Stockholms universitet

Large Dimensional Analysis of Covariance Matrix Estimators in the Framework of General Minimum Variance Portfolio (GMVP) based on Random Matrix Theory (RMT)

Mete Indere

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2023:21
Matematisk statistik
September 2023

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2023:21**
http://www.math.su.se

Large Dimensional Analysis of Covariance Matrix Estimators in the Framework of General Minimum Variance Portfolio (GMVP) based on Random Matrix Theory (RMT)

Mete Indere*

September 2023

Abstract

The General Minimum Variance Portfolio (GMVP) has the smallest variance of all the portfolios. The weight of this portfolio depends on the inverse of the population covariance matrix, which is an unknown object in practice and must be replaced by an estimator. Several estimators of the GMV portfolio weights exist in the literature. The preliminary aim of the thesis is their comparison with respect to the out-of-sample performance and their asymptotic behaviors based on Random Matrix Theory (RMT). For numerical simulations synthetic and real data are used. Asymptotic behaviors are analyzed when the number of assets p and the sample size n are going together to infinity at the same convergence rate p/n, which is called in the literature as double-limit regime or high-dimensional asymptotic. The different estimators we are interested in are based on the Sample Covariance Matrix (SCM) and Tyler's robust M-estimator in non-regularized and regularized (shrinkage) forms. The following four approaches are considered: 1-Frahm and Memmel (2010) [2] They treat the case of the linear shrinkage estimator based on the sample covariance estimator and a non-random target under the assumption of serially independent and identically normally distributed asset returns. 2-Bodnar et al. (2018) [1] They improve the estimator of Frahm and Memmel and suggest shrinking the sample estimator for the portfolio weights directly and not the whole sample covariance matrix, which is dominant but not necessarily optimal. A new estimator for the GMV portfolio based on Random Matrix Theory is derived, which is optimal and distribution-free. 3-Rubio et al. (2012) [3] They regularize the SCM estimator, where the shrinkage target is a nonrandom positive-definite matrix. They invert the sum of the SCM estimator and the target matrix and in GMVP implementation find the portfolio weights. To find the minimum realized variance the shrinkage intensity is optimized. 4-Yang et al. (2015) [4] They are in principle following the same procedure as in Rubio et al. using instead for the SCM estimator the Tyler´s M-estimator.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: inderemete26@hotmail.com. Supervisor: Taras Bodnar.

Acknowledgements

I would like to acknowledge Prof. Taras Bodnar my supervisor for his help having let me use his R-codes in his paper [2] and his patience during the time writing this thesis. I thank as well Buket Coşkun, Research Assistant at the Department of Electronics at Yeditepe University in İstanbul for her help in type writing this thesis using LaTeX program.

Contents

List of Figures

List of Tables

1. Introduction

The classical mean-variance portfolio optimization is based on the frame work studied by Warkowitz [6]. In the analysis of this portfolio one is confronted with the estimator errors of the mean and covariance matrix, which depend on the asset returns. It is shown by Merton [7], that the estimates of the covariance matrix are more accurate than the expected returns. For this reason one has turned to the performance study of the global minimum variance portfolio (GMVP), which depends only on the covariance matrix estimator with the intention to reduce portfolio risk.

The Global Minimum Variance Portfolio (GMVP) is based on the solution of a quadratic optimization problem. Using the portfolio variance (risk)

$$\underset{\mathbf{w}}{min}\, \sigma^2(\mathbf{w}) = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \tag{1.1}$$

subject to the linear constraint

$$\mathbf{w}^T \mathbf{1} p = 1 \tag{1.2}$$

where $\mathbf{w} = (\mathbf{w}_1, ..., \mathbf{w}_p)^T$ are the portfolio weights and $\boldsymbol{\Sigma}$ is the covariance matrix of the asset returns. And $\mathbf{1}_p$ is the p-dimensional vector of ones. The solution is known as

$$\mathbf{w}_{GMVP} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}_p}{\mathbf{1}_p^T \boldsymbol{\Sigma}^{-1}\mathbf{1}_p} \tag{1.3}$$

The corresponding portfolio risk is then

$$\sigma^2(\mathbf{w}_{GMVP}) = \frac{1}{\mathbf{1}_p^T \boldsymbol{\Sigma}^{-1}\mathbf{1}_p} \tag{1.4}$$

which represents the minimum risk bound.

But not knowing the population covariance matrix, instead in Equation (1.3) a plug-in estimator is used, which we denote by $\hat{\Sigma}(\alpha)$ and α stands for the shrinkage parameter. Based on this the GMVP weight is given by

$$\hat{\mathbf{w}}_{GMVP}(\alpha) = \frac{\hat{\boldsymbol{\Sigma}}^{-1}(\alpha)\mathbf{1}_p}{\mathbf{1}_p^T \hat{\boldsymbol{\Sigma}}^{-1}(\alpha)\mathbf{1}_p} \tag{1.5}$$

The performance of this kind of an estimator is measured by the achieved out-of-sample portfolio variance:

$$\sigma^2(\hat{\mathbf{w}}_{GMVP}(\alpha)) = \frac{\mathbf{1}_p^T \hat{\boldsymbol{\Sigma}}^{-1}(\alpha)\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}(\alpha)\mathbf{1}_p}{(\mathbf{1}_p^T \hat{\boldsymbol{\Sigma}}^{-1}(\alpha)\mathbf{1}_p)^2} \tag{1.6}$$

If we choose in Equation (1.6) the Sample Covariance Matrix (SCM) $\mathbf{S}_n$ as an estimator of $\mathbf{\Sigma}$, then we get the out of sample variance of the traditional estimator:

$$\sigma_s^2 = \frac{\mathbf{1}_p^T \mathbf{S}_n^{-1} \mathbf{\Sigma} \mathbf{S}_n^{-1} \mathbf{1}_p}{(\mathbf{1}_p^T \mathbf{S}_n^{-1} \mathbf{1}_p)^2} \tag{1.7}$$

The performance of it has been analyzed in [8]. It is known that this realized variance underestimate the risk and hereby leads to optimistic investment decisions. Classical asymptotic theory requires the dimension p of the portfolio to be fixed and the number of samples n goes to infinity n » p.

However in the double limit regime it is of more practical and empirical importance to analyze the case when $n \simeq p$, which is done based on Random Matrix Theory such as in [9] and [10].

In Equation (1.6) the covariance matrix estimator is the only input parameter in the GMVP framework. We consider the returns of the assets over the consecutive investment periods to be independent and identically distributed. We use the data matrix (p x n)

$$\mathbf{Y}_n = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n) \tag{1.8}$$

consisting of n vectors of p row asset returns. Then the Sample Covariance Matrix (SCM) is obtained as

$$\mathbf{S}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{\Lambda}_n \mathbf{Y}_n^T \tag{1.9}$$

where

$$\mathbf{\Lambda}_n = (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \tag{1.10}$$

is the centering matrix.

As assumed in [2], [5] and [4] we use the following stochastic decomposition for the data matrix where $\boldsymbol{\mu}$ is the mean vector of asset returns.

$$\mathbf{Y}_n = \boldsymbol{\mu} \mathbf{1}_n^T + \mathbf{\Sigma}^{1/2} \mathbf{X}_n \tag{1.11}$$

$\mathbf{X}_n$ is a pxn random matrix, of which the entries are assumed to be independent and identically distributed (i.i.d) having mean zero and variance one. $\mathbf{\Sigma}^{1/2}$ stands for the square root of positive definite population matrix $\mathbf{\Sigma}$, $\mathbf{\Sigma} = \mathbf{\Sigma}^{1/2} (\mathbf{\Sigma}^{1/2})^T$. Then the stochastic decomposition implies for SCM in Equation 1.9:

$$\mathbf{S}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{\Lambda}_n \mathbf{Y}_n^T \overset{d}{=} \frac{1}{n} \mathbf{\Sigma}^{1/2} \mathbf{X}_n \mathbf{\Lambda}_n \mathbf{X}_n^T \mathbf{\Sigma}^{1/2} \tag{1.12}$$

having used as in [5]

$$(\boldsymbol{\mu} \mathbf{1}_n^T + \mathbf{\Sigma}_n^{1/2} \mathbf{X}_n) \mathbf{\Lambda}_n = \mathbf{\Sigma}^{1/2} \mathbf{X}_n \mathbf{\Lambda}_n \tag{1.13}$$

The stochastic decomposition in matrix form in Equation (1.11) could be as well written in stochastic decomposition vector form as

$$\mathbf{y}_i \overset{d}{=} \boldsymbol{\mu} + \sqrt{T_i} \mathbf{\Sigma}^{1/2} \mathbf{x}_i \quad i = 1, 2, ..., n \tag{1.14}$$

This model includes the class of elliptical distributions and as special cases the multivariate normal and the multivariate student-t distributions. T_i is a real, positive random variable and independent of $\mathbf{x}_i$'s. For a multivariate student-t distribution $\sqrt{T_i} = \sqrt{(d-2)/d}\sqrt{d/\chi_d^2}$. χ_d^2 is a Chi-square random variable with d degree of freedom.

For a symmetric random vector $\mathbf{y} \in \mathbb{R}^p$ our interest lies in the centered elliptical symmetric distribution, which is characterized by [10; 11]

$$\mu_p(\mathbf{y}) = C(g_p)det(\mathbf{\Sigma_p})^{-1/2}g_p(\mathbf{y}^T\mathbf{\Sigma}_p\mathbf{y}) \tag{1.15}$$

where $\mathbf{\Sigma}_p$ is as positive definite matrix in $\mathbb{R}^{pxp}$, $g_p : [0,\infty) \to [0,\infty)$ is fixed function called the density generator fulfilling

$$\int_0^\infty g_p(x)x^{p-1} < \infty \tag{1.16}$$

and $C(g_p > 0)$ is a normalizing constant that only depends on g_p, such that $\mu_p(y)$ integrates to one. We denote this case by $y \sim \mathcal{E}_p(\mathbf{0},\mathbf{\Sigma}_p,g_p)$. When $\mathbf{\Sigma}_p$ is identity matrix, the distribution is isotropic and μ_p is then called the spherically symmetric distribution [10].

$$\mu_p(\mathbf{y}) = C(g_p)g_p(\|y\|^2) \tag{1.17}$$

and we denote this case by $y \sim \mathcal{E}_p(\mathbf{0},\mathbf{I}_p,g_p)$.

In the normal distribution case $y \sim N_p(\mathbf{0},\mathbf{\Sigma}_p)$ the p.d.f. is [11]

$$\mu_p(\mathbf{y}) = \pi^{-p/2}det(\mathbf{\Sigma}_p)^{-1/2}exp(-\frac{1}{2}\mathbf{y}^T\mathbf{\Sigma}_p^{-1}\mathbf{y}) \tag{1.18}$$

and the density generator is $g(t) = exp(-t/2)$. In the t-distribution case with v degrees of freedom $y \sim t_{p,v}(\mathbf{0},\mathbf{\Sigma}_p)$ the p.d.f. is [11]

$$\mu_{p,v} \propto det(\mathbf{\Sigma}_p)^{-1/2}(1+\mathbf{y}^T\mathbf{\Sigma}_p^{-1}\mathbf{y}/v)^{-(p+v)/2}, v > 0 \tag{1.19}$$

and the density generator is $g_v(t) = (1+t/v)^{-(p+v)/2}$.

The aim of the thesis is the study of different estimators in the GMVP framework. Their performances and asymptotic behaviors are studied by means of random matrix theory. For this purpose we use the double limit regime, where p the number of variables and n the number of samples both together $p,n \to \infty$ with $c_p = p/n \to c \in (0,\infty)$. We consider both the classical case $n >> p$ and the case $n \simeq p$ when n and p are of similar order. In considering the performances we use Monte/Carlo simulations with normally and student-t distributed samples. Further we consider daily real data simulations for stock returns from the Hong Kong's Hang Sen Index (HSI).

