DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 2/08

Dominating Estimators for the Global Minimum Variance Portfolio

by

Gabriel Frahm
Christoph Memmel

2nd version
November 6, 2008

DISKUSSIONSBEITRÄGE ZUR STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

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

This page intentionally left blank.

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 2/08

Dominating Estimators for the Global Minimum Variance Portfolio

by

Gabriel Frahm[1]
Christoph Memmel[2]

2$^{\text{nd}}$ version
November 6, 2008

Abstract

In this paper, we derive two shrinkage estimators for the global minimum variance portfolio that dominate the traditional estimator with respect to the out-of-sample variance of the portfolio return. The presented results hold for any number of observations $n \geq d + 2$ and number of assets $d \geq 4$. The small-sample properties of the shrinkage estimators as well as their large-sample properties for fixed d but $n \to \infty$ as well as $n, d \to \infty$ but $n/d \to q \leq \infty$ are investigated. Furthermore, we present a small-sample test for the question of whether it is better to completely ignore time series information in favor of naive diversification.

Keywords: Covariance matrix estimation, Global minimum variance portfolio, James-Stein estimation, Naive diversification, Shrinkage estimator.

JEL classification: C13, G11.

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

[2]Deutsche Bundesbank, Banking and Financial Supervision Department, Wilhelm-Epstein-Strasse 14, D-60431 Frankfurt, Germany. Phone: +49 69 9566-8531, email: christoph.memmel@bundesbank.de.

This page intentionally left blank.

Dominating Estimators for the Global Minimum Variance Portfolio*

Gabriel Frahm†
University of Cologne
Chair for Statistics & Econometrics
Department of Economic and Social Statistics
Albertus-Magnus-Platz, D-50923 Cologne, Germany

Christoph Memmel‡
Deutsche Bundesbank
Banking and Financial Supervision Department
Wilhelm-Epstein-Strasse 14, D-60431 Frankfurt, Germany

Abstract

In this paper, we derive two shrinkage estimators for the global minimum variance portfolio that dominate the traditional estimator with respect to the out-of-sample variance of the portfolio return. The presented results hold for any number of observations $n \geq d + 2$ and number of assets $d \geq 4$. The small-sample properties of the shrinkage estimators as well as their large-sample properties for fixed d but $n \to \infty$ as well as $n, d \to \infty$ but $n/d \to q \leq \infty$ are investigated. Furthermore, we present a small-sample test for the question of whether it is better to completely ignore time series information in favor of naive diversification.

JEL classification: C13, G11.

Keywords: Covariance matrix estimation, Global minimum variance portfolio, James-Stein estimation, Naive diversification, Shrinkage estimator.

---

*We would like to thank Alexander Kempf and Julia Nasev for their helpful comments on the manuscript. The opinions expressed in this paper are those of the authors and do not necessarily reflect the opinions of the Deutsche Bundesbank.

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

‡Phone: +49 69 9566-8531, email: christoph.memmel@bundesbank.de.

1 Introduction

When implementing portfolio optimization according to Markowitz (1952), one needs to estimate the expected asset returns as well as the corresponding variances and covariances. If the parameter estimates are based only on time series information, the suggested portfolio tends to be far removed from the optimum. For this reason, there is a broad literature which addresses the question of how to reduce estimation risk in portfolio optimization. In a recent study, DeMiguel et al. (2007) compare portfolio strategies which differ in the treatment of estimation risk. It turns out that none of the strategies suggested in the literature is significantly better than naive diversification, i.e. taking the equally weighted portfolio. Further, the study conducted by DeMiguel et al. (2007) confirms that the considered strategies perform better than the traditional implementation of Markowitz optimization, which means replacing the unknown parameters by their sample counterparts.

The global minimum variance portfolio (GMVP) has been frequently advocated in the literature (Frahm, 2008; Jagannathan and Ma, 2003; Kempf and Memmel, 2006; Ledoit and Wolf, 2003) because it is completely independent of the expected asset returns, which have been found to be the principal source of estimation risk (Chopra and Ziemba, 1993; Merton, 1980). We present two estimators for the GMVP which *dominate* the traditional estimator with respect to the out-of-sample variance of the portfolio return. Due to the arguments set forth by Frahm (2008), the same conclusion can be drawn for estimating local minimum variance portfolios, i.e. minimum variance portfolios where the portfolio weights are subject to other linear equality constraints besides the budget constraint.

Okhrin and Schmid (2006), Kempf and Memmel (2006) and Frahm (2008) all explore the properties of the traditional GMVP estimator by assuming jointly normally distributed asset returns. They derive the small-sample distribution of the estimated portfolio weights and give a closed-form expression for the out-of-sample variance of the portfolio return. In contrast, Bayesian and shrinkage approaches have a long tradition in the implementation of modern portfolio optimization. Jobson and Korkie (1979) and Jorion (1986) introduce shrinkage estimators for the expected returns. Frost and Savarino (1986) generalize these estimators by also including the variances and covariances. Furthermore, DeMiguel et al. (2007), Garlappi et al. (2007), Golosnoy and Okhrin (2007) as well as Kan and Zhou (2007) present some shrinkage estimators for the weights of mean-variance optimal portfolios, whereas Ledoit and Wolf (2003) introduce a shrinkage estimator for the covariance matrix of stock returns and apply their results to the estimation of the GMVP.

Our work is related to these shrinkage approaches. However, it differs in two important aspects. First, we derive *feasible* estimators, and our dominance results turn out to be valid even in small samples. The shrinkage approaches presented by the aforementioned authors can only be justified for a large number of observations. As pointed out by Frahm (2008), large-sample results can be misleading in the context of portfolio optimization since, even if the sample size is large, the number of observations can be small compared to the number of assets. Second, in contrast to Ledoit and Wolf (2003) we do not seek to obtain a better covariance matrix estimator but instead to reduce the out-of-sample variance of the portfolio return, which seems to be the major goal when searching for a minimum variance portfolio.

Another method of alleviating the impact of estimation risk is to impose certain restrictions on the estimated covariance matrix or portfolio weights. Examples for restrictions on the covariance matrix are the single index model of Sharpe (1963) and the constant correlation model suggested by Elton and Gruber (1973). Jagannathan and Ma (2003) show that imposing short-sales constraints on the GMVP is equivalent to assuming a special structure of the covariance matrix. Frahm (2008) analyzes linear equality constraints on the portfolio weights and proves that linear restrictions reduce estimation risk. All these approaches have in common the fact that the restrictions may be binding and so the true GMVP does not need to be attained if the length of the time series approaches infinity. Nevertheless, in an empirical study presented by Chan et al. (1999) it has been shown that the reduction of estimation risk typically outweighs the loss caused by applying 'wrong' restrictions. Shrinkage estimators reduce the estimation risk as well. However, in addition they have the appealing property of converging towards the optimal portfolio weights as the sample size grows to infinity.

Our contribution to the literature is threefold. First, we derive two shrinkage estimators for the GMVP that dominate the traditional estimator with respect to the out-of-sample variance of the portfolio return. Second, we present not only the small-sample properties of the shrinkage estimators and some related quantities, but also their large-sample properties for fixed d and $n \to \infty$ as well as $n, d \to \infty$ and $n/d \to q \leq \infty$. The latter kind of asymptotic behavior becomes relevant when analyzing the estimators in large asset universes. Third, backed by the results of DeMiguel et al. (2007), we derive a small-sample test for the *naive diversification hypothesis*, i.e. for deciding the question of whether or not it is better to completely ignore time series information in favor of naive diversification.

2 Preliminaries

2.1 Notation and Assumptions

Suppose that the investment universe consists of d assets and the investor is searching for a buy-and-hold portfolio which will be liquidated after one period. We will consider the asset excess returns $R_t = (R_{1t}, \ldots, R_{dt})$ for $t = 1, \ldots, n$,[1] i.e. the asset returns minus the corresponding risk-free interest rates. Nevertheless we will drop the prefix 'excess' for convenience and make the following assumptions:

A1. The asset returns are jointly normally distributed, i.e. $R_t \sim \mathcal{N}_d(\mu, \Sigma)$ for $t = 1, \ldots, n$ with $\mu \in \mathbb{R}^d$ and positive-definite matrix $\Sigma \in \mathbb{R}^{d \times d}$.

A2. The mean vector μ and the covariance matrix Σ are unknown.

A3. The asset returns are serially independent.

A4. The sample size exceeds the number of assets, more precisely $n \geq d + 2$.

A5. There exist at least four assets, i.e. $d \geq 4$.

The GMVP w is defined as the solution of the minimization problem

$$\min_{v \in \mathbb{R}^d} v'\Sigma\, v\,, \qquad \text{s.t.}\; v\,'\mathbf{1} = 1\,. \tag{1}$$

Here $\mathbf{1}$ denotes a vector of ones. Since Σ is positive-definite, the GMVP is unique and the solution of this minimization problem corresponds to $w = \Sigma^{-1}\mathbf{1}/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$. The traditional estimator $\hat{w}_{\mathrm{T}}$ for the GMVP consists in replacing the unknown covariance matrix Σ with the sample covariance matrix $\widehat{\Sigma}$, i.e.

$$\widehat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}\big(R_t - \bar{R}\big)\big(R_t - \bar{R}\big)'\,, \tag{2}$$

where $\bar{R} = 1/n\sum_{t=1}^{n} R_t$ represents the sample mean vector of $R_1, \ldots, R_n$. The variance of the GMVP return corresponds to $\sigma^2 = w'\Sigma\, w = 1/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$ and its traditional estimator is given by $\hat{\sigma}^2_{\mathrm{T}} = \hat{w}'_{\mathrm{T}}\widehat{\Sigma}\,\hat{w}_{\mathrm{T}} = 1/(\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1})$.

[1] In the following '$(x_1, \ldots, x_d)$' indicates a d-tuple, i.e. a d-dimensional column vector.

Since the portfolio weights always add up to 1, it is possible to omit one element of the portfolio weights vector without losing information. We choose to omit the first element and define $w^{\text{ex}} := (w_2, \ldots, w_d)$. For convenience we introduce the $(d-1) \times d$ matrix $\Delta := [\,\mathbf{1}\ -I_{d-1}\,]$. By using the operator Δ, we can easily switch between the two notations. For instance, note that $(v_1 - v_2) = -\Delta'(v_1^{\text{ex}} - v_2^{\text{ex}})$ for all vectors $v_1, v_2 \in \mathbb{R}^d$ whose elements add up to 1. Moreover, the following relationship will be useful in the subsequent discussion:

$$(v_1 - v_2)' A \,(v_1 - v_2) = (v_1^{\text{ex}} - v_2^{\text{ex}})' B \,(v_1^{\text{ex}} - v_2^{\text{ex}}) \tag{3}$$

with $B := \Delta A \Delta'$ for any $d \times d$ matrix A. A key note of the present work is that

$$v' \Sigma\, v = \sigma^2 + (v - w)' \Sigma \,(v - w) = \sigma^2 + (v^{\text{ex}} - w^{\text{ex}})' \Omega \,(v^{\text{ex}} - w^{\text{ex}}) \tag{4}$$

for every vector $v \in \mathbb{R}^d$ with $v'\mathbf{1} = 1$, where Ω is defined as $\Omega := \Delta \Sigma \Delta'$. The first equality in (4) can be obtained by noting that $\Sigma w = \mathbf{1}/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$ and thus $v'\Sigma\, w = 1/(\mathbf{1}'\Sigma^{-1}\mathbf{1}) = \sigma^2$. The second equality follows from the arguments given above.

