AP 2015–01

Chair for Applied Stochastics and Risk Management

Helmut Schmidt Universität
Universität der Bundeswehr Hamburg
Faculty of Economics and Social Sciences
Department of Mathematics/Statistics

Working Paper

Tyler's M-Estimator in High-Dimensional Financial-Data Analysis

Gabriel Frahm and Uwe Jaekel

March 16, 2015

Working Paper AP 2015–01
March 16, 2015

Tyler's M-Estimator in High-Dimensional Financial-Data Analysis

Gabriel Frahm

Helmut Schmidt University
Faculty of Economics and Social Sciences
Department of Mathematics/Statistics
Chair for Applied Stochastics and Risk Management
Holstenhofweg 85, D-22043 Hamburg, Germany

URL: www.hsu-hh.de/stochastik
Phone: +49 (0)40 6541-2791
E-mail: frahm@hsu-hh.de

Uwe Jaekel

University of Applied Sciences Koblenz, RheinAhrCampus
Faculty for Economics and Social Sciences
Department of Mathematics and Technology
Südallee 2, D-53424 Remagen, Germany

URL: www.hs-koblenz.de/profile/jaekel
Phone: +49 (0)2642 932-334
E-mail: jaekel@hs-koblenz.de

Supervised by: Prof. Dr. Gabriel Frahm
Chair for Applied Stochastics and
Risk Management

URL: www.hsu-hh.de/stochastik

Tyler’s M-Estimator in High-Dimensional Financial-Data Analysis

Gabriel Frahm*
Helmut Schmidt University
Department of Mathematics / Statistics
Chair for Applied Stochastics and
Risk Management

Uwe Jaekel†
University of Applied Sciences Koblenz, RheinAhrCampus
Department of Mathematics and Technology

March 16, 2015

Abstract

Standard methods of random matrix theory have been often applied to high-dimensional financial data. We discuss the fundamental results and potential shortcomings of random matrix theory by taking the stylized facts of empirical finance into consideration. In particular, the Marčenko-Pastur law generally fails when analyzing the spectral distribution of the sample covariance matrix if the data are generalized spherically distributed and heavy tailed. We propose Tyler’s M-estimator as an alternative. Substituting the sample covariance matrix by Tyler’s M-estimator resolves the typical difficulties that occur in financial-data analysis. In particular, the Marčenko-Pastur law remains valid. This holds even if the data are not generalized spherically distributed but independent and identically distributed.

Keywords: Generalized elliptical distribution, Marčenko-Pastur law, principal-components analysis, random matrix theory, semicircle law, tail dependence, Tyler’s M-estimator.

JEL Subject Classification: C46, C58

*Phone: +49 40 6541-2791, e-mail: frahm@hsu-hh.de.
†Phone: +49 2642 932-334, e-mail: jaekel@hs-koblenz.de.

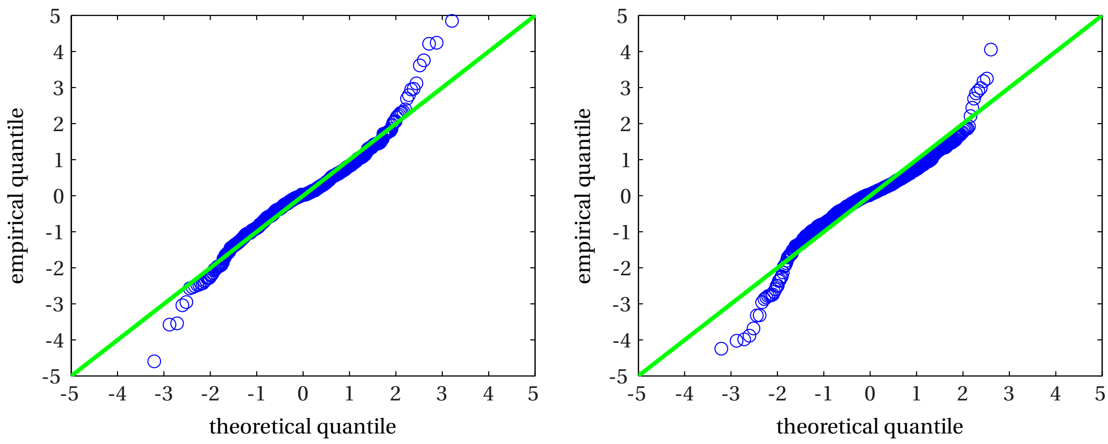

Figure 1: Normal Q-Q plots of daily log-returns on OMX Helsinki 25 (left) and DAX 30 (right) from 2007-01-03 to 2009-12-31 ($n = 756$).

1. Motivation

The distribution of short-term asset returns usually exhibits heavy tails or at least leptokurtosis, tail dependence, skewness, volatility clusters or even long memory, etc. Moreover, high-frequency data generally are non-stationary, have jumps, and are strongly dependent. These stylized facts can be observed in particular for stocks, stock indices, and foreign exchange rates. Indeed, the literature on this topic is overwhelming (Bouchaud et al., 1997, Breymann et al., 2003, Ding et al., 1993, Eberlein and Keller, 1995, Embrechts et al., 1997, Engle, 1982, Fama, 1965, Junker and May, 2005, Mandelbrot, 1963, McNeil et al., 2005, Mikosch, 2003, etc.).

Figure 1 shows normal Q-Q plots of daily log-returns on the OMX Helsinki 25 and DAX 30 from 2007-01-03 to 2009-12-31.[1] Hence, the chosen period covers the financial crisis 2007–2009. During this period we can observe $n = 756$ log-returns and the given Q-Q plots clearly indicate that the normal-distribution hypothesis is inappropriate. More precisely, the probability of extremes is much higher than suggested by the normal distribution.

Figure 2 shows the joint distribution of the log-returns considered above. We can observe the following effects in the scatter plot:

1. The central region of the distribution seems to be elliptically contoured.
2. In the margins, the joint distribution of asset returns is asymmetric.
3. There is a number of outliers or extreme values.
4. Extremes typically occur simultaneously.

The fact that extreme asset returns typically occur simultaneously is denoted by tail dependence. This is part of copula theory as well as multivariate extreme value theory. A profound treatment of copula theory can be found, e.g., in Joe (1997) and Nelsen (2006). Moreover,

[1] The particular choice of the stock indices shall symbolize the nice and fruitful collaboration between Hannu Oja (Finland) and Gabriel Frahm (Germany). Nonetheless, the empirical phenomena that can be observed in the figures occur worldwide for most other stocks and stock indices. The data used in this work have been obtained from VWD (Vereinigte Wirtschaftsdienste GmbH).

