DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 1/10

An Analytical Investigation of Estimators for Expected Asset Returns from the Perspective of Optimal Asset Allocation

by

Gabriel Frahm

1st version
March 4, 2010

DISKUSSIONSBEITRÄGE ZUR STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

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

This page intentionally left blank.

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 1/10

An Analytical Investigation of Estimators for Expected Asset Returns from the Perspective of Optimal Asset Allocation

by

Gabriel Frahm[1]

1st version
March 4, 2010

Abstract

In the present work I derive the risk functions of 5 standard estimators for expected asset returns which are frequently advocated in the literature, viz the sample mean vector, the James-Stein and Bayes-Stein estimator, the minimum-variance estimator, and the CAPM estimator. I resolve the question why it is meaningful to study the risk function in the context of optimal asset allocation. Further, I derive the quantities which determine the risks of the different expected return estimators and show which estimators are preferable with respect to optimal asset allocation. Finally, I discuss the question whether it pays to strive for the optimal portfolio by using time series information. It turns out that in many practical situations it is better to renounce parameter estimation altogether and pursue some trivial strategy such as the totally risk-free investment.

Keywords: Asset allocation, Bayes-Stein estimator, CAPM estimator, James-Stein estimator, Minimum-variance estimator, Naive diversification, Out-of-sample performance, Risk function, Shrinkage estimation.

JEL classification: C13, G11.

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

This page intentionally left blank.

An Analytical Investigation of Estimators for Expected Asset Returns from the Perspective of Optimal Asset Allocation*

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

Abstract

In the present work I derive the risk functions of 5 standard estimators for expected asset returns which are frequently advocated in the literature, viz the sample mean vector, the James-Stein and Bayes-Stein estimator, the minimum-variance estimator, and the CAPM estimator. I resolve the question why it is meaningful to study the risk function in the context of optimal asset allocation. Further, I derive the quantities which determine the risks of the different expected return estimators and show which estimators are preferable with respect to optimal asset allocation. Finally, I discuss the question whether it pays to strive for the optimal portfolio by using time series information. It turns out that in many practical situations it is better to renounce parameter estimation altogether and pursue some trivial strategy such as the totally risk-free investment.

JEL classification: C13, G11.

Keywords: Asset allocation, Bayes-Stein estimator, CAPM estimator, James-Stein estimator, Minimum-variance estimator, Naive diversification, Out-of-sample performance, Risk function, Shrinkage estimation.

*I would like to thank Christoph Memmel for his very helpful comments on the manuscript and Tobias Wickern for kindly providing the data which has been used in the empirical study.

†Phone: +49 221 470-4267, e-mail: frahm@statistik.uni-koeln.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. The estimation of expected asset returns is the Achilles' heel in the implementation of portfolio optimization (Chopra and Ziemba, 1993; Merton, 1980). If the estimates are based only on time series information, it is well-known that 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.[1]

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. The same conclusion might be drawn for any value-weighted stock-index portfolio. In my opinion the underlying problem is twofold:

(i) Even if the number n of observations is large, the impact of errors due to covariance matrix estimation can be substantial if the number d of assets is large, too, i.e. if the quantity n/d is small (Frahm, 2008; Jagannathan and Ma, 2003; Kempf and Memmel, 2006; Wolf, 2007). However, errors which are due to covariance matrix estimation can be principally reduced by taking short-term data, such as daily or weekly asset returns, into consideration.[2]

(ii) Unfortunately, increasing the sampling frequency does not work for estimating expected asset returns. This argument has been already pointed out by Merton (1980). See also Jorion (1985) as well as Chopra and Ziemba (1993) for a discussion in the context of optimal asset allocation.

These fundamental problems are still a matter of debate. In the following I will concentrate on the estimation of expected returns. During the past three decades several alternatives to

[1]See, e.g., Bade et al. (2008); DeMiguel et al. (2009); Frahm and Memmel (2010); Frost and Savarino (1986, 1988); Garlappi et al. (2007); Jobson and Korkie (1979); Jorion (1986); Kan and Zhou (2007); Kempf and Memmel (2006); Ledoit and Wolf (2003) and the references therein.