In the following $\chi^2_k(\lambda)$ denotes a noncentral χ^2-distributed random variable with $k \in \mathbb{N}$ degrees of freedom and noncentrality parameter $\lambda \geq 0$. This means $\chi^2_k(\lambda) \sim X'X$ with $X \sim \mathcal{N}_k(\theta, I_k)$ and $\theta \in \mathbb{R}^k$, where the noncentrality parameter is defined as $\lambda := \theta'\theta/2$. By contrast, χ^2_k stands for a central χ^2-distributed random variable (i.e. $\lambda = 0$) and we also define $\chi^r_k(\lambda) := \left\{\chi^2_k(\lambda)\right\}^{r/2}$ for any $r \in \mathbb{Z}$. Moreover, let $\chi^2_{k_1}(\lambda)$ and $\chi^2_{k_2}$ with $k_1, k_2 \in \mathbb{N}$ be stochastically independent. Then $F_{k_1,k_2}(\lambda) \sim (k_2/k_1)\big(\chi^2_{k_1}(\lambda)/\chi^2_{k_2}\big)$ has a noncentral F-distribution with k_1 and k_2 degrees of freedom as well as noncentrality parameter $\lambda \geq 0$. Now suppose that $X_1, \ldots, X_m$ are m independent copies of $X \sim \mathcal{N}_q(\mathbf{0}, \Sigma)$, where $\mathbf{0}$ denotes a vector of zeros and Σ is a positive-definite $q \times q$ matrix. Then the $q \times q$ random matrix $W_q(\Sigma, m) \sim \sum_{i=1}^m X_i X_i'$ possesses a q-dimensional Wishart distribution with covariance matrix Σ and m degrees of freedom. Furthermore, $x^+ := \max\{x, 0\}$ denotes the positive part and $x^- := -\min\{x, 0\}$ the negative part of $x \in \mathbb{R}$. Let A be some positive-definite $q \times q$ matrix. Then $A^{\frac{1}{2}}$ represents the unique symmetrical $q \times q$ matrix such that $A = A^{\frac{1}{2}} A^{\frac{1}{2}}$. Finally, $x \propto y$ means 'x is proportional to y' and $\|\cdot\|$ denotes the Euclidean norm.

2.2 Important Theorems

Let us now provide some important theorems which will come in handy in the following sections. First, we present some elementary small-sample properties of the traditional estimator for the GMVP and its related quantities. A proof can be found in Kempf and Memmel (2006).

Lemma 1 (Kempf and Memmel (2006))

Under assumptions A1 to A3 and $n > d$, the sample covariance matrix $\widehat{\Omega}$ of ΔR, the traditional estimator $\hat{w}_{\mathrm{T}}^{\mathrm{ex}}$ for the GMVP (except for the first portfolio weight), and the traditional estimator $\hat{\sigma}_{\mathrm{T}}^2$ for the minimum variance σ^2 satisfy the following properties:

P1. $n\,\widehat{\Omega} \sim W_{d-1}(\Omega, n-1)$, *where* $\widehat{\Omega} := \frac{1}{n}\sum_{t=1}^{n}\big(\Delta R - \Delta\bar{R}\big)\big(\Delta R - \Delta\bar{R}\big)'$.

P2. $\hat{w}_{\mathrm{T}}^{\mathrm{ex}} \,|\, \widehat{\Omega} \sim \mathcal{N}_{d-1}\big(w^{\mathrm{ex}}, \sigma^2\widehat{\Omega}^{-1}/n\big)$.

P3. $n\hat{\sigma}_{\mathrm{T}}^2/\sigma^2 \sim \chi_{n-d}^2$.

P4. $\hat{\sigma}_{\mathrm{T}}^2$ *is stochastically independent of* $\widehat{\Omega}$ *and* $\hat{w}_{\mathrm{T}}^{\mathrm{ex}}$.

The following theorem will play the central role in the development of the shrinkage estimator and its dominance property.

Theorem 1

Consider a $q \times q$ random matrix $W \sim W_q\big(\Omega, m\big)$, where Ω is a positive-definite $q \times q$ matrix, $q \geq 3$ and $m \geq q+2$, a q-dimensional random vector X with $X \,|\, W \sim \mathcal{N}_q\big(\omega, W^{-1}\big)$, where $\omega \in \mathbb{R}^q$ is an unknown parameter, and a random variable $\chi^2 \sim \chi_k^2$ with $k \geq 2$, which is stochastically independent of W and X. Furthermore, consider a non-stochastic vector $x \in \mathbb{R}^q$. For all $0 < c < 2\,(q-2)/(k+2)$, the shrinkage estimator

$$X_{\mathrm{S}} = x + \left(1 - \frac{c\,\chi^2}{(X-x)'W(X-x)}\right)(X-x)$$

dominates the estimator X with respect to the loss function

$$\mathcal{L}_{\omega,\Omega}\big(\hat{\omega}\big) = \big(\hat{\omega} - \omega\big)'\Omega\,\big(\hat{\omega} - \omega\big)\,, \tag{5}$$

i.e. $\mathrm{E}\big\{(X_{\mathrm{S}} - \omega)'\Omega\,(X_{\mathrm{S}} - \omega)\big\} < \mathrm{E}\big\{(X - \omega)'\Omega\,(X - \omega)\big\}$. *In case $x = \omega$ the expected loss of the shrinkage estimator becomes minimal if and only if $c = (q-2)/(k+2)$.*

Proof: See the appendix.

Note that Theorem 1 coincides with the well-known result developed by Stein (1956) if W is substituted by the identity matrix I_q. Other extensions of Stein's theorem, which can be found in the literature, require that W correspond to a non-stochastic but observable matrix Ω, or at least that W be stochastically independent of X where Ω is unobservable (Judge and Bock (1978, p. 177), Srivastava and Bilodeau (1989), and Press (2005, p. 189)). By contrast, we allow X to depend on a Wishart-distributed random matrix W, but the matrix Ω given in Theorem 1 remains unobservable.

Theorem 1 also clarifies why the shrinkage constant $c = (q-2)/(k+2)$ is a natural choice. Although any constant within the interval given in Theorem 1 would lead to a dominant estimator, only $c = (q-2)/(k+2)$ turns out to be the best choice if the reference vector x corresponds to the unknown parameter ω. The same value for c remains optimal in the variants of Stein's theorem where W is non-stochastic or stochastically independent of X.

2.3 Out-of-Sample Variance

The out-of-sample variance of the return of a stochastic portfolio $\hat{v}$ is defined as

$$\mathrm{Var}\big(\hat{v}\,'R\big) = \mathrm{E}\big\{\mathrm{Var}(\hat{v}\,'R\,|\,\hat{v})\big\} + \mathrm{Var}\big\{\mathrm{E}(\hat{v}\,'R\,|\,\hat{v})\big\} = \mathrm{E}\big(\hat{v}'\Sigma\,\hat{v}\big) + \mu'\mathrm{Var}\big(\hat{v}\big)\mu\,.$$

This means the total variance of the portfolio $\hat{v}$ can be split into a within variance $\mathrm{E}\big(\hat{v}'\Sigma\,\hat{v}\big)$ and a between variance $\mu'\mathrm{Var}\big(\hat{v}\big)\mu$. Due to (4), it holds that

$$\mathrm{Var}\big(\hat{v}\,'R\big) = \sigma^2 + \mathrm{E}\big\{(\hat{v}-w)'\Sigma\,(\hat{v}-w)\big\} + \mu'\mathrm{Var}\big(\hat{v}\big)\mu\,. \tag{6}$$

Hence, the minimum variance σ^2 is a lower bound for the out-of-sample variance of any given portfolio $\hat{v}$. Interestingly, the between variance $\mu'\mathrm{Var}\big(\hat{v}\big)\mu$ vanishes whenever the expected asset returns are equal to each other, i.e. $\mu = \eta\,\mathbf{1}$ for any $\eta \in \mathbb{R}$. This can be seen by noting that $\mathrm{Var}(\hat{v}) = \Delta'\mathrm{Var}(\hat{v}^{\mathrm{ex}})\Delta$ and $\Delta\mu = \mathbf{0}$ if $\mu = \eta\,\mathbf{1}$.

Kempf and Memmel (2006) showed that – concerning the traditional estimator $\hat{w}$ for the GMVP – the second part of (6) corresponds to

$$\mathrm{E}\big\{(\hat{w}_{\mathrm{T}}-w)'\Sigma\,(\hat{w}_{\mathrm{T}}-w)\big\} = \frac{d-1}{n-d-1}\cdot\sigma^2\,.$$

The factor $(d-1)/(n-d-1)$ is large whenever the sample size n is small compared to the number of assets d. For $n, d \to \infty$ but $n/d \to q$ with $1 < q \leq \infty$, this factor tends to $1/(q-1)$. Hence even in large samples the contribution of the estimation risk to the out-of-sample variance is not negligible if the 'effective sample size' q is small. For instance, given an investment universe with $d = 50$ assets and a history of $n = 100$ monthly observations, the additional variance caused by the estimation risk is $1/(100/50 - 1) = 100\%$.

From the small-sample distribution of $\hat{w}$ presented by Frahm (2008), it follows that the third part of (6) corresponds to

$$\mu' \mathrm{Var}\big(\hat{w}_{\mathrm{T}}\big)\mu = \frac{r_{\max}^2 - r_{\mathrm{GMVP}}^2}{n-d-1} \cdot \sigma^2 \,,$$

where $r_{\max}$ denotes the Sharpe ratio of the tangential portfolio $\Sigma^{-1}\mu\,/(\mathbf{1}'\Sigma^{-1}\mu)$ and r_{GMVP} the Sharpe ratio of the GMVP.[2] This means it holds that

$$\mathrm{Var}\big(\hat{w}_{\mathrm{T}}' R\big) = \left(1 + \frac{d-1}{n-d-1} + \frac{r_{\max}^2 - r_{\mathrm{GMVP}}^2}{n-d-1}\right) \cdot \sigma^2 \,.$$

In most practical situations the difference of $r_{\max}^2$ and r_{GMVP}^2 turns out to be much smaller than the numerator $d - 1$ (and even vanishes if $\mu = \eta\,\mathbf{1}$).

Generally, in real-world asset markets the expected returns presumably do not differ so greatly in the cross-section; the between variance is therefore very small compared to the within variance. Hence we believe that the between variance $\mu'\mathrm{Var}\big(\hat{v}\big)\mu$ for any portfolio $\hat{v}$ is negligible in most practical situations and will concentrate in the following on reducing the within variance $\mathrm{E}\big(\hat{v}'\Sigma\,\hat{v}\big)$. Note that each realization of $\hat{v}'\Sigma\,\hat{v}$ represents the *actual variance* of the return belonging to the portfolio $\hat{v}$, which has been chosen on the basis of historical observations, for instance. Then due to (4), each realization of $(\hat{v} - w)'\Sigma\,(\hat{v} - w)$ can be interpreted as that part of the actual variance which is caused by estimation risk. In the subsequent analysis this quantity will be referred to as the *loss* of $\hat{v}$.

3 The Dominant Estimators

3.1 Small-Sample Properties

We now present the shrinkage estimator for the GMVP that dominates the traditional estimator. Kempf and Memmel (2006) show that the traditional estimator is the best unbiased estimator in the case of jointly normally distributed asset returns.[3] However, as

[2]The Sharpe ratio of a portfolio is the expected excess return divided by the standard deviation.

[3]An estimator is called *best* if its covariance matrix attains the Rao-Cramér lower bound.

already discussed earlier, this estimator can lead to a huge out-of-sample variance of the portfolio return compared to σ^2, i.e. the smallest of all possible portfolio return variances.

In this section we will use the following notation. Let $\hat{w}_{\mathrm{A}}$ be an arbitrary portfolio. Then $\sigma^2_{\mathrm{A}} = \hat{w}'_{\mathrm{A}}\Sigma\,\hat{w}_{\mathrm{A}}$ is the actual variance of the portfolio return, whereas $\hat{\sigma}^2_{\mathrm{A}} = \hat{w}'_{\mathrm{A}}\widehat{\Sigma}\,\hat{w}_{\mathrm{A}}$ denotes the corresponding estimator. This notation will be used both for stochastic and non-stochastic portfolios, i.e. if w_{A} is a non-stochastic portfolio, it holds that $\sigma^2_{\mathrm{A}} = w'_{\mathrm{A}}\Sigma\,w_{\mathrm{A}}$ and $\hat{\sigma}^2_{\mathrm{A}} = w'_{\mathrm{A}}\widehat{\Sigma}\,w_{\mathrm{A}}$.

Theorem 2

Suppose that the assumptions A1 to A5 are satisfied. Let $\hat{w}_{\mathrm{T}}$ be the traditional estimator for the GMVP w, whereas $w_{\mathrm{R}} \in \mathbb{R}^d$ with $w'_{\mathrm{R}}\mathbf{1} = 1$ denotes an arbitrary reference portfolio. Consider the shrinkage estimator

$$\hat{w}_{\mathrm{S}} = \kappa_{\mathrm{S}} w_{\mathrm{R}} + \big(1 - \kappa_{\mathrm{S}}\big)\hat{w}_{\mathrm{T}} \tag{7}$$

with

$$\kappa_{\mathrm{S}} = \frac{d-3}{n-d+2}\cdot\frac{1}{\hat{\tau}_{\mathrm{R}}}\,,$$

where $\hat{\tau}_{\mathrm{R}} = \big(\hat{\sigma}^2_{\mathrm{R}} - \hat{\sigma}^2_{\mathrm{T}}\big)/\hat{\sigma}^2_{\mathrm{T}}$ is the estimated relative loss of the reference portfolio w_{R}. The shrinkage estimator $\hat{w}_{\mathrm{S}}$ dominates $\hat{w}_{\mathrm{T}}$ with respect to the loss function $\mathcal{L}_{w,\Sigma}(\hat{v}) = (\hat{v} - w)'\Sigma\,(\hat{v} - w)$, i.e.

$$\mathrm{E}\big\{(\hat{w}_{\mathrm{S}} - w)'\Sigma\,(\hat{w}_{\mathrm{S}} - w)\big\} < \mathrm{E}\big\{(\hat{w}_{\mathrm{T}} - w)'\Sigma\,(\hat{w}_{\mathrm{T}} - w)\big\}\,.$$

Proof: See the appendix.

