



# Dominating Estimators for Minimum-Variance Portfolios

Gabriel Frahm, Christoph Memmel

## ► To cite this version:

Gabriel Frahm, Christoph Memmel. Dominating Estimators for Minimum-Variance Portfolios. *Econometrics*, 2010, 159 (2), pp.289. 10.1016/j.jeconom.2010.07.007 . hal-00741629

**HAL Id: hal-00741629**

**<https://hal.science/hal-00741629>**

Submitted on 15 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Dominating estimators for minimum-variance portfolios

Gabriel Frahm, Christoph Memmel

PII: S0304-4076(10)00159-4

DOI: [10.1016/j.jeconom.2010.07.007](https://doi.org/10.1016/j.jeconom.2010.07.007)

Reference: ECONOM 3394

To appear in: *Journal of Econometrics*

Received date: 6 November 2008

Revised date: 24 July 2010

Accepted date: 31 July 2010

Please cite this article as: Frahm, G., Memmel, C., Dominating estimators for minimum-variance portfolios. *Journal of Econometrics* (2010), doi:10.1016/j.jeconom.2010.07.007

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Dominating Estimators for Minimum-Variance Portfolios\*

Gabriel Frahm<sup>†</sup>

University of Cologne

Chair for Statistics & Econometrics

Department of Economic and Social Statistics

Albertus-Magnus-Platz, D-50923 Cologne, Germany

Christoph Memmel<sup>‡</sup>

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 minimum-variance portfolios 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 assets  $d \geq 4$  and number of observations  $n \geq d + 2$ . The small-sample properties of the shrinkage estimators as well as their large-sample properties for fixed  $d$  but  $n \rightarrow \infty$  and  $n, d \rightarrow \infty$  but  $n/d \rightarrow q \leq \infty$  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, Minimum-variance portfolio, James-Stein estimation, Naive diversification, Shrinkage estimator.

---

\*We would like to thank André Güttler, Alexander Kempf, Julia Nasev, and Michael Wolf for their helpful comments on the manuscript. Special thanks are due to Arndt Musche, who conducted the first run of the moving-blocks bootstrap which is presented at the end of this work. The opinions expressed in this paper are those of the authors and do not necessarily reflect the opinions of the Deutsche Bundesbank.

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

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

## 1 Motivation

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. (2009) 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. (2009) confirms that the considered strategies perform better than the traditional “plug-in” implementation of Markowitz optimization, which means replacing the unknown parameters by their sample counterparts.

Constrained and unconstrained minimum-variance portfolios have been frequently advocated in the literature (Frahm, 2008; Jagannathan and Ma, 2003; Kempf and Memmel, 2006; Ledoit and Wolf, 2003) because they are 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). In fact, many empirical studies indicate that minimum-variance portfolios in general lead to a better out-of-sample performance than stock index portfolios (Haugen, 1990; Haugen and Baker, 1991, 1993; Winston, 1993).

The portfolio which minimizes the portfolio return variance only with respect to the budget constraint is called the global minimum-variance portfolio. We present two estimators for the global minimum-variance portfolio which *dominate* the traditional estimator with respect to the out-of-sample variance of the portfolio return. However, 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 (Frahm, 2008). This means local minimum-variance portfolios and even mean-variance optimal portfolios can be seen as “global” minimum-variance portfolios after a suitable transformation of the asset universe (Frahm, 2008; Memmel, 2004, p. 121). The latter can be achieved by expressing the profitability constraint of the mean-variance optimal portfolio as a linear equality constraint. Hence, in the following discussion we will omit the adjective “global” and use only the term “minimum-variance portfolio” (MVP),

since our methodology can be applied to a much broader spectrum of portfolio optimization problems.

Okhrin and Schmid (2006), Kempf and Memmel (2006) as well as Frahm (2008) all explore the properties of the traditional MVP 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. By 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 including also the variances and covariances. Furthermore, DeMiguel et al. (2009), 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 MVP.

Our work is related to these shrinkage approaches but it differs in three important aspects:

- (i) our estimators are *feasible* by construction,
- (ii) our dominance results are valid in *small* samples, and
- (iii) we focus on the problem of minimizing the *out-of-sample variance*.

**ad i.** The shrinkage estimators presented by the aforementioned authors are determined by unknown quantities, which have to be estimated (see, e.g., Kan and Zhou, 2007). For estimating the weights of an optimal portfolio, i.e., for making these shrinkage estimators feasible, the unknown quantities are substituted by the corresponding estimates. By contrast, our estimators do not contain unknown quantities.

**ad ii.** The potential benefit of shrinking the portfolio weights might be eventually destroyed by estimating the unknown quantities. From a theoretical point of view it is not clear to what extent the methods which have been presented in the literature outperform the traditional “plug-in” approach or any other strategy. Substituting the unknown quantities by some estimates can only be justified for a large number of observations (provided the given estimators are consistent). However, the dominance results of Stein-type estimation theory are derived under a small-sample assumption. Indeed, in large samples

the shrinkage estimates are not substantially different from the corresponding maximum-likelihood estimates.<sup>1</sup> Moreover, as pointed out by Frahm (2008), large-sample results can be misleading in the context of portfolio optimization. Even if the sample size ( $n$ ) is large, the number of observations relative to the number of assets ( $d$ ) can be small. In fact, the results presented here suggest that large-sample approximations work only if  $n/d$  is a large number. Our shrinkage estimators are not affected by the aforementioned drawbacks, since the presented dominance results hold for almost every combination of  $n$  and  $d$ .

**ad iii.** 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. This seems to be the major goal when searching for an MVP. Thus our approach focuses on the portfolio optimization problem itself rather than solving another problem, where the optimal portfolio weights appear only as a by-product. However, that does not stop us from deriving shrinkage estimators for the covariance matrix as a by-product of our shrinkage estimators for the MVP.

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-selling constraints on the MVP 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 MVP 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” constraints. Shrinkage estimators reduce the estimation risk as well. However, in addition they have the appealing property of converging towards the optimal portfolio as the sample size grows to infinity.

Above we justified the analysis of the MVP – instead of the tangency portfolio or some other mean-variance optimal portfolio – by its out-of-sample performance in empirical studies. In addition, there are some more reasons in favor of this portfolio: (i) If all expected returns

---

<sup>1</sup>Maximum-likelihood estimators are asymptotically efficient under the usual regularity conditions.

were the same in the cross section, the tangency portfolio and the MVP would coincide. In fact, the null hypothesis that all expected returns are equal often *cannot* be rejected in empirical studies (see, e.g., Memmel, 2004, p. 89). We believe that in a large part the typical variation in the cross section of historical means is due to estimation errors but not to actual differences in expected returns. (ii) In practice, expected returns are often derived from fundamental analysis, whereas time series data are only used for the estimation of return variances and covariances. Due to the arguments set forth by Frahm (2008), our shrinkage approach of estimating the MVP can be extended to incorporate both sources of information. (iii) As already mentioned, after a suitable transformation of the asset universe, every mean-variance optimal portfolio can be expressed as an MVP under the condition of a given estimate for the expected asset returns.

It is worth pointing out that we make the assumption of jointly normally distributed and serially independent asset returns, although there exist by far more advanced time series models, especially for high-frequency data. However, we choose the normal distribution assumption for two reasons. (i) Generally it is difficult to derive small-sample results without the normal distribution assumption. This might be the reason why other researchers have made the same assumption (see, e.g., Frost and Savarino, 1986; Jorion, 1986; Kan and Zhou, 2007; Okhrin and Schmid, 2006). (ii) In asset allocation it is common to use low-frequency data, for example monthly asset returns. For these data, the assumption of normality and serial independence is not so farfetched. For instance, McNeil et al. (2005, p. 122) write that “As we progressively increase the interval of the returns by moving from daily to weekly, monthly, quarterly and yearly data, [...] volatility clustering decreases and returns begin to look both more i.i.d. and less heavy-tailed.”. Moreover, Aparicio and Estrada (2001) conclude that “normality does seem a reasonable assumption for monthly stock returns.”.<sup>2</sup>

Our contribution to the literature is threefold. First, we derive two shrinkage estimators for the MVP 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  $n \rightarrow \infty$  as well as  $n, d \rightarrow \infty$  and  $n/d \rightarrow q \leq \infty$ . The latter kind

---

<sup>2</sup>The aforementioned authors refer to asset log-returns. Similar conclusions can be drawn for discrete asset returns if the investment horizon is not too large.

of asymptotic behavior becomes relevant when analyzing the estimators in large asset universes. Third, backed by the results of DeMiguel et al. (2009), 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  $R_t = (R_{1t}, \dots, R_{dt})$  for  $t = 1, \dots, n$ ,<sup>3</sup> 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  $t = 1, \dots, n$  with  $\mu \in \mathbb{R}^d$  and positive-definite matrix  $\Sigma \in \mathbb{R}^{d \times d}$ .
- A2.** The mean vector  $\mu$  and the covariance matrix  $\Sigma$  are unknown.
- A3.** The asset returns are serially independent.
- A4.** There exist at least four assets, i.e.,  $d \geq 4$ .
- A5.** The sample size exceeds the number of assets, more precisely  $n \geq d + 2$ .