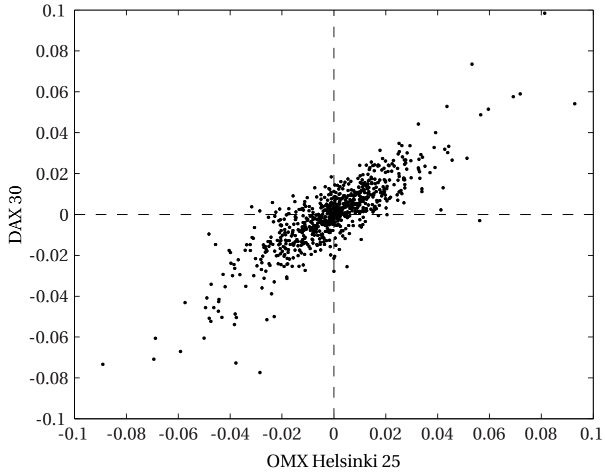

Figure 2: Daily log-returns on OMX Helsinki 25 vs. DAX 30 from 2007-01-03 to 2009-12-31.

Mikosch (2003, Ch. 4) gives a nice overview of extreme value theory. The (lower) tail-dependence coefficient of a pair of random variables X and Y or, equivalently, of their joint distribution, is defined as

$$\lambda(X,Y) := \lim_{t\searrow 0} \mathbb{P}\big(F_Y(Y) \le t \mid F_X(X) \le t\big) = \lim_{t\searrow 0} \frac{C(t,t)}{t},$$

where C is the copula of (X,Y), F_X is the marginal cumulative distribution function (c.d.f.) of X, and F_Y is the marginal c.d.f. of Y. There exist various ways to extend the concept of tail dependence to the multivariate case (De Luca and Rivieccio, 2012, Ferreira and Ferreira, 2012, Frahm, 2006).

2. Elliptical and Generalized Elliptical Distributions

2.1. Elliptical Distributions

It is well-known that the multivariate normal distribution neither allows for heavy tails nor for tail dependence. To overcome this problem, members of the traditional class of elliptical distributions (Cambanis et al., 1981, Fang et al., 1990, Kelker, 1970) are often proposed for the modeling of asset returns (cf., e.g., Bingham and Kiesel, 2002, Eberlein and Keller, 1995, McNeil et al., 2005, Ch. 3).

In the following $\mathscr{S}^{k-1} := \{\mathbf{u} \in \mathbb{R}^k : \|\mathbf{u}\| = 1\}$ represents the unit hypersphere, i.e., $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^k$.

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

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

$$\mathbf{X} = \boldsymbol{\mu} + \mathscr{R}\boldsymbol{\Lambda}\mathbf{U}.$$

The random vector $\mathbf{X}$ *is said to be spherically distributed if and only if* $\mathbf{X} = \mathscr{R}\mathbf{U}$.

We will assume that the location vector $\boldsymbol{\mu}$ is known and so we set $\boldsymbol{\mu} = \mathbf{0}$ without loss of generality. Further, we will call $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$ the dispersion matrix of $\mathbf{X}$ and $\mathscr{R}$ its generating variate.

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 $\boldsymbol{\Sigma}$ and

2. *nonlinear* dependencies imposed by the generating variate $\mathscr{R}$.

For example, consider a bivariate elliptically distributed random vector with components X and Y. Further, suppose that the generating variate $\mathscr{R}$ is regularly varying. This means we have that

$$\mathbb{P}(\mathscr{R} > x) = f(x)\, x^{-\alpha}, \qquad \forall\ x > 0,$$

where f is a slowly varying function, i.e., $f(tx)/f(x) \to 1$ as $x \to \infty$ for every $t > 0$. The number $\alpha > 0$ represents the tail index of $\mathscr{R}$ (Mikosch, 2003). Thus $\mathbb{P}(\mathscr{R} > x)$ tends to a power law for $x \to \infty$ and $\mathscr{R}$ is said to be "heavy tailed." It is intuitively clear that in this case the two components X and Y are heavy tailed, too. In fact, as is shown by Frahm et al. (2003), the tail-dependence coefficient of X and Y amounts to

$$\lambda = 2\bar{t}_{\alpha+1}\left(\sqrt{\alpha+1}\sqrt{\frac{1-\rho}{1+\rho}}\right),$$

where $\bar{t}_\nu$ denotes the survival function of Student's t-distribution with $\nu > 0$ degrees of freedom and ρ is the linear correlation coefficient of X and Y. Hence, the tail dependence is essentially determined by the tail index, α, of $\mathscr{R}$. In particular, the components X and Y can highly depend on each other in a nonlinear way even if they are uncorrelated, i.e., if $\rho = 0$ but $\mathscr{R}$ is regularly varying. The same conclusion can be drawn in the multivariate case (Frahm, 2006). Without regular variation, the most evident example, where the components of $\mathbf{X}$ are uncorrelated but (strongly) dependent, is the uniform distribution on a sphere.

2.2. Generalized Elliptical Distributions

Elliptical distributions inherit many nice properties from the Gaussian distribution. For example, they are closed under affine-linear transformations, the marginal distributions are also elliptical, and even the conditional distributions remain elliptical. Many elliptical distributions are infinitely divisible, which is an appealing property in the context of financial-data analysis (Bingham and Kiesel, 2002). Further, due to the simple stochastic representation of elliptical distributions, they are appropriate for the modeling of high-dimensional financial data. Nevertheless, they suffer from the property of symmetry. For this reason we will bear on the class of generalized elliptical distributions (Frahm, 2004, Ch. 3).

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

1. *a k-dimensional random vector* $\mathbf{U}$, *uniformly distributed on* $\mathscr{S}^{k-1}$,

2. *a random variable* $\mathscr{R}$,

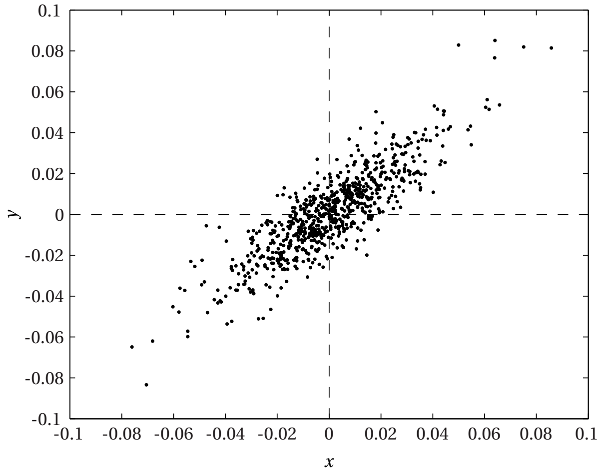

Figure 3: Simulated generalized elliptically distributed daily log-returns ($n = 756$).