The estimator suggested in Theorem 2 exhibits the typical structure of James-Stein-type shrinkage estimators. It is a weighted average of a given reference portfolio and the traditional estimator for the GMVP. The better the reference portfolio fits the actual GMVP, the smaller the out-of-sample variance of the shrinkage estimator will be. When it comes to portfolio diversification without any subjective or empirical information as well as restrictions on the portfolio weights, the *naive portfolio* $w_{\mathrm{N}} := \mathbf{1}/d$ can be viewed as a natural choice for the reference portfolio. Due to the arguments given by DeMiguel et al. (2007), there are even doubts as to whether time series information can add useful information at all, and so $w_{\mathrm{R}} = w_{\mathrm{N}}$ might serve as a rule. We will come back to this point in Section 4.

Theorem 3

Under the assumptions of Theorem 2, the distribution of the relative loss

$$\tau_{\mathrm{S}} = \frac{\sigma_{\mathrm{S}}^2 - \sigma^2}{\sigma^2}$$

of the shrinkage estimator for the GMVP given by (7) depends only on the number of observations n, *the number of assets* d, *and the relative loss* $\tau_{\mathrm{R}} = (\sigma_{\mathrm{R}}^2 - \sigma^2)/\sigma^2$ *of the reference portfolio. More precisely,* τ_{S} *can be represented stochastically by*

$$\tau_{\mathrm{S}} = \left\|\kappa_{\mathrm{S}}\theta - \big(1-\kappa_{\mathrm{S}}\big)V^{-\frac{1}{2}}\xi\right\|^2, \tag{8}$$

with any $\theta \in \mathbb{R}^{d-1}$ *such that* $\theta'\theta = \tau_{\mathrm{R}}$, $\xi \sim \mathcal{N}_{d-1}(\mathbf{0}, I_{d-1})$, $V \sim W_{d-1}(I_{d-1}, n-1)$, *and*

$$\kappa_{\mathrm{S}} = \frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}}{\big(\theta + V^{-\frac{1}{2}}\xi\big)'V\big(\theta + V^{-\frac{1}{2}}\xi\big)}\,.$$

Here ξ, V, *and* χ^2_{n-d} *are supposed to be mutually independent.*

Proof: See the appendix.

Due to Theorem 2, the shrinkage estimator is dominant in the sense that $\mathrm{E}\big(\tau_{\mathrm{S}}\big) < \mathrm{E}\big(\tau_{\mathrm{T}}\big)$, where $\tau_{\mathrm{T}} = (\sigma_{\mathrm{T}}^2 - \sigma^2)/\sigma^2$ represents the relative loss of the traditional estimator for the GMVP. It can be shown that the expected relative loss of the shrinkage estimator is a strictly increasing function of τ_{R} and its infimum is attained if and only if $\tau_{\mathrm{R}} = 0$. Note that $\tau_{\mathrm{R}} = 0$ or, equivalently, $\theta = \mathbf{0}$ holds if and only if $w_{\mathrm{R}} = w$, since Σ is positive-definite. In that case it turns out that

$$\mathrm{E}\big(\tau_{\mathrm{S}}\big) = \left(1 - \frac{d-3}{d-1} \cdot \frac{n-d}{n-d+2}\right)\frac{d-1}{n-d-1}\,.$$

By contrast, $\mathrm{E}\big(\tau_{\mathrm{S}}\big) \to \mathrm{E}(\tau_{\mathrm{T}})$ for $\tau_{\mathrm{R}} \to \infty$.

Following the arguments given by Judge and Bock (1978, p. 182), we can try to reduce the out-of-sample variance of the suggested estimator by restricting κ_{S} to values smaller than or equal to 1, i.e. by taking $\kappa_{\mathrm{M}} := \min\{\kappa_{\mathrm{S}}, 1\}$ instead of κ_{S}. Then the corresponding shrinkage estimator is given by

$$\hat{w}_{\mathrm{M}} := \kappa_{\mathrm{M}} w_{\mathrm{R}} + \big(1 - \kappa_{\mathrm{M}}\big)\hat{w}_{\mathrm{T}}\,. \tag{9}$$

The shrinkage constant κ_{M} can only attain values between 0 and 1, which prevents $\hat{w}_{\mathrm{M}}$ from having the opposite sign of $\hat{w}_{\mathrm{T}}$ whenever $\hat{\tau}_{\mathrm{R}}$ is small, i.e. whenever the traditional estimate of the GMVP is close to the reference portfolio. The next theorem states that the modified shrinkage estimator does, in fact, lead to a better out-of-sample performance.

Theorem 4

Under the assumptions of Theorem 2 and given the notation of Theorem 3, the distribution of the relative loss

$$\tau_{\mathrm{M}} = \frac{\sigma_{\mathrm{M}}^2 - \sigma^2}{\sigma^2}$$

of the modified shrinkage estimator for the GMVP given by (9) depends only on the number of observations n *, the number of assets* d *, and the relative loss* τ_{R} *of the reference portfolio. More precisely,* τ_{M} *can be represented stochastically by*

$$\tau_{\mathrm{M}} = \left\| \kappa_{\mathrm{M}}\theta - \left(1 - \kappa_{\mathrm{M}}\right)V^{-\frac{1}{2}}\xi \right\|^2 , \tag{10}$$

with $\kappa_{\mathrm{M}} = \min\{\kappa_{\mathrm{S}}, 1\}$*, and it holds that*

$$\mathrm{E}\big(\tau_{\mathrm{M}}\big) < \mathrm{E}\big(\tau_{\mathrm{S}}\big) < \mathrm{E}\big(\tau_{\mathrm{T}}\big) .$$

Proof: See the appendix.

The stochastic representations (8) and (10) can be used, for instance, for evaluating the out-of-sample performances of the presented shrinkage estimators by Monte Carlo simulation. Theorem 4 asserts that the modified shrinkage estimator dominates not only the traditional estimator but also the simple shrinkage estimator given by (7). Moreover, it can be shown that the expected relative loss of $\hat{w}_{\mathrm{M}}$ corresponds to

$$\mathrm{E}\big(\tau_{\mathrm{M}}\big) = \mathrm{E}\left[\left\{\left(1 - \frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}}{\chi^2_{d+1}}\right)^{+}\right\}^2\right] \frac{d-1}{n-d-1}$$

in the event that $\tau_{\mathrm{R}} = 0$.

Our results about the superiority of the presented shrinkage estimators require the asset universe to consist of at least four assets. By contrast, if there are only two or three assets, one should draw on the traditional estimator. It is worth pointing out that the methodology presented here can be easily applied to the estimation of local minimum variance portfolios. As has been shown by Frahm (2008), any d-dimensional asset universe can be transformed into a $(d-q)$-dimensional asset universe such that q linear equality constraints (besides the budget constraint) are implicitly satisfied for each portfolio of the $d - q$ available assets. In that case assumptions A4 and A5 have to be changed to $n \geq d - q + 2$ and $d \geq q + 4$. Furthermore, the chosen reference portfolio must satisfy the given linear restrictions.

3.2 Large-Sample Properties

In the previous section, we investigated the small-sample properties of the relative losses of the shrinkage estimators $\hat{w}_{\mathrm{S}}$ and $\hat{w}_{\mathrm{M}}$. Due to Theorem 3 and Theorem 4, it can be seen that the expected relative losses of the shrinkage estimators as well as the traditional estimator tend to zero if the number of assets d is fixed but $n \to \infty$. However, that does not mean that the presented shrinkage estimators are always asymptotically equivalent to the traditional estimator. This is confirmed by the next theorem.

Theorem 5

Under assumptions A1 to A3 it holds that

$$\sqrt{n}\begin{bmatrix}\hat{w}_{\mathrm{T}} - w \\ \hat{w}_{\mathrm{S}} - w \\ \hat{w}_{\mathrm{M}} - w\end{bmatrix} \xrightarrow{\mathrm{d}} \begin{bmatrix} 1 \\ \mathbb{1}_{\{\tau_{\mathrm{R}}=0\}}\Big(1 - \frac{d-3}{\xi'\xi}\Big) + \mathbb{1}_{\{\tau_{\mathrm{R}}>0\}} \\ \mathbb{1}_{\{\tau_{\mathrm{R}}=0\}}\Big(1 - \frac{d-3}{\xi'\xi}\Big)^{+} + \mathbb{1}_{\{\tau_{\mathrm{R}}>0\}}\end{bmatrix}\Lambda\xi\,, \qquad n \longrightarrow \infty\,,$$

where Λ *is a* $d \times (d-1)$ *matrix such that* $\Lambda\Lambda' = \sigma^2\Sigma^{-1} - ww'$ *and* $\xi \sim \mathcal{N}_{d-1}\big(\mathbf{0}, I_{d-1}\big)$.

Proof: See the appendix.

For instance, from the last theorem it follows that

$$\sqrt{n}\,\big(\hat{w}_{\mathrm{T}} - w\big) \xrightarrow{\mathrm{d}} \mathcal{N}_d\big(\mathbf{0}, \sigma^2\Sigma^{-1} - ww'\big)\,, \qquad n \longrightarrow \infty\,,$$

and the shrinkage estimators are asymptotically equivalent to the traditional estimator, i.e.

$$\sqrt{n}\,\big(\hat{w}_{\mathrm{T}} - \hat{w}_{\mathrm{S}}\big) \xrightarrow{\mathrm{p}} \mathbf{0} \qquad \text{and} \qquad \sqrt{n}\,\big(\hat{w}_{\mathrm{T}} - \hat{w}_{\mathrm{M}}\big) \xrightarrow{\mathrm{p}} \mathbf{0}\,, \qquad n \longrightarrow \infty\,, \tag{11}$$

only if $w_{\mathrm{R}} \neq w$.[4] The last theorem also implies that if $w_{\mathrm{R}} = w$ and the sample size is large (compared to the number of assets), the modified shrinkage estimate corresponds to the true GMVP roughly with probability $F_{\chi^2_{d-1}}\big(d-3\big)$. Admittedly, this might be regarded as purely theoretical, since it has to be assumed that $w_{\mathrm{R}} \neq w$ in most practical situations, with $\hat{w}_{\mathrm{M}}$ then being asymptotically equivalent to $\hat{w}_{\mathrm{T}}$ in the sense given above.

So far we have focused on the expected relative losses of the estimators for the GMVP but, as already mentioned, these quantities vanish if the sample size tends to infinity. However,

[4] Actually, the proof of Theorem 5 reveals that (11) can be even strengthened to almost sure convergence.

due to the next theorem it is possible to make statements about the relative loss itself if d is fixed but n tends to infinity.

Theorem 6

Under assumptions A1 to A3 it holds that

$$n \begin{bmatrix} \tau_{\mathrm{T}} \\ \tau_{\mathrm{S}} \\ \tau_{\mathrm{M}} \end{bmatrix} \xrightarrow{\mathrm{d}} \begin{bmatrix} 1 \\ \mathbb{1}_{\{\tau_{\mathrm{R}}=0\}} \left(1 - \frac{d-3}{\chi^2_{d-1}}\right)^2 + \mathbb{1}_{\{\tau_{\mathrm{R}}>0\}} \\ \mathbb{1}_{\{\tau_{\mathrm{R}}=0\}} \left\{\left(1 - \frac{d-3}{\chi^2_{d-1}}\right)^+\right\}^2 + \mathbb{1}_{\{\tau_{\mathrm{R}}>0\}} \end{bmatrix} \chi^2_{d-1}\,, \qquad n \longrightarrow \infty\,.$$

Proof: See the appendix.