In sections 2, 3, 4, 5 we introduce the study of our different regularized SCM (RSCM) estimators. In section 6 we study the Tyler's M-Estimator and the regularized Tyler's M-Estimator. In section 7 we compare the performances and asymptotic behaviors between the Traditional and Tyler's M-Estimator using synthetic data simulations.

In section 8 we study the different performances and asymptotic behaviors of our RSCM estimators and the regularized Tyler's M-Estimator, using synthetic data simulations.
In section 9 we proceed in analyzing the annualized standard deviations of all our non-regularized and regularized estimator using real market data simulations.
In section 10 we conclude by analyzing our results.
In Appendix we are interested in spectral norms of different matrix estimators in the double limit regime, which are known in the literature.

Notations

The vectors are column vectors. The dependence of vectors and matrices on their dimension p are sometimes added sometimes neglected, due to clarity. Because p is dependent on the sample size n sometimes n is emphasized. $\mathbf{I}_p$ denotes the identity matrix. $\mathbf{1}_p$ denotes the p-dimensional vector with all the entries to be one. tr(.) stands for the matrix trace operator. $|a-b| \xrightarrow{a.s} 0$ stands for almost sure convergence, and the asymptotic equivalence of the quantities a and b. $|.|$ denotes the absolute value. $\|.\|$ stands for the Euclidean norm of a vector and the spectral norm of a matrix. The ordered eigenvalues of a symmetric matrix A of size pxp are denoted by $\lambda_1 \leq ... \leq \lambda_p(A)$. $\delta\lambda_i(\mathbf{\Sigma}_p)$ denotes the Dirac measure of $\lambda_i(\mathbf{\Sigma}_p)$. The spectral norm is defined as the square root of the maximum eigenvalue of $\mathbf{A}^H\mathbf{A}$ as $(\lambda_{max}(\mathbf{A}^H\mathbf{A}))^{1/2}$, where $\mathbf{A}^H$ is the conjugate transpose of the square matrix $\mathbf{A}$.

2. Oracle Estimator [2]

The shrinkage estimator for the weights of the GMVP is defined as a linear combination of the plug-in sample estimator in Equation (1.5) and a target portfolio $\mathbf{b}_n$ with $\mathbf{b}_n^T \mathbf{1} = 1$.

$$\hat{\mathbf{w}}_n(\alpha_n) = \alpha_n \frac{S_n^{-1}\mathbf{1}}{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}} + (1-\alpha_n)\mathbf{b}_n \tag{2.1}$$

The out-of-sample variance is

$$\sigma^2(\hat{\mathbf{w}}_n(\alpha_n)) = \hat{\mathbf{w}}_n^T(\alpha_n)\mathbf{\Sigma}_n \hat{\mathbf{w}}_n(\alpha_n) \tag{2.2}$$

which has to be optimized with respect to the shrinkage parameter α_n. Rewriting Equation (2.2) by using Equation (2.1) result in

$$\sigma^2(\hat{\mathbf{w}}_n(\alpha_n)) = \alpha_n^2 \frac{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{S}_n^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1})^2} + 2\alpha_n(1-\alpha_n) \frac{\mathbf{1}^T \mathbf{S}_n^{-1}}{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}} \mathbf{\Sigma}_n \mathbf{b}_n + (1-\alpha_n)^2 \mathbf{b}_n^T \mathbf{\Sigma}_n \mathbf{b}_n \tag{2.3}$$

where $\sigma_{\mathbf{b}_n}^2 = \mathbf{b}_n^T \mathbf{\Sigma}_n \mathbf{b}_n$ is the variance of the target portfolio.

In [2] the optimal shrinkage intensity α_n^*, which minimizes $\sigma^2(\hat{\mathbf{w}}_n(\alpha_n))$ is proven to be

$$\alpha_n^* = \frac{\mathbf{b}_n^T \mathbf{\Sigma}_n \mathbf{b}_n - \frac{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{b}_n}{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}}}{\frac{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{S}_n^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1})^2} - 2 \frac{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{b}_n}{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}} + \mathbf{b}_n^T \mathbf{\Sigma}_n \mathbf{b}_n} \tag{2.4}$$

In the double-asymptotic regime the asymptotic behavior of the terms in Equation (2.3) and (2.4) are of interest. In [2], using Theorem 1 in [3] they have proven the following asymptotics (see Appendix in [2]).

$$\left| \mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1} - (1-c)^{-1} \mathbf{1}^T \mathbf{\Sigma}_n^{-1} \mathbf{1} \right| \overset{a.s}{\rightarrow} 0 \tag{2.5}$$

$$\left| \mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{S}_n^{-1} \mathbf{1} - (1-c)^{-3} \mathbf{1}^T \mathbf{\Sigma}_n^{-1} \mathbf{1} \right| \overset{a.s}{\rightarrow} 0 \tag{2.6}$$

$$\left| \mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{b}_n - (1-c)^{-1} \right| \overset{a.s}{\rightarrow} 0 \tag{2.7}$$

Using Equations (2.5), (2.6), and (2.7) leads to the following asymptotics, which are used in finding the asymptotic behavior of α_n^* and $\sigma^2(\hat{\mathbf{w}}_n(\alpha_n^*))$.

$$\left| \frac{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{S}_n^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1})^2} - (1-c)^{-1} \sigma_{GMVP}^2 \right| \overset{a.s}{\rightarrow} 0 \tag{2.8}$$

$$\left| \frac{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{b}_n}{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}} - \sigma_{GMVP}^2 \right| \stackrel{a.s}{\to} 0 \tag{2.9}$$

In Equation (2.8) the term

$$\sigma_s^2 = \frac{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{\Sigma}_n \mathbf{S}_n^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1})^2} \tag{2.10}$$

is the out of sample variance of the traditional estimator. Using σ_s^2 in Equation (2.8) and the relative loss of the traditional estimator $R_s = \frac{\sigma_s^2 - \sigma_{GMVP}^2}{\sigma_{GMVP}^2}$ we get as in Corollary 2.1 of [2].

$$\left| R_s - \frac{c}{1-c} \right| \stackrel{a.s.}{\to} 0 \tag{2.11}$$

In [2] the assumption $0 < M_l \leq \sigma_{GMVP}^2 \leq \sigma_{\mathbf{b}_n}^2 \leq M_u < \infty$ is made, such that the variance of the target portfolio converges to a limit σ_b^2. The relative loss of the target portfolio

$$R_{\mathbf{b}_n} = (\sigma_{\mathbf{b}_n}^2 - \sigma_{GMVP}^2)/\sigma_{GMVP}^2 = \mathbf{b}_n^T \mathbf{\Sigma}_n \mathbf{b}_n . \mathbf{1}^T \mathbf{\Sigma}_n^{-1} \mathbf{1} \tag{2.12}$$

is then assumed to converge to the *limit* $R_{\mathbf{b}} = \frac{\sigma_{\mathbf{b}}^2 - \sigma_{GMVP}^2}{\sigma_{GMVP}^2}$ such as

$$|R_{\mathbf{b}_n} - R_{\mathbf{b}}| \stackrel{a.s.}{\to} 0 \tag{2.13}$$

Using Equations (2.8), (2.9) and (2.11) in Equation (2.4) leads as in Theorem (2.1) of [2] to the asymptotic behavior of the optimal shrinkage intensity α_n^*

$$|\alpha_n^* - \alpha^*| \stackrel{a.s.}{\to} 0 \tag{2.14}$$

where

$$\alpha^* = \frac{(1-c)R_{\mathbf{b}}}{c + (1-c)R_{\mathbf{b}}} \tag{2.15}$$

is a nonrandom quantity. Our interest lies in the optimized relative loss

$$R(\alpha_n^*) = \frac{\hat{\mathbf{w}}_n^T(\alpha_n^*) \mathbf{\Sigma}_n \hat{\mathbf{w}}_n(\alpha_n^*) - \sigma_{GMVP}^2}{\sigma_{GMVP}^2} \tag{2.16}$$

and its asymptotic behavior.

Replacing in Equation (2.3) α_n by α_n^*, and using Equations (2.8), (2.9), (2.12), (2.13) we get for $R(\alpha_n^*)$ the asymptotic equivalence as in the Corollary 2.1 of [2].

$$\left| R(\alpha_n^*) - \left\{ (\alpha^*)^2 \frac{c}{1-c} + (1-\alpha^*)^2 R_{\mathbf{b}} \right\} \right| \stackrel{a.s.}{\to} 0 \tag{2.17}$$

Further by using α^* in Equation (2.15)

$$\left| R(\alpha_n^*) - \frac{(1-c)cR_{\mathbf{b}}^2 + c^2 R_{\mathbf{b}}}{(c + (1-c)R_{\mathbf{b}})^2} \right| \stackrel{a.s.}{\to} 0 \tag{2.18}$$

3. Bona Fide Estimator [2]

The Oracle estimator cannot be constructed in practice, since it depends on the unknown Σ_n. The bona fide estimator for the weights of the GMVP is expressed as

$$\hat{\mathbf{w}}_n(\hat{\alpha}_n^*) = \hat{\alpha}_n^* \frac{\mathbf{S}_n^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{S}_n^{-1}\mathbf{1}} + (1-\hat{\alpha}_n^*)\mathbf{b}_n \tag{3.1}$$

where

$$\hat{\alpha}_n^* = \frac{(1-p/n)\hat{R}_{\mathbf{b}_n}}{p/n+(1-p/n)\hat{R}_{\mathbf{b}_n}} \tag{3.2}$$

and $\hat{R}_{\mathbf{b}_n}$ is a consistent estimator of $R_{\mathbf{b}_n}$ (2.12) with

$$\hat{R}_{\mathbf{b}_n} = (1-p/n).\mathbf{b}_n^T\mathbf{S}_n\mathbf{b}_n.\mathbf{1}^T\mathbf{S}_n^{-1}\mathbf{1} - 1 \tag{3.3}$$

which is to be understood in the following asymptotic equivalence sense

$$\left|\hat{R}_{\mathbf{b}_n} - R_{\mathbf{b}_n}\right| \overset{a.s.}{\to} 0 \tag{3.4}$$

Equations (3.3) and (3.4) make up Theorem 2.3 in [2]. From Equations (3.2) and (3.4) and using Equations (2.13), (2.15) we get

$$|\hat{\alpha}_n^* - \alpha^*| \overset{a.s.}{\to} 0 \tag{3.5}$$

For this reason $\hat{\mathbf{w}}_n(\hat{\alpha}_n^*)$ presents an optimal shrinkage estimator, as it is stated in [2].