3. *a vector* $\boldsymbol{\mu} \in \mathbb{R}^d$, *and a matrix* $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times k}$ *such that*

$$\mathbf{X} = \boldsymbol{\mu} + \mathscr{R}\boldsymbol{\Lambda}\mathbf{U}.$$

The random vector $\mathbf{X}$ *is said to be generalized spherically distributed if and only if* $\mathbf{X} = \mathscr{R}\mathbf{U}$.

All components of elliptical distributions, i.e., the location vector $\boldsymbol{\mu}$, the linear operator $\boldsymbol{\Lambda}$, and the generating variate $\mathscr{R}$, are preserved in Definition 2. The only difference is that $\mathscr{R}$ can be negative and even more it may depend on $\mathbf{U}$. This means the radial part of $\mathbf{X}$ may depend on its angular part. This allows us to control for tail dependence and asymmetry. A more detailed discussion regarding the practical implementation of generalized elliptical distributions can be found in Frahm (2004, Sec. 3.4) and Kring et al. (2009).

It is worth pointing out that the class of generalized elliptical distributions does not only include the traditional class of elliptical distributions, but also the class of skew-elliptical distributions (Branco and Dey, 2001, Liu and Dey, 2004). The latter 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).

By fitting Student's t-distribution to daily log-returns on stocks, several authors come to the conclusion that the number of degrees of freedom, ν, typically lies between 3 and 7 (see, e.g., McNeil et al., 2005, p. 85). Hence, the t-distribution seems to provide a fairly good fit to daily log-returns. Figure 3 contains simulated generalized elliptically distributed daily log-returns. The simulation is based on the idea that ν depends on the direction of the data. More precisely, the log-returns have been simulated as follows:

1. We calculated the eigenvector, $\mathbf{v}$, associated with the larger eigenvalue of the sample covariance matrix of the $n = 756$ daily log-returns depicted in Figure 2. This means we applied a principal-components analysis.

2. Then we simulated $n = 756$ i.i.d. random vectors $\mathbf{X}_t = \mathscr{R}_t\boldsymbol{\Lambda}\mathbf{U}_t$ ($t = 1,2,\ldots,n$), where $\boldsymbol{\Lambda}$

denotes the lower triangular Cholesky root of the sample covariance matrix.[2] Moreover, the generating variate is given by

$$\mathscr{R}_t = \sqrt{\frac{\chi^2_{t,2}}{\chi^2_{t,\nu}/\nu}}$$

with $\nu = 5 + 95\big(\min\{\angle(\mathbf{\Lambda U}_t, \mathbf{v}), \angle(\mathbf{\Lambda U}_t, -\mathbf{v})\}/(\pi/2)\big)^2$, where $\angle(\mathbf{a}, \mathbf{b})$ denotes the angle between $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ and $\chi^2_{t,2}$ is independent of $\chi^2_{t,\nu}$ and $\mathbf{U}_t$.

Hence, if $\mathbf{X}$ tends to its first principal component (or to the opposite direction), $\mathscr{R}_t$ has a tail index of $\nu = 5$ and can be considered as heavy tailed. By contrast, if it tends to its second principal component, the tail index amounts to $\nu = 100$ and so $\mathscr{R}_t$ is close to the generating variate of a normal distribution. This demonstrates that the class of generalized elliptical distributions is able to reproduce the aforementioned observations regarding the daily log-returns on the stock indices during the financial crisis 2007–2009 (cf. Figure 2).

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

1. The class of generalized elliptical distributions is sufficiently rich. In particular, it includes the class of elliptical and skew-elliptical distributions.

2. The stylized facts of empirical finance can be reproduced by the class of generalized elliptical distributions.

3. This class of distributions seems to be an appropriate model for financial data when investigating standard methods of random matrix theory (RMT).

The problem is that there exists a tremendous amount of generalized elliptical distribution families that could be considered for the modeling of financial data. Later on we will see that the results given by standard methods of RMT heavily depend on the underlying assumptions concerning the dependence structure of the data and this is essentially determined by the generating variate $\mathscr{R}$. Thus we aim at finding a *distribution-free* approach such that standard methods of RMT can be applied irrespective of the generating variate $\mathscr{R}$.

3. Random Matrix Theory

RMT has its origin in nuclear physics, where it has been developed for the modeling of the energy levels of complex nuclei. A contemporary overview of RMT can be found, e.g., in Bai and Silverstein (2010) and Debashis and Aue (2014). During the last years, this topic becomes increasingly important in statistics, particularly in financial-data analysis. For example, Bai et al. (2009), Glombek (2012) and Karoui (2010, 2013) investigate problems of Markowitz portfolio optimization. Moreover, Bouchaud et al. (2003), Laloux et al. (1999) and Plerou et al. (1999, 2002) discuss the application of RMT in the context of principal-components analysis, whereas Bai (2003) as well as Bai and Ng (2002, 2007) refer to factor analysis.

The spectral distribution of a random matrix $\mathbf{M}$ is defined as follows.

[2] The sample means of the daily log-returns on the OMX Helsinki 25 and DAX 30 are $-4.7489 \cdot 10^{-4}$ and $-1.3487 \cdot 10^{-4}$, respectively. For this reason, we can simply ignore the location.

Definition 3 (Spectral distribution)**.** *Let* $\mathbf{M}$ *be a* $d \times d$ *symmetric random matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_d$*. Then the function*

$$F_{\mathbf{M}}(\lambda) = \frac{1}{d} \sum_{i=1}^{d} \mathbf{1}_{\lambda_i \leq \lambda}, \qquad \forall\, \lambda \in \mathbb{R},$$

is called the spectral distribution of $\mathbf{M}$.