This theorem asserts that the relative losses are super-consistent. It is worth pointing out that, even if the expected relative losses of the shrinkage estimators presented here are always smaller than the expected loss of the traditional estimator (which follows from Theorem 3 and Theorem 4), a given realization of τ_{S} may turn out to be greater than τ_{T}. Surprisingly, Theorem 6 implies that, only if $w_{\mathrm{R}} = w$, the probability of this event does not vanish (even asymptotically) but tends to $F_{\chi^2_{d-1}}\big\{(d-3)/2\big\} > 0$. For example, if there exist $d = 5$ assets, this adverse effect occurs with a probability of approximately 9%. However, the same theorem confirms that $\tau_{\mathrm{M}} > \tau_{\mathrm{T}}$ is asymptotically impossible. This is another advantage of the modified shrinkage estimator over the simple one.

As already discussed earlier, it might be criticized that in many practical applications of portfolio theory the number of assets is large compared to the number of observations. In the following we will investigate the asymptotic distribution of the relative loss assuming that $n, d \to \infty$ but $n/d \to q$ with $1 < q \leq \infty$. Here the relative loss of the reference portfolio is assumed to be constant; recall that the number q can be interpreted as the effective sample size. The following theorem asserts that if the asset universe is large, the relative losses of all GMVP estimators are no longer super-consistent.

Theorem 7

Under assumptions A1 to A3 it holds that

$$\tau_{\mathrm{T}} \xrightarrow{\mathrm{a.s.}} \frac{1}{q-1}$$

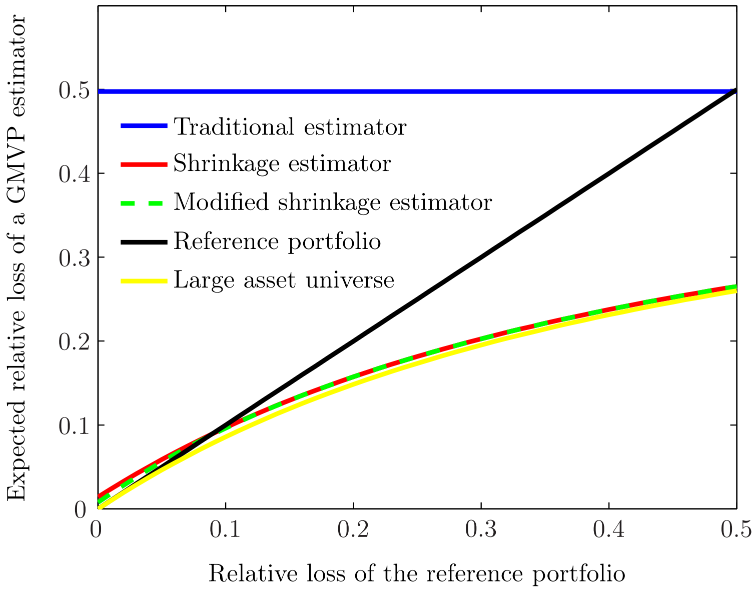

Figure 1: Expected relative losses of the traditional (blue), simple (red) and modified (dashed green) shrinkage estimator for $n = 300$ and $d = 100$ as well as the relative loss of the reference portfolio (black) and the asymptotic loss function $L(\tau_{\mathrm{R}}\,, 3)$ (yellow).

as $n, d \to \infty$ *but* $n/d \to q$ *with* $1 < q \leq \infty$ *. Moreover, concerning the shrinkage estimators for the GMVP it holds that*

$$\kappa_{\mathrm{S}}\,, \kappa_{\mathrm{M}} \xrightarrow{\text{a.s.}} \frac{1}{1 + q\tau_{\mathrm{R}}}$$

as well as

$$\tau_{\mathrm{S}}\,, \tau_{\mathrm{M}} \xrightarrow{\text{a.s.}} L\big(\tau_{\mathrm{R}}, q\big) := \frac{\tau_{\mathrm{R}}}{(1 + q\tau_{\mathrm{R}})^2} + \left(1 - \frac{1}{1 + q\tau_{\mathrm{R}}}\right)^2 \frac{1}{q - 1}$$

as $n, d \to \infty$ *but* $n/d \to q$ *with* $1 < q \leq \infty$ *.*

Proof: See the appendix.

It can be shown that the *asymptotic loss function* L is increasing in τ_{R}, and it holds that $L\big(\tau_{\mathrm{R}}, q\big) < 1/(q - 1)$ whenever $q < \infty$, i.e. the shrinkage estimators dominate the traditional estimator with respect to the asymptotic loss if not only the number of observations but also the number of assets tend to infinity and the effective sample size remains finite. Moreover, it turns out that $L\big(\tau_{\mathrm{R}}, q\big) > \tau_{\mathrm{R}}$ if and only if

$$\tau_{\mathrm{R}} < \frac{1}{q} \cdot \frac{2 - q}{q - 1}\,. \tag{12}$$

	$n \to \infty\,,\ d < \infty$		$n \to \infty\,,\ d \to \infty\,,\ n/d \to q$		
	$q = \infty$		$q < \infty$		$q = \infty$
	$\tau_{\mathrm{R}} = 0$	$\tau_{\mathrm{R}} > 0$	$\tau_{\mathrm{R}} = 0$	$\tau_{\mathrm{R}} > 0$	$\tau_{\mathrm{R}} \geq 0$
τ_{T}	0	0	$\frac{1}{q-1} > 0$	$\frac{1}{q-1} > 0$	0
τ_{S}	0	0	0	$0 < L(\tau_{\mathrm{R}}, q) < \frac{1}{q-1}$	0
τ_{M}	0	0	0	$0 < L(\tau_{\mathrm{R}}, q) < \frac{1}{q-1}$	0
$n\tau_{\mathrm{T}}$	χ^2_{d-1}	χ^2_{d-1}	∞	∞	∞
$n\tau_{\mathrm{S}}$	$\left(1 - \frac{d-3}{\chi^2_{d-1}}\right)^2 \chi^2_{d-1}$	χ^2_{d-1}	0	∞	∞
$n\tau_{\mathrm{M}}$	$\left\{\left(1 - \frac{d-3}{\chi^2_{d-1}}\right)^+\right\}^2 \chi^2_{d-1}$	χ^2_{d-1}	0	∞	∞

Table 1: Large-sample properties of the relative losses of $\hat{w}_{\mathrm{T}}$, $\hat{w}_{\mathrm{S}}$, and $\hat{w}_{\mathrm{M}}$.

Therefore, the shrinkage estimators dominate the reference portfolio *uniformly* if $q \geq 2$ (see Figure 1). Conversely, in terms of the asymptotic loss they become uniformly worse than w_{R} as q tends to 1 from above, since the right-hand side of (12) then tends to infinity. The large-sample properties of the relative losses of the GMVP estimators $\hat{w}_{\mathrm{T}}$, $\hat{w}_{\mathrm{S}}$, and $\hat{w}_{\mathrm{M}}$ are summarized in Table 1.

3.3 The Link to Covariance Matrix Estimation

Jagannathan and Ma (2003) analyze short-sales constraints as a means of lessening the impact of estimation errors on the sample covariance matrix. They show that using short-sales constraints is equivalent to transforming the sample covariance matrix and taking this quantity for calculating the GMVP on the basis of the unconstrained traditional estimator for the GMVP. The following theorem states that a similar argument holds for the shrinkage estimators presented earlier.

Theorem 8

For any reference portfolio w_{R} there exists a positive-definite $d \times d$ matrix Σ_{R}^{-1} such that $w_{\mathrm{R}} \propto \Sigma_{\mathrm{R}}^{-1}\mathbf{1}$ as well as $\mathbf{1}'\Sigma_{\mathrm{R}}^{-1}\mathbf{1} = \mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}$, where $\widehat{\Sigma}$ is the sample covariance matrix given by Eq. 2 and it is assumed that $n > d$. The shrinkage estimators for the GMVP can be calculated by using

$$\widehat{\Sigma}_{\mathrm{S}}^{-1} := \kappa_{\mathrm{S}}\Sigma_{\mathrm{R}}^{-1} + \left(1 - \kappa_{\mathrm{S}}\right)\widehat{\Sigma}^{-1} \qquad \text{and} \qquad \widehat{\Sigma}_{\mathrm{M}}^{-1} := \kappa_{\mathrm{M}}\Sigma_{\mathrm{R}}^{-1} + \left(1 - \kappa_{\mathrm{M}}\right)\widehat{\Sigma}^{-1}$$

for the traditional GMVP estimator, i.e.

$$\hat{w}_{\mathrm{S}} = \frac{\widehat{\Sigma}_{\mathrm{S}}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\Sigma}_{\mathrm{S}}^{-1}\mathbf{1}} \qquad \textit{and} \qquad \hat{w}_{\mathrm{M}} = \frac{\widehat{\Sigma}_{\mathrm{M}}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\Sigma}_{\mathrm{M}}^{-1}\mathbf{1}}\,.$$

Proof: See the appendix.

The random matrices $\widehat{\Sigma}_{\mathrm{S}}$ and $\widehat{\Sigma}_{\mathrm{M}}$ can be interpreted as shrinkage estimators for the unknown covariance matrix Σ. However, $\widehat{\Sigma}_{\mathrm{M}}$ is positive-definite, a trait that does not hold for $\widehat{\Sigma}_{\mathrm{S}}$ in general. Any other matrix which is proportional to $\widehat{\Sigma}_{\mathrm{S}}$ or $\widehat{\Sigma}_{\mathrm{M}}$ would lead to the same shrinkage estimators for the GMVP, but the expressions given in Theorem 8 satisfy a convenient scaling condition, i.e. $\mathbf{1}'\widehat{\Sigma}_{\mathrm{S}}^{-1}\mathbf{1} = \mathbf{1}'\widehat{\Sigma}_{\mathrm{M}}^{-1}\mathbf{1} = \mathbf{1}'\Sigma_{\mathrm{R}}^{-1}\mathbf{1} = \mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1} = 1/\hat{\sigma}_{\mathrm{T}}^2$.

Similar shrinkage estimators for the covariance matrix have been already suggested by Ledoit and Wolf (2001, 2003). However, the estimators given in Theorem 8 differ from the estimators introduced by Ledoit and Wolf in two aspects:

1. Their shrinkage constants depend on unobservable quantities which have to be estimated from empirical data. Even if the suggested covariance matrix estimators dominate the sample covariance matrix asymptotically, it is not clear why the dominance result should be valid in small samples. By contrast, our shrinkage approach focuses on the small-sample properties of the resulting portfolio weights.

2. Ledoit and Wolf shrink the covariance matrix itself, whereas our approach is based on shrinking its inverse. By shrinking the covariance matrix, it is possible to allow for $n \leq d$, i.e. the aforementioned authors are able to apply their approach to asset universes where the number of assets exceed the number of observations.

So far our methodology consists of shrinking the traditional GMVP estimator towards some non-stochastic reference portfolio w_{R}. However, all the presented results remain valid if w_{R} is a stochastic portfolio satisfying the budget constraint and being stochastically independent of the historical observations which are used for calculating $\hat{w}_{\mathrm{T}}$.[5] Nevertheless, in the following we will concentrate on the special case $w_{\mathrm{R}} = w_{\mathrm{N}} = \mathbf{1}/d$.

[5]For example, w_{R} could be interpreted as a portfolio which has been suggested by a layman.

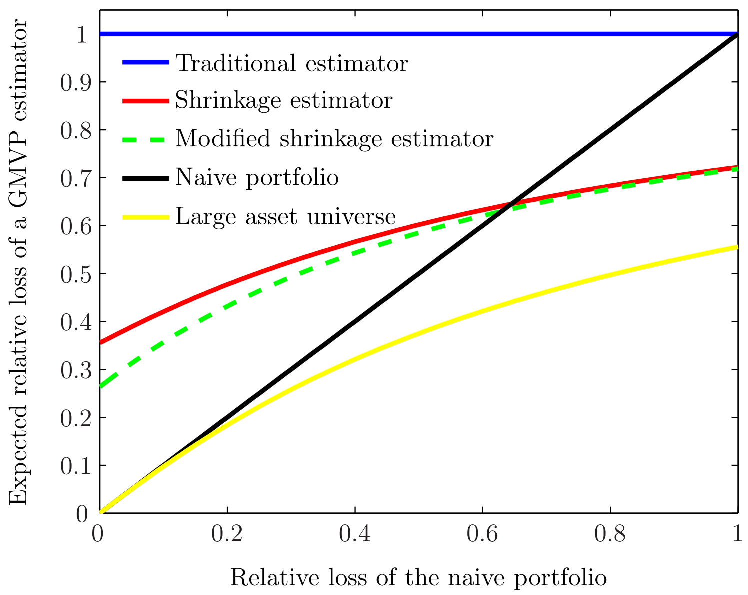

Figure 2: Expected relative losses of the traditional (blue), simple (red) and modified shrinkage (dashed green) estimator for $n = 20$ and $d = 10$ as well as the relative loss of the naive portfolio (black) and the asymptotic loss function $L(\tau_{\mathrm{R}}, q)$ with $q = 2$ (yellow).