The relative loss of the Bona Fide Estimator is:

$$R(\hat{\alpha}_n^*) = \frac{\hat{\mathbf{w}}_n^T(\hat{\alpha}_n^*)\mathbf{\Sigma}_n\hat{\mathbf{w}}_n(\hat{\alpha}_n^*) - \sigma_{GMVP}^2}{\sigma_{GMVP}^2} \tag{3.6}$$

We can conclude then

$$\left|R(\hat{\alpha}_n^*) - \left\{(\alpha^*)^2\frac{c}{1-c} + (1-\alpha^*)^2 R_{\mathbf{b}}\right\}\right| \overset{a.s.}{\to} 0 \tag{3.7}$$

or written as in Equation (2.18) for the Oracle Estimator gives

$$\left|R(\hat{\alpha}_n^*) - \frac{(1-c)cR_{\mathbf{b}}^2 + c^2R_{\mathbf{b}}}{(c+(1-c)R_{\mathbf{b}})^2}\right| \overset{a.s.}{\to} 0 \tag{3.8}$$

as we have

$$|\alpha_n^* - \hat{\alpha}_n^*| \overset{a.s.}{\to} 0 \tag{3.9}$$

We conclude that the relative loss of the Oracle Estimator in Equation (2.16) and the relative loss of the Bona Fide Estimator in Equation (3.6) has to be understood in the sense of asymptotic equivalence

$$|R(\alpha_n^*) - R(\hat{\alpha}_n^*)| \overset{a.s.}{\rightarrow} 0 \tag{3.10}$$

4. Dominating Estimator [1]

It is given by

$$\hat{\mathbf{w}}_{n;FM} = \hat{\alpha}_{n;FM} \frac{\mathbf{S}_n^{-1}}{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}} + (1 - \hat{\alpha}_{n;FM}) \mathbf{b}_n \tag{4.1}$$

with

$$\hat{\alpha}_{n;FM} = 1 - \frac{p-3}{n-p+2} . \frac{1}{\hat{R}_{FM}} \tag{4.2}$$

where

$$\hat{R}_{FM} = \frac{\mathbf{b}_n^T \mathbf{S}_n \mathbf{b}_n - \sigma_{\mathbf{S}_n}^2}{\sigma_{\mathbf{S}_n}{}^2} \tag{4.3}$$

with

$$\sigma_{\mathbf{S}_n}{}^2 = \frac{1}{\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}} \tag{4.4}$$

which leads to

$$\hat{\alpha}_{n;FM} = 1 - \frac{p-3}{n-p+2} (\mathbf{b}_n^T \mathbf{S}_n \mathbf{b}_n . \mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1} - 1)^{-1} \tag{4.5}$$

The relative loss is given as

$$R_{FM}(\hat{\alpha}_{n;FM}) = \frac{\hat{\mathbf{w}}_{n;FM}^T \mathbf{\Sigma}_n \hat{\mathbf{w}}_{n;FM} - \sigma_{GMVP}^2}{\sigma_{GMVP}^2} \tag{4.6}$$

We are interested in the asymptotic behavior and following the same procedure as in the case of the Oracle Estimator of [2], we only need to be concerned to find the asymptotic of the $\hat{\alpha}_{n;FM}$ in Equation (4.5). The factor $\frac{p-3}{n-p+2}$ tends to $\frac{c}{1-c}$ for $n, p \to \infty$ and $n/p \to c$. The asymptotic behavior of $\mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1}$ is as earlier in Equation (2.5).

$$\left| \mathbf{1}^T \mathbf{S}_n^{-1} \mathbf{1} - (1-c)^{-1} \mathbf{1}^T \mathbf{\Sigma}_n^{-1} \mathbf{1} \right| \xrightarrow{a.s.} 0 \tag{2.5}$$

The term $\mathbf{b}_n^T \mathbf{S}_n \mathbf{b}_n$ has the asymptotic behavior as proven in [2] [eq.A.38]

$$\left| \mathbf{b}_n^T \mathbf{S}_n \mathbf{b}_n - \mathbf{b}^T \mathbf{\Sigma}_n \mathbf{b} \right| \xrightarrow{a.s.} 0 \tag{4.7}$$

Using the asymptotic behaviors in (2.5) and (4.7) in (4.5) leads to the asymptotic non random limit of $\hat{\alpha}_{n;FM}$

$$\alpha_{FM} = 1 - \frac{c}{1-c} \left[\mathbf{b}^T \mathbf{\Sigma}_n \mathbf{b} . (1-c)^{-1} . \mathbf{1}^T \mathbf{\Sigma}_n^{-1} \mathbf{1} - 1 \right]^{-1} \tag{4.7}$$

Rewriting Equation (4.7) gives

$$\alpha_{FM} = 1 - \frac{c}{1-c}\left[(1-c)^{-1}(R_{\mathbf{b}}+1) - 1\right]^{-1} \tag{4.8}$$

and rearranging the terms

$$\alpha_{FM} = \frac{R_{\mathbf{b}}}{c + R_{\mathbf{b}}} \tag{4.9}$$

As in the Oracle case in Equation (2.17) we know the limit of relative loss $R_{FM}(\hat{\alpha}_n; FM)$ in (4.6) will have the form

$$R_{FM} = \alpha_{FM}^2 \frac{c}{1-c} + (1-\alpha_{FM})^2 R_{\mathbf{b}} \tag{4.10}$$

And using α_{FM} in Equation (4.9) gives

$$R_{FM} = \frac{(c/1-c)R_{\mathbf{b}}^2 + c^2 R_{\mathbf{b}}}{(c+R_{\mathbf{b}})^2} \tag{4.11}$$

Last we can write

$$\left| R_{FM}(\hat{\alpha}_n; FM) - \frac{(c/1-c)R_{\mathbf{b}}^2 + c^2 R_{\mathbf{b}}}{(c+R_{\mathbf{b}})^2} \right| \xrightarrow{a.s.} 0 \tag{4.12}$$

5. Regularized Sample Covariance Matrix Estimator [5]

In [5] the Sample Covariance Matrix (SCM) is regularized such as[1] for $\rho \in [max\left\{0, 1 - \frac{n}{p}\right\}, 1]$

$$\hat{\boldsymbol{\Sigma}}_{RMP}(\rho) = (1-\rho)\mathbf{S}_n + \rho\mathbf{I} \tag{5.1}$$

The GMVP implementation based on this estimator is given by

$$\hat{\mathbf{w}}_{RMP}(\rho) = \frac{\hat{\boldsymbol{\Sigma}}^{-1}_{RMP}(\rho)\mathbf{1}}{\mathbf{1}^T\hat{\boldsymbol{\Sigma}}^{-1}_{RMP}(\rho)\mathbf{1}} \tag{5.2}$$

The corresponding out-of-sample variance is

$$\sigma^2(\hat{\mathbf{w}}_{RMP}(\rho)) = \frac{\mathbf{1}^T\hat{\boldsymbol{\Sigma}}^{-1}_{RMP}(\rho)\boldsymbol{\Sigma}_n\hat{\boldsymbol{\Sigma}}^{-1}_{RMP}(\rho)\mathbf{1}}{(\mathbf{1}^T\hat{\boldsymbol{\Sigma}}^{-1}_{RMP}(\rho)\mathbf{1})^2} \tag{5.3}$$

then

$$\sigma^2(\hat{\mathbf{w}}_{RMP}(\rho)) = \frac{\mathbf{1}^T((1-\rho)\mathbf{S}_n + \rho\mathbf{I})^{-1}\boldsymbol{\Sigma}_n((1-\rho)\mathbf{S}_n + \rho\mathbf{I})^{-1}\mathbf{1}}{(\mathbf{1}^T((1-\rho)\mathbf{S}_n + \rho\mathbf{I})^{-1}\mathbf{1})^2} \tag{5.4}$$

In Equation (5.4) the denominator is given by observable data, whereas in the numerator the unknown population covariance matrix $\boldsymbol{\Sigma}_n$ appears.+

They point at, that there is a need for a consistent estimator of

$$b = \mathbf{1}^T\hat{\boldsymbol{\Sigma}}^{-1}_{RMP}(\rho)\boldsymbol{\Sigma}_n\hat{\boldsymbol{\Sigma}}^{-1}_{RMP}(\rho)\mathbf{1} \tag{5.5}$$

in the double-limit regime $p, n \to \infty, p/n = c_p \to c \in (0, \infty)$ which is proven in their following theorem.

Theorem 2 (*Generalized Consistent Estimator*):

$$\left|b - \hat{b}\right| \overset{a.s.}{\to} 0 \tag{5.6}$$

where $\hat{b}$ is given as:

$$\hat{b} = \hat{\alpha}\hat{\beta} \tag{5.7}$$

[1] In[5] we have choosen the identity matrix as the target matrix

In Equation (5.7) $\hat{\alpha}$ and $\hat{\beta}$ are defined as

$$\hat{\alpha} = \frac{1}{\frac{1}{n}tr[(1-\rho)\mathbf{\Lambda}_n(\mathbf{I}_n + \hat{\delta}(1-\rho)\mathbf{\Lambda}_n)^{-2}]} \tag{5.8}$$

and

$$\hat{\beta} = (1-\rho)\mathbf{1}^T((1-\rho)\mathbf{S}_n + \rho\mathbf{I})^{-1}\mathbf{S}_n((1-\rho)\mathbf{S}_n + \rho\mathbf{I})^{-1}\mathbf{1} \tag{5.9}$$

The proof is given in Appendix iii [5].
Then we conclude the following Corollary of Theorem 2.
Corollary 1.

$$\left|\sigma^2(\hat{\mathbf{w}}_{RMP}(\rho)) - \hat{\sigma}^2(\hat{\mathbf{w}}_{RMP}(\rho))\right| \overset{a.s.}{\to} 0 \tag{5.10}$$

where

$$\hat{\sigma}^2(\hat{\mathbf{w}}_{RMP}(\rho)) = \hat{\alpha} \times \frac{\hat{\beta}}{(\mathbf{1}^T((1-\rho)\mathbf{S}_n + \rho\mathbf{I})^{-1}\mathbf{1})^2} \tag{5.11}$$

In the optimization of the out-of-sample variance we base our finding on the optimized shrinkage parameter by a numerical grid search optimization:

$$\rho^* = \underset{\rho\in\left[max\left\{0,1-C_p^{-1}\right\},1\right]}{arg\,min} \left\{\hat{\sigma}^2(\hat{\mathbf{w}}_{RMP}(\rho))\right\} \tag{5.12}$$

Introducing the following relative losses as defined

$$R_{RMP}(\hat{\mathbf{w}}_{RMP}(\rho)) = \frac{\sigma^2(\hat{\mathbf{w}}_{RMP}(\rho)) - \sigma^2_{GMVP}}{\sigma^2_{GMVP}} \tag{5.13}$$

and

$$\hat{R}_{RMP}(\hat{\mathbf{w}}_{RMP}(\rho^*)) = \frac{\hat{\sigma}^2(\hat{\mathbf{w}}_{RMP}(\rho^*)) - \sigma^2_{GMVP}}{\sigma^2_{GMVP}} \tag{5.14}$$

we get for the asymptotic equivalence of the estimator for the relative loss

$$\left|R_{RMP}(\hat{\mathbf{w}}_{RMP}(\rho)) - \hat{R}_{RMP}(\hat{\mathbf{w}}_{RMP}(\rho^*))\right| \overset{a.s.}{\to} 0 \tag{5.15}$$

To find the estimator $\hat{\sigma}^2(\hat{\mathbf{w}}_{RMP}(\rho))$ in Equation (5.11) we still are in need to find the generalized estimator $\hat{\delta}$ which appears in $\hat{\alpha}$ Equation (5.8). The parameter δ to be estimated is introduced in [5], when they define the following matrices

$$\begin{gathered}\mathbf{E} = \mathbf{\Sigma}(\tilde{\delta}\mathbf{\Sigma} + \rho\mathbf{I})^{-1} \\ \tilde{\mathbf{E}} = (1-\rho)\mathbf{\Lambda}_n(\mathbf{I}_n + \delta(1-\rho)\mathbf{\Lambda}_n)^{-1}\end{gathered} \tag{5.16}$$

with

$$\begin{gathered}\tilde{\delta} = \tfrac{1}{n}tr[\tilde{\mathbf{E}}] \\ \delta = \tfrac{1}{n}tr[\mathbf{E}]\end{gathered} \tag{5.17}$$

Then the following system of equations have to be solved which are unique and positive introduced in [12] in Proposition 1:

$$
\begin{aligned}
\tilde{\delta} &= \tfrac{1}{n} tr[(1-\rho)\mathbf{\Lambda}_n(\mathbf{I}_n + \delta(1-\rho)\mathbf{\Lambda}_n)^{-1}] \\
\delta &= \tfrac{1}{n} tr[\mathbf{\Sigma}(\tilde{\delta}\mathbf{\Sigma} + \rho\mathbf{I}_p)^{-1}]
\end{aligned}
\tag{5.18}
$$

To find the generalized consistent estimator of δ [5] they use their Proposition 1.
Proposition 1:

$$
\hat{a} := tr[(1-\rho)\mathbf{S}_n((1-\rho)\mathbf{S}_n + \rho\mathbf{I})^{-1}] = \delta tr[(1-\rho)\mathbf{\Lambda}_n(\mathbf{I} + \delta(1-\rho)\mathbf{\Lambda}_n)^{-1}] \tag{5.19}
$$

We are using on the right hand side of Equation 5.19 the matrix inversion lemma

$$
(\mathbf{I} + \delta(1-\rho)\mathbf{\Lambda}_n)^{-1} = \mathbf{I} - \frac{\delta(1-\rho)}{1+\delta(1-\rho)(n-1)}\mathbf{\Lambda}_n \tag{5.20}
$$

and get for the generalized consistent estimator of δ:

$$
\hat{\delta} \cong \frac{1}{1-\rho}.\frac{1}{n-1}.\hat{a} \tag{5.21}
$$

Then $\hat{\delta}$ being given, $\hat{b}$ in Equation (5.7) turns out to be a strongly consistent estimator of b.