The MVP  $w$  is defined as the solution of the minimization problem

$$\min_{v \in \mathbb{R}^d} v' \Sigma v, \quad \text{s.t. } v' \mathbf{1} = 1. \quad (1)$$

Here  $\mathbf{1}$  denotes a vector of ones. Since  $\Sigma$  is positive definite, the MVP 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  $\hat{w}_T$  for the MVP consists in replacing the unknown covariance matrix  $\Sigma$  with the sample covariance matrix  $\hat{\Sigma}$ , i.e.,

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

---

<sup>3</sup>In the following “ $(x_1, \dots, x_d)$ ” indicates a  $d$ -tuple, i.e., a  $d$ -dimensional column vector.



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

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  $w^{\text{ex}} := (w_2, \dots, w_d)$ . For convenience we introduce the  $(d-1) \times d$  matrix  $\Delta := [\mathbf{1} \ -I_{d-1}]$ . By using the operator  $\Delta$ , 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:

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

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

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

for every vector  $v \in \mathbb{R}^d$  with  $v' \mathbf{1} = 1$ , where  $\Omega$  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  $\chi_k^2(\lambda)$  denotes a noncentral  $\chi^2$ -distributed random variable with  $k \in \mathbb{N}$  degrees of freedom and noncentrality parameter  $\lambda \geq 0$ . This means  $\chi_k^2(\lambda) \sim X'X$  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,  $\chi_k^2$  stands for a central  $\chi^2$ -distributed random variable (that is  $\lambda = 0$ ) and we also define  $\chi_k^r(\lambda) := \{\chi_k^2(\lambda)\}^{r/2}$  for any  $r \in \mathbb{Z}$ .

Moreover, let  $\chi_{k_1}^2(\lambda)$  and  $\chi_{k_2}^2(k_1, k_2 \in \mathbb{N})$  be stochastically independent. Then  $F_{k_1, k_2}(\lambda) \sim (k_2/k_1)(\chi_{k_1}^2(\lambda)/\chi_{k_2}^2)$  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  $X_1, \dots, X_m$  are  $m$  independent copies of  $X \sim \mathcal{N}_q(\mathbf{0}, \Sigma)$ , where  $\mathbf{0}$  denotes a vector of zeros and  $\Sigma$  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  $\Sigma$  and  $m$  degrees of freedom.

In the following  $t_k(a, B, \nu)$  stands for the  $k$ -variate  $t$ -distribution with  $\nu > 0$  degrees of freedom, location vector  $a$  ( $k \times 1$ ), and positive-definite dispersion matrix  $B$  ( $k \times k$ ). This means  $a + \zeta/\sqrt{\chi_\nu^2/\nu} \sim t_k(a, B, \nu)$ , where  $\zeta \sim \mathcal{N}_k(0, B)$  is stochastically independent of  $\chi_\nu^2$ . 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 MVP and its related quantities. The proofs can be found in Frahm (2008) as well as Kempf and Memmel (2006).

### Lemma 1 (Frahm (2008); Kempf and Memmel (2006))

Under the assumptions **A1** to **A3** and  $n > d$ , the sample covariance matrix  $\hat{\Omega}$  of  $\Delta R$ , the traditional estimator  $\hat{w}_T^{\text{ex}}$  for the MVP (except for the first portfolio weight), and the traditional estimator  $\hat{\sigma}_T^2$  for the minimum variance  $\sigma^2$  satisfy the following properties:

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

**P2.**  $\hat{w}_T^{\text{ex}} | \hat{\Omega} \sim \mathcal{N}_{d-1}(w^{\text{ex}}, \sigma^2 \hat{\Omega}^{-1}/n)$ .

**P3.**  $\hat{w}_T^{\text{ex}} \sim t_{d-1}(w^{\text{ex}}, \sigma^2 \Omega^{-1}/(n-d+1), n-d+1)$ .

**P4.**  $n\hat{\sigma}_T^2/\sigma^2 \sim \chi_{n-d}^2$ .

**P5.**  $\hat{\sigma}_T^2$  is stochastically independent of  $\hat{\Omega}$  and  $\hat{w}_T^{\text{ex}}$ .

The following theorem will play the central role in the development of the shrinkage estimators and their dominance properties.

### Theorem 1

Consider a  $q \times q$  random matrix  $W \sim W_q(\Omega, m)$ , where  $\Omega$  is a positive-definite  $q \times q$  matrix,  $q \geq 3$  and  $m \geq q+2$ , a  $q$ -dimensional random vector  $X$  with  $X | W \sim \mathcal{N}_q(\omega, W^{-1})$ , 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

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

dominates the estimator  $X$  with respect to the loss function

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

i.e.,  $E\{(X_S - \omega)' \Omega (X_S - \omega)\} < E\{(X - \omega)' \Omega (X - \omega)\}$ . 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.

Theorem 1 states that  $X$  is *inadmissible* (Judge and Bock, 1978, p. 14), since the expected loss of  $X_S$  is smaller than the expected loss of  $X$ . 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  $W$  corresponds to a non-stochastic but observable matrix  $\Omega$ , or that  $W$  is stochastic but independent of  $X$ , where  $\Omega$  is unobservable (Judge and Bock (1978, p. 177); Srivastava and Bilodeau (1989); Press (2005, p. 189)). By contrast, we allow  $X$  to depend on a Wishart-distributed random matrix  $W$ , but the matrix  $\Omega$  given in Theorem 1 remains unobservable. To the best of our knowledge this is a novel result and might be useful in its own right.

For example, consider the linear regression model  $y_i = \alpha + \beta' x_i + u_i$  ( $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{R}^{m-1}$ , and  $i = 1, \dots, n \geq m + 2$ ), where  $(y_i, x_i)$  is a normally distributed  $m$ -dimensional random vector with  $m \geq 4$  and the covariance matrix of  $x_i$  is denoted by  $\Omega$ . It is well-known that, under the standard assumptions of linear regression theory,  $\hat{\beta}_{OLS} | \mathbf{X} \sim \mathcal{N}_{m-1}(\beta, \sigma^2 W^{-1})$  with  $\sigma^2 = \text{Var}(u_i) > 0$  and  $W \sim W_{m-1}(\Omega, n-1)$  is  $n$  times the sample covariance matrix of  $x_i$ . Here  $\mathbf{X}$  denotes the  $n \times m$  regressor matrix. It follows that  $\hat{\beta}_{OLS} \sim t_{m-1}(\beta, \sigma^2 \Omega^{-1}/(n-m+1), n-m+1)$ . Finally, the sum of squared residuals is given by  $\hat{u}'\hat{u} \sim \sigma^2 \chi_{n-m}^2$ , which is stochastically independent of  $W$  and  $\hat{\beta}_{OLS}$ . A natural choice for the loss function of an estimator  $\hat{\beta}$  for  $\beta$  is  $\mathcal{L}_{\beta, \Omega}(\hat{\beta}) = (\hat{\beta} - \beta)' \Omega (\hat{\beta} - \beta)$ . Theorem 1 states that for any  $b \in \mathbb{R}^{m-1}$  the shrinkage estimator

$$\hat{\beta}_S = b + \left( 1 - \frac{m-3}{n-m+2} \cdot \frac{\hat{u}'\hat{u}}{(\hat{\beta}_{OLS} - b)' W (\hat{\beta}_{OLS} - b)} \right) (\hat{\beta}_{OLS} - b)$$

dominates the OLS estimator with respect to the chosen loss function.

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  $\omega$ . 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

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

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

$$\text{Var}(\hat{v}'R) = \sigma^2 + \text{E}\{(\hat{v} - w)'\Sigma(\hat{v} - w)\} + \mu'\text{Var}(\hat{v})\mu. \quad (6)$$

Hence, the minimum variance  $\sigma^2$  is a lower bound for the out-of-sample variance of any given portfolio  $\hat{v}$ . Interestingly, the between variance  $\mu'\text{Var}(\hat{v})\mu$  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  $\text{Var}(\hat{v}) = \Delta'\text{Var}(\hat{v}^{\text{ex}})\Delta$  and  $\Delta\mu = \mathbf{0}$  if  $\mu = \eta \mathbf{1}$ .

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

$$\text{E}\{(\hat{w}_T - w)'\Sigma(\hat{w}_T - w)\} = \frac{d-1}{n-d-1} \cdot \sigma^2.$$

The factor  $(d-1)/(n-d-1)$  is large whenever the sample size  $n$  is small relative to the number  $d$  of assets. For  $n, d \rightarrow \infty$  but  $n/d \rightarrow q$  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 the traditional MVP estimator  $\hat{w}_T$  presented by Frahm (2008), it follows that the third part of (6) corresponds to

$$\mu'\text{Var}(\hat{w}_T)\mu = \frac{r_{\text{TP}}^2 - r_{\text{MVP}}^2}{n-d-1} \cdot \sigma^2,$$

where  $r_{TP}$  denotes the Sharpe ratio of the tangency portfolio  $\Sigma^{-1}\mu/(\mathbf{1}'\Sigma^{-1}\mu)$  and  $r_{MVP}$  the Sharpe ratio of the MVP.<sup>4</sup> This means it holds that

$$\text{Var}(\hat{w}'_T R) = \left(1 + \frac{d-1}{n-d-1} + \frac{r_{TP}^2 - r_{MVP}^2}{n-d-1}\right) \cdot \sigma^2.$$

Hence, the additional out-of-sample variance due to estimation errors amounts to

$$\frac{\sigma^2}{n-d-1} \cdot \{(d-1) + (r_{TP}^2 - r_{MVP}^2)\}. \quad (7)$$

The tangency portfolio is the portfolio with the maximal Sharpe ratio. In the literature, the Sharpe ratio of this portfolio is often approximated by the empirical Sharpe ratio of a well diversified stock index, using long estimation horizons (up to 100 years). For the period 1900 to 2002, Dimson et al. (2003) find for the world stock portfolio a realized annualized equity premium of 5.7% and a return standard deviation of 16.5%, yielding an annualized Sharpe ratio of about 35%.<sup>5</sup> After dividing it by  $\sqrt{4} = 2$ , the quarterly Sharpe ratio becomes about 0.173. Even if it is assumed that the Sharpe ratio of the MVP is zero, the second summand in Eq. 7 is only about 0.030, compared with the first summand of, say,  $d-1 = 9$  in case of a relatively small portfolio of ten stocks.