In multivariate analysis it is usually assumed that the number of dimensions, i.e., d, is fixed. By contrast, in RMT we have that $d \to \infty$ as $n \to \infty$. This makes it possible to derive asymptotic results for high-dimensional data.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a sequence of d-dimensional random vectors with zero mean. More precisely, we have that $\mathbf{X}_t = (X_{1t}, X_{2t}, \ldots, X_{dt})$ with $\mathrm{E}(X_{it}) = 0$ for each sample element X_{it} $(i, t = 1, 2, \ldots, d, n)$. The sample covariance matrix is given by

$$\mathbf{S} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{X}_t \mathbf{X}_t'.$$

In the following $\mathbf{I}_d$ denotes the $d \times d$ identity matrix and x^+ represents the positive part of $x \in \mathbb{R}$, i.e., $x^+ = \max\{0, x\}$. In RMT it is typically assumed that the sample elements are independent and identically distributed (i.i.d.).

Theorem 1 (Bai and Yin (1988))**.** *Suppose that the sample elements are i.i.d., have zero mean, unit variance, and finite fourth moment. Consider the random matrix* $\mathbf{M} = \sqrt{n/d}\,(\mathbf{S} - \mathbf{I}_d)$*. Then for all* $\lambda \in \mathbb{R}$ *the spectral distribution* $F_{\mathbf{M}}(\lambda)$ *converges almost surely to* $F_{\mathrm{W}}(\lambda) = \int_{-\infty}^{\lambda} f_{\mathrm{W}}(x)\, dx$ *with*

$$f_{\mathrm{W}}(x) = \frac{1}{2\pi} \left(4 - x^2\right)^+$$

as $n, d \to \infty$ *and* $n/d \to \infty$.

This theorem guarantees that the eigenspectrum of the sample covariance matrix of a sequence of i.i.d. data converges to Wigner's semicircle law (Wigner, 1955, 1958) as $n, d \to \infty$ and $n/d \to \infty$.[3] Indeed, this is a remarkable result, but in many practical applications, the number of observations, n, is not large enough compared to the number of dimensions, d. The following theorem only requires that $n/d \to q$ with $q \in \,]0, \infty[$ and so the effective sample size n/d can be a small number.

Theorem 2 (Bai and Silverstein (2010))**.** *Suppose that the sample elements are i.i.d., have zero mean, and unit variance. Then for all* $\lambda \in \mathbb{R}$ *the spectral distribution* $F_{\mathbf{S}}(\lambda)$ *converges almost surely to* $F_{\mathrm{MP}}(\lambda) = \int_{-\infty}^{\lambda} f_{\mathrm{MP}}(x)\, dx$ *with*

$$f_{\mathrm{MP}}(x) = \frac{q}{2\pi} \cdot \frac{\sqrt{(\lambda_+ - x)^+ \,(x - \lambda_-)^+}}{x}, \qquad \lambda_{\pm} = \left(1 \pm \frac{1}{\sqrt{q}}\right)^2,$$

as $n, d \to \infty$ *and* $n/d \to q$ *with* $1 \leq q < \infty$*. In case* $0 < q < 1$ *the limiting density is a mixture of a point mass at* 0 *and* $f_{\mathrm{MP}}(x)$ *with weights* $1 - q$ *and* q*, respectively.*

The limiting distribution that is given by Theorem 2 is known as the Marčenko-Pastur law (Marčenko and Pastur, 1967). It implies that all eigenvalues outside its support $\left[(1 - 1/\sqrt{q})^2, (1 + 1/\sqrt{q})^2\right]$ vanish asymptotically.

[3]The semicircle law implies that all eigenvalues outside its support $[-2, 2]$ vanish asymptotically.

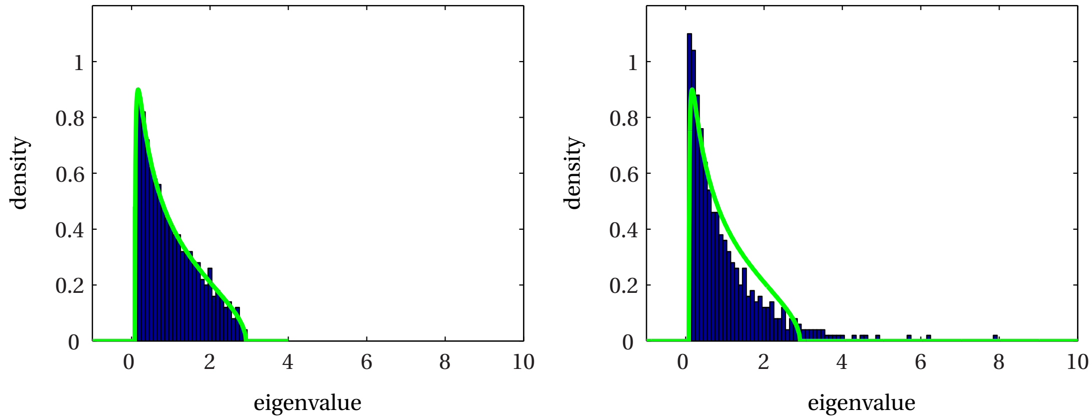

Figure 4: Eigenspectra of $\mathbf{S}$ based on non-elliptically, i.e., independent (left), vs. elliptically, i.e., uncorrelated (right), multivariate t-distributed data ($n = 1000, d = 500$) with 5 degrees of freedom. The green curve represents the density function of the Marčenko-Pastur law for $q = 2$.

4. Pitfall and Alternative

4.1. Sample Covariance Matrix

Consider a sample of 500-dimensional random vectors with sample size $n = 1000$, where the vector components are mutually independent and possess a standardized univariate t-distribution with 5 degrees of freedom. In the subsequent discussion this is said to be a multivariate *non-elliptical* t-distribution. The left-hand side of Figure 4 contains the eigenspectrum obtained by the sample covariance matrix. Obviously, this is consistent with the Marčenko-Pastur law. By contrast, suppose that the random vectors have a standardized multivariate *elliptical* t-distribution with 5 degrees of freedom. More precisely, it is supposed that the vector components are uncorrelated, i.e., $\mathbf{\Sigma} \propto \mathbf{I}_{500}$, but not independent. In this case the Marčenko-Pastur law is clearly violated.

More precisely, we find 27 spurious eigenvalues exceeding the Marčenko-Pastur upper bound $\lambda_+ = (1 + 1/\sqrt{2})^2 = 2.91$ and the largest eigenvalue corresponds to 14.0528. In the physics literature, the exceeding eigenvalues are often considered as "signals" or "information" (see, e.g., Bouchaud et al., 2003, Laloux et al., 1999, Plerou et al., 1999, 2002). In terms of principal-components analysis, the exceeding eigenvalues could be interpreted as the contribution of the first principal components to the total variation of the data. Figure 4 demonstrates that in this case we would seriously overestimate the systematic risk of asset returns.