6. Tyler's M-Estimator and the Regularized Tyler's M-Estimator [4]

Tyler introduced the following M-estimator [13; 14], which is the unique solution to the fixed-point equation

$$\hat{\mathbf{C}} = \frac{1}{n} \textstyle\sum_{i=1}^{n} \frac{\tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^T}{\frac{1}{p} \tilde{\mathbf{y}}_i \hat{\mathbf{C}}^{-1} \tilde{\mathbf{y}}_i} \\ where\ \tilde{\mathbf{y}}_i = \mathbf{y}_i - \tfrac{1}{n} \textstyle\sum_{t=1}^{n} \mathbf{y}_t \tag{6.1}$$

Concerning the regularized Tyler's M-estimator introduced in [14; 15], we have the unique solution to the fixed-point equation for $\rho \epsilon (max \left\{0, 1 - \frac{n}{p}\right\}, 1]$:

$$\hat{\mathbf{C}}_{reg}(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^T}{\frac{1}{p}\tilde{\mathbf{y}}_i^T \hat{\mathbf{C}}_{reg}^{-1}(\rho)\tilde{\mathbf{y}}_i} + \rho \mathbf{I} \tag{6.2}$$

which is based on the Tyler's M-estimator [13] and the shrinkage estimator of Ledoit-Wolf [16].

The GMVP portfolio weights in the case of Equation (6.1) are given by

$$\hat{\mathbf{w}}_{YMC} = \frac{\hat{\mathbf{C}}^{-1}\mathbf{1}}{\mathbf{1}^T \hat{\mathbf{C}}^{-1}\mathbf{1}} \tag{6.3}$$

and in the case of Equation (6.2) by

$$\hat{\mathbf{w}}_{YMC}(\rho) = \frac{\hat{\mathbf{C}}_{reg}^{-1}(\rho)\mathbf{1}}{\mathbf{1}^T \hat{\mathbf{C}}_{reg}^{-1}(\rho)\mathbf{1}} \tag{6.4}$$

To gain insight into their performances, the out-of-sample variances have to be considered

$$\sigma^2(\hat{\mathbf{w}}_{YMC}) = \frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{\Sigma}_n \hat{\mathbf{C}}^{-1}\mathbf{1}}{(\mathbf{1}^T \hat{\mathbf{C}}^{-1}\mathbf{1})^2} \tag{6.5}$$

and

$$\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho)) = \frac{\mathbf{1}^T \hat{\mathbf{C}}_{reg}^{-1}(\rho) \mathbf{\Sigma}_n \hat{\mathbf{C}}_{reg}^{-1}(\rho)\mathbf{1}}{(\mathbf{1}^T \hat{\mathbf{C}}_{reg}^{-1}(\rho)\mathbf{1})^2} \tag{6.6}$$

In Equation 6.6 the goal is to find the optimized value of the shrinkage parameter ρ, such that $\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho))$ is minimized. Because in Equation 6.6 the unobservable population matrix $\mathbf{\Sigma}_n$ is appearing this optimization cannot be solved. In this case as it is done in [5] based on the double asymptotic regime they derive in their Theorem 1 first a deterministic asymptotic equivalence of the realized portfolio risk and thereafter they find a consistent asymptotic estimator for $\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho))$ in their Theorem 2. To avoid their lengthy results in their Theorem 1, we mention only their Assumption 1, which they use in their asymptotic analysis and are of value to be used in their Theorem 2.

Assumption 1:

(a) As $p,n \to \infty, p/n = c_p \to c \in (0,\infty)$

(b) Denoting $0 < \lambda_1 \leq ... \leq \lambda_p$ the ordered eigenvalues of $\mathbf{\Sigma}_p$ as $p,n \to \infty, \nu_p \triangleq \frac{1}{p}\sum_{i=1}^{p} \boldsymbol{\delta}_{\lambda_i}$ satisfies $\nu_p \to \nu$ weakly with $\nu \neq \delta_0$ almost everywhere, where δ_χ is the Dirac measure at χ. In addition $lim\,sup_p\,\lambda_p < \infty$.

In their asymptotic analysis they define for $\rho \in (max(0, 1-c^{-1}), 1]$ the unique positive solution to the equation in γ

$$1 = \int \frac{t}{\gamma\rho + (1-\rho)t}\nu(dt) \tag{6.7}$$

and

$$\kappa = \int t\nu(dt) \tag{6.8}$$

The asymptotic consistent estimator of the scaled out-of-sample variance $\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho))$ by $1/\kappa$ is denoted as $\hat{\sigma}^2_{sc}(\rho)$ and given in their following theorem.
Theorem 2: For $\varepsilon \in (0, min\{1, c^{-1}\})$, define $R_\varepsilon = [\varepsilon + max\{0, 1-c^{-1}\}, 1]$
Then as $p,n \to \infty$,

$$\sup_{\rho \in R_\varepsilon} \left| \hat{\sigma}^2_{sc}(\rho) - \frac{1}{\kappa}\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho)) \right| \overset{a.s.}{\to} 0 \tag{6.9}$$

where $\hat{\sigma}^2_{sc}(\rho)$ is defined in Equation 6.10.
Note: κ is independent of ρ, therefore same ρ minimizes $\sigma^2(\hat{w}_{YMC}(\rho))$ and $\sigma^2(\hat{w}_{YMC}(\rho))/\kappa$.

$$\hat{\sigma}^2_{sc}(\rho) = \frac{\hat{\gamma}_{sc}}{(1-\rho)-(1-\rho)^2 c_p} . \frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1}_{reg}(\rho)(\hat{\mathbf{C}}_{reg}(\rho) - \rho\mathbf{I})\hat{\mathbf{C}}^{-1}_{reg}(\rho)\mathbf{1}}{(\mathbf{1}^T \hat{\mathbf{C}}^{-1}_{reg}(\rho)\mathbf{1})^2} \tag{6.10}$$

Further of use is their Lemma 1, of which the proof is given in Appendix C in [4], when they use their Theorem 1.
Lemma1: Under the settings of Theorem 1, as $p,n \to \infty$

$$\sup_{\rho \in R_\varepsilon} |\hat{\gamma}_{sc} - \gamma/\kappa| \overset{a.s.}{\to} 0 \tag{6.11}$$

where

$$\hat{\gamma}_{sc} = \frac{1}{1-(1-\rho)c_p}\frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{\mathbf{y}}_i^T \hat{\mathbf{C}}^{-1}_{reg}(\rho)\tilde{\mathbf{y}}_i}{\|\tilde{\mathbf{y}}_i\|^2} \tag{6.12}$$

Then in the Appendix E [4], they derive the following corollary of Theorem 2.

Corollary 1:

When ρ° stands for the minimizer of $\sigma^2_{SC}(\rho)$ and ρ^* for the minimizer of the unobservable $\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho))$over R_ε respectively, then there exist an asymptotically equivalence of the form, as $p, n \to \infty$,.

$$\left|\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho^\circ)) - \sigma^2(\hat{\mathbf{w}}_{YMC}(\rho^*))\right| \xrightarrow{a.s.} 0 \tag{6.13}$$

They conclude, that the GMVP optimization in the case of the regularized Tyler's estimator can be handled by the minimization of $\sigma^2_{SC}(\rho)$.

As stated in their ***Algorithm 1***, the optimization parameter ρ° which minimizes $\hat{\sigma}^2_{SC}(\rho)$ has to be found by a numerical optimizing grid search of the form:

$$\rho^\circ = \underset{\rho \in [\varepsilon + max\left\{0, 1 - c_p^{-1}\right\}, 1]}{arg\,min} \left\{\hat{\sigma}^2_{SC}(\rho)\right\} \tag{6.14}$$

Thereafter the unique solution to the fixed-point equation is sought:

$$\hat{\mathbf{C}}_{reg}(\rho^\circ) = (1 - \rho^\circ)\frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^T}{\frac{1}{p}\tilde{\mathbf{y}}_i^T \hat{\mathbf{C}}_{reg}^{-1}(\rho^\circ)\tilde{\mathbf{y}}_i} + \rho^\circ \mathbf{I} \tag{6.15}$$

And the portfolio weights constructed

$$\hat{\mathbf{w}}_{YMC}(\rho^\circ) = \frac{\hat{\mathbf{C}}_{reg}^{-1}(\rho^\circ)\mathbf{1}}{\mathbf{1}^T \hat{\mathbf{C}}_{reg}^{-1}(\rho^\circ)\mathbf{1}} \tag{6.16}$$

Further we are in need of the out-of-sample variance

$$\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho^\circ)) = \frac{\mathbf{1}^T \hat{\mathbf{C}}_{reg}^{-1}(\rho^\circ)\mathbf{\Sigma}_n \hat{\mathbf{C}}_{reg}^{-1}(\rho^\circ)\mathbf{1}}{(\mathbf{1}^T \hat{\mathbf{C}}_{reg}^{-1}(\rho^\circ)\mathbf{1})^2} \tag{6.17}$$

As we are interested in the out-of-sample relative loss then from Corollary 1, Equation (6.13) we receive

$$\left|R_{reg}(\rho^\circ) - R_{reg}(\rho^*)\right| \xrightarrow{a.s.} 0 \tag{6.18}$$

where

$$R_{reg}(\rho^\circ) = \frac{\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho^\circ)) - \sigma^2_{GMVP}}{\sigma^2_{GMVP}} \tag{6.19}$$

and

$$R_{reg}(\rho^*) = \frac{\sigma^2(\hat{\mathbf{w}}_{YMC}(\rho^*)) - \sigma^2_{GMVP}}{\sigma^2_{GMVP}} \tag{6.20}$$

7. The Traditional and Tyler's M-Estimator in Synthetic Data Simulations

Here we describe our results for the performance and asymptotic behaviors of the Traditional and Tyler's M-Estimator based on their relative losses and variances under the setting of the normally and student-t distributed data simulations.

In [10] it is proven in Theorem 3.2 that in normally distributed data the scaled Tyler's M-Estimator $p\hat{\mathbf{C}}$ converges to the sample covariance estimator in operator norm (see Appendix, Equation 1). Further in Theorem 3.5[10], their analysis is based on sampling from elliptical distribution and they receive the property for Tyler's M-Estimator as in Equation (3) in operator norm (see Appendix).