In other words, in the last example the within part of the additional out-of-sample variance is roughly 300 times as large as the between part. For data of higher frequency (for instance monthly data) or larger portfolios, the proportion of the between part would be even smaller. Note that these analytical results are derived for the traditional MVP estimator with normally, serially independent and identically distributed asset returns. However, there is no reason to assume qualitatively different results if one uses other estimators or if the assumptions about the return distribution is different.

Hence we believe that the between variance  $\mu'\text{Var}(\hat{v})\mu$  for any portfolio  $\hat{v}$  is negligible in most practical situations and thus we will concentrate in the following on reducing the within variance  $E(\hat{v}'\Sigma\hat{v})$ .<sup>6</sup> Note that each realization of  $\hat{v}'\Sigma\hat{v}$  represents the actual

---

<sup>4</sup>The Sharpe ratio of a portfolio is the expected excess return divided by the standard deviation.

<sup>5</sup>The corresponding estimates for the UK and the US are 29% and 36%, respectively. Using US data for the period 1872 to 2002, Cogley and Sargent (2008) get an estimate of the US Sharpe Ratio of 24%. Kan and Zhou (2007) use in their study 20% and 40%.

<sup>6</sup>Kan and Zhou (2007) a priori forgo the between variance in their analytical expressions.

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

$$(\hat{v} - w)' \Sigma (\hat{v} - w) = \hat{v}' \Sigma \hat{v} - \sigma^2$$

represents that part of the actual variance which is caused by estimation risk. This quantity will be referred to as the *loss* of  $\hat{v}$ . Now, it turns out that reducing the within variance is equivalent to reducing the expected loss  $E\{(\hat{v} - w)' \Sigma (\hat{v} - w)\}$  of the portfolio  $\hat{v}$ . This leads immediately to an application of Stein-type estimation theory.

### 3 The Dominant Estimators

#### 3.1 Small-Sample Properties

We now present the shrinkage estimators for the MVP that dominate 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.<sup>7</sup> However, as already discussed earlier, this estimator can lead to a huge out-of-sample variance of the portfolio return compared to  $\sigma^2$ , i.e., the smallest of all possible portfolio return variances.

In this section we will use the following notation. Let  $\hat{w}_A$  be an arbitrary portfolio. Then  $\sigma_A^2 = \hat{w}_A' \Sigma \hat{w}_A$  is the actual variance of the portfolio return, whereas  $\hat{\sigma}_A^2 = \hat{w}_A' \hat{\Sigma} \hat{w}_A$  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  $\sigma_A^2 = w_A' \Sigma w_A$  and  $\hat{\sigma}_A^2 = w_A' \hat{\Sigma} w_A$ .

#### Theorem 2

*Suppose that the assumptions **A1** to **A5** are satisfied. Let  $\hat{w}_T$  be the traditional estimator for the MVP  $w$ , whereas  $w_R \in \mathbb{R}^d$  with  $w_R' \mathbf{1} = 1$  denotes an arbitrary reference portfolio. Consider the shrinkage estimator*

$$\hat{w}_S = \kappa_S w_R + (1 - \kappa_S) \hat{w}_T \quad (8)$$

with

$$\kappa_S = \frac{d-3}{n-d+2} \cdot \frac{1}{\hat{\tau}_R},$$

---

<sup>7</sup>An estimator is called *best* if its covariance matrix attains the Rao-Cramér lower bound.

where  $\hat{\tau}_R = (\hat{\sigma}_R^2 - \hat{\sigma}_T^2)/\hat{\sigma}_T^2$  is the estimated relative loss of the reference portfolio  $w_R$ . The shrinkage estimator  $\hat{w}_S$  dominates  $\hat{w}_T$  with respect to the loss function  $\mathcal{L}_{w,\Sigma}(\hat{v}) = (\hat{v} - w)' \Sigma (\hat{v} - w)$ , i.e.,

$$E\{(\hat{w}_S - w)' \Sigma (\hat{w}_S - w)\} < E\{(\hat{w}_T - w)' \Sigma (\hat{w}_T - w)\}.$$

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 MVP. The better the reference portfolio fits the actual MVP, 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*  $w_N := \mathbf{1}/d$  can be viewed as a natural choice for the reference portfolio. Due to the arguments given by DeMiguel et al. (2009), there are even doubts as to whether time series information can add useful information at all, and so  $w_R = w_N$  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

$$\tau_S = \frac{\sigma_S^2 - \sigma^2}{\sigma^2}$$

of the shrinkage estimator for the MVP given by (8) depends only on the number of observations  $n$ , the number of assets  $d$ , and the relative loss  $\tau_R = (\sigma_R^2 - \sigma^2)/\sigma^2$  of the reference portfolio. More precisely,  $\tau_S$  can be represented stochastically by

$$\tau_S = \|\kappa_S \theta - (1 - \kappa_S) V^{-\frac{1}{2}} \xi\|^2, \quad (9)$$

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

$$\kappa_S = \frac{d-3}{n-d+2} \cdot \frac{\chi_{n-d}^2}{(\theta + V^{-\frac{1}{2}} \xi)' V (\theta + V^{-\frac{1}{2}} \xi)}.$$

Here  $\xi$ ,  $V$ , and  $\chi_{n-d}^2$  are supposed to be mutually independent.

Proof: See the appendix.

Due to Theorem 2, the shrinkage estimator is dominant in the sense that  $E(\tau_S) < E(\tau_T)$ , where  $\tau_T = (\sigma_T^2 - \sigma^2)/\sigma^2$  represents the relative loss of the traditional estimator for the MVP. It can be shown that the expected relative loss of the shrinkage estimator is a strictly increasing function of  $\tau_R$  and its infimum is attained if and only if  $\tau_R = 0$ . Note that  $\tau_R = 0$  or, equivalently,  $\theta = \mathbf{0}$  holds if and only if  $w_R = w$ , since  $\Sigma$  is positive definite. In that case it turns out that

$$E(\tau_S) = \left(1 - \frac{d-3}{d-1} \cdot \frac{n-d}{n-d+2}\right) \frac{d-1}{n-d-1}.$$

By contrast,  $E(\tau_S) \rightarrow E(\tau_T)$  for  $\tau_R \rightarrow \infty$ .

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  $\kappa_S$  to values smaller than or equal to 1, i.e., by taking the shrinkage weight  $\kappa_M := \min\{\kappa_S, 1\}$  instead of  $\kappa_S$ . Then the corresponding shrinkage estimator is given by

$$\hat{w}_M := \kappa_M w_R + (1 - \kappa_M) \hat{w}_T. \quad (10)$$

The shrinkage weight  $\kappa_M$  can only attain values between 0 and 1, which prevents  $\hat{w}_M$  from having the opposite sign of  $\hat{w}_T$  whenever  $\hat{\tau}_R$  is small, i.e., whenever the traditional estimate of the MVP 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 with the notation of Theorem 3, the distribution of the relative loss*

$$\tau_M = \frac{\sigma_M^2 - \sigma^2}{\sigma^2}$$

*of the modified shrinkage estimator for the MVP given by (10) depends only on the number of observations  $n$ , the number of assets  $d$ , and the relative loss  $\tau_R$  of the reference portfolio. More precisely,  $\tau_M$  can be represented stochastically by*

$$\tau_M = \|\kappa_M \theta - (1 - \kappa_M) V^{-\frac{1}{2}} \xi\|^2, \quad (11)$$

*with  $\kappa_M = \min\{\kappa_S, 1\}$ , and it holds that*

$$E(\tau_M) < E(\tau_S) < E(\tau_T).$$



Proof: See the appendix.

The stochastic representations (9) and (11) 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 (8). Moreover, it can be shown that the expected relative loss of  $\hat{w}_M$  corresponds to

$$E(\tau_M) = E \left[ \left\{ \left( 1 - \frac{d-3}{n-d+2} \cdot \frac{\chi_{n-d}^2}{\chi_{d+1}^2} \right)^+ \right\}^2 \right] \frac{d-1}{n-d-1}$$

in the event that  $\tau_R = 0$ .

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  $\hat{w}_S$  and  $\hat{w}_M$ . 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  $n \rightarrow \infty$ . 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 the assumptions **A1** to **A4** and  $\tau_R > 0$  it holds that*

$$\sqrt{n} \begin{bmatrix} \hat{w}_T - w \\ \hat{w}_S - w \\ \hat{w}_M - w \end{bmatrix} \xrightarrow{d} \Lambda \xi, \quad n \rightarrow \infty,$$

where  $\Lambda$  is a  $d \times (d-1)$  matrix such that  $\Lambda\Lambda' = \sigma^2\Sigma^{-1} - ww'$  and  $\xi \sim \mathcal{N}_{d-1}(\mathbf{0}, I_{d-1})$ . In case  $\tau_R = 0$  it turns out that

$$\sqrt{n} \begin{bmatrix} \hat{w}_T - w \\ \hat{w}_S - w \\ \hat{w}_M - w \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Lambda\xi \\ \left(1 - \frac{d-3}{\xi'\xi}\right)\Lambda\xi \\ \left(1 - \frac{d-3}{\xi'\xi}\right)^+ \Lambda\xi \end{bmatrix}, \quad n \longrightarrow \infty.$$

Proof: See the appendix.