Theorem 2 does not require any specific distributional assumption. In the context of elliptical and generalized elliptical distributions this is a potential fallacy. It is well-known that the multivariate normal is the only elliptical distribution that allows for independent components. Hence, the problem is that the components of a spherically distributed random vector $\mathbf{X} = \mathcal{R}\mathbf{U}$ are not independent unless $\mathcal{R} \propto \chi_d$, i.e., if $\mathbf{X}$ has a multivariate normal distribution. For example, if asset returns follow a multivariate elliptical t-distribution, they might be uncorrelated but never independent in the cross section. More precisely, short-term asset returns are tail dependent. This is the reason why the Marčenko-Pastur law in general does not work for spherically distributed data.[4] We often have observations suggesting that the vector components are highly correlated. The smaller the tail index of the generating variate $\mathcal{R}$, i.e., the heavier the tails of $\mathbf{X}$, the more spurious eigenvalues occur.

4.2. Tyler's M-Estimator

Since daily asset returns follow a leptokurtic or heavy-tailed distribution, it seems natural to use a robust covariance matrix as an alternative to the sample covariance matrix. In the following discussion we focus on Tyler's M-estimator (Tyler, 1987a). Its many nice properties have been established, e.g., by Adrover (1998), Dümbgen (1998), Dümbgen and Tyler (2005), Frahm (2009), Frahm and Glombek (2012), Kent and Tyler (1988, 1991), Maronna and Yohai (1990), Paindaveine (2008), Tyler (1987b), etc. We do not take other estimators into consideration, because Tyler's M-estimator turns out to be a canonical choice in the context of financial time series. This will become clear by the end of this section.

If the log-returns are elliptically distributed and the second moment of $\mathcal{R}$ is finite, we have that $\mathrm{Var}(\mathbf{X}) = \mathrm{E}(\mathcal{R}^2)/k \cdot \mathbf{\Sigma}$. However, in many applications of multivariate data analysis we need to know only the shape matrix of $\mathbf{X}$, i.e., $\mathbf{\Omega} = \mathbf{\Sigma}/\sigma^2(\mathbf{\Sigma})$, where σ^2 is any scale function, i.e., a positive homogeneous function of degree 1 such that $\sigma^2(\mathbf{I}_d) = 1$.[5] The shape matrix $\mathbf{\Omega}$ reflects the linear dependence structure of $\mathbf{X}$. Since the covariance matrix of $\mathbf{X}$ is proportional to $\mathbf{\Sigma}$, $\mathbf{S}/\sigma^2(\mathbf{S})$ represents a consistent estimator for $\mathbf{\Omega}$. In general, this is not satisfied if $\mathcal{R}$ depends on $\mathbf{U}$, i.e., if $\mathbf{X}$ is not elliptically distributed. In the subsequent discussion it is shown that Tyler's M-estimator is a canonical estimator for the linear dependence structure of any generalized elliptically distributed random vector $\mathbf{X}$.

We assume that $\boldsymbol{\mu} = \mathbf{0}$, $\mathbf{\Lambda} \in \mathbb{R}^{d\times k}$ with $\mathrm{rk}\mathbf{\Lambda} = d$, and $\mathbb{P}(\mathcal{R} = 0) = 0$, i.e., $\mathbf{X}$ has no point mass at its origin. Due to the stochastic representation of $\mathbf{X}$ given by Definition 2, the following relations hold:

$$\frac{\mathbf{X}}{\|\mathbf{X}\|} = \frac{\mathcal{R}\mathbf{\Lambda}\mathbf{U}}{\|\mathcal{R}\mathbf{\Lambda}\mathbf{U}\|} = \mathrm{sgn}(\mathcal{R})\frac{\mathbf{\Lambda}\mathbf{U}}{\|\mathbf{\Lambda}\mathbf{U}\|} = \mathrm{sgn}(\mathcal{R})\mathbf{V}, \qquad \mathbf{V} := \mathbf{\Lambda}\mathbf{U}. \tag{1}$$

The unit random vector $\mathrm{sgn}(\mathcal{R})\mathbf{V}$ does not depend on the absolute value of $\mathcal{R}$. In particular, it is invariant under the occurrence of extreme values of $\mathcal{R}$. Nonetheless, $\mathrm{sgn}(\mathcal{R})$ cannot be cancelled out and indeed $\mathrm{sgn}(\mathcal{R})$ may depend on $\mathbf{U}$.

Suppose for the moment that $\mathrm{sgn}(\mathcal{R})$ was known for each realization of $\mathcal{R}$, so that we can easily calculate every realization of $\mathbf{V}$, i.e., $\mathbf{V}_t = \mathrm{sgn}(\mathcal{R}_t)\mathbf{X}_t/\|\mathbf{X}_t\|$ for $t = 1, 2, \ldots, n$. The distribution of $\mathbf{V}$ depends on $\mathbf{\Lambda}$ only through $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^\top$ and thus we can estimate $\mathbf{\Sigma}$ by maximum likelihood. Interestingly, for this purpose we need no assumption about the generating variate $\mathcal{R}$. Even the dependence structure of $\mathcal{R}$ and $\mathbf{U}$ is not relevant. Hence, the resulting estimator is distribution-free.

For deriving the corresponding ML-estimator, we have to calculate the density function of $\mathbf{V}$ and search for $\mathbf{T} = \mathbf{\Upsilon}\mathbf{\Upsilon}^\top$ with

$$\mathbf{\Upsilon} = \arg\max_{\mathbf{\Lambda}} \prod_{t=1}^{n} \psi(\mathbf{V}_t\,;\mathbf{\Lambda}),$$

[4] We know only two exceptions, i.e., (i) $\mathcal{R} = \chi_d$ and (ii) $\mathcal{R} = \sqrt{d}$ (Marčenko and Pastur, 1967).

[5] For example, we could choose $\sigma^2(\mathbf{\Sigma}) = (\mathrm{tr}\mathbf{\Sigma})/d$ so that $\mathrm{tr}\mathbf{\Omega} = d$ or $\sigma^2(\mathbf{\Sigma}) = (\det\mathbf{\Sigma})^{1/d}$ so that $\det\mathbf{\Sigma} = 1$ (Frahm, 2009, Paindaveine, 2008).