In Corollary 3.7 [10] it is proven under the sampling from spherically symmetric distribution, that the empirical spectral density (e.s.d.) of $p\hat{\mathbf{C}}$ converges to the Marchenko-Pastur distribution [17]. As shown for $\mathbf{B} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$ in [17] the e.s.d converges to the same limiting distribution. For this reason $p\hat{\mathbf{C}}$ and $\mathbf{B}$ are asymptotically equivalent.

In Corollary 3.9 [10] for elliptical distribution, the e.s.d of $tr(\mathbf{\Sigma}_p)\hat{\mathbf{C}}$ converges to a limiting density of which the Stieltjes transform has the form as in Equation (5) (See Appendix). But this is the same Stieltjes transform as in Theorem 1, Silverstein (1995) [18] for the limiting distribution of the sample covariance matrix $\mathbf{B} = \frac{1}{n}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^T\mathbf{\Sigma}^{\frac{1}{2}}$. We conclude that the SCM $\mathbf{B}$ and $tr(\mathbf{\Sigma}_p)\hat{\mathbf{C}}$ are asymptotically equivalent.

From these theoretical results we conjecture that there exist asymptotic equivalence between the Traditional and Tyler's M-Estimator for the different convergence rates in the case for normally and student-t distributed data.

Our results, which are based on GMVP framework confirm these theoretical findings in the case for relative loss as in Figure 7.1 and Figure 7.2 and for variance data in Table 7.1 and Table 7.2.

Concerning the performances at the convergence rate $c = 0.9$ when $p \simeq n$ the Tyler's M-Estimator slightly out-performs the Traditional estimator for the normal and student-t distribution in Figure 7.1 and 7.2.

For $c = 0.5$ there are almost no differences in performances for the Tyler's M-Estimator and the Traditional estimator as in 7.1 for the normally distributed and in Figure 7.2 for the student-t distributed data. For $c = 0.1$ in Figure 7.1 and 7.2 the Traditional Estimator outperforms the Tyler's M-Estimator. This reflects the large n case, despite even p is changing and not of constant values.

From these considerations concerning the performances of the Tyler's M-Estimator and the Traditional Estimator the convergence rate has a direct impact on the results. We point at that the Dominating estimator outperforms the Traditional and Tyler's

M-Estimator as in Figure 7.1 and Figure 7.2 for the relative loss and in Table 7.1 and Table 7.2 for the variance measure.

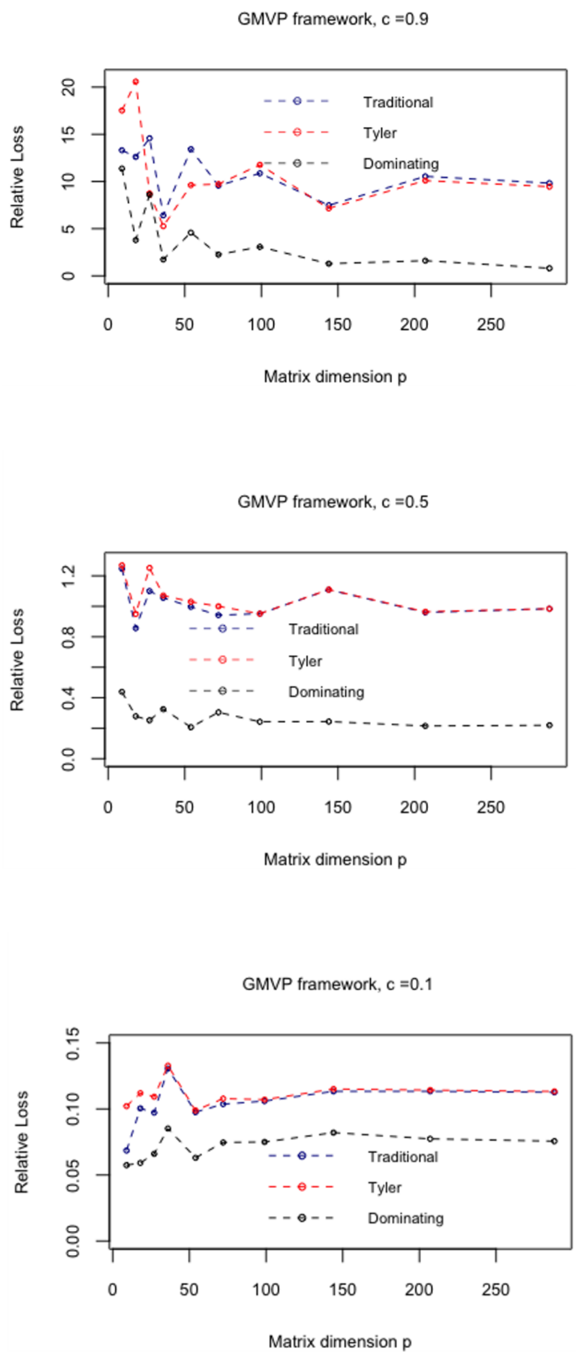

Figure 7.1: Simulation results for normally distributed data in the case of c = (0.9,0.5,0.1)

Table 7.1: Simulation results for normally distributed data in the case of c = (0.9,0.5,0.1)

p	Traditional	Tyler	Dominating	Traditional-Tyler
9	13.5546647	17.5478020	11.71873853	-3.9931373
18	6.1136104	9.6941739	2.15411127	-3.5805635
27	4.2790146	2.6726175	2.63487532	1.6063971
36	1.3799321	1.1663630	0.50940721	0.2135691
54	1.6524064	1.2164466	0.63966062	0.4359598
72	0.8275216	0.8404607	0.25519016	-0.0129391
99	0.6550977	0.7045928	0.22490137	-0.0494951
144	0.3279908	0.3153696	0.08868296	0.0126212
207	0.3388997	0.3257346	0.07658475	0.0131651
288	0.2131356	0.2058218	0.03553428	0.0073138

p	Tyler	Traditional	Dominating	Tyler-Traditional
9	2.14877188	2.12858757	1.36285457	0.02018431
18	0.87552168	0.83380404	0.57425159	0.04171764
27	0.61844475	0.57680239	0.34397835	0.04164236
36	0.38550044	0.38270985	0.24665432	0.00279059
54	0.23261050	0.22887736	0.13833496	0.00373314
72	0.15673794	0.15218561	0.10223421	0.00455233
99	0.10772424	0.10773146	0.06858013	0.00000722
144	0.08161804	0.08157693	0.04811871	0.00004111
207	0.05766335	0.05754882	0.03566776	0.00011453
288	0.03905404	0.03904288	0.02400289	0.00001116

p	Traditional	Tyler	Dominating	Tyler-Traditional
9	1.01224868	1.04405250	1.00171379	0.03180382
18	0.49421939	0.49942659	0.47567704	0.0052072
27	0.30135697	0.30467058	0.29278346	0.00331361
36	0.21048328	0.21089176	0.20203564	0.00040848
54	0.12580194	0.12594448	0.12183838	0.00014254
72	0.08650063	0.08682755	0.08422620	0.00032692
99	0.06105162	0.06111273	0.05934233	0.00006111
144	0.04307651	0.04314448	0.04186647	0.00006797
207	0.03268785	0.03271517	0.03163168	0.00002732
288	0.02189992	0.02191370	0.02117194	0.00001378

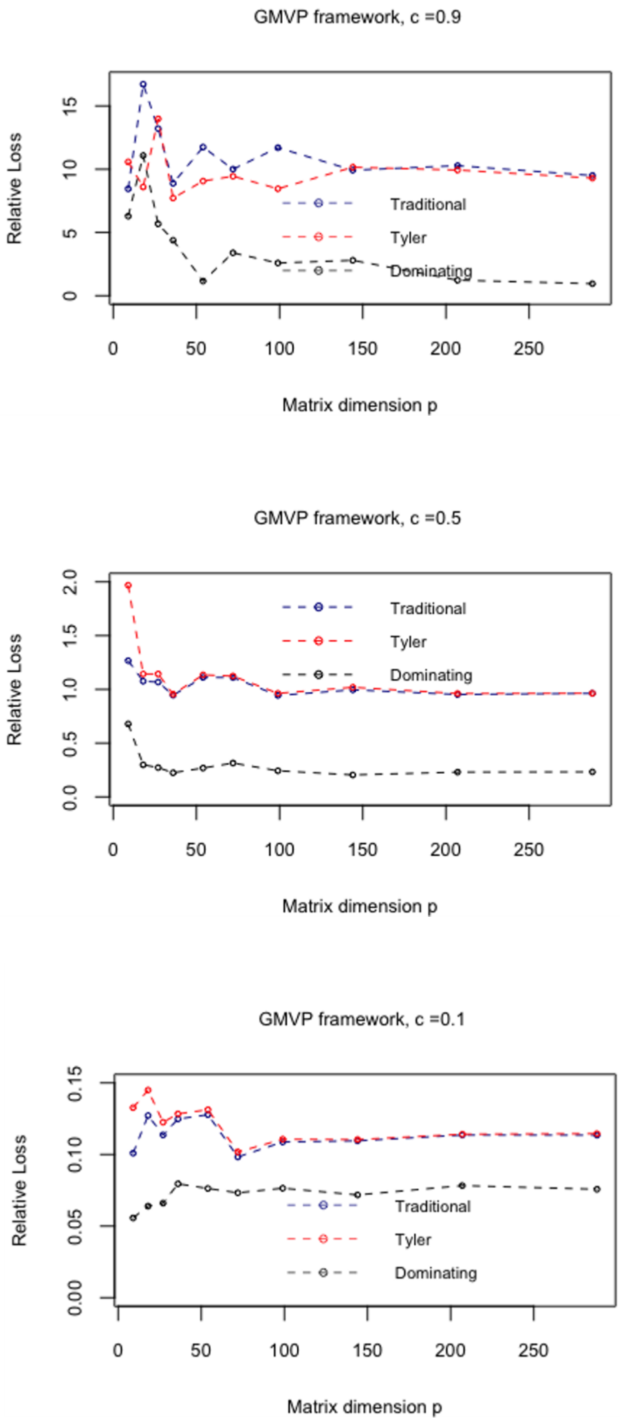

Figure 7.2: Simulation results for student-t distributed data with 5 degrees of freedom in the case of c = (0.9,0.5,0.1)

Table 7.2: Simulation results for student-t distributed data with 5 degrees of freedom in the case of c = (0.9,0.5,0.1)

p	Traditional	Tyler	Dominating	Traditional-Tyler
9	8.9534710	10.9685926	6.90775176	-2.0151216
18	7.9588469	4.3110919	5.42465915	3.6477550
27	3.9040361	4.1135582	1.83510074	-0.2095221
36	1.8405034	1.6244985	1.00244878	0.2160049
54	1.4608603	1.1533409	0.24759407	0.3075194
72	0.8623609	0.8186027	0.34415092	0.0437582
99	0.7011578	0.5220861	0.19795376	0.1790717
144	0.4226358	0.4316890	0.14700240	-0.0090532
207	0.3311891	0.3210649	0.06542999	0.0101242
288	0.2067173	0.2027040	0.03841953	0.0040133

p	Tyler	Traditional	Dominating	Tyler-Traditional
9	2.81193855	2.14717757	1.59022084	0.66476098
18	0.96241907	0.93241086	0.58324533	0.03000821
27	0.58870819	0.56788558	0.34950602	0.02082261
36	0.36396269	0.36219831	0.22794760	0.00176438
54	0.24467459.	0.24203511	0.14537727	0.00263948
72	0.16643199	0.16535967.	0.10302973	0.00107232
99	0.10835379	0.10727003	0.06862263	0.00108376
144	0.07811830	0.07720750	0.04658911	0.00091080
207	0.05758098	0.05731993	0.03612357	0.00026105
288	0.03864983	0.03863484	0.02425704	0.00001499

p	Traditional	Tyler	Dominating	Traditional-Tyler
9	8.9534710	10.9685926	6.90775176	-2.0151216
18	7.9588469	4.3110919	5.42465915	3.6477550
27	3.9040361	4.1135582.	1.83510074	-0.2095221
36	1.8405034	1.6244985	1.00244878	0.2160049
54	1.4608603	1.1533409	0.24759407	0.3075194
72	0.8623609	0.8186027.	0.34415092	0.0437582
99	0.7011578	0.5220861	0.19795376	0.1790717
144	0.4226358	0.4316890	0.14700240	-0.0090532
207	0.3311891	0.3210649	0.06542999	0.0101242
288	0.2067173	0.2027040	0.03841953	0.0040133

8. The Regularized Tyler's M-Estimator and the Regularized SCM Estimators in Synthetic Data Simulations

We describe our results for the performances and asymptotic behaviors of these estimators analyzing their relative loss and variance under the setting of the normally and student-t distributions using synthetic data simulations. In [9] Couillet and McKay have shown in Theorem 1 (see Appendix) that the Tyler's M-Estimator $\hat{\mathbf{C}}_p(\rho)$ in Equation (6) converges to the following regularized SCM estimator in Equation (8):

$$\hat{\mathbf{S}}_p(\rho) = \frac{1}{\gamma(\rho)} . \frac{1-\rho}{1-(1-\rho)c} \mathbf{S}_p + \rho \mathbf{I}_p$$

in spectral norm almost surely as in Equation (7).