For instance, from the last theorem it follows that

$$\sqrt{n}(\hat{w}_T - w) \xrightarrow{d} \mathcal{N}_d(\mathbf{0}, \sigma^2\Sigma^{-1} - ww'), \quad n \longrightarrow \infty,$$

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

$$\sqrt{n}(\hat{w}_T - \hat{w}_S) \xrightarrow{p} \mathbf{0} \quad \text{and} \quad \sqrt{n}(\hat{w}_T - \hat{w}_M) \xrightarrow{p} \mathbf{0}, \quad n \longrightarrow \infty, \quad (12)$$

only if  $w_R \neq w$ .<sup>8</sup> The last theorem also implies that if  $w_R = w$  and the sample size is large (relative to the number of assets), the modified shrinkage estimate corresponds to the true MVP roughly with probability  $F_{\chi_{d-1}^2}(d-3)$ . Admittedly, this might be regarded as purely theoretical, since it has to be assumed that  $w_R \neq w$  in most practical situations, with  $\hat{w}_M$  then being asymptotically equivalent to  $\hat{w}_T$  in the sense given above.

So far we have focused on the expected relative losses of the estimators for the MVP but, as already mentioned, these quantities vanish if the sample size tends to infinity. However, 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 the assumptions **A1** to **A4** and  $\tau_R > 0$  it holds that

$$n \begin{bmatrix} \tau_T \\ \tau_S \\ \tau_M \end{bmatrix} \xrightarrow{d} \chi_{d-1}^2, \quad n \longrightarrow \infty.$$

<sup>8</sup>Actually, the proof of Theorem 5 reveals that (12) can be even strengthened to almost sure convergence.

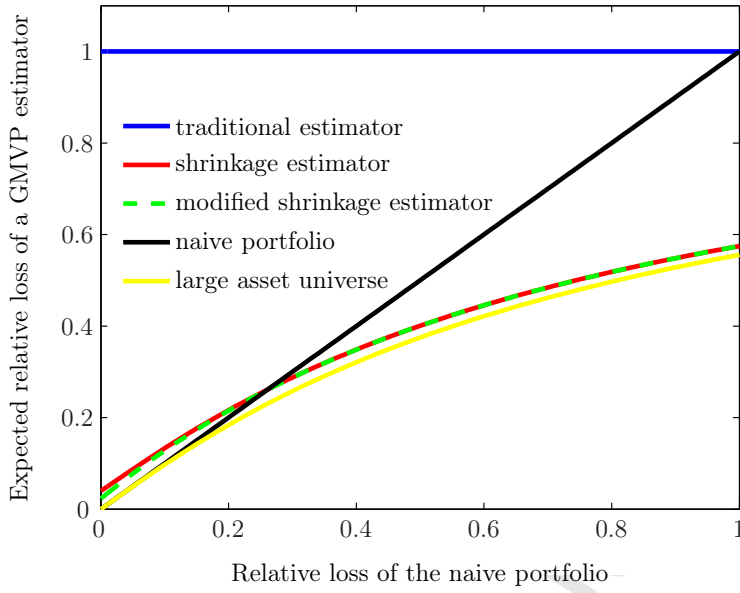


Figure 1: Expected relative losses of the traditional (blue), simple (red) and modified (dashed green) shrinkage estimator for  $n = 200$  and  $d = 100$  as well as the relative loss of the reference portfolio (black) and the asymptotic loss function  $L(\tau_R; 2)$  (yellow).

In case  $\tau_R = 0$  it turns out that

$$n \begin{bmatrix} \tau_T \\ \tau_S \\ \tau_M \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \chi_{d-1}^2 \\ \left(1 - \frac{d-3}{\chi_{d-1}^2}\right)^2 \chi_{d-1}^2 \\ \left\{ \left(1 - \frac{d-3}{\chi_{d-1}^2}\right)^+ \right\}^2 \chi_{d-1}^2 \end{bmatrix}, \quad n \rightarrow \infty.$$

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  $\tau_S$  may turn out to be greater than  $\tau_T$ . Surprisingly, Theorem 6 implies that, only if  $w_R = w$ , the probability of this event does not vanish (even asymptotically) but tends to  $F_{\chi_{d-1}^2}\{(d-3)/2\} > 0$ . For example, if there exist  $d = 5$  assets, this adverse effect occurs with a probability of approximately 9%. However, the same theorem confirms that  $\tau_M > \tau_T$  is asymptotically impossible. This is another advantage of the modified shrinkage estimator over the simple one.

In many practical applications of portfolio theory, the number of observations is small relative to the number of assets. In the following we will investigate the asymptotic distribution of the relative loss assuming that  $n, d \rightarrow \infty$  but  $n/d \rightarrow q$  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 MVP estimators are no longer super-consistent.

**Theorem 7**

Under the assumptions **A1** to **A3** it holds that

$$\tau_T \xrightarrow{\text{a.s.}} \frac{1}{q-1}$$

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

$$\kappa_S, \kappa_M \xrightarrow{\text{a.s.}} \frac{1}{1+q\tau_R}$$

as well as

$$\tau_S, \tau_M \xrightarrow{\text{a.s.}} L(\tau_R; q) := \frac{\tau_R}{(1+q\tau_R)^2} + \left(1 - \frac{1}{1+q\tau_R}\right)^2 \frac{1}{q-1}$$

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

Proof: See the appendix.

It can be shown that the *asymptotic loss function*  $L(\cdot; q)$  is strictly increasing in  $\tau_R$  and it holds that  $L(\tau_R; q) < 1/(q-1)$  whenever  $q < \infty$ . This means the traditional estimator is inadmissible (with respect to the asymptotic loss) even if the number of observations tends to infinity but the effective sample size remains finite.

Moreover, it turns out that  $L(\tau_R; q) < \tau_R$  if and only if

$$\tau_R > \frac{1}{q} \cdot \frac{2-q}{q-1}. \quad (13)$$

Hence, the shrinkage estimators dominate the reference portfolio *uniformly* if  $q \geq 2$  (see Figure 1). This means not only the traditional estimator but also any arbitrary reference portfolio  $w_R \neq w$  becomes inadmissible if  $n, d \rightarrow \infty$  with  $n/d \rightarrow q \geq 2$ .

Conversely, in terms of the asymptotic loss,  $\hat{w}_S$  and  $\hat{w}_M$  become uniformly worse than  $w_R$  as  $q$  tends to 1 from above, since the right-hand side of (13) then tends to infinity. The large-sample properties of the relative losses of the MVP estimators  $\hat{w}_T$ ,  $\hat{w}_S$ , and  $\hat{w}_M$  are summarized in Table 1.

	$n \rightarrow \infty, d < \infty$		$n \rightarrow \infty, d \rightarrow \infty, n/d \rightarrow q$		
	$q = \infty$		$q < \infty$		$q = \infty$
	$\tau_R = 0$	$\tau_R > 0$	$\tau_R = 0$	$\tau_R > 0$	$\tau_R \geq 0$
$\tau_T$	0	0	$\frac{1}{q-1} > 0$	$\frac{1}{q-1} > 0$	0
$\tau_S$	0	0	0	$0 < L(\tau_R, q) < \frac{1}{q-1}$	0
$\tau_M$	0	0	0	$0 < L(\tau_R, q) < \frac{1}{q-1}$	0
$n\tau_T$	$\chi_{d-1}^2$	$\chi_{d-1}^2$	$\infty$	$\infty$	$\infty$
$n\tau_S$	$\left(1 - \frac{d-3}{\chi_{d-1}^2}\right)^2 \chi_{d-1}^2$	$\chi_{d-1}^2$	0	$\infty$	$\infty$
$n\tau_M$	$\left\{\left(1 - \frac{d-3}{\chi_{d-1}^2}\right)^+\right\}^2 \chi_{d-1}^2$	$\chi_{d-1}^2$	0	$\infty$	$\infty$

Table 1: Large-sample properties of the relative losses of  $\hat{w}_T$ ,  $\hat{w}_S$ , and  $\hat{w}_M$ .

### 3.3 Covariance Matrix Estimation

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

#### Theorem 8

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

$$\hat{\Sigma}_S^{-1} := \kappa_S \Sigma_R^{-1} + (1 - \kappa_S) \hat{\Sigma}^{-1} \quad \text{and} \quad \hat{\Sigma}_M^{-1} := \kappa_M \Sigma_R^{-1} + (1 - \kappa_M) \hat{\Sigma}^{-1}$$

for the traditional MVP estimator, i.e.,

$$\hat{w}_S = \frac{\hat{\Sigma}_S^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}_S^{-1} \mathbf{1}} \quad \text{and} \quad \hat{w}_M = \frac{\hat{\Sigma}_M^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}_M^{-1} \mathbf{1}}.$$

Proof: See the appendix.

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

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:

- (i) Their shrinkage weights 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.
- (ii) 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 exceeds the number of observations.

So far our methodology consists of shrinking the traditional MVP 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  $\hat{w}_T$ .<sup>9</sup> Nevertheless, in the following we will concentrate on the special case  $w_R = w_N = \mathbf{1}/d$ .