4 Naive Diversification vs. Portfolio Optimization

4.1 A Small-Sample Simulation Study

DeMiguel et al. (2007) raise the question of whether optimizing a portfolio using time series information is worthwhile to begin with. They do not even refer to the fact that asset returns typically exhibit structural breaks, serial correlations in the higher moments, and heavy tails. According to these authors, the estimation error outweighs the potential gain of portfolio optimization, even if the asset returns are normally distributed and serially independent. In this section we address a similar question: Does it pay to strive for the GMVP by using time series information or is it better to renounce parameter estimation altogether and put the money straight away into the naive portfolio?

In order to revisit this question, we may focus on the expected relative loss which is caused by a given GMVP estimator. Due to Theorem 4 and the arguments given in Section 3.2, we will concentrate on the modified shrinkage estimator $\hat{w}_{\mathrm{M}}$ and choose the naive portfolio w_{N} as a reference portfolio. Although closed-form expressions for τ_{M} in large samples and asset universes have been already presented in Section 3.2, the relative loss can only be simulated, e.g. by using Equations 8 and 10, if the sample is small. Figure 2 contains the expected relative losses of the four different portfolio strategies, i.e. naive diversification, traditional estimation, as well as simple and modified shrinkage estimation for $n = 20$ observations and $d = 10$ assets. The x-axis denotes the relative loss τ_{N} of the naive portfolio, whereas the y-axis accounts for the expected relative losses of the different portfolio strategies depending on τ_{N}. Note that (according to Theorem 3) the expected relative loss of the traditional estimator does not depend on τ_{N} but only on the number n of observations and the number d of assets.

It can be seen that the expected relative loss of the traditional estimator corresponds to 100%. Due to Theorem 3 and Theorem 4 it is clear that the expected relative losses of the shrinkage estimators are always below the expected relative loss of the traditional estimator. This is also confirmed by Figure 2. Particularly if τ_{N} is small, i.e. the true GMVP does not differ too greatly from the naive portfolio (which serves as an anchor point for $\hat{w}_{\mathrm{S}}$ and $\hat{w}_{\mathrm{M}}$), the shrinkage estimators are more favorable than the traditional estimator.

Figure 2 also indicates the *critical relative loss* τ_{N}^* of the naive portfolio with respect to the modified shrinkage estimator $\hat{w}_{\mathrm{M}}$. This is that point on the x-axis where the modified shrinkage estimator leads to the same expected relative loss as naive diversification. As indicated by Figure 2, this critical value is about 63%. For example if there are 5 years of quarterly asset returns and 10 stocks on the market, naive diversification would be better as long as $\tau_{\mathrm{N}} < 63\%$. Suppose that the standard deviation of the GMVP return corresponds to $\sigma = 10\%$, whereas its counterpart related to the naive portfolio amounts to 11% (per quarter). In that case, the relative loss of naive diversification is $\tau_{\mathrm{N}} = (0.11/0.10)^2 - 1 = 21\%$, whereas the expected relative loss caused by the modified shrinkage estimator roughly amounts to $\mathrm{E}(\tau_{\mathrm{M}}) = 43\%$. Therefore, it would not pay to use the modified shrinkage estimator in that case. In contrast, if the naive portfolio leads to a standard deviation of 13%, it holds that $\tau_{\mathrm{N}} = (0.13/0.10)^2 - 1 = 69\% > \tau_{\mathrm{N}}^*$ and so the modified shrinkage estimator is slightly better than the naive portfolio. Note that traditional estimation is always worse than naive diversification in all such cases.

Table 2 lists some critical relative losses of naive diversification for different combinations of n and d. For example, if 10 years of monthly asset return observations are available (i.e. $n = 120$) and the stock market consists of $d = 50$ assets, one should use the modified

$n \setminus d$	5	10	25	50	100
12	52% (550%)	847% (99261%)	—	—	—
24	16% (111%)	40% (334%)	—	—	—
36	9% (59%)	19% (132%)	152% (1809%)	—	—
60	5% (30%)	9% (58%)	28% (209%)	420% (7806%)	—
120	2% (13%)	4% (24%)	8% (57%)	21% (161%)	377% (5202%)

Table 2: Critical relative losses of the naive portfolio with respect to the modified shrinkage estimator for different combinations of n and d. The parentheses under the critical relative losses contain the critical thresholds of $\hat{\tau}_{\mathrm{N}}$ for testing the naive diversification hypothesis at a significance level of $\alpha = 5\%$.

shrinkage estimator if and only if the variance of the naive portfolio return is at least 21% greater than the variance of the GMVP return. Depending on the length of the time series and the number of assets, the modified shrinkage estimator is able to reduce the relative loss of naive diversification. However, the table also indicates that, if the number of assets is large compared to the number of observations, naive diversification is apparently the best strategy, which reconfirms the naive diversification hypothesis of DeMiguel et al. (2007).

4.2 Testing the Naive Diversification Hypothesis

For applying the decision rule discussed above, one needs two numbers, i.e.

1. the critical relative loss of the naive portfolio with respect to the modified shrinkage estimator and
2. the relative loss of the naive portfolio.

The critical relative loss can be calculated by Monte Carlo simulation (as it was done to obtain Table 2), whereas the actual relative loss of the naive portfolio is not observable and needs to be estimated from the history. The next theorem provides the distribution of its empirical counterpart $\hat{\tau}_{\mathrm{N}}$ or, more generally, $\hat{\tau}_{\mathrm{R}}$ (see also Theorem 2).

Theorem 9

Under assumptions A1 to A3 and $n > d$ *, the estimator* $\hat{\tau}_{\mathrm{R}} = \left(\hat{\sigma}^2_{\mathrm{R}} - \hat{\sigma}^2_{\mathrm{T}}\right)/\hat{\sigma}^2_{\mathrm{T}}$ *for the relative loss of the reference portfolio is conditionally noncentrally* F*-distributed, more precisely*

$$\hat{\tau}_{\mathrm{R}} \sim \frac{d-1}{n-d} \cdot F_{d-1,n-d}\left(\tau_{\mathrm{R}}\chi^2_{n-1}/2\right).$$

Proof: See the appendix.

With Theorem 9, it is possible to test whether one should invest in the naive portfolio or to apply a GMVP estimator, i.e.

$H_0\colon\ \tau_{\mathrm{N}} \leq \tau^*_{\mathrm{N}}$ vs.

$H_1\colon\ \tau_{\mathrm{N}} > \tau^*_{\mathrm{N}}$.

The test statistic is given by $\hat{\tau}_{\mathrm{N}} = \left(\hat{\sigma}^2_{\mathrm{N}} - \hat{\sigma}^2_{\mathrm{T}}\right)/\hat{\sigma}^2_{\mathrm{T}}$ and according to Theorem 9, H_0 can be rejected whenever the realization of $\hat{\tau}_{\mathrm{N}}$ exceeds the upper α-quantile ($0 < \alpha < \frac{1}{2}$) of the cumulative distribution function of

$$\frac{d-1}{n-d} \cdot F_{d-1,n-d}\left(\tau^*_{\mathrm{N}}\chi^2_{n-1}/2\right),$$

which can be also calculated by Monte Carlo simulation.[6]

Critical thresholds for this hypothesis test at a significance level of $\alpha = 5\%$ are presented in Table 2. For instance, suppose that the asset universe consists of 50 assets and the investor can observe 10 years of monthly asset returns. Then the naive diversification hypothesis can be only rejected if $\hat{\tau}_{\mathrm{N}} > 161\%$. Note that this is by far greater than the theoretical value of the critical relative loss $\tau^*_{\mathrm{N}} = 21\%$, since the distribution of $\hat{\tau}_{\mathrm{N}}$ is considerably skewed to the right.

We consciously formulate the hypothesis test in such a way that the naive portfolio has to be rejected but not the portfolio based on some GMVP estimator. Therefore, for typical significance levels like $\alpha = 1\%, 5\%, 10\%$, our decision rule favors naive diversification. More precisely, if H_0 can be rejected, the considered GMVP estimator significantly leads to a better out-of-sample performance but if H_0 is not rejected, from a statistical point of view it cannot be assumed that naive diversification is better. However, in that case the naive

[6] This hypothesis test can be adapted to any GMVP estimator if its expected relative loss $\mathrm{E}(\tau) < \infty$ depends only on n, d, and τ_{N} and provided $\tau_{\mathrm{N}} \mapsto \mathrm{E}(\tau)$ has only one intersection point with $\tau_{\mathrm{N}} \mapsto \tau_{\mathrm{N}}$.

portfolio can be justified either empirically, e.g. because of the well-known stylized facts of financial data, or due to the arguments given by DeMiguel et al. (2007). In other words: if it is not possible to guarantee that a statistical method will lead to a better result but it is likely that the outcome will become worse, the naive portfolio can be justified by the principle of insufficient reason (against naive diversification).

5 Conclusion

We present two shrinkage estimators for the GMVP that dominate the traditional estimator under the assumption of serially independent and identically normally distributed asset returns. Their small-sample and their large-sample properties alike have been investigated. The presented shrinkage estimators considerably reduce the out-of-sample variance of the portfolio return compared to the traditional estimator, especially if the asset universe is large. In addition, we provide a hypothesis test to decide whether one should invest in a portfolio based on estimators for the GMVP or in the naive portfolio. This decision depends only on three quantities: the number of observations, the number of assets, and the relative loss (compared to the GMVP) caused by naive diversification. Further research could include, for instance, an empirical investigation of the presented shrinkage estimators.

Appendix

Lemma 2

For any $\lambda \geq 0$ it holds that

$$\mathrm{E}\Big\{\chi_q^{-2}(\lambda)\Big\} = q\,\mathrm{E}\Big\{\chi_{q+2}^{-4}(\lambda)\Big\} + 2\lambda\,\mathrm{E}\Big\{\chi_{q+4}^{-4}(\lambda)\Big\}\,, \tag{13}$$

and if $q \geq 3$,

$$(q-2)\,\mathrm{E}\Big\{\chi_q^{-2}(\lambda)\Big\} = (q-2\lambda)\,\mathrm{E}\Big\{\chi_{q+2}^{-2}(\lambda)\Big\} + 2\lambda\,\mathrm{E}\Big\{\chi_{q+4}^{-2}(\lambda)\Big\}\,. \tag{14}$$

Proof: Eq. 13 follows immediately from Theorem 2 in Judge and Bock (1978, p. 322) by setting $\phi(x) = x^{-2}$, $A = I_q$, and $\theta \in \mathbb{R}^q$ such that $\lambda = \theta'\theta/2$. Similarly, with $\phi(x) = x^{-1}$,

$$1 = q\,\mathrm{E}\Big\{\chi_{q+2}^{-2}(\lambda)\Big\} + 2\lambda\,\mathrm{E}\Big\{\chi_{q+4}^{-2}(\lambda)\Big\} = (q-2)\,\mathrm{E}\Big\{\chi_q^{-2}(\lambda)\Big\} + 2\lambda\,\mathrm{E}\Big\{\chi_{q+2}^{-2}(\lambda)\Big\}$$

for any $q \geq 3$, which leads to (14). Q.E.D.