In [9] it is mentioned that the statistical characteristics of $\hat{\mathbf{S}}_p(\rho)$ are well studied by Marchenko-Pastur(1967) and are simpler then the ones for $\hat{\mathbf{C}}_p(\rho)$. Further in [9] in Corollary 1 it is proven that the empirical spectral distribution of the regularized Tyler's M-Estimator $\hat{\mathbf{C}}_p(\rho)$ converge to the limiting spectral distribution of $\hat{\mathbf{S}}_p(\rho)$. From these theoretical results we conjecture that there should exist asymptotic equivalence between the regularized SCM estimators we have considered and the regularized Tyler's M-Estimator for the different convergence rates. Our results which are based on GMVP framework confirm these theoretical findings in the case for relative loss as shown in Figure 8.1 to Figure 8.2 and in the case for variance data in Table 8.1 to 8.2. Concerning the performances for the convergence rate $c = 0.9$ when $p \simeq n$ the regularized Tyler's M-Estimator performs best in Figure 8.1 for normally and in Figure 8.2 for student-t distributions. Whereas the Bona Fide estimator performs worst, even valid for the variances in Table 8.1 and Table 8.2.

For $c = 0.5$ there are slight differences in performances between the estimators as seen in Figure 8.1 and Table 8.1 for normally and in Figure 8.2 and Table 8.2 for student-t distributions.

In case $c = 0.1$, when n is large and p is small $n >> p$, even when p is not constant the Tyler's regularized M-Estimator performs worst, whereas the Bona Fide estimator performs best as in Figure 8.1 and Figure 8.2 and as well in Table 8.1 and Table 8.2. Concerning the performances the converge rate plays a distinct role, as we would have expected based on theoretical considerations.

We kept the graphs for the relative loss and the variance data for the Dominating estimator separate from the optimized shrinkage estimators.

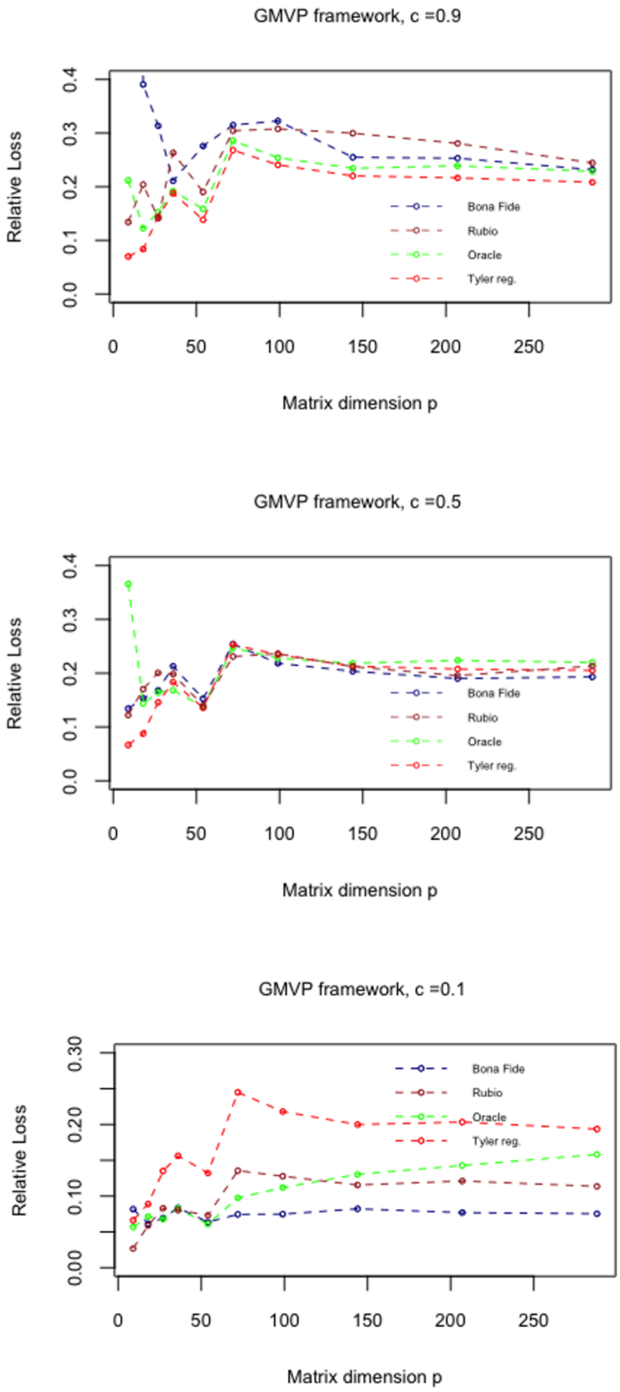

Figure 8.1: Simulation results for normally distributed data in the case of c = (0.9,0.5,0.1)

Table 8.1: Simulation results for normally distributed data in the case of c = (0.9,0.5,0.1)

p	Bona Fide	Rubio	Oracle	Tyler reg	Bona Fide-Tyler reg
9	4.08178878	1.07430912	1.14799448	1.01345225	3.06833653
18	0.62458461	0.54071681	0.50410291	0.48675588.	0.13782873
27	0.36079012	0.31342402	0.31662911	0.31417463.	0.04661549
36	0.22545512	0.23515514	0.22189471	0.22114407	0.00431105
54	0.14623494	0.13642877	0.13279122	0.13047164.	0.01576330
72	0.10305926	0.10222143	0.10077011	0.09939092.	0.00366834
99	0.07300678	0.07219243	0.06922180	0.06849047	0.00451631
144	0.04855209	0.05028696	0.04776063	0.04720797	0.00134412
207	0.03679009	0.03760040	0.03637268	0.03571429	0.00107580
288	0.02424431	0.02449399	0.02418399	0.02378314	0.00046117

p	Bona Fide	Rubio	Oracle	Tyler reg.	Bona Fida-Tylerreg
9	1.07429862	1.06315473	1.29390640	1.01037024	0.06392838
18	0.51789605	0.52563841	0.51357199	0.48851473	0.02938132
27	0.32078548	0.32977316	0.31973103	0.31468502	0.00610046
36	0.22579827	0.22311287	0.21751726	0.22032850	0.00546977
54	0.13209345	0.13059950	0.13041298	0.13016334	0.00193011
72	0.09828684	0.09648343	0.09772825	0.09819627	0.00009057
99	0.06726127	0.06824591	0.06779729	0.06814057	-0.00087930
144	0.04656598	0.04689275	0.04713495	0.04693028	-0.00036430
207	0.03494074	0.03510185	0.03593032	0.03545593	-0.00051519
288	0.02348580	0.02388381	0.02401391	0.02372171	-0.00023591

p	Tyler reg.	Oracle	Rubio	Bona Fide	Tyler reg.-Bona Fide
9	1.01004401	1.00093009	0.97269210	1.02485001	-0.01480600
18	0.48907843	0.48116093	0.47563528.	0.47674048	0.01233795
27	0.31173244	0.29325679	0.29744335	0.29371714	0.01801530
36	0.21526571	0.20153445	0.20114816	0.20174446	0.01352125
54	0.12976997	0.12157329	0.12300741	0.12191604	0.00785393
72	0.09757281	0.08602127	0.08899255	0.08421351	0.01335930
99	0.06724043	0.06137966	0.06225437	0.05933165	0.00790878
144	0.04642515	0.04373157	0.04315827	0.04186993	0.00455522
207	0.03533096	0.03355145	0.03291386.	0.03161827	0.00371269
288	0.02349356	0.02279750	0.02191858.	0.02116584	0.00232772

For $c = 0.9$ when Table 7.1 is compared with Table 8.1 the Dominating estimator performs worst relative to the other estimators for normally distributed data. Furthermore there is no asymptotic equivalence between the Dominating estimator and the other estimators. The same findings apply for the student-t distribution when Table 7.2 is compared with Table 8.2. If we would have incorporated the results for the relative loss of the Dominating estimator in Figure 8.1 and Figure 8.2 then due to scaling the graphs of the other estimators would have been compressed and became indistinguishable.

For the case $c = 0.1$ in Table 8.1 the Bona Fide estimator is performing best and as seen in Table 7.1 the Dominating estimator performs equivalently well in normal distribution and for student-t distribution, when we compare Table 7.2 with Table 8.2. Then the optimization in Bona Fide estimator doesn't cause any advantage compared with the Dominating estimator. There is no asymptotic equivalence in the sense of double-limit regime.

For $c = 0.5$ comparing the variance of the Dominating estimator in Table 7.1 with the other estimators in Table 8.1 for normal distribution and Table 7.2 with Table 8.2 for student-t distribution the Dominating estimator performs worst.