## 4 Naive Diversification vs. Portfolio Optimization

### 4.1 A Small-Sample Simulation Study

DeMiguel et al. (2009) 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

---

<sup>9</sup>For example,  $w_R$  could be interpreted as a portfolio which has been suggested by somebody who completely refuses statistical methods and whose decision is independent of historical observations.

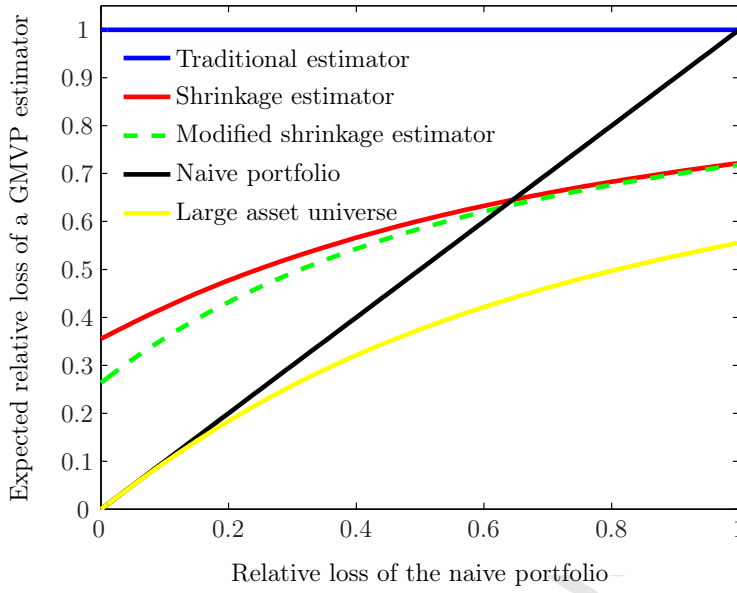


Figure 2: Expected relative losses of the traditional (blue), simple (red) and modified shrinkage (dashed green) estimator for  $n = 20$  and  $d = 10$  as well as the relative loss of the naive portfolio (black) and the asymptotic loss function  $L(\tau_R; 2)$  (yellow).

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 MVP 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 MVP estimator. Due to Theorem 4 and the arguments given in Section 3.2, we will concentrate on the modified shrinkage estimator  $\hat{w}_M$  and choose the naive portfolio  $w_N$  as a reference portfolio. Although closed-form expressions for  $\tau_M$  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 9 and 11, if the sample is small. Figure 2 contains the expected relative losses of the four different portfolio strategies, i.e., traditional estimation, simple and modified shrinkage estimation as well as naive diversification for  $n = 20$  observations and  $d = 10$  assets. The  $x$ -axis denotes the relative loss  $\tau_N$  of the naive portfolio, whereas the  $y$ -axis accounts for the expected relative losses of the different portfolio strategies

depending on  $\tau_N$ . Note that (according to Theorem 3) the expected relative loss of the traditional estimator does not depend on  $\tau_N$  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  $\tau_N$  is small, i.e., the true MVP does not differ too greatly from the naive portfolio (which serves as an anchor point for  $\hat{w}_S$  and  $\hat{w}_M$ ), the shrinkage estimators are more favorable than the traditional estimator.

Figure 2 also indicates the *critical relative loss*  $\tau_N^*$  of the naive portfolio with respect to the modified shrinkage estimator  $\hat{w}_M$ . 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  $\tau_N < 63\%$ . Suppose that the standard deviation of the MVP 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  $\tau_N = (0.11/0.10)^2 - 1 = 21\%$ , whereas the expected relative loss caused by the modified shrinkage estimator roughly amounts to  $E(\tau_M) = 43\%$ . Therefore, it would not pay to use the modified shrinkage estimator in that case. By contrast, if the naive portfolio leads to a standard deviation of 13%, it holds that  $\tau_N = (0.13/0.10)^2 - 1 = 69\% > \tau_N^*$  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 (that is  $n = 120$ ) and the stock market consists of  $d = 50$  assets, one should use the modified shrinkage estimator if and only if the variance of the naive portfolio return is at least 21% greater than the variance of the MVP 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 observations is *very* small relative to the number of assets (see the entries in Table 2 where  $n/d < 2$ ), naive diversification is apparently the best strategy, which reconfirms the



$n \setminus d$	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  $\hat{\tau}_N$  for testing the naive-diversification hypothesis on a significance level of  $\alpha = 5\%$ .

naive-diversification hypothesis of DeMiguel et al. (2009).

## 4.2 Testing for 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  $\hat{\tau}_N$  or, more generally,  $\hat{\tau}_R$  (see also Theorem 2).

### Theorem 9

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

$$\hat{\tau}_R \sim \frac{d-1}{n-d} \cdot F_{d-1, n-d}(\tau_R \chi_{n-1}^2/2).$$

Proof: See the appendix.

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

$$H_0: \tau_N \leq \tau_N^* \text{ vs.}$$

$$H_1: \tau_N > \tau_N^* .$$

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

$$\frac{d-1}{n-d} \cdot F_{d-1, n-d}(\tau_N^* \chi_{n-1}^2/2) ,$$

which can be also calculated by Monte Carlo simulation.<sup>10</sup>

Critical thresholds for this hypothesis test on 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  $\hat{\tau}_N > 161\%$ . Note that this is by far greater than the theoretical value of the critical relative loss  $\tau_N^* = 21\%$ , since the distribution of  $\hat{\tau}_N$  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 MVP 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 MVP 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 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. (2009). 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).

---

<sup>10</sup>This hypothesis test can be adapted to any MVP estimator if its expected relative loss  $E(\tau) < \infty$  depends only on  $n$ ,  $d$ , and  $\tau_N$  and provided  $\tau_N \mapsto E(\tau)$  has only one intersection point with  $\tau_N \mapsto \tau_N$ .

## 5 Empirical Study

For the following empirical study we use monthly historical returns of 7 MSCI stock market indices, namely for USA, UK, Germany, France, Italy, Canada, and Japan. The indices are adjusted for dividends, splits, etc. We use the monthly closing prices quoted in US \$ from September 1970 to August 2009. From these data we calculate the excess returns with respect to the average monthly interest rate of the 3-month US treasury bill. Finally, we obtain  $N = 468$  monthly excess returns.

First of all we apply the naive-diversification hypothesis test which has been derived in Section 4 to the empirical data. Note that the naive hypothesis test consists of two parts. The first one is the determination of the critical relative loss  $\tau_N^*$  of the naive portfolio. In our case we choose the parameters  $n = 60$  and  $d = 7$ . This means we assume that the investor calculates the modified shrinkage estimator on the basis of the past 5 years of observations. The second one is about the distribution of the test statistic. In contrast to Section 4, we choose different parameter values for the number of observations in the two parts: as stated above,  $n = 60$  for the first part but  $n = N = 468$  for the second part, where the distribution of the test statistic is approximated by Monte Carlo simulation. We do so, because we want to make use of the full sample of  $N = 468$  observations and, at the same time, we want to be in line with our empirical study which follows below.

We find that the critical relative loss of the naive portfolio is  $\tau_N^* = 0.064$  and the critical threshold on a significance level of  $\alpha = 0.05$  amounts to 0.125 (for  $\alpha = 0.01$  the critical threshold is given by 0.149). By contrast, the estimated relative loss of the naive portfolio corresponds to  $\hat{\tau}_N = 0.273$ . This means we can clearly reject the naive-diversification hypothesis and hence the modified shrinkage estimator turns out to be significantly better (with respect to the out-of-sample variance) than the naive portfolio if the investor makes use of 5 years of historical observations.

Now we apply a circular moving-blocks bootstrap (Politis and Romano, 1992) to validate our results on the shrinkage estimators. Our goal is to show that the presented shrinkage estimators indeed reduce the out-of-sample variance compared to the traditional estimator and the naive portfolio. We also compute the short-selling constrained MVP estimator ( $\hat{w}_C$ ), i.e., the vector  $v = (v_1, \dots, d)$  which minimizes  $v' \hat{\Sigma} v$  under the additional constraint  $v_1, \dots, v_d \geq 0$ . Moreover, we take the “three-fund optimal portfolio” estimator ( $\hat{w}_{KZ}$ ) into consideration, which has been proposed by Kan and Zhou (2007). The aforementioned

authors allow for the risk-free investment. So we have to choose a normalized version where the portfolio weights sum up to 1,<sup>11</sup> which makes the different investment strategies comparable.

The estimators are calculated on rolling windows with length  $n = 60$  (that is 5 years). For example, the portfolios are constructed at first in time  $t = 60$  on the basis of the 60 monthly returns contained in  $R_1, \dots, R_{60}$ . These portfolios are held constant up to time  $t = 61$  and the corresponding portfolio returns are calculated with  $R_{61}$ . Then the window is switched to  $R_2, \dots, R_{61}$  and the out-of-sample returns of the different strategies are calculated with  $R_{62}$ , etc., until the end of the available data is reached. Therefore we obtain  $N - n = 408$  out-of-sample returns for each of the considered strategies. These out-of-sample returns are used to estimate the out-of-sample variances of each strategy.

Because the observation periods are overlapping, each time series of out-of-sample returns exhibits a serial dependence structure. Hence, traditional hypothesis tests which require stochastically independent data would be inadequate. This is the reason why we apply the circular moving-blocks bootstrap.<sup>12</sup> Block-bootstrap procedures guarantee that the considered hypothesis tests remain asymptotically valid even for serially dependent time series (Politis, 2003) and have become popular instruments in financial data analysis (Ledoit and Wolf, 2008). These procedures are non-parametric and so they work under mild regularity conditions. This means no parametric assumption regarding the stationary distribution or the serial dependence structure of the considered time series is necessary. Hence, the results of our bootstrap study are not subject to model misspecification, which could be the normality assumption or the serial independence assumption.