Lemma 3

Consider a $q \times q$ *random matrix* $V \sim W_q(I_q, m)$ *with* $q \geq 3$ *and* $m \geq q+2$. *Further, define* $\lambda := \theta'\theta/2$ *and* $\hat{\lambda} := \theta' V \theta/2$ *for some* $\theta \in \mathbb{R}^q$. *Then it holds that*

$$\mathrm{E}\left[\left(\operatorname{tr} V^{-1} - \frac{\lambda}{\hat{\lambda}} \cdot q\right) \mathrm{E}\left\{\chi_{q+2}^{-2}(\hat{\lambda}) \,|\, V\right\}\right] = \frac{q-1}{m-q-1} \cdot \mathrm{E}\left[(q-2) \cdot \frac{\lambda}{\hat{\lambda}} \cdot \mathrm{E}\left\{\chi_q^{-2}(\hat{\lambda}) \,|\, V\right\}\right]$$

and

$$\begin{aligned}\mathrm{E}\left[\left(\operatorname{tr} V^{-1} - \frac{\lambda}{\hat{\lambda}} \cdot q\right) \mathrm{E}\left\{\chi_{q+2}^{-4}(\hat{\lambda}) \,|\, V\right\}\right] &= \frac{q-1}{m-q-1} \cdot \mathrm{E}\left[\frac{\lambda}{\hat{\lambda}} \cdot \mathrm{E}\left\{\chi_q^{-2}(\hat{\lambda}) \,|\, V\right\}\right] - \\ &\quad \frac{q-1}{m-q-1} \cdot \mathrm{E}\left[2\lambda\, \mathrm{E}\left\{\chi_{q+2}^{-4}(\hat{\lambda}) \,|\, V\right\}\right].\end{aligned}$$

Proof: Consider the function $h(2\hat{\lambda}) = \mathrm{E}\{\chi_{q+2}^{-2}(\hat{\lambda}) \,|\, V\}$ and note that, after rotating θ, it holds that $2\hat{\lambda} = \theta'\theta\chi^2$ for some random variable $\chi^2 \sim \chi_m^2$. Then, due to Theorem 6 in Judge and Bock (1978, p. 324),

$$\mathrm{E}\left\{\left(\operatorname{tr} V^{-1}\right) h(2\hat{\lambda})\right\} = \frac{q\,(m-2)}{m-q-1} \cdot \mathrm{E}\left\{\frac{h(2\hat{\lambda})}{\chi^2}\right\} + \frac{2\,(q-1)}{m-q-1} \cdot \mathrm{E}\left\{\theta'\theta\, h'(2\hat{\lambda})\right\},$$

where h' denotes the first derivative of h with respect to $2\hat{\lambda}$. Since $\lambda/\hat{\lambda} = 1/\chi^2$,

$$\mathrm{E}\left\{\left(\operatorname{tr} V^{-1} - \frac{\lambda}{\hat{\lambda}} \cdot q\right) h(2\hat{\lambda})\right\} = \frac{q-1}{m-q-1} \cdot \left[q\, \mathrm{E}\left\{\frac{h(2\hat{\lambda})}{\chi^2}\right\} + 2\,\theta'\theta\, \mathrm{E}\left\{h'(2\hat{\lambda})\right\}\right], \qquad (15)$$

where

$$h'(2\hat{\lambda}) = \frac{1}{2} \cdot \frac{\mathrm{dE}\{\chi_{q+2}^{-2}(\hat{\lambda}) \,|\, V\}}{\mathrm{d}\hat{\lambda}} = \frac{1}{2} \cdot \left[\mathrm{E}\{\chi_{q+4}^{-2}(\hat{\lambda}) \,|\, V\} - \mathrm{E}\{\chi_{q+2}^{-2}(\hat{\lambda}) \,|\, V\}\right],$$

which follows from the derivative rule on page 327 in Judge and Bock (1978). After substituting $h'(2\hat{\lambda})$ in (15) and some re-arrangement, we obtain

$$\begin{aligned}&\mathrm{E}\left[\left(\operatorname{tr} V^{-1} - \frac{\lambda}{\hat{\lambda}} \cdot q\right) \mathrm{E}\left\{\chi_{q+2}^{-2}(\hat{\lambda}) \,|\, V\right\}\right] = \\ &\qquad \frac{q-1}{m-q-1} \cdot \mathrm{E}\left[\frac{\lambda}{\hat{\lambda}}\left[(q-2\hat{\lambda})\, \mathrm{E}\left\{\chi_{q+2}^{-2}(\hat{\lambda}) \,|\, V\right\} + 2\hat{\lambda}\mathrm{E}\left\{\chi_{q+4}^{-2}(\hat{\lambda}) \,|\, V\right\}\right]\right].\end{aligned}$$

Now the first statement of the lemma appears immediately after applying (14). Similarly, by allowing for the function $h(2\hat{\lambda}) = \mathrm{E}\{\chi_{q+2}^{-4}(\hat{\lambda}) \,|\, V\}$ and using (13), the second statement of the lemma becomes valid. Q.E.D.

Proof of Theorem 1

The loss function $\mathcal{L}_{\omega,\Omega}$ can be re-formulated as

$$\mathcal{L}_{\omega,\Omega}(\hat{\omega}) = (\hat{\omega}-\omega)'\Omega\,(\hat{\omega}-\omega) = (\hat{\theta}-\theta)'(\hat{\theta}-\theta) = \mathcal{L}_{\theta}(\hat{\theta})\,,$$

where $\hat{\theta} := \Omega^{\frac{1}{2}}(\hat{\omega}-x)$ and $\theta := \Omega^{\frac{1}{2}}(\omega-x)$. Accordingly, the random vector X is transformed into $Y := \Omega^{\frac{1}{2}}(X-x)\,|\,V \sim \mathcal{N}_q\big(\theta, V^{-1}\big)$ with $V := \Omega^{-\frac{1}{2}}W\Omega^{-\frac{1}{2}} \sim W_q\big(I_q, m\big)$ and similarly

$$Y_{\mathrm{S}} := \Omega^{\frac{1}{2}}(X_{\mathrm{S}}-x) = \Big(1-\frac{c\,\chi^2}{Y'VY}\Big)Y\,.$$

After some elementary transformations, it turns out that

$$\mathcal{L}_{\theta}(Y_S) = \mathcal{L}_{\theta}(Y) - \left\{2c\chi^2\cdot\frac{Y'(Y-\theta)}{Y'VY} - c^2\chi^4\cdot\frac{Y'Y}{(Y'VY)^2}\right\}.$$

This means the random variable Y_S dominates Y if and only if

$$\mathrm{E}\big\{\mathcal{L}_{\theta}(Y)-\mathcal{L}_{\theta}(Y_S)\big\} = 2ck\mathcal{E}_1 - c^2k\,(k+2)\,\mathcal{E}_2 > 0\,, \tag{16}$$

where

$$\mathcal{E}_1 := \mathrm{E}\left\{\frac{Y'(Y-\theta)}{Y'VY}\right\} \qquad \text{and} \qquad \mathcal{E}_2 := \mathrm{E}\left\{\frac{Y'Y}{(Y'VY)^2}\right\}.$$

Hence, the dominance result is satisfied for all c with $0 < c < 2/(k+2)\cdot\mathcal{E}_1/\mathcal{E}_2$ and, to prove the theorem, it has to be shown that $\mathcal{E}_1/\mathcal{E}_2 \geq (q-2)$. Now we define $Z := V^{\frac{1}{2}}Y$ and $\zeta := V^{\frac{1}{2}}\theta$ so that $Z\,|\,V \sim \mathcal{N}_q(\zeta, I_q)$. Then it holds that

$$\frac{Y'(Y-\theta)}{Y'VY}\,|\,V \sim \frac{Z'V^{-1}(Z-\zeta)}{Z'Z}\,|\,V \qquad \text{and} \qquad \frac{Y'Y}{(Y'VY)^2}\,|\,V \sim \frac{Z'V^{-1}Z}{(Z'Z)^2}\,|\,V\,.$$

By setting $\phi(x) = x^{-1}$ in Theorem 1 and Theorem 2 of Judge and Bock (1978, pp. 321–322) and allowing for $\lambda = \theta'\theta/2$ and $\hat{\lambda} = \theta'V\theta/2$ it follows that

$$\mathrm{E}\left\{\frac{Y'(Y-\theta)}{Y'VY}\,|\,V\right\} = \big(\mathrm{tr}\,V^{-1}\big)\mathrm{E}\big\{\chi^{-2}_{q+2}(\hat{\lambda})\,|\,V\big\} + 2\lambda\,\mathrm{E}\big\{\chi^{-2}_{q+4}(\hat{\lambda})\,|\,V\big\} - 2\lambda\,\mathrm{E}\big\{\chi^{-2}_{q+2}(\hat{\lambda})\,|\,V\big\}.$$

Similarly, by setting $\phi(x) = x^{-2}$ in Theorem 2 given by Judge and Bock (1978, p. 322), we find that

$$\mathrm{E}\left\{\frac{Y'Y}{(Y'VY)^2}\,|\,V\right\} = \big(\mathrm{tr}\,V^{-1}\big)\mathrm{E}\big\{\chi^{-4}_{q+2}(\hat{\lambda})\,|\,V\big\} + 2\lambda\,\mathrm{E}\big\{\chi^{-4}_{q+4}(\hat{\lambda})\,|\,V\big\}.$$

After some re-arrangement and an application of (14) we obtain

$$\mathrm{E}\left(\frac{Y'(Y-\theta)}{Y'VY}\,|\,V\right) = (q-2)\cdot\frac{\lambda}{\hat{\lambda}}\cdot\mathrm{E}\big\{\chi^{-2}_{q}(\hat{\lambda})\,|\,V\big\} + \Big(\mathrm{tr}\,V^{-1}-\frac{\lambda}{\hat{\lambda}}\cdot q\Big)\,\mathrm{E}\big\{\chi^{-2}_{q+2}(\hat{\lambda})\,|\,V\big\}.$$

Moreover, with an application of (13) it also turns out that

$$\mathrm{E}\left(\frac{Y'Y}{(Y'VY)^2}\,|\,V\right) = \frac{\lambda}{\hat\lambda}\cdot \mathrm{E}\left\{\chi_q^{-2}(\hat\lambda)\,|\,V\right\} + \left(\operatorname{tr} V^{-1} - \frac{\lambda}{\hat\lambda}\cdot q\right)\mathrm{E}\left\{\chi_{q+2}^{-4}(\hat\lambda)\,|\,V\right\}.$$

Now, from Lemma 3 it follows that $\mathcal{E}_1 = (q-2)\,\mathcal{E}_2 + \varepsilon$ with

$$\varepsilon := \frac{(q-1)(q-2)}{m-q-1}\cdot 2\lambda\,\mathrm{E}\left[\mathrm{E}\left\{\chi_{q+2}^{-4}(\hat\lambda)\,|\,V\right\}\right] \geq 0\,.$$

Since $\mathcal{E}_1 \geq (q-2)\,\mathcal{E}_2$ with $\mathcal{E}_2 > 0$ it follows that $\mathcal{E}_1/\mathcal{E}_2 \geq (q-2)$. For $x = \omega$ it holds that $\lambda = 0$ and thus $\mathcal{E}_1 = (q-2)\,\mathcal{E}_2$. This means the optimal constant c of the quadratic function given by (16) does not depend on $\mathcal{E}_1$ or $\mathcal{E}_2$. Further, it is unique and corresponds to $c = (q-2)/(k+2)$. Q.E.D.

Proof of Theorem 2

Lemma 1 and Theorem 1 can be brought together by the following substitutions: $m = n-1$, $q = d-1$, $W = n\,\widehat\Omega/\sigma^2$, $X = \hat w_{\mathrm{T}}^{\mathrm{ex}}$, $\chi^2 = n\,\hat\sigma_{\mathrm{T}}^2/\sigma^2$, $k = n-d$, and $x = w_{\mathrm{R}}^{\mathrm{ex}}$. Then the constant

$$c = \frac{q-2}{k+2} = \frac{d-3}{n-d+2}$$

leads to a dominating shrinkage estimator $\hat w_{\mathrm{S}}^{\mathrm{ex}}$ for w^{ex}, viz

$$\hat w_{\mathrm{S}}^{\mathrm{ex}} = w_{\mathrm{R}}^{\mathrm{ex}} + \left(1 - \frac{d-3}{n-d+2}\cdot\frac{\hat\sigma_{\mathrm{T}}^2}{(\hat w_{\mathrm{T}}^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}})'\widehat\Omega\,(\hat w_{\mathrm{T}}^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}})}\right)\big(\hat w_{\mathrm{T}}^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}}\big)\,.$$

Note that

$$(\hat w_{\mathrm{T}}^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}})'\widehat\Omega\,(\hat w_{\mathrm{T}}^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}}) = (\hat w_{\mathrm{T}} - w_{\mathrm{R}})'\widehat\Sigma\,(\hat w_{\mathrm{T}} - w_{\mathrm{R}})$$

and thus

$$\frac{\hat\sigma_{\mathrm{T}}^2}{(\hat w_{\mathrm{T}}^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}})'\widehat\Omega\,(\hat w_{\mathrm{T}}^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}})} = \frac{\hat\sigma_{\mathrm{T}}^2}{(\hat w_{\mathrm{T}} - w_{\mathrm{R}})'\widehat\Sigma\,(\hat w_{\mathrm{T}} - w_{\mathrm{R}})} = \frac{\hat\sigma_{\mathrm{T}}^2}{\hat\sigma_{\mathrm{R}}^2 - \hat\sigma_{\mathrm{T}}^2} = \frac{1}{\hat\tau_{\mathrm{R}}}\,.$$

Due to $\hat w_{\mathrm{S}} = \mathbf{e}_1 - \Delta'\hat w_{\mathrm{S}}^{\mathrm{ex}}$ it follows that

$$\hat w_{\mathrm{S}} = w_{\mathrm{R}} + \left(1 - \frac{d-3}{n-d+2}\cdot\frac{1}{\hat\tau_{\mathrm{R}}}\right)\big(\hat w_{\mathrm{T}} - w_{\mathrm{R}}\big) = \kappa_{\mathrm{S}}\,w_{\mathrm{R}} + \big(1-\kappa_{\mathrm{S}}\big)\,\hat w_{\mathrm{T}}\,.$$

Q.E.D.