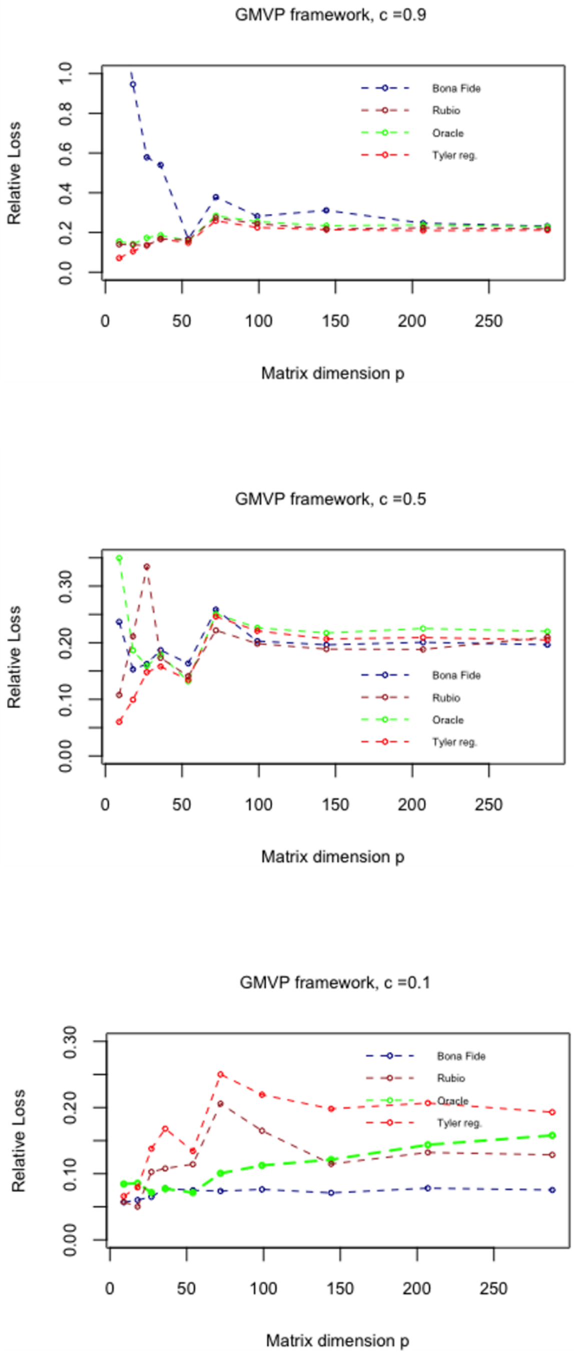

Figure 8.2: Simulation results for student-t distributed data

Table 8.2: Simulation results for student-t distributed data with 5 degrees of freedom in the case of c = (0.9,0.5,0.1)

p	Bona Fide	Rubio	Oracle	Tyler reg	Bona Fide-Tyler reg
9	2.30567350	1.08010099	1.09434730	1.01496034	1.29071316
18	0.87412240	0.51117505	0.51294215	0.49646301	0.37765939
27	0.43361364	0.31164999	0.32213546	0.31222905	0.12138459
36	0.28674840	0.21737124	0.22083098	0.21736361	0.06938479
54	0.13367616	0.13324179	0.13278974	0.13161257	0.00206359
72	0.10800923	0.09986017	0.10065469	0.09869612	0.00931311
99	0.07076372	0.06866098.	0.06925810	0.06763322	0.00313050
144	0.05073993	0.04710060	0.04775329	0.04698520	0.00375473
207	0.03660035	0.03592598	0.03636112.	0.03551563	0.00108472
288	0.02425151	0.02398028	0.02418357.	0.02385389	0.00039762

p	Bona Fide	Rubio	Oracle	Tyler reg.	Bona Fide-Tyler reg.
9	1.17166364	1.04937829	1.27847948.	1.00441340.	0.16725024
18	0.51782748	0.54390687	0.53294061	0.49374568	0.02408180
27	0.31927393	0.36644980	0.31841933	0.31529992	0.00397401
36	0.22096800	0.21842125	0.21938211	0.21557137	0.00539663
54	0.13334240	0.13075494	0.12973983	0.13001776	0.00332464
72	0.09862209	0.09576301	0.09803059	0.09767171	0.00095038
99	0.06640040	0.06614937	0.06769920	0.06739667	-0.00099627
144	0.04628242	0.04599063	0.04709199	0.04668642	-0.00040400
207	0.03525223.	0.03488433	0.03597014	0.03550068	-0.00024845
288	0.02355080	0.02381625	0.02401543	0.02371727	-0.00016647

p	Bona Fide	Oracle	Tyler reg	Rubio	Tyler reg-Bona Fide
9	1.00113020	1.02735528	1.00977428	1.00164650	0.00864408
18	0.47616419	0.48756823	0.48476516	0.47166446	0.00860097
27	0.29248692	0.29431832	0.31248925	0.30284483	0.02000233
36	0.20061766	0.20056577	0.21745393	0.20626829	0.01683627
54	0.12319845	0.12279641	0.13000567	0.12771050	0.00680722
72	0.08415056	0.08626404.	0.09798220	0.09450744	0.01383164
99	0.05941188	0.06141362	0.06731534	0.06430885	0.00790346
144	0.04143847	0.04338370	0.04635478	0.04313180	0.00491631
207	0.03165103	0.03357844	0.03543206	0.03323798	0.00378103
288	0.02116680	0.02279259	0.02348341	0.02221303	0.00231661

9. Real Market Data Simulations

In this section we study the performances of the different GMVP estimators using the real market data. We base the analysis on 45 constituents from the Hang Seng Index (HSI). The dividend adjusted closing prices are down loaded from the Yahoo! Finance (http://finance.yahoo.com). Than converted into logarithmic returns over $T = 739$ working days from Aug. 1, 2019 to July 31, 2022. As a performance measure we use out-of-sample variance in terms of rolling-window approach as described by De Miguel et al [19]. The portfolio weights are estimated for an estimation window length $n < T$. The covariance and weights at a certain day t are estimated by using the previous n days as an estimation window. This procedure is followed until the day index reaches the last day T of trading. The out-of-sample variance has the following form [2].

$$\hat{\sigma}^2_{out} = \frac{1}{T-1-n} \sum_{t=n}^{T-1} (\hat{\mathbf{w}}_t^T \mathbf{r}_{t+1} - \hat{\boldsymbol{\mu}}_t)^2 \tag{9.1}$$

with

$$\hat{\boldsymbol{\mu}}_t = \frac{1}{T-n} \sum_{t=n}^{T-1} \hat{\mathbf{w}}_t^T \mathbf{r}_{t+1} \tag{9.2}$$

We are concerned in annualized realized risk, which are obtained when the standard deviations are multiplied by $\sqrt{250}$. In Figure 9.1 we compare the performances of the regularized Tyler's and Rubio et al. estimator.

It is seen that the regularized Tyler's estimator outperforms Rubio et al. estimator. When the estimation window length is n=230 the lowest risk is achieved for both. For estimation windows greater than n=230 their performances start to degrade increasingly. Similar degrading happened as well in [4]. They claim this presumably could be due to the lack of data stationarity, when such long durations are taken into considerations. In Figure 9.2 we are considering the Bona Fide, Dominating, Traditional and Tyler's estimator. And the Rubio and Tyler's regularized estimator. The estimation window lengths in Figure 9.1 and Figure 9.2 are ranging from $n = 50$ to $n = 450$ by steps of 20. Than, because we are considering a fixed $p = 45$ our rate of convergence $c_p = p/n$ lies in the range 0.9 to 0.1. This is of importance as we used in the simulation studies the set (0.9, 0.5, 0.1) and therefore should allow us to compare the performances of different estimators in synthetic data simulations and real market data simulations. We see from Figure 9.2 for all window lengths from $n = 50$ to $n = 450$ the Bona Fide estimator outperforms the Dominating and the Dominating outperforms the Traditional.

In the range of for $n = 50$ to around $n = 270$ the Tyler's estimator outperforms the Bona Fide and hereby the other ones too. But around $n = 270$ to $n = 450$ the Tyler's

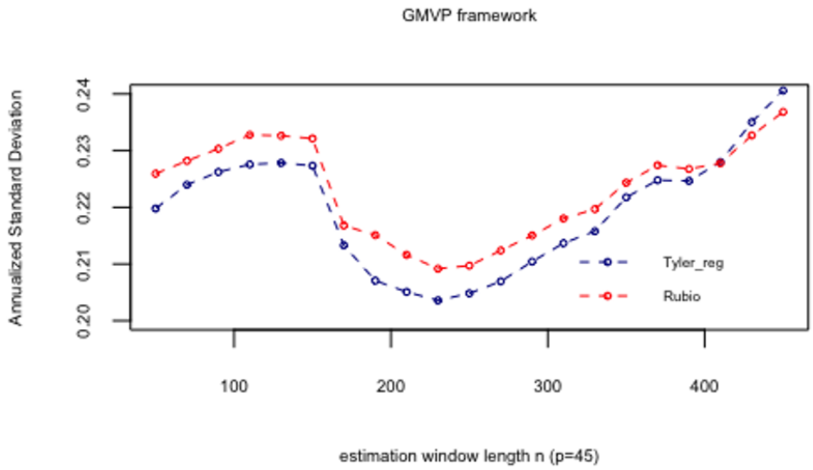

Figure 9.1: Out-of sample portfolio risks over 739 days chosen from HSI real market data from Aug. 1, 2019 to July 31, 2022

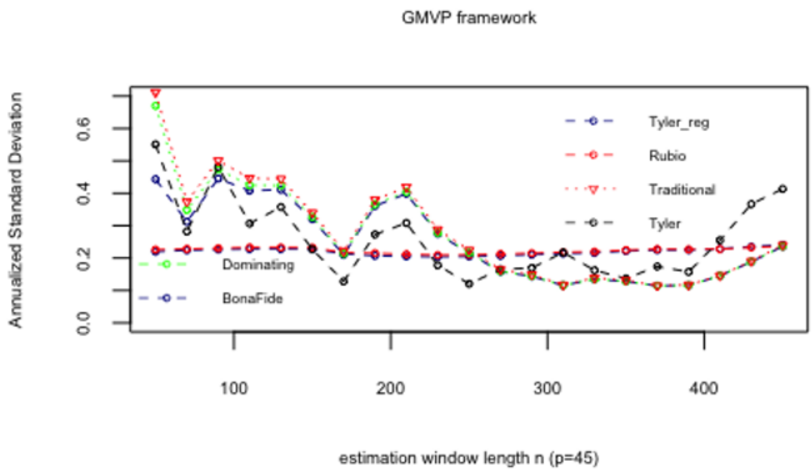

Figure 9.2: Out-of sample portfolio risks over 739 days chosen from HSI real market data from Aug. 1, 2019 to July 31, 2022

estimator performs worst compared to the rest of estimators. From Figure 9.2 it could be seen, that in the range from 50 to almost $n = 250$ the regularized Tyler's and the Rubio et al. estimator perform better than the Bona Fide estimator. But as we know already from $n = 230$ on they start to degrade in their performances and the comparison with the other estimators are not useful.

10. Conclusion

In this thesis we considered several estimators of the GMV portfolio weights which exist in the literature. The preliminary aim was their comparison with respect to their out-of-sample performance of the GMV portfolio based on Random Matrix Theory. We were as well interested in their behavior in high dimensional asymptotic, when the number of assets p and the sample size n together were going to infinity, whereas the convergence rate p/n remained finite.

We analyzed the unregularized Tyler's M-estimator and the unregularized SCM estimator. In the regularized cases we analyzed Tyler's M-estimator and different SCM estimators. The shrinkage target was the non-random identity matrix and the asymptotic of the optimal shrinkage intensities were found and estimated consistently. We used different convergence rates which have an impact on the performances of the estimators and their asymptotic behaviors.

Our results have shown that there do exist asymptotic equivalences between the different estimators in agreement with the theoretical results. In the case of unregularized estimators [10] have proven that a scaled Tyler's M-estimator converges to the SCM estimator in operator norm using normal and elliptical distributions. Further the empirical spectral density of a scaled Tyler's M-estimator converge to the limiting spectral distribution of the SCM estimator.

In the case of regularized estimators [9] have proven a convergence based on spectral norm between Tyler's M-estimator and a specific SCM estimator. Even they have shown that the empirical spectral density of the regularized Tyler's M-estimator converges to the limiting spectral distribution of a specific regularized SCM estimator.

Concerning the performances of the different estimators we applied different convergence rates. For $p/n = 0.9$ when the number of assets p is close to the sample size n, considered as the high dimensional asymptotic case the unregularized and the regularized Tyler's M-estimator outperforms the other estimators which were based on unregularized SCM estimator and the different regularized SCM estimators. In the case $p/n = 0.5$ there are similar performances between the unregularized Tyler's M-estimator and the SCM estimator. This applies even for regularized estimators. For $p/n = 0.1$ when $p < n$ we can consider it as an approximative standard asymptotic case, even when the number of assets p is not fixed. The unregularized SCM estimator outperforms the unregularized Tyler's M-estimator and the different regularized SCM estimators outperform the regularized Tyler's M-estimator. From these considerations it is seen that from $p/n = 0.9$ on towards $p/n = 0.5$ the Tyler's M-estimators perform better than the SCM estimators. From $p/n = 0.5$ on towards $p/n = 0.1$ the SCM estimators perform better than the Tyler's M-estimators. For this reason, at $p/n = 0.5$ when there are slight differences in performances between these estimators, we can consider the case $c = 0.5$ as a transition zone.