Table 3 gives the estimated out-of-sample variances and their differences between each strategy. For example, the differences between the estimated out-of-sample variance of the traditional estimator and the shrinkage estimators amount to  $1.301 \cdot 10^{-5}$ .<sup>13</sup> The given one-sided  $p$ -values are always understood to be in favor of the strategy which produces the smaller out-of-sample variance. This means, e.g., the null hypotheses that the true out-of-sample variance of the traditional estimator is smaller than the true out-of-sample variances

---

<sup>11</sup>So the Kan-Zhou estimator reduces to a “two-fund portfolio estimator” and represents a linear combination of the traditional “plug-in” estimator for the tangential portfolio and the traditional MVP estimator.

<sup>12</sup>The number of bootstrap replications in our empirical study is 10 000 and the block length corresponds to 120 (Politis and Romano, 1992).

<sup>13</sup>In our study the realizations of  $\hat{w}_S$  and  $\hat{w}_M$  were identical in each rolling window.

Strategy	$\hat{w}_T$	$\hat{w}_S$	$\hat{w}_M$	$\hat{w}_C$	$\hat{w}_{KZ}$
Variance	$1.595 \cdot 10^{-3}$	$1.582 \cdot 10^{-3}$	$1.582 \cdot 10^{-3}$	$1.676 \cdot 10^{-3}$	$4.293 \cdot 10^{-2}$
Standard error	$1.426 \cdot 10^{-4}$	$1.344 \cdot 10^{-4}$	$1.344 \cdot 10^{-4}$	$1.165 \cdot 10^{-4}$	$2.245 \cdot 10^{-2}$
$\hat{w}_S$	$1.301 \cdot 10^{-5}$ (0.1798)	—	—	—	—
$\hat{w}_M$	$1.301 \cdot 10^{-5}$ (0.1798)	0 (0.5000)	—	—	—
$\hat{w}_C$	$-8.145 \cdot 10^{-5}$ (0.2231)	$-9.447 \cdot 10^{-5}$ (0.1574)	$-9.447 \cdot 10^{-5}$ (0.1598)	—	—
$\hat{w}_{KZ}$	$-4.133 \cdot 10^{-2}$ (0.0330)	$-4.135 \cdot 10^{-2}$ (0.0327)	$-4.135 \cdot 10^{-2}$ (0.0333)	$-0.413 \cdot 10^{-2}$ (0.0320)	—
$w_N$	$-5.502 \cdot 10^{-4}$ (0.0103)	$-5.632 \cdot 10^{-4}$ (0.0062)	$-5.632 \cdot 10^{-4}$ (0.0065)	$-4.687 \cdot 10^{-4}$ (0.0003)	$0.408 \cdot 10^{-2}$ (0.0347)

Table 3: Estimated out-of-sample variances and their differences between each strategy. The indices “T”, “S”, “M”, “C”, “KZ”, and “N” denote the traditional estimator, the shrinkage estimator, the modified shrinkage estimator, the constrained estimator for the MVP as well as the Kan-Zhou estimator and the naive portfolio, respectively. The one-sided  $p$ -values are provided in parentheses in favor of the respective strategy which produces the smaller out-of-sample variance.

of each shrinkage estimator could be rejected on a significance level of 20%. In particular, the out-of-sample variance of the naive portfolio is significantly larger than the out-of-sample variances of the shrinkage estimators and all other strategies – except for the Kan-Zhou estimator – on a significance level of almost 1%. The constrained estimator for the MVP is outperformed by the unconstrained traditional MVP estimator and the shrinkage estimators, but the results are not very significant. By contrast, the constrained estimator significantly outperforms the Kan-Zhou estimator and the naive portfolio. Finally, the Kan-Zhou estimator is significantly outperformed by all other investment strategies. However, it is worth emphasizing that the Kan-Zhou estimator is not designed for minimizing the out-of-sample variance but for maximizing the out-of-sample Sharpe ratio (Kan and Zhou, 2007) and so this result is not surprising.

Table 3 reveals that the shrinkage estimators outperform all other strategies. However, the results are significant on a level of 5% only with respect to the Kan-Zhou estimator and the naive portfolio. Admittedly it is difficult to obtain always *significant* results in the context of financial time series all the more because non-parametric block-bootstrap procedures per se require a large number of observations.

## 6 Conclusion

We present two shrinkage estimators for the MVP 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 are 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 MVP 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 true MVP) caused by naive diversification. An empirical study for 7 major stock markets demonstrates the superiority of the shrinkage estimators with respect to naive diversification, the Kan-Zhou estimator, and – to some extent – with respect to the constrained and unconstrained traditional MVP estimators.

## Appendix

### Lemma 2

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

$$\mathbb{E}\left\{\chi_q^{-2}(\lambda)\right\} = q \mathbb{E}\left\{\chi_{q+2}^{-4}(\lambda)\right\} + 2\lambda \mathbb{E}\left\{\chi_{q+4}^{-4}(\lambda)\right\}, \quad (14)$$

and if  $q \geq 3$ ,

$$(q-2) \mathbb{E}\left\{\chi_q^{-2}(\lambda)\right\} = (q-2\lambda) \mathbb{E}\left\{\chi_{q+2}^{-2}(\lambda)\right\} + 2\lambda \mathbb{E}\left\{\chi_{q+4}^{-2}(\lambda)\right\}. \quad (15)$$

Proof: Eq. 14 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  $\phi(x) = x^{-1}$ ,

$$1 = q \mathbb{E}\left\{\chi_{q+2}^{-2}(\lambda)\right\} + 2\lambda \mathbb{E}\left\{\chi_{q+4}^{-2}(\lambda)\right\} = (q-2) \mathbb{E}\left\{\chi_q^{-2}(\lambda)\right\} + 2\lambda \mathbb{E}\left\{\chi_{q+2}^{-2}(\lambda)\right\}$$

for any  $q \geq 3$ , which leads to (15). 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

$$\mathbb{E}\left[\left(\text{tr } V^{-1} - \frac{\lambda}{\hat{\lambda}} \cdot q\right) \mathbb{E}\left\{\chi_{q+2}^{-2}(\hat{\lambda}) \mid V\right\}\right] = \frac{q-1}{m-q-1} \cdot \mathbb{E}\left[(q-2) \cdot \frac{\lambda}{\hat{\lambda}} \cdot \mathbb{E}\left\{\chi_q^{-2}(\hat{\lambda}) \mid V\right\}\right]$$

and

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

Proof: Consider the function  $h(2\hat{\lambda}) = \mathbb{E} \left\{ \chi_{q+2}^{-2}(\hat{\lambda}) \mid V \right\}$  and note that, after rotating  $\theta$ , 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),

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

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

$$\mathbb{E} \left\{ \left( \text{tr } V^{-1} - \frac{\lambda}{\hat{\lambda}} \cdot q \right) h(2\hat{\lambda}) \right\} = \frac{q-1}{m-q-1} \cdot \left[ q \mathbb{E} \left\{ \frac{h(2\hat{\lambda})}{\chi^2} \right\} + 2 \theta' \theta \mathbb{E} \left\{ h'(2\hat{\lambda}) \right\} \right], \quad (16)$$

where

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

which follows from the derivative rule on page 327 in Judge and Bock (1978). After substituting  $h'(2\hat{\lambda})$  in (16) and some re-arrangement, we obtain

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

Now the first statement of the lemma appears immediately after applying (15). Similarly, by allowing for the function  $h(2\hat{\lambda}) = \mathbb{E} \left\{ \chi_{q+2}^{-4}(\hat{\lambda}) \mid V \right\}$  and using (14), 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  $Y := \Omega^{\frac{1}{2}}(X - x) \mid V \sim \mathcal{N}_q(\theta, V^{-1})$  with  $V := \Omega^{-\frac{1}{2}} W \Omega^{-\frac{1}{2}} \sim W_q(I_q, m)$  and similarly

$$Y_S := \Omega^{\frac{1}{2}}(X_S - x) = \left( 1 - \frac{c\chi^2}{Y' V Y} \right) Y.$$

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

$$\mathbb{E}\{\mathcal{L}_\theta(Y) - \mathcal{L}_\theta(Y_S)\} = 2ck\mathcal{E}_1 - c^2k(k+2)\mathcal{E}_2 > 0, \quad (17)$$

where

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

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

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

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

$$\mathbb{E}\left\{\frac{Y'(Y - \theta)}{Y'VY} | V\right\} = (\text{tr } V^{-1})\mathbb{E}\left\{\chi_{q+2}^{-2}(\hat{\lambda}) | V\right\} + 2\lambda\mathbb{E}\left\{\chi_{q+4}^{-2}(\hat{\lambda}) | V\right\} - 2\lambda\mathbb{E}\left\{\chi_{q+2}^{-2}(\hat{\lambda}) | V\right\}.$$

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

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

After some re-arrangement and an application of (15) we obtain

$$\begin{aligned} \mathbb{E}\left(\frac{Y'(Y - \theta)}{Y'VY} | V\right) &= (q-2) \cdot \frac{\lambda}{\hat{\lambda}} \cdot \mathbb{E}\left\{\chi_q^{-2}(\hat{\lambda}) | V\right\} + \\ &\quad \left(\text{tr } V^{-1} - \frac{\lambda}{\hat{\lambda}} \cdot q\right) \mathbb{E}\left\{\chi_{q+2}^{-2}(\hat{\lambda}) | V\right\}. \end{aligned}$$