Proof of Theorem 3

After some calculations we find that

$$\tau_{\mathrm{S}} = \tau_{\mathrm{R}} - 2\big(1-\kappa_{\mathrm{S}}\big)a + \big(1-\kappa_{\mathrm{S}}\big)^2 b\,,$$

where

$$\kappa_{\mathrm{S}} = \frac{d-3}{n-d+2}\cdot\frac{n\hat{\sigma}_{\mathrm{T}}^2/\sigma^2}{(\hat{w}_{\mathrm{T}}^{\mathrm{ex}}-w_{\mathrm{R}}^{\mathrm{ex}})'(n\widehat{\Omega}/\sigma^2)(\hat{w}_{\mathrm{T}}^{\mathrm{ex}}-w_{\mathrm{R}}^{\mathrm{ex}})}\,,$$

$$a = \frac{(\hat{w}_{\mathrm{T}}^{\mathrm{ex}}-w_{\mathrm{R}}^{\mathrm{ex}})'\Omega\,(w^{\mathrm{ex}}-w_{\mathrm{R}}^{\mathrm{ex}})}{\sigma^2} \qquad \text{and} \qquad b = \frac{(\hat{w}_{\mathrm{T}}^{\mathrm{ex}}-w_{\mathrm{R}}^{\mathrm{ex}})'\Omega\,(\hat{w}_{\mathrm{T}}^{\mathrm{ex}}-w_{\mathrm{R}}^{\mathrm{ex}})}{\sigma^2}\,.$$

With $\theta = \Omega^{\frac{1}{2}}/\sigma\,(w^{\mathrm{ex}}-w_{\mathrm{R}}^{\mathrm{ex}})$, $\xi \sim \mathcal{N}_{d-1}(\mathbf{0}, I_{d-1})$, and $V \sim W_{d-1}(I_{d-1}, n-1)$, the shrinkage constant κ_{S} can be represented by

$$\kappa_{\mathrm{S}} = \frac{d-3}{n-d+2}\cdot\frac{\chi^2_{n-d}}{\big(\theta+V^{-\frac{1}{2}}\xi\big)'V\big(\theta+V^{-\frac{1}{2}}\xi\big)}$$

as well as $a = \theta'\big(\theta+V^{-\frac{1}{2}}\xi\big)$ and $b = \big(\theta+V^{-\frac{1}{2}}\xi\big)'\big(\theta+V^{-\frac{1}{2}}\xi\big)$, where ξ , V, and χ^2_{n-d} are mutually independent. Hence, τ_{S} is equal to the expression given on the right hand side of (8). Moreover, it holds that

$$\tau_{\mathrm{S}} = \big\|\mathcal{O}\big\{\kappa_{\mathrm{S}}\theta - \big(1-\kappa_{\mathrm{S}}\big)V^{-\frac{1}{2}}\xi\big\}\big\|^2 = \big\|\kappa_{\mathrm{S}}\eta - \big(1-\kappa_{\mathrm{S}}\big)\mathcal{O}V^{-\frac{1}{2}}\xi\big\|^2$$

with $\eta := \mathcal{O}\theta$ for any orthogonal $(d-1)\times(d-1)$ matrix $\mathcal{O}$; note also that κ_{S} is a function of $V^{-\frac{1}{2}}\xi$ only through the quadratic form

$$\big(\theta+V^{-\frac{1}{2}}\xi\big)'V\big(\theta+V^{-\frac{1}{2}}\xi\big) = \big(\eta+\mathcal{O}V^{-\frac{1}{2}}\xi\big)'\big(\mathcal{O}V\mathcal{O}'\big)\big(\eta+\mathcal{O}V^{-\frac{1}{2}}\xi\big)\,.$$

The random matrix V has a radial distribution, i.e. $\mathcal{O}V\mathcal{O}' \sim V$ as well as $\mathcal{O}V^{-1}\mathcal{O}' \sim V^{-1}$. Similarly, ξ has a spherical distribution, i.e. $\mathcal{O}\xi \sim \xi$. It follows that $\mathcal{O}V^{-\frac{1}{2}}\mathcal{O}' \sim V^{-\frac{1}{2}}$ and thus $\mathcal{O}V^{-\frac{1}{2}}\xi \sim V^{-\frac{1}{2}}\xi$. This means for any rotation η of θ it holds that

$$\tau_{\mathrm{S}} \sim \big\|\kappa_{\mathrm{S}}\eta - \big(1-\kappa_{\mathrm{S}}\big)V^{-\frac{1}{2}}\xi\big\|^2\,.$$

Ergo, the distribution of τ_{S} depends only on n, d, and $\tau_{\mathrm{R}} = \theta'\theta$. Q.E.D.

Proof of Theorem 4

From the proof of Theorem 3 it follows that the distribution of τ_{M}, too, is only a function of n, d, and τ_{R}. To prove that $\mathrm{E}(\tau_{\mathrm{M}}) < \mathrm{E}(\tau_{\mathrm{S}})$, the relative loss of the simple shrinkage estimator can be written as

$$\tau_{\mathrm{S}} = \tau_{\mathrm{R}} - 2\,\theta' V^{-\frac{1}{2}}\big(1-\kappa_{\mathrm{S}}\big)\big(V^{\frac{1}{2}}\theta + \xi\big) + \big(1-\kappa_{\mathrm{S}}\big)^2\,\|V^{\frac{1}{2}}\theta + \xi\|_V^2\,.$$

Since $\big(1-\kappa_{\mathrm{S}}\big) = \big(1-\kappa_{\mathrm{S}}\big)^+ - \big(1-\kappa_{\mathrm{S}}\big)^-$, the relative loss of the modified shrinkage estimator becomes

$$\tau_{\mathrm{M}} = \tau_{\mathrm{S}} - 2\,\theta' V^{-\frac{1}{2}}\big(1-\kappa_{\mathrm{S}}\big)^-\big(V^{\frac{1}{2}}\theta + \xi\big) - \big\{\big(1-\kappa_{\mathrm{S}}\big)^-\big\}^2\,\|V^{\frac{1}{2}}\theta + \xi\|_V^2\,.$$

Here it holds that

$$\mathrm{E}\Big[\big\{\big(1-\kappa_{\mathrm{S}}\big)^-\big\}^2\,\|V^{\frac{1}{2}}\theta + \xi\|_V^2\Big] > 0$$

and from Theorem 1 given by Judge and Bock (1978, pp. 321) it follows that

$$\mathrm{E}\Big\{\theta' V^{-\frac{1}{2}}\big(1-\kappa_{\mathrm{S}}\big)^-\big(V^{\frac{1}{2}}\theta + \xi\big)\Big\} = \tau_{\mathrm{R}}\mathrm{E}\left[\left\{1 - \frac{d-3}{n-d+2}\cdot\frac{\chi^2_{n-d}}{\chi^2_{d+1}(\tau_{\mathrm{R}}\chi^2_{n-1}/2)}\right\}^-\right] \geq 0\,.$$

That means $\mathrm{E}\big(\tau_{\mathrm{M}}\big) < \mathrm{E}\big(\tau_{\mathrm{S}}\big)$. The second inequality $\mathrm{E}\big(\tau_{\mathrm{S}}\big) < \mathrm{E}\big(\tau_{\mathrm{T}}\big)$ is a direct consequence of Theorem 2. Q.E.D.

Proof of Theorem 5

The traditional estimator for the GMVP without the first portfolio weight can be represented by $\hat{w}_{\mathrm{T}}^{\mathrm{ex}} = w^{\mathrm{ex}} + \sigma\,\Omega^{-\frac{1}{2}}V^{-\frac{1}{2}}\xi$, where $V \sim W_{d-1}(I_{d-1}, n-1)$ is stochastically independent of $\xi \sim \mathcal{N}_{d-1}(\mathbf{0}, I_{d-1})$. Since $\sqrt{n}\,V^{-\frac{1}{2}} = \big(V/n\big)^{-\frac{1}{2}} \overset{\text{a.s.}}{\to} I_{d-1}$ as $n \to \infty$, it holds that

$$\sqrt{n}\,\big(\hat{w}_{\mathrm{T}}^{\mathrm{ex}} - w^{\mathrm{ex}}\big) \overset{\text{a.s.}}{\longrightarrow} \sigma\,\Omega^{-\frac{1}{2}}\xi\,, \qquad n \longrightarrow \infty\,.$$

The presented expression for the asymptotic normality of $\hat{w}_{\mathrm{T}} = \mathbf{e}_1 - \Delta'\hat{w}_{\mathrm{T}}^{\mathrm{ex}}$ follows from the relationship $\sigma^2\Delta'\Omega^{-1}\Delta = \sigma^2\Sigma^{-1} - ww'$ (Frahm, 2008). Further, the shrinkage estimator can be represented by

$$\hat{w}_{\mathrm{S}}^{\mathrm{ex}} = w_{\mathrm{R}}^{\mathrm{ex}} + \left\{1 - \frac{d-3}{n-d+2}\cdot\frac{\chi^2_{n-d}}{\big(\theta + V^{-\frac{1}{2}}\xi\big)'V\big(\theta + V^{-\frac{1}{2}}\xi\big)}\right\}\Big\{\big(w^{\mathrm{ex}} - w_{\mathrm{R}}^{\mathrm{ex}}\big) + \sigma\,\Omega^{-\frac{1}{2}}V^{-\frac{1}{2}}\xi\Big\}\,,$$

where $\theta = \Omega^{\frac{1}{2}}/\sigma\,(w^{\rm ex} - w_{\rm R}^{\rm ex})$ and $\theta'\theta = \tau_{\rm R}$. Following the proof of Theorem 3 it can be assumed that $\theta = \big(\sqrt{\tau_{\rm R}}\,,\mathbf{0}\big)$ without loss of generality. Since

$$\frac{\theta' V \theta}{n} = \tau_{\rm R} \cdot \frac{\chi^2_{n-1}}{n} \xrightarrow{\text{a.s.}} \tau_{\rm R}\,, \qquad \frac{2\theta' V^{\frac{1}{2}}\xi}{n} = 2\theta'(V/n)^{\frac{1}{2}}\xi/\sqrt{n} \xrightarrow{\text{a.s.}} 0\,, \qquad \frac{\xi'\xi}{n} \xrightarrow{\text{a.s.}} 0\,, \quad n \longrightarrow \infty\,,$$

it follows that $\big(\theta + V^{-\frac{1}{2}}\xi\big)'V\big(\theta + V^{-\frac{1}{2}}\xi\big)/n \overset{\text{a.s.}}{\rightarrow} \tau_{\rm R}$ as well as $\chi^2_{n-d}/n \overset{\text{a.s.}}{\rightarrow} 1$ as $n \rightarrow \infty$. Hence, in the event that $\tau_{\rm R} > 0$ it holds that

$$\sqrt{n} \cdot \frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}/n}{\big(\theta + V^{-\frac{1}{2}}\xi\big)'V\big(\theta + V^{-\frac{1}{2}}\xi\big)/n} \cdot \big(w_{\rm R}^{\rm ex} - w^{\rm ex}\big) \xrightarrow{\text{a.s.}} \mathbf{0}\,, \qquad n \longrightarrow \infty\,.$$