Appendix

Theorem 3.2 [10]

Suppose that $\{y_i\}_{i=1}^n$ are i.i.d. sampled from $N(\mathbf{O},\mathbf{I}), p, n \to \infty$ and $p/n = c$, where $0 < c < 1$, and $y_i \sim N(\mathbf{0},\mathbf{I})$ for all $1 \leq i \leq n$, then a scaled Tyler's M-estimator converges to the sample covariance estimator in operator norm almost surely, and there exist $C, c, c' > 0$ such that for any $\varepsilon < c'$,

$$\Pr\left(\left\|p\hat{\mathbf{C}} - \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^T\right\| \leq \varepsilon\right) \geq 1 - Cne^{-c\varepsilon^2 n} \tag{1}$$

In Theorem 3.5 they extend Theorem 3.2 to the case of elliptical distributions.

They base their arguments by using Theorem 3.2 and two properties of Tyler's M-estimator based on the fixed-point equation

$$\hat{\mathbf{C}} = \sum_{i=1}^{n}\frac{\mathbf{y}_i\mathbf{y}_i^T}{\mathbf{y}_i^T\hat{\mathbf{C}}^{-1}\mathbf{y}_i} \tag{2}$$

The first property is achieved when in (2) $\{\mathbf{y}_i\}_{i=1}^n$ are replaced by $\{c_i\mathbf{y}_i\}_{i=1}^n$, where $\{c_i\}_{i=1}^n$ are arbitrary numbers in $\mathbb{R}$, then the solution $\hat{\mathbf{C}}$ in equation (2) remains the same. The second property is achieved when in (2) $\{\mathbf{y}_i\}_{i=1}^n$ are replaced by $\left\{\boldsymbol{\Sigma}_p^{-1/2}\mathbf{y}_i\right\}_{i=1}^n$ and $\hat{\mathbf{C}}$ by $\boldsymbol{\Sigma}_p^{-1/2}\hat{\mathbf{C}}\boldsymbol{\Sigma}_p^{-1/2}/tr(\boldsymbol{\Sigma}_p^{-1/2}\hat{\mathbf{C}}\boldsymbol{\Sigma}_p^{-1/2})$, then the fixed-point equation (2) still holds.

And taking further steps in Section 3.2.1 [10], they prove their Theorem 3.5.

Theorem 3.5 [10]

If $\{\mathbf{y}_i\}_{i=1}^n$ are i.i.d. sampled from the centered elliptical distribution $\mu_p(\mathbf{y}) = C(g_p)det(\boldsymbol{\Sigma}_p)^{-1/2}g_p(\mathbf{y}^T\boldsymbol{\Sigma}_p^{-1}\mathbf{y})$, then there is the following property for Tyler's M-estimator: there exists $c, C, c' > 0$ such that for any $\varepsilon < c'$,

$$\Pr\left(\left\|\frac{p\boldsymbol{\Sigma}_p^{-1/2}\hat{\mathbf{C}}\boldsymbol{\Sigma}_p^{-1/2}}{tr(\boldsymbol{\Sigma}_{\boldsymbol{p}}^{\mathbf{-1/2}}\hat{\mathbf{C}}\boldsymbol{\Sigma}_p^{-1/2})} - \frac{p}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^T\right\| \leq \varepsilon\right) \geq 1 - Cne^{-c\varepsilon^2 n}$$
$$for\ \ \mathbf{z}_i = \boldsymbol{\Sigma}_p^{-1/2}\mathbf{y}_i / \left\|\boldsymbol{\Sigma}_p^{-1/2}\mathbf{y}_i\right\| \tag{3}$$

Theorem 1 [18]

For the Sample Covariance Matrix (SCM):

$$\mathbf{B} = \frac{1}{n}\mathbf{\Sigma}_n^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{\Sigma}_n^{1/2} \tag{4}$$

Silverstein (1995) has proven, assuming that there exists an asymptotic eigenvalue distribution function for $\mathbf{\Sigma}_n$ as $H(t)$ and the empirical spectral density of SCM converges almost surely to a limiting spectral distribution, then the Stieltjes transform of it satisfies for $z \in \mathbf{C}^+$

$$m_{\mathbf{B}(z)} = \int \frac{1}{t(1-c-czm_{\mathbf{B}}(z))-z}dH(t) \tag{5}$$

Theorem 1 [9]

They consider n sample vectors $\mathbf{y}_1,...,\mathbf{y}_n \in \mathbb{R}^p$ with the following characteristics:
Assumption 1 (Growth Rate): $c_p = p/n, c_p \to c \in (0,\infty)\ as\ p \to \infty$
Assumption 2 (Population Model): The vectors $\mathbf{y}_1,...,\mathbf{y}_n \in \mathbb{R}^p$ are independent with (a) $\mathbf{y}_i = \mathbf{\Sigma_p^{1/2}}\mathbf{x}_i$, where $\mathbf{x}_i \in \mathbb{R}^p$ is a random zero mean unitarily invariant vector with norm $\|\mathbf{x}_i\|^2 = p$
(b) $\mathbf{\Sigma}_p = \mathbf{\Sigma}_p^{1/2}(\mathbf{\Sigma}_p^{1/2})^T$ is nonnegative definite, with trace $\frac{1}{p}tr\mathbf{\Sigma}_p = 1$ and spectral norm satisfying $lim\ sup_p \|\mathbf{\Sigma}_p\| < \infty$.
(c) $\nu_p = \frac{1}{p}\sum_{i=1}^p \boldsymbol{\delta}_{\lambda_i(\mathbf{\Sigma_p})}$ satisfies $\nu_p \to \nu$ weakly with $\nu \neq \boldsymbol{\delta}_o$ almost everywhere.

Based on their Assumptions 1 and 2 their Theorem 1 is as follows:
For $\varepsilon \in (0, min\{1, c^{-1}\})$, define $R_\varepsilon = [\varepsilon + max\{0, 1-c^{-1}\}, 1]$.

For each $\rho \in (max\left\{0, 1-c_p^{-1}\right\}, 1]$, let $\hat{\mathbf{C}}_p(\rho)$ be the unique solution to

$$\hat{\mathbf{C}}_p(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{y}_i\mathbf{y}_i^T}{\frac{1}{p}\mathbf{y}_i^T\hat{\mathbf{C}}_p(\rho)^{-1}\mathbf{y}_i} + \rho\mathbf{I}p \tag{6}$$

Then as $p \to \infty$,

$$\sup_{\rho \in R_\varepsilon}\left\|\hat{\mathbf{C}}_p(\rho) - \hat{\mathbf{S}}_p(\rho)\right\| \xrightarrow{a.s.} 0 \tag{7}$$

where

$$\hat{\mathbf{S}}_p(\rho) = \frac{1}{\gamma(\rho)}.\frac{1-\rho}{1-(1-\rho)c}\frac{1}{n}\sum_{i=1}^n \mathbf{y}_i\mathbf{y}_i^T + \rho\mathbf{I}p \tag{8}$$

and $\gamma(\rho)$ is the unique positive solution to the equation.

$$1 = \int \frac{t}{\gamma\rho + (1-\rho)t}\nu(dt) \tag{9}$$

where, the function $\rho \to \gamma(\rho)$ is continuous on $(0,1]$. The proof is given in Section 5.1 [9].

References

[1] Gabriel Frahm and Christoph Memmel. **Dominating estimators for minimum-variance portfolios**. *Journal of Econometrics*, **159**(2):289–302, 2010. i, iv, 15

[2] Taras Bodnar, Nestor Parolya, and Wolfgang Schmid. **Estimation of the global minimum variance portfolio in high dimensions**. *European Journal of Operational Research*, **266**(1):371–390, 2018. The R-codes in this reference were used by the permission of T. Bodnar. i, iii, iv, 7, 11, 12, 13, 15, 34

[3] Francisco Rubio and Xavier Mestre. **Spectral convergence for a general class of random matrices**. *Statistics & Probability Letters*, **81**(5):592–602, 2011. i, 11

[4] Liusha Yang, Romain Couillet, and Matthew R. McKay. **A Robust Statistics Approach to Minimum Variance Portfolio Optimization**. *IEEE Transactions on Signal Processing*, **63**(24):6684–6697, 2015. ii, iv, 7, 20, 21, 22, 34

[5] Francisco Rubio, Xavier Mestre, and Daniel P. Palomar. **Performance Analysis and Optimal Selection of Large Minimum Variance Portfolios Under Estimation Risk**. *IEEE Journal of Selected Topics in Signal Processing*, **6**(4):337–350, 2012. iv, 7, 17, 18, 19, 21

[6] Harry Markowitz. **Portfolio Selection**. *The Journal of Finance*, **7**(1):77–91, 1952. 6

[7] Robert C. Merton. **On estimating the expected return on the market: An exploratory investigation**. *Journal of Financial Economics*, **8**(4):323–361, 1980. 6

[8] Jean-Philippe Bouchaud and Marc Potters. **Financial applications of random matrix theory, A short review, in Chater 40 of Handbook of Random Matrix Theory**. Chapter 40 of Handbook of Random Matrix Theory, G. Akemann, J. Baik, and P. D. Francesko Eds. Oxford, U.K. , Oxford University Press, 2011. 7

[9] Romain Couillet and Matthew McKay. **Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators**. *Journal of Multivariate Analysis*, **131**:99–120, 2014. 7, 28, 37, xl

[10] Teng Zhang, Xiuyuan Cheng, and Amit Singer. **Marchenko-Pastur Law for Tyler's M-estimator**, 2014. 7, 8, 23, 37, xxxix

[11] Robb J Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics. Wiley, 1982. 8

[12] Walid Hachem, Oleksiy Khorunzhiy, Philippe Loubaton, Jamal Najim, and Leonid Pastur. **A new approach for capacity analysis of large dimensional multi-antenna channels**. *IEEE Trans. Inf. Theory*, **54**(9):3987–4004, 2008. 19

[13] David E. Tyler. **A Distribution-Free M-Estimator of Multivariate Scatter**. *The Annals of Statistics*, **15**(1):234 – 251, 1987. 20

[14] Esa Ollila. *Regularized M-Estimator of the Covariance Matrix, Summer School, Rüdesheim.* Aalto University, 2016. The R-codes in this reference were used for the Tyler's M-Estimator and the regularized M-Estimator. 20

[15] Esa Ollila and David E. Tyler. **Regularized M-Estimators of Scatter Matrix**. *IEEE Transactions on Signal Processing*, **62**(22):6059–6070, 2014. 20

[16] Olivier Ledoit and Michael Wolf. **A well-conditioned estimator for large-dimensional covariance matrices**. *Journal of Multivariate Analysis*, **88**(2):365–411, 2004. 20

[17] Vladimir Alexandrovich Marchenko and Leonid Andreevich Pastur. **Distribution of eigenvalues for some sets of random matrices**. *Matematicheskii Sbornik*, **114**(4):507–536, 1967. 23

[18] Jack W Silverstein. **Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices**. *Journal of Multivariate Analysis*, **55**(2):331–339, 1995. 23, xl

[19] Victor DeMiguel, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal. **A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms**. *Management science*, **55**(5):798–812, 2009. 34