Moreover, with an application of (14) it also turns out that

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

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

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

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 (17) 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:  $m = n-1$ ,  $q = d-1$ ,  $W = n\hat{\Omega}/\sigma^2$ ,  $X = \hat{w}_T^{\text{ex}}$ ,  $\chi^2 = n\hat{\sigma}_T^2/\sigma^2$ ,  $k = n-d$ , and  $x = w_R^{\text{ex}}$ . Then the constant

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

leads to a dominating shrinkage estimator  $\hat{w}_S^{\text{ex}}$  for  $w^{\text{ex}}$ , viz

$$\hat{w}_S^{\text{ex}} = w_R^{\text{ex}} + \left(1 - \frac{d-3}{n-d+2} \cdot \frac{\hat{\sigma}_T^2}{(\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})' \hat{\Omega} (\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})}\right) (\hat{w}_T^{\text{ex}} - w_R^{\text{ex}}).$$

Note that

$$(\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})' \hat{\Omega} (\hat{w}_T^{\text{ex}} - w_R^{\text{ex}}) = (\hat{w}_T - w_R)' \hat{\Sigma} (\hat{w}_T - w_R)$$

and thus

$$\frac{\hat{\sigma}_T^2}{(\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})' \hat{\Omega} (\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})} = \frac{\hat{\sigma}_T^2}{(\hat{w}_T - w_R)' \hat{\Sigma} (\hat{w}_T - w_R)} = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2 - \hat{\sigma}_T^2} = \frac{1}{\hat{\tau}_R}.$$

Due to  $\hat{w}_S = \mathbf{e}_1 - \Delta' \hat{w}_S^{\text{ex}}$  it follows that

$$\hat{w}_S = w_R + \left(1 - \frac{d-3}{n-d+2} \cdot \frac{1}{\hat{\tau}_R}\right) (\hat{w}_T - w_R) = \kappa_S w_R + (1 - \kappa_S) \hat{w}_T.$$

Q.E.D.

### Proof of Theorem 3

After some calculations we find that

$$\tau_S = \tau_R - 2(1 - \kappa_S)a + (1 - \kappa_S)^2 b,$$

where

$$\kappa_S = \frac{d-3}{n-d+2} \cdot \frac{n\hat{\sigma}_T^2/\sigma^2}{(\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})' (n\hat{\Omega}/\sigma^2) (\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})},$$

$$a = \frac{(\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})' \Omega (w^{\text{ex}} - w_R^{\text{ex}})}{\sigma^2} \quad \text{and} \quad b = \frac{(\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})' \Omega (\hat{w}_T^{\text{ex}} - w_R^{\text{ex}})}{\sigma^2}.$$

With  $\theta = \Omega^{\frac{1}{2}}/\sigma (w^{\text{ex}} - w_R^{\text{ex}})$ ,  $\xi \sim \mathcal{N}_{d-1}(\mathbf{0}, I_{d-1})$ , and  $V \sim W_{d-1}(I_{d-1}, n-1)$ , the shrinkage weight  $\kappa_S$  can be represented by

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

as well as  $a = \theta'(\theta + V^{-\frac{1}{2}}\xi)$  and  $b = (\theta + V^{-\frac{1}{2}}\xi)'(\theta + V^{-\frac{1}{2}}\xi)$ , where  $\xi$ ,  $V$ , and  $\chi_{n-d}^2$  are mutually independent. Hence,  $\tau_S$  is equal to the expression given on the right hand side of (9). Moreover, it holds that

$$\tau_S = \|\mathcal{O}\{\kappa_S\theta - (1 - \kappa_S)V^{-\frac{1}{2}}\xi\}\|^2 = \|\kappa_S\eta - (1 - \kappa_S)\mathcal{O}V^{-\frac{1}{2}}\xi\|^2$$

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

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

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,  $\xi$  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  $\eta$  of  $\theta$  it holds that

$$\tau_S \sim \|\kappa_S\eta - (1 - \kappa_S)V^{-\frac{1}{2}}\xi\|^2.$$

Ergo, the distribution of  $\tau_S$  depends only on  $n$ ,  $d$ , and  $\tau_R = \theta'\theta$ .

Q.E.D.

#### Proof of Theorem 4

From the proof of Theorem 3 it follows that the distribution of  $\tau_M$ , too, is only a function of  $n$ ,  $d$ , and  $\tau_R$ . To prove that  $E(\tau_M) < E(\tau_S)$ , the relative loss of the simple shrinkage estimator can be written as

$$\tau_S = \tau_R - 2\theta'V^{-\frac{1}{2}}(1 - \kappa_S)(V^{\frac{1}{2}}\theta + \xi) + (1 - \kappa_S)^2\|V^{\frac{1}{2}}\theta + \xi\|_V^2.$$

Since  $(1 - \kappa_S) = (1 - \kappa_S)^+ - (1 - \kappa_S)^-$ , the relative loss of the modified shrinkage estimator becomes

$$\tau_M = \tau_S - 2\theta'V^{-\frac{1}{2}}(1 - \kappa_S)^-(V^{\frac{1}{2}}\theta + \xi) - \{(1 - \kappa_S)^-\}^2\|V^{\frac{1}{2}}\theta + \xi\|_V^2.$$

Here it holds that

$$E\left[\{(1 - \kappa_S)^-\}^2\|V^{\frac{1}{2}}\theta + \xi\|_V^2\right] > 0$$

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

$$E\left\{\theta'V^{-\frac{1}{2}}(1 - \kappa_S)^-(V^{\frac{1}{2}}\theta + \xi)\right\} = \tau_R E\left[\left\{1 - \frac{d-3}{n-d+2} \cdot \frac{\chi_{n-d}^2}{\chi_{d+1}^2(\tau_R\chi_{n-1}^2/2)}\right\}^-\right] \geq 0.$$

This means  $E(\tau_M) < E(\tau_S)$ . The second inequality  $E(\tau_S) < E(\tau_T)$  is a direct consequence of Theorem 2.

Q.E.D.

### Proof of Theorem 5

The traditional estimator for the MVP without the first portfolio weight can be represented by  $\hat{w}_T^{\text{ex}} = w^{\text{ex}} + \sigma \Omega^{-\frac{1}{2}} V^{-\frac{1}{2}} \xi$ , 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  $\sqrt{n} V^{-\frac{1}{2}} = (V/n)^{-\frac{1}{2}} \xrightarrow{\text{a.s.}} I_{d-1}$  as  $n \rightarrow \infty$ , it holds that

$$\sqrt{n} (\hat{w}_T^{\text{ex}} - w^{\text{ex}}) \xrightarrow{\text{a.s.}} \sigma \Omega^{-\frac{1}{2}} \xi, \quad n \rightarrow \infty.$$

The presented expression for the asymptotic normality of  $\hat{w}_T = \mathbf{e}_1 - \Delta' \hat{w}_T^{\text{ex}}$  follows from the relationship  $\sigma^2 \Delta' \Omega^{-1} \Delta = \sigma^2 \Sigma^{-1} - w w'$  (Frahm, 2008). Further, the shrinkage estimator can be represented by

$$\hat{w}_S^{\text{ex}} = w_R^{\text{ex}} + \left\{ 1 - \frac{d-3}{n-d+2} \cdot \frac{\chi_{n-d}^2}{(\theta + V^{-\frac{1}{2}} \xi)' V (\theta + V^{-\frac{1}{2}} \xi)} \right\} \left\{ (w^{\text{ex}} - w_R^{\text{ex}}) + \sigma \Omega^{-\frac{1}{2}} V^{-\frac{1}{2}} \xi \right\},$$

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

$$\frac{\theta' V \theta}{n} = \tau_R \cdot \frac{\chi_{n-1}^2}{n} \xrightarrow{\text{a.s.}} \tau_R, \quad \frac{2\theta' V^{\frac{1}{2}} \xi}{n} = 2\theta' (V/n)^{\frac{1}{2}} \xi / \sqrt{n} \xrightarrow{\text{a.s.}} 0, \quad \frac{\xi' \xi}{n} \xrightarrow{\text{a.s.}} 0, \quad n \rightarrow \infty,$$

it follows that  $(\theta + V^{-\frac{1}{2}} \xi)' V (\theta + V^{-\frac{1}{2}} \xi) / n \xrightarrow{\text{a.s.}} \tau_R$  as well as  $\chi_{n-d}^2 / n \xrightarrow{\text{a.s.}} 1$  as  $n \rightarrow \infty$ . Hence, in the event that  $\tau_R > 0$  it holds that

$$\sqrt{n} \cdot \frac{d-3}{n-d+2} \cdot \frac{\chi_{n-d}^2 / n}{(\theta + V^{-\frac{1}{2}} \xi)' V (\theta + V^{-\frac{1}{2}} \xi) / n} \cdot (w_R^{\text{ex}} - w^{\text{ex}}) \xrightarrow{\text{a.s.}} \mathbf{0}, \quad n \rightarrow \infty.$$

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

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

By contrast, if  $\tau_R = 0$  and thus  $\theta = \mathbf{0}$  as well as  $w^{\text{ex}} = w_R^{\text{ex}}$ ,

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

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

$$\sqrt{n} (\hat{w}_S^{\text{ex}} - w^{\text{ex}}) \xrightarrow{\text{a.s.}} \left( 1 - \frac{d-3}{\xi' \xi} \right) \sigma \Omega^{-\frac{1}{2}} \xi, \quad n \rightarrow \infty.$$

Similar arguments hold for the modified shrinkage estimator, since

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

if  $\tau_R > 0$  and otherwise

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

Q.E.D.