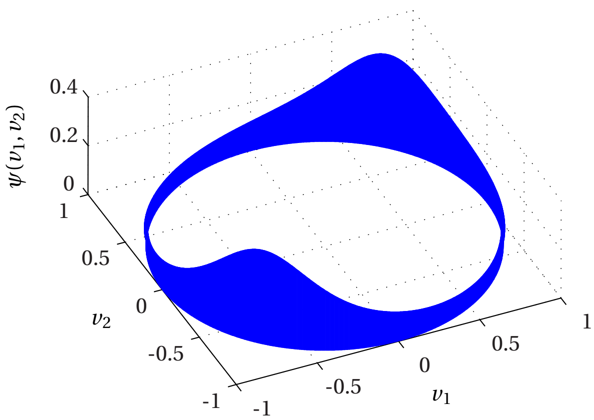

Figure 5: Density of the angular central Gaussian distribution of a 2-dimensional unit random vector generated by $\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{22} \propto 1$ and $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21} \propto 0.7$.

where $\psi(\mathbf{v})$ represents the density of $\mathbf{V}$ at $\mathbf{v} \in \mathscr{S}^{d-1}$. In the following theorem it is assumed without loss of generality that $\det \boldsymbol{\Lambda} = \det \boldsymbol{\Sigma} = 1$.[6]

Theorem 3. *Let* $\boldsymbol{\Lambda}$ *be a* $d \times k$ *matrix with* $\text{rk}\boldsymbol{\Lambda} = d$ *and* $\det \boldsymbol{\Lambda} = 1$. *Further, consider the matrix* $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$. *The density of the unit random vector* $\mathbf{V} = \boldsymbol{\Lambda}\mathbf{U}/\|\boldsymbol{\Lambda}\mathbf{U}\|$ *with respect to the uniform measure on* $\mathscr{S}^{d-1}$ *corresponds to*

$$\psi(\mathbf{v}) = \frac{\Gamma(d/2)}{2\pi^{d/2}} \cdot \sqrt{\mathbf{v}^\top \boldsymbol{\Sigma}^{-1} \mathbf{v}}^{\,-d}$$

for all $\mathbf{v} \in \mathscr{S}^{d-1}$.

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

The distribution given by Theorem 3 is the angular central Gaussian distribution (Tyler, 1987b, Watson, 1983). Due to the Courant-Fischer Theorem, the density function ψ has a local extremum at $\mathbf{w} \in \mathscr{S}^{d-1}$ if and only if $\mathbf{w}$ is an eigenvector of $\boldsymbol{\Sigma}$ and we have that $\psi(\mathbf{w}) \propto \lambda^{d/2}$, where λ is the eigenvalue associated with $\mathbf{w}$.

Figure 5 exemplifies the density function of the angular central Gaussian. Note that ψ is symmetric, i.e., $\psi(\mathbf{v}) = \psi(-\mathbf{v})$, and thus it is not necessary to know the sign of $\mathscr{R}$ for calculating the ML-estimator based on the density function of the angular central Gaussian. This means our previous assumption is superfluous.

Now, consider a sample of generalized elliptically distributed observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. As noted by Tyler (1987b) and Frahm (2004, Sec. 4.2.2),[7] the desired ML-estimator is given by the fixed-point equation

$$\mathbf{T} = \frac{d}{n} \sum_{t=1}^{n} \frac{\mathbf{V}_t \mathbf{V}_t'}{\mathbf{V}_t' \mathbf{T}^{-1} \mathbf{V}_t}. \tag{2}$$

Actually, this corresponds to Tyler's M-estimator (Tyler, 1987a), i.e.,

$$\mathbf{T} = \frac{d}{n} \sum_{t=1}^{n} \frac{\mathbf{X}_t \mathbf{X}_t^\top}{\mathbf{X}_t^\top \mathbf{T}^{-1} \mathbf{X}_t}.$$

[6] This sort of normalization can be considered as canonical (Paindaveine, 2008).

[7] Tyler (1987b) refers only to elliptical distributions, whereas Frahm (2004) observes that the same result applies as well to generalized elliptical distributions.

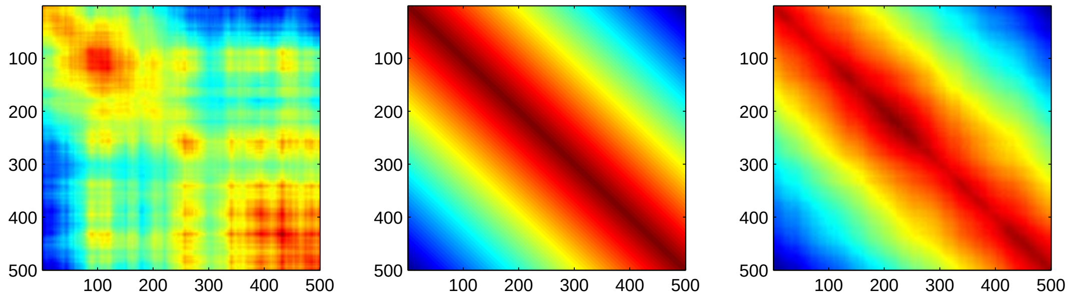

Figure 6: True dispersion matrix (middle), sample covariance matrix (left) and Tyler's M-estimator (right). The estimates are based on a sample of multivariate elliptically t-distributed observations with sample size $n = 1000$, $d = 500$ dimensions, and $\nu = 2$ degrees of freedom.

If the solution of this fixed-point equation exists, it is unique only up to a scaling constant. This means $\mathbf{T}$ must be normalized in any way and in the following we assume that $\mathrm{tr}\mathbf{T} = d$.

The right-hand side of Figure 6 contains a realization of Tyler's M-estimator $\mathbf{T}$, based on a simulated sample of $n = 1000$ multivariate elliptically t-distributed observations with $d = 500$ dimensions and $\nu = 2$ degrees of freedom. The true dispersion matrix $\mathbf{\Sigma}$ is a symmetric Toeplitz matrix, which can be seen in the middle of Figure 6. The corresponding realization of the sample covariance matrix $\mathbf{S}$ is depicted on the left-hand side of Figure 6. Obviously, $\mathbf{T}$ is a robust alternative to $\mathbf{S}$.