Further, as already mentioned above, $\sqrt{n}\,\sigma\,\Omega^{-\frac{1}{2}}V^{-\frac{1}{2}}\xi \overset{\text{d}}{\rightarrow} \sigma\,\Omega^{-\frac{1}{2}}\xi$ and so

$$\left\{1 - \frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}/n}{\big(\theta + V^{-\frac{1}{2}}\xi\big)'V\big(\theta + V^{-\frac{1}{2}}\xi\big)/n}\right\}\sqrt{n}\,\sigma\,\Omega^{-\frac{1}{2}}V^{-\frac{1}{2}}\xi \xrightarrow{\text{a.s.}} \sigma\,\Omega^{-\frac{1}{2}}\xi$$

as $n \rightarrow \infty$. By contrast, if $\tau_{\rm R} = 0$ and thus $\theta = \mathbf{0}$ as well as $w^{\rm ex} = w_{\rm R}^{\rm ex}$,

$$\frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}}{\big(\theta + V^{-\frac{1}{2}}\xi\big)'V\big(\theta + V^{-\frac{1}{2}}\xi\big)} = \frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}}{\xi'\xi}$$

and since $\chi^2_{n-d}/(n-d+2) \overset{\text{a.s.}}{\rightarrow} 1$ as $n \rightarrow \infty$,

$$\sqrt{n}\,\big(\hat{w}_{\rm S}^{\rm ex} - w^{\rm ex}\big) \xrightarrow{\text{a.s.}} \left(1 - \frac{d-3}{\xi'\xi}\right)\sigma\,\Omega^{-\frac{1}{2}}\xi\,, \qquad n \longrightarrow \infty\,.$$

Similar arguments hold for the modified shrinkage estimator, since

$$\min\left\{\sqrt{n} \cdot \frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}/n}{\big(\theta + V^{-\frac{1}{2}}\xi\big)'V\big(\theta + V^{-\frac{1}{2}}\xi\big)/n}\,, \sqrt{n}\right\} \xrightarrow{\text{a.s.}} 0\,, \qquad n \longrightarrow \infty\,,$$

if $\tau_{\rm R} > 0$ and otherwise

$$\min\left\{\frac{d-3}{n-d+2} \cdot \frac{\chi^2_{n-d}}{\xi'\xi}\,, 1\right\} \xrightarrow{\text{a.s.}} \min\left\{\frac{d-3}{\xi'\xi}\,, 1\right\}, \qquad n \longrightarrow \infty\,.$$

Q.E.D.

Proof of Theorem 6

Due to Eq. 3 it will suffice to concentrate on the GMVP estimators without the first portfolio weight for calculating the relative losses, e.g.

$$n\tau_{\rm T} = \frac{\sqrt{n}\,(\hat{w}_{\rm T}^{\rm ex} - w^{\rm ex})'\Omega\sqrt{n}\,(\hat{w}_{\rm T}^{\rm ex} - w^{\rm ex})}{\sigma^2}\,.$$

Now the theorem follows immediately by applying the Continuous Mapping Theorem to the results which are given in the proof of Theorem 5 and noting that

$$\Big[\mathbb{1}_{\{\tau_{\rm R}=0\}}X + \mathbb{1}_{\{\tau_{\rm R}>0\}}\Big]^2 = \mathbb{1}_{\{\tau_{\rm R}=0\}}X^2 + \mathbb{1}_{\{\tau_{\rm R}>0\}}$$

for any random variable X. Q.E.D.

Proof of Theorem 7

Due to the proof of Theorem 5 it holds that

$$\tau_{\mathrm{T}} = \frac{(\hat{w}_{\mathrm{T}}^{\mathrm{ex}} - w^{\mathrm{ex}})'\Omega\,(\hat{w}_{\mathrm{T}}^{\mathrm{ex}} - w^{\mathrm{ex}})}{\sigma^2} = \xi' V^{-1}\xi = \frac{\chi^2_{d-1}}{\chi^2_{n-d+1}}$$

with $\chi^2_{d-1} := \xi'\xi$ and $\chi^2_{n-d+1} := \chi^2_{d-1}/\xi' V^{-1}\xi$. Note that $(n-d) \to \infty$ as $n, d \to \infty$ and $n/d \to q$. That means

$$\tau_{\mathrm{T}} = \frac{d}{n-d} \cdot \frac{\chi^2_{d-1}/d}{\chi^2_{n-d+1}/(n-d)} \xrightarrow{\text{a.s.}} \frac{1}{q-1}\,, \qquad n, d \longrightarrow \infty\,,\; n/d \longrightarrow q\,.$$

For proving the almost sure convergence of the shrinkage constants κ_{S} and κ_{M} , consider $\theta = \big(\sqrt{\tau_{\mathrm{R}}}\,, \mathbf{0}\big)$ and suppose that $V^{\frac{1}{2}}$ is the Cholesky root of V, i.e.

$$\theta' V^{\frac{1}{2}}\xi = \sqrt{\tau_{\mathrm{R}}}\,\chi_{n-1}\xi_1\,.$$

Furthermore, note that $(d-3)/(n-d+2) \to 1/(q-1)$, $\chi^2_{n-d}/(n-d) \overset{\text{a.s.}}{\to} 1$,

$$\frac{\theta' V\theta}{n-d} = \tau_{\mathrm{R}} \cdot \frac{\chi^2_{n-1}}{n} \cdot \frac{n}{n-d} \xrightarrow{\text{a.s.}} \frac{q\tau_{\mathrm{R}}}{q-1}\,, \qquad \frac{2\theta' V^{\frac{1}{2}}\xi}{n-d} = 2\sqrt{\tau_{\mathrm{R}}} \cdot \frac{\chi_{n-1}\xi_1}{n-d} \xrightarrow{\text{a.s.}} 0$$

as well as

$$\frac{\xi'\xi}{n-d} = \frac{\xi'\xi}{d} \cdot \frac{d}{n-d} \xrightarrow{\text{a.s.}} \frac{1}{q-1}\,, \qquad n, d \longrightarrow \infty\,,\; n/d \longrightarrow q\,.$$

Now, by applying the Continuous Mapping Theorem, we obtain $\kappa_{\mathrm{S}}\,, \kappa_{\mathrm{M}} \overset{\text{a.s.}}{\to} 1/(1+q\tau_{\mathrm{R}})$ as $n, d \to \infty$ and $n/d \to q$. Similarly, note that

$$2\theta' V^{-\frac{1}{2}}\xi = 2\sqrt{\tau_{\mathrm{R}}} \cdot \frac{\xi_1}{\chi_{n-d+1}} = 2\sqrt{\tau_{\mathrm{R}}} \cdot \frac{n-d}{\chi_{n-d+1}} \cdot \frac{\xi_1}{n} \cdot \frac{n}{n-d} \xrightarrow{\text{a.s.}} 0$$

and $\xi' V^{-1}\xi \overset{\text{a.s.}}{\to} 1/(q-1)$ as $n, d \to \infty$ and $n/d \to q$. By relying on (8) and (10) it turns out that

$$\tau_{\mathrm{S}}\,, \tau_{\mathrm{M}} \xrightarrow{\text{a.s.}} \frac{\tau_{\mathrm{R}}}{1+q\tau_{\mathrm{R}}} - \left(1 - \frac{1}{1+q\tau_{\mathrm{R}}}\right)\tau_{\mathrm{R}} + \left(1 - \frac{1}{1+q\tau_{\mathrm{R}}}\right)^2 \left(\tau_{\mathrm{R}} + \frac{1}{q-1}\right).$$

After a little calculation it can be found that the limit corresponds to the asymptotic loss function $L\big(\tau_{\mathrm{R}}, q\big)$ which is given in the theorem. Q.E.D.

Proof of Theorem 8

Since $w'_{\mathrm{R}}\mathbf{1} = 1 > 0$, the angle between w_{R} and $\mathbf{1}$ is acute. Therefore, there exists an orthogonal $d \times d$ matrix $\mathcal{O}$ such that both $\mathcal{O}w_{\mathrm{R}}$ and $\mathcal{O}\mathbf{1}$ belong to the set $\{x \in \mathbb{R}^d : x > \mathbf{0}\}$. That means there also exists a positive-definite diagonal $d \times d$ matrix Λ such that $\mathcal{O}\mathbf{1} = \Lambda\mathcal{O}w_{\mathrm{R}}$, i.e. $w_{\mathrm{R}} = A\mathbf{1}$ with $A := \mathcal{O}'\Lambda^{-1}\mathcal{O}$ being positive-definite. The matrix Σ_{R}^{-1} can be obtained by re-scaling A such that the condition $\mathbf{1}'\Sigma_{\mathrm{R}}^{-1}\mathbf{1} = \mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1} > 0$ is satisfied. Now the rest of the theorem can be verified by substituting $\widehat{\Sigma}^{-1}$ by the given expressions for $\widehat{\Sigma}_{\mathrm{S}}^{-1}$ and $\widehat{\Sigma}_{\mathrm{M}}^{-1}$ within the traditional GMVP estimator. Q.E.D.

Proof of Theorem 9

Due to the proof of Theorem 3 it can be seen that

$$\hat{\tau}_{\mathrm{R}} = \frac{\big(V^{\frac{1}{2}}\theta + \xi\big)'\big(V^{\frac{1}{2}}\theta + \xi\big)}{\chi^2_{n-d}};$$

note that $\theta' V \theta = \tau_{\mathrm{R}}\chi^2_{n-1}$. Q.E.D.

References

L.K.C. Chan, J. Karceski, and J. Lakonishok (1999), 'On portfolio optimization: Forecasting covariances and choosing the risk model', *Review of Financial Studies* **12**, pp. 937–974.

V.K. Chopra and W.T. Ziemba (1993), 'The effect of errors in means, variances, and covariances on optimal portfolio choice', *Journal of Portfolio Management* **19**, pp. 6–11.

V. DeMiguel, L. Garlappi, and R. Uppal (2007), 'Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy?', *Review of Financial Studies*, URL: http://rfs.oxfordjournals.org/cgi/content/abstract/hhm075v1.

E.J. Elton and M. Gruber (1973), 'Estimating the dependence structure of share prices - Implications for portfolio selection', *Journal of Finance* **28**, pp. 1203–1232.

G. Frahm (2008), 'Linear statistical inference for global and local minimum variance portfolios', *Statistical Papers*, DOI: 10.1007/s00362-008-0170-z.

P.A. Frost and J.E. Savarino (1986), 'An empirical Bayes approach to efficient portfolio selection', *Journal of Financial and Quantitative Analysis* **21**, pp. 293–305.

L. Garlappi, R. Uppal, and T. Wang (2007), 'Portfolio selection with parameter and model uncertainty: a multi-prior approach', *Review of Financial Studies* **20**, pp. 41–81.

V. Golosnoy and Y. Okhrin (2007), 'Multivariate shrinkage for optimal portfolio weights', *The European Journal of Finance* **13**, pp. 441–458.

R. Jagannathan and T. Ma (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**, pp. 1651–1683.

J.D. Jobson and B. Korkie (1979), 'Improved estimation for Markowitz portfolios using James-Stein type estimators', in: 'Proceedings of the American Statistical Association (Business and Economic Statistics)', volume 1, pp. 279–284.

P. Jorion (1986), 'Bayes-Stein estimation for portfolio analysis', *Journal of Financial and Quantitative Analysis* **21**, pp. 279–292.

G.G. Judge and M.E. Bock (1978), *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, North-Holland Publishing Company.

R. Kan and G. Zhou (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**, pp. 621–656.

A. Kempf and C. Memmel (2006), 'Estimating the global minimum variance portfolio', *Schmalenbach Business Review* **58**, pp. 332–348.

O. Ledoit and M. Wolf (2001), 'A well-conditioned estimator for large-dimensional covariance matrices', *Journal of Multivariate Analysis* **88**, pp. 365–411.

O. Ledoit and M. Wolf (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of Empirical Finance* **10**, pp. 603–621.

H.M. Markowitz (1952), 'Portfolio selection', *Journal of Finance* **7**, pp. 77–91.

R.C. Merton (1980), 'On estimating the expected return on the market: An exploratory investigation', *Journal of Financial Economics* **8**, pp. 323–361.

Y. Okhrin and W. Schmid (2006), 'Distributional properties of portfolio weights', *Journal of Econometrics* **134**, pp. 235–256.

S.J. Press (2005), *Applied Multivariate Analysis*, Dover Publications, second edition.

W.F. Sharpe (1963), 'A simplified model for portfolio analysis', *Management Science* **9**, pp. 277–293.

M.S. Srivastava and M. Bilodeau (1989), 'Stein estimation under elliptical distributions', *Journal of Multivariate Analysis* **28**, pp. 247–259.

C. Stein (1956), 'Inadmissability of the usual estimator for the mean of a multivariate normal distribution', in: 'Proceedings of the 3rd Berkeley Symposium on Probability and Statistics', volume 1, pp. 197–206.