### Proof of Theorem 6

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

$$n\tau_T = \frac{\sqrt{n}(\hat{w}_T^{\text{ex}} - w^{\text{ex}})' \Omega \sqrt{n}(\hat{w}_T^{\text{ex}} - w^{\text{ex}})}{\sigma^2}.$$

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

$$\left[ \mathbb{1}_{\{\tau_R=0\}}X + \mathbb{1}_{\{\tau_R>0\}} \right]^2 = \mathbb{1}_{\{\tau_R=0\}}X^2 + \mathbb{1}_{\{\tau_R>0\}}$$

for any random variable  $X$ .

Q.E.D.

### Proof of Theorem 7

Due to the proof of Theorem 5 it holds that

$$\tau_T = \frac{(\hat{w}_T^{\text{ex}} - w^{\text{ex}})' \Omega (\hat{w}_T^{\text{ex}} - w^{\text{ex}})}{\sigma^2} = \xi' V^{-1} \xi = \frac{\chi_{d-1}^2}{\chi_{n-d+1}^2}$$

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

$$\tau_T = \frac{d}{n-d} \cdot \frac{\chi_{d-1}^2/d}{\chi_{n-d+1}^2/(n-d)} \xrightarrow{\text{a.s.}} \frac{1}{q-1}, \quad n, d \longrightarrow \infty, n/d \longrightarrow q.$$

For proving the almost sure convergence of the shrinkage weights  $\kappa_S$  and  $\kappa_M$ , consider  $\theta = (\sqrt{\tau_R}, \mathbf{0})$  and suppose that  $V^{\frac{1}{2}}$  is the Cholesky root of  $V$ , i.e.,

$$\theta' V^{\frac{1}{2}} \xi = \sqrt{\tau_R} \chi_{n-1} \xi_1.$$

Furthermore, note that  $(d-3)/(n-d+2) \rightarrow 1/(q-1)$ ,  $\chi_{n-d}^2/(n-d) \xrightarrow{\text{a.s.}} 1$ ,

$$\frac{\theta' V \theta}{n-d} = \tau_R \cdot \frac{\chi_{n-1}^2}{n} \cdot \frac{n}{n-d} \xrightarrow{\text{a.s.}} \frac{q\tau_R}{q-1}, \quad \frac{2\theta' V^{\frac{1}{2}} \xi}{n-d} = 2\sqrt{\tau_R} \cdot \frac{\chi_{n-1} \xi_1}{n-d} \xrightarrow{\text{a.s.}} 0$$

as well as

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

Now, by applying the continuous mapping theorem, we obtain  $\kappa_S, \kappa_M \xrightarrow{\text{a.s.}} 1/(1+q\tau_R)$  as  $n, d \rightarrow \infty$  and  $n/d \rightarrow q$ . Similarly, note that

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

and  $\xi' V^{-1} \xi \xrightarrow{\text{a.s.}} 1/(q-1)$  as  $n, d \rightarrow \infty$  and  $n/d \rightarrow q$ . By relying on (9) and (11) it turns out that

$$\tau_S, \tau_M \xrightarrow{\text{a.s.}} \frac{\tau_R}{1+q\tau_R} - \left(1 - \frac{1}{1+q\tau_R}\right) \tau_R + \left(1 - \frac{1}{1+q\tau_R}\right)^2 \left(\tau_R + \frac{1}{q-1}\right).$$

After a little calculation it can be found that the limit corresponds to the asymptotic loss function  $L(\tau_R; q)$  which is given in the theorem. Q.E.D.

### Proof of Theorem 8

Since  $w_R' \mathbf{1} = 1 > 0$ , the angle between  $w_R$  and  $\mathbf{1}$  is acute. Therefore, there exists an orthogonal  $d \times d$  matrix  $\mathcal{O}$  such that  $\mathcal{O}w_R$  and  $\mathcal{O}\mathbf{1}$  belong to the set  $\{x \in \mathbb{R}^d : x > \mathbf{0}\}$ . This means there also exists a positive-definite diagonal  $d \times d$  matrix  $\Lambda$  such that  $\mathcal{O}\mathbf{1} = \Lambda \mathcal{O}w_R$ , i.e.,  $w_R = A\mathbf{1}$  with  $A := \mathcal{O}'\Lambda^{-1}\mathcal{O}$  being positive definite. The matrix  $\Sigma_R^{-1}$  can be obtained by re-scaling  $A$  such that the condition  $\mathbf{1}'\Sigma_R^{-1}\mathbf{1} = \mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1} > 0$  is satisfied. Now the rest of the theorem can be verified by substituting  $\hat{\Sigma}^{-1}$  by the given expressions for  $\hat{\Sigma}_S^{-1}$  and  $\hat{\Sigma}_M^{-1}$  within the traditional MVP estimator. Q.E.D.

### Proof of Theorem 9

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

$$\hat{\tau}_R = \frac{(V^{\frac{1}{2}}\theta + \xi)'(V^{\frac{1}{2}}\theta + \xi)}{\chi_{n-d}^2}$$

and note that  $\theta' V \theta = \tau_R \chi_{n-1}^2$ . Q.E.D.

## References

- Aparicio, F., Estrada, J., 2001. Empirical distributions of stock returns: European securities markets, 1990–95. *European Journal of Finance* 7, 1–21.
- Chan, L., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: forecasting covariances and choosing the risk model. *Review of Financial Studies* 12, 937–974.
- Chopra, V., Ziemba, W., 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* 19, 6–11.
- Cogley, T., Sargent, T., 2008. The market price of risk and the equity premium: A legacy of the great depression. *Journal of Monetary Economics* 55, 454–476.
- DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the  $1/n$  portfolio strategy? *Review of Financial Studies* 22, 1915–1953.
- Dimson, E., Marsh, P., Staunton, M., 2003. Global evidence on the equity risk premium. *Journal of Applied Corporate Finance* 15, 27–38.
- Elton, E., Gruber, M., 1973. Estimating the dependence structure of share prices – Implications for portfolio selection. *Journal of Finance* 28, 1203–1232.
- Frahm, G., 2008. Linear statistical inference for global and local minimum variance portfolios. *Statistical Papers*, DOI: 10.1007/s00362-008-0170-z.
- Frost, P., Savarino, J., 1986. An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis* 21, 293–305.
- Garlappi, L., Uppal, R., Wang, T., 2007. Portfolio selection with parameter and model uncertainty: a multi-prior approach. *Review of Financial Studies* 20, 41–81.
- Golosnoy, V., Okhrin, Y., 2007. Multivariate shrinkage for optimal portfolio weights. *The European Journal of Finance* 13, 441–458.
- Haugen, R., 1990. Building a better index: Cap-weighted benchmarks are inefficient vehicles. *Pension and Investments* 18, 1–3.
- Haugen, R., Baker, N., 1991. The efficient market inefficiency of capitalization-weighted stock portfolios. *Journal of Portfolio Management* 17, 35–40.

- Haugen, R., Baker, N., 1993. Interpreting the evidence on risk and expected return: comment. *Journal of Portfolio Management* 19, 36–43.
- Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance* 58, 1651–1683.
- Jobson, J., Korkie, B., 1979. Improved estimation for Markowitz portfolios using James-Stein type estimators. In: *Proceedings of the American Statistical Association (Business and Economic Statistics)*. Vol. 1. pp. 279–284.
- Jorion, P., 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21, 279–292.
- Judge, G., Bock, M., 1978. *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. North-Holland Publishing Company.
- Kan, R., Zhou, G., 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42, 621–656.
- Kempf, A., Memmel, C., 2006. Estimating the global minimum variance portfolio. *Schmalenbach Business Review* 58, 332–348.
- Ledoit, O., Wolf, M., 2001. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88, 365–411.
- Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10, 603–621.
- Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance* 15, 850–859.
- Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7, 77–91.
- McNeil, A., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management*. Princeton University Press.
- Memmel, C., 2004. *Schätzrisiken in der Portfoliotheorie*. Ph.D. thesis, University of Cologne, Department of Economic and Social Statistics, Germany.
- Merton, R., 1980. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* 8, 323–361.

- Okhrin, Y., Schmid, W., 2006. Distributional properties of portfolio weights. *Journal of Econometrics* 134, 235–256.
- Politis, D., 2003. The impact of bootstrap methods on time series analysis. *Statistical Science* 18, 219–230.
- Politis, D., Romano, J., 1992. A circular block-resampling procedure for stationary data. In: LePage, R., Billard, L. (Eds.), *Exploring the Limits of Bootstrap*. Wiley, pp. 263–270.
- Press, S., 2005. *Applied Multivariate Analysis*, 2nd Edition. Dover Publications.
- Sharpe, W., 1963. A simplified model for portfolio analysis. *Management Science* 9, 277–293.
- Srivastava, M., Bilodeau, M., 1989. Stein estimation under elliptical distributions. *Journal of Multivariate Analysis* 28, 247–259.
- Stein, C., 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*. Vol. 1. pp. 197–206.
- Winston, K., 1993. The efficient index and prediction of portfolio variance. *Journal of Portfolio Management* 19, 27–34.