At the beginning of this section we claimed that $\mathbf{T}$ is a canonical choice when dealing with financial time-series data. Asset returns typically exhibit nonlinear dependencies both in the cross section and in time. We already showed that the tail-dependence coefficients of the components of an elliptically distributed random vector $\mathbf{X}$ essentially depend on the tail index of its generating variate. Now, suppose that the time series $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ is such that $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$ are serially independent, but in contrast $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n$ have a serial dependence structure. For example, the log-returns could be conditionally heteroscedastic. Our key note is that $\mathbf{T}$ depends only on $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$, i.e., the angular part, but not on $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n$, i.e., the radial part of the data. This can be seen by re-writing Eq. 2:

$$\mathbf{T} = \frac{d}{n} \sum_{t=1}^{n} \frac{\mathbf{\Lambda} \mathbf{U}_t \mathbf{U}_t^\top \mathbf{\Lambda}^\top}{\mathbf{U}_t^\top \mathbf{\Lambda}^\top \mathbf{T}^{-1} \mathbf{\Lambda} \mathbf{U}_t}.$$

Hence, the solution of the fixed-point equation does not depend on $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n$ and so it does not matter how the sequence $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ is driven by the generating variates both in the cross section or in time. This means the asset returns might depend on each other through their generating variates in an *arbitrary* way. Even the finite-sample distribution of $\mathbf{T}$ is not influenced by $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n$. This makes $\mathbf{T}$ highly favorable for heavy-tailed financial time series, irrespective of whether the sample size, n, is large or small or the number of dimensions, d, is high or low.

5. Spectral Properties of Tyler's M-Estimator

In virtue of the aforementioned results we can expect that $\mathbf{T}$ is an appropriate alternative to $\mathbf{S}$ in the context of RMT. This is confirmed by the next theorem.

Theorem 4 (Frahm and Glombek (2012)). *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a sequence of d-dimensional generalized spherically distributed random vectors whose angular parts $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$ are mutually independent. Consider the random matrix $\mathbf{M} = \sqrt{n/d}\,(\mathbf{T} - \mathbf{I}_d)$ with $\operatorname{tr}\mathbf{T} = d$. Then for all $\lambda \in \mathbb{R}$ the spectral distribution $F_{\mathbf{M}}(\lambda)$ converges in probability to $F_{\mathrm{W}}(\lambda) = \int_{-\infty}^{\lambda} f_{\mathrm{W}}(x)\,dx$ with*

$$f_{\mathrm{W}}(x) = \frac{1}{2\pi}\left(4 - x^2\right)^+$$

as $n, d \to \infty$ and $n/d \to \infty$.

Hence, after an appropriate normalization, the spectral distribution of Tyler's M-estimator converges in probability to Wigner's semicircle law as $n, d \to \infty$ but $n/d \to \infty$.[8] Hence, in contrast to Theorem 1, the components of $\mathbf{X}$ are not required to be independent. For other results related to Tyler's M-estimator in the case $n, d \to \infty$ and $n/d \to \infty$ see Dümbgen (1998).

The remaining question is whether the spectral distribution of $\mathbf{T}$ converges to the Marčenko-Pastur law in case $n/d \to q < \infty$. This is formalized by the following conjecture.

Conjecture. *Suppose that one of the following conditions is satisfied:*

1. *The sample elements are i.i.d., have zero mean, finite variance, and a continuous distribution.*

2. *$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ is a sequence of d-dimensional generalized spherically distributed random vectors whose angular parts $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$ are mutually independent.*

Then for all $\lambda \in \mathbb{R}$ the spectral distribution $F_{\mathbf{T}}(\lambda)$ converges in probability to

$$F_{\mathrm{MP}}(\lambda) = \int_{-\infty}^{\lambda} f_{\mathrm{MP}}(x)\,dx$$

with

$$f_{\mathrm{MP}}(x) = \frac{q}{2\pi} \cdot \frac{\sqrt{(\lambda_+ - x)^+ (x - \lambda_-)^+}}{x}, \qquad \lambda_{\pm} = \left(1 \pm \frac{1}{\sqrt{q}}\right)^2,$$

as $n, d \to \infty$ and $n/d \to q$ with $1 \le q < \infty$. In case $0 < q < 1$ the limiting density is a mixture of a point mass at 0 and $f_{\mathrm{MP}}(x)$ with weights $1 - q$ and q, respectively.

Figure 7 demonstrates our conjecture. In contrast to Figure 6 we can see that the spectral distribution of $\mathbf{T}$ converges to the Marčenko-Pastur law both if the data are independent and if they are only uncorrelated.

Our arguments can be understood as follows. Suppose that the first condition of our conjecture is satisfied. Then we have that $\mathrm{E}(\mathbf{X}\mathbf{X}^{\top}) = \sigma^2 \mathbf{I}_d$, where $\sigma^2 > 0$ is the variance of any sample

[8] In contrast to Theorem 1 and Theorem 2, Theorem 4 only states that the spectral distribution converges in probability but not almost surely. More details on the technical difficulties related to the proof of strong consistency can be found at the end of Frahm and Glombek (2012).

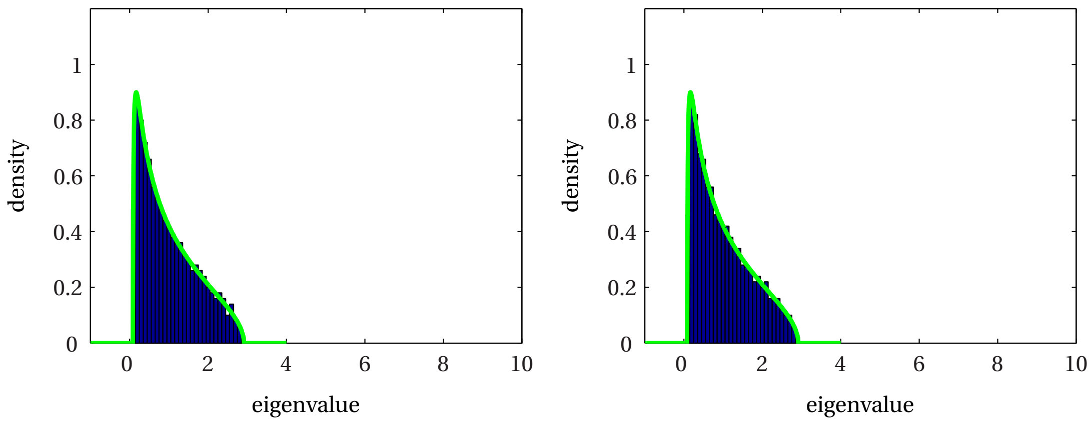

Figure 7: Eigenspectra of $\mathbf{T}$ based on non-elliptically, i.e., independent (left), vs. elliptically, i.e., uncorrelated (right), multivariate t-distributed data ($n = 1000, d = 500$) with 5 degrees of freedom. The green curve represents the density function of the Marčenko-Pastur law for $q = 2$.

