1990), *Symmetric Multivariate and Related Distributions*, Chapman & Hall.

G. Frahm (2004), *Generalized Elliptical Distributions: Theory and Applications*, Ph.D. thesis, University of Cologne, Department of Economic and Social Statistics, Germany.

G. Frahm (2006), 'On the extremal dependence coefficient of multivariate distributions', *Statistics and Probability Letters* **76**, pp. 1470–1481.

G. Frahm, M. Junker, and A. Szimayer (2003), 'Elliptical copulas: applicability and limitations', *Statistics and Probability Letters* **63**, pp. 275–286.

R. Gnanadesikan and J.R. Kettenring (1972), 'Robust estimates, residuals, and outlier detection with multiresponse data', *Biometrika* **28**, pp. 81–124.

U. Grenander (1963), *Probabilities on Algebraic Structures*, John Wiley.

H. Hult and F. Lindskog (2002), 'Multivariate extremes, aggregation and dependence in elliptical distributions', *Advances in Applied Probability* **34**, pp. 587–608.

H. Joe (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall.

M. Junker and A. May (2005), 'Measurement of aggregate risk with copulas', *Econometrics Journal* **8**, pp. 428–454.

D. Kelker (1970), 'Distribution theory of spherical distributions and a location-scale parameter generalization', *Sankhya A* **32**, pp. 419–430.

J.T. Kent and D.E. Tyler (1988), 'Maximum likelihood estimation for the wrapped Cauchy distribution', *Journal of Applied Statistics* **15**, pp. 247–254.

J.T. Kent and D.E. Tyler (1991), 'Redescending M-estimates of multivariate location and scatter', *Annals of Statistics* **19**, pp. 2102–2119.

L. Laloux, P. Cizeau, J.P. Bouchaud, et al. (1999), 'Noise dressing of financial correlation matrices', *Physical Review Letters* **83**, pp. 1467–1470.

F. Lindskog, A.J. McNeil, and U. Schmock (2003), 'Kendall's tau for elliptical distributions', in: G. Bol, G. Nakhaeizadeh, S.T. Rachev, et al., eds., 'Credit Risk - Measurement, Evaluation and Management', Physica.

J. Liu and D.K. Dey (2004), 'Skew-Elliptical Distributions', in: M.G. Genton, ed., 'Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality', chapter 3, Chapman & Hall.

H.P. Lopuhaä (1989), 'On the relation between S-estimators and M-estimators of multivariate location and covariance', *Annals of Statistics* **17**, pp. 1662–1683.

B. Mandelbrot (1963), 'The variation of certain speculative prices', *Journal of Business* **36**, pp. 394–419.

R. Maronna and V. Yohai (1990), 'The maximum bias of robust covariances', *Communications in Statistics: Theory and Methods* **19**, pp. 3925–3933.

R.A. Maronna (1976), 'Robust M-estimators of multivariate location and scatter', *Annals of Statistics* **4**, pp. 51–67.

V.A. Marčenko and L.A. Pastur (1967), 'Distribution of eigenvalues for some sets of random matrices', *Mathematics of the USSR Sbornik* **72**, pp. 457–483.

A.J. McNeil, R. Frey, and P. Embrechts (2005), *Quantitative Risk Management*, Princeton University Press.

M.L. Mehta (1991), *Random Matrices*, Academic Press, second edition.

T. Mikosch (2003), 'Modeling dependence and tails of financial time series', in: B. Finkenstaedt and H. Rootzén, eds., 'Extreme Values in Finance, Telecommunications, and the Environment', Chapman & Hall.

R.B. Nelsen (2006), *An Introduction to Copulas*, Springer, second edition.

H. Oja (2003), 'Multivariate M-estimates of location and shape', in: R. Höglund, M. Jäntti, and G. Rosenqvist, eds., 'Statistics, Econometrics and Society. Essays in Honor of Leif Nordberg.', Statistics Finland.

L.A. Pastur (1972), 'On the spectrum of random matrices', *Theoretical Mathematical Physics* **10**, pp. 67–74.

L.A. Pastur (1973), 'Spectra of random self-adjoint operators', *Russian Mathematical Surveys* **28**, pp. 1–67.

V. Plerou, P. Gopikrishnan, B. Rosenow, et al. (1999), 'Universal and nonuniversal properties of cross correlations in financial time series', *Physical Review Letters* **83**, pp. 1471–1474.

V. Plerou, P. Gopikrishnan, B. Rosenow, et al. (2002), 'Random matrix approach to cross correlations in financial data', *Physical Review E* **65**, art.-no. 066126.

B.M.S. van Praag and B.M. Wesselman (1989), 'Elliptical multivariate analysis', *Journal of Econometrics* **41**, pp. 189–203.

K. Prause (1999), *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*, Ph.D. thesis, University of Freiburg, Department of Mathematical Stochastics, Germany.

R. Schmidt (2002), 'Tail dependence for elliptically contoured distributions', *Mathematical Methods of Operations Research* **55**, pp. 301–327.

W.A. Stahel (1981), 'Breakdown of covariance estimators', Research report 31, ETH Zurich, Fachgruppe für Statistik, Switzerland.

D.E. Tyler (1983), 'Robustness and efficiency properties of scatter matrices', *Biometrika* **70**, pp. 411–420.

D.E. Tyler (1987a), 'A distribution-free M-estimator of multivariate scatter', *Annals of Statistics* **15**, pp. 234–251.

D.E. Tyler (1987b), 'Statistical analysis for the angular central Gaussian distribution on the sphere', *Biometrika* **74**, pp. 579–589.

E.P. Wigner (1955), 'Characteristic vectors of bordered matrices with infinite dimensions', *Annals of Mathematics* **62**, pp. 548–564.

E.P. Wigner (1958), 'On the distributions of the roots of certain symmetric matrices', *Annals of Mathematics* **67**, pp. 325–327.

Y.Q. Yin (1986), 'Limiting spectral distribution for a class of random matrices', *Journal of Multivariate Analysis* **20**, pp. 50–68.