element. Suppose that d is large. From the Law of Large Numbers we conclude that $\mathbf{X}^\top\mathbf{X}/d \approx \sigma^2$ and thus $\mathrm{E}\big[(\mathbf{X}\mathbf{X}^\top)/(\mathbf{X}^\top\mathbf{X}/d)\big] \approx \mathbf{I}_d$. For this reason, Tyler's M-estimator

$$\mathbf{T} = \frac{1}{n}\sum_{t=1}^{n}\frac{\mathbf{X}_t\mathbf{X}_t^\top}{\mathbf{X}_t^\top\mathbf{T}^{-1}\mathbf{X}_t/d}$$

converges almost surely to a $d \times d$ matrix that is close to $\mathbf{I}_d$ as $n \to \infty$ (Tyler, 1987a). This means for all $t = 1, 2, \ldots, n$ we expect that $\mathbf{X}_t^\top\mathbf{T}^{-1}\mathbf{X}_t/d \to_{\text{a.s.}} \sigma^2$ as $n, d \to \infty$ and thus $\mathbf{T} \approx \mathbf{S}$ if d and n are large. Here $\mathbf{S}$ denotes the sample covariance matrix of $\mathbf{X}_1/\sigma, \mathbf{X}_2/\sigma, \ldots, \mathbf{X}_n/\sigma$, which satisfies the global i.i.d. assumption of Theorem 2.

By contrast, if the second condition is satisfied, Tyler's M-estimator corresponds to

$$\mathbf{T} = \frac{d}{n}\sum_{t=1}^{n}\frac{\mathbf{U}_t\mathbf{U}_t^\top}{\mathbf{U}_t^\top\mathbf{T}^{-1}\mathbf{U}_t} = \frac{d}{n}\sum_{t=1}^{n}\frac{(\chi_{d,t}\mathbf{U}_t)(\chi_{d,t}\mathbf{U}_t^\top)}{(\chi_{d,t}\mathbf{U}_t^\top)\mathbf{T}^{-1}(\chi_{d,t}\mathbf{U}_t)} = \frac{1}{n}\sum_{t=1}^{n}\frac{\mathbf{Y}_t\mathbf{Y}_t^\top}{\mathbf{Y}_t^\top\mathbf{T}^{-1}\mathbf{Y}_t/d},$$

where $\mathbf{Y}_t = \chi_{d,t}\mathbf{U}_t$ $(t = 1, 2, \ldots, n)$ is a sequence of independent standard normally distributed random vectors. Now, due to the same arguments we obtain $\mathbf{T} \approx \mathbf{S}$, where $\mathbf{S}$ denotes the sample covariance matrix of $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$. It is clear that also the latter sample satisfies the global i.i.d. assumption of Theorem 2. Thus our intuition tells us that $F_{\mathbf{T}}$ and $F_{\mathbf{S}}$ converge to the same limit, i.e., to the Marčenko-Pastur law, if $n, d \to \infty$ with $n/d \to q < \infty$. Nonetheless, the proof of our conjecture is formidable and, to the best of our knowledge, still missing in the literature.

Acknowledgement

We thank David Tyler for his valuable and encouraging comments as well as Karl Mosler for his important suggestions. We also thank Marc Hallin, Frank Hampel, Ingo Klein, Hans-Rudolf Künsch, Alexander McNeil, Davy Paindaveine, and Bernd Rosenow for our fruitful discussions. Special thanks go to Hannu Oja. Without his scientific contributions this paper might have never been accomplished.

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.

J. Bai (2003), 'Inferential theory for factor models for large dimensions', *Econometrica* **71**, pp. 135–171.

J. Bai and S. Ng (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**, pp. 191–221.

J. Bai and S. Ng (2007), 'Determining the number of primitive shocks in factor models', *Journal of Business & Economic Statistics* **25**, pp. 52–60.

Z.D. Bai, H. Liu, and W.-K. Wong (2009), 'On the Markowitz mean-variance analysis of self-financing portfolios', *Risk and Decision Analysis* **1**, pp. 35–42.

Z.D. Bai and J.W. Silverstein (2010), *Spectral Analysis of Large Dimensional Random Matrices*, Springer, 2 Edition.

Z.D. Bai and Y.Q. Yin (1988), 'Convergence to the semicircle law', *Annals of Probability* **16**, pp. 863–875.

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, M. Mézard, and M. Potters (2003), *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, 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.

G. De Luca and G. Rivieccio (2012), 'Multivariate tail dependence coefficients for Archimedean copulae', in: A. Di Ciaccio, M. Coli, and J.M.A. Ibanez, eds., 'Advanced Statistical Methods for the Analysis of Large Data-Sets', Studies in Theoretical and Applied Statistics, pp. 287–296, Springer.

P. Debashis and A. Aue (2014), 'Random matrix theory in statistics: A review', *Journal of Statistical Planning and Inference* **150**, pp. 1–29.

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 heteroscedasticity 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.

H. Ferreira and M. Ferreira (2012), 'On extremal dependence of block vectors', *Kybernetika* **48**, pp. 988–1006.

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

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

G. Frahm (2009), 'Asymptotic distributions of robust shape matrices and scales', *Journal of Multivariate Analysis* **100**, pp. 1329–1337.

G. Frahm and K. Glombek (2012), 'Semicircle law of Tyler's M-estimator for scatter', *Statistics and Probability Letters* **82**, pp. 959–964.

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

K. Glombek (2012), *High-Dimensionality in Statistics and Portfolio Optimization*, Ph.D. thesis, University of Cologne.

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.

N.E. Karoui (2010), 'High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: risk underestimation', *Annals of Statistics* **38**, pp. 3487–3566.

N.E. Karoui (2013), 'On the realized risk of high-dimensional Markowitz portfolios', *SIAM Journal on Financial Mathematics* **4**, pp. 737–783.

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.

S. Kring, S.T. Rachev, M. Höchstötter, et al. (2009), 'Multi-tail generalized elliptical distributions for asset returns', *The Econometrics Journal* **12**, pp. 272–291.

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

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.

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.

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.

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, 2nd Edition.

D. Paindaveine (2008), 'A canonical definition of shape', *Statistics and Probability Letters* **78**, pp. 2240–2247.

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.

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.

G.S. Watson (1983), *Statistics on Spheres*, Wiley.

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.