[2]When increasing the sampling frequency, asset returns cannot be assumed to be normally distributed or serially independent. In that case alternative methods for estimating the covariance matrix, such as robust covariance matrix estimation (see, e.g., Frahm, 2009) should be applied.

the usual sample mean estimator have been proposed. In the present work I will investigate the following standard estimators, which are frequently advocated in the literature:

- The sample mean estimator,
- the James-Stein estimator and the Bayes-Stein estimator,
- the minimum-variance estimator, and
- the estimator based on the Capital Asset Pricing Model.

This work contributes to the literature by providing exact *analytical* results. By contrast, the results which can be found in the literature are typically based on Monte Carlo simulations, re-sampling methods such as (block-)bootstrapping, or empirical investigation. These procedures suffer from well-known drawbacks. For instance, it is hardly possible to generalize results which have been obtained by Monte Carlo simulation. Re-sampling procedures – especially block-bootstrap procedures which take the serial dependence structure of (out-of-sample) returns into consideration – require very long periods of observed asset returns. Finally, empirical studies often suffer from insignificant results even if the sample sizes are large.

I will study the so-called *risk functions* of the different expected return estimators. This instrument is well-known from statistical decision theory. Jorion (1986) states that 'the computation of the risk function is an arduous task.'. However, the main advantage of using that kind of statistical analysis is that it becomes possible to make general statements without relying on specific parametrical assumptions for numerical simulation or struggling with insignificant results from empirical data. Many important insights can be found by analytical investigation only, in particular understanding the circumstances under which quantitative methods of portfolio optimization will outperform naive diversification or any other trivial strategy.

I will try to provide answers to the following questions:

(i) Why should it be meaningful at all to study the risk function, i.e. a standard instrument of statistical decision theory, in the context of optimal asset allocation?

(ii) Which quantities determine the risk of the different expected return estimators?

(iii) Which estimator is preferable with respect to optimal asset allocation?

(iv) Does it pay to strive for the optimal portfolio by using time series information or is it better to renounce parameter estimation altogether and pursue some trivial strategy such as the totally risk-free investment?

2 Notation and Assumptions

Let $R_t = (R_{1t}, \ldots, R_{dt})$ be a d-dimensional vector of asset returns at time $t = 1, \ldots, n$.[3] More precisely, R_t denotes a vector of excess returns with respect to the risk-free interest rate, but the prefix 'excess' will be dropped for convenience unless otherwise stated. In the following $\mathbf{0}$ denotes a vector of zeros and $\mathbf{1}$ is a vector of ones. Further, I_d is the d-dimensional identity matrix.

In the following $\mathcal{N}_k(\omega, \Omega)$ denotes a k-variate normal distribution with mean vector $\omega \in \mathbb{R}^k$ and positive-definite covariance matrix $\Omega \in \mathbb{R}^{k \times k}$. Further, $\chi^2_k(\lambda)$ denotes a noncentral χ^2-distribution with $k \in \mathbb{N}$ degrees of freedom and noncentrality parameter $\lambda \geq 0$. This means $\chi^2_k(\lambda) \sim X'X$ with $X \sim \mathcal{N}_k(\theta, I_k)$ and $\theta \in \mathbb{R}^k$, where the noncentrality parameter is defined as $\lambda = \theta'\theta$. By contrast, χ^2_k stands for a central χ^2-distribution (i.e. $\lambda = 0$) with k degrees of freedom. The symbols χ^2_k and $\chi^2_k(\lambda)$ are also used to indicate some random variables which are correspondingly distributed.

Suppose that the following assumptions are satisfied:

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

A2. The asset returns are serially independent.

A3. It holds that $\mathbf{1}'\Sigma^{-1}\mu > 0$.

I make the assumption of jointly normally distributed and serially independent asset returns even though there exist by far more advanced time series models (especially for high-frequency data). However, Assumptions **A1** and **A2** are made for three reasons:

(i) These are the standard assumptions in the finance literature.

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

(ii) In workaday portfolio management it is common to use low-frequency data, for example monthly asset returns. For these data, the assumption of joint normality and serial independence can be readily justified (Aparicio and Estrada, 2001; McNeil et al., 2005, p. 122).

(iii) Generally it is difficult to derive small-sample results without the normal distribution assumption but for understanding the impact of estimation errors on optimal asset allocation, studying the small-sample properties of the estimators is crucial (Frahm and Memmel, 2010).

Every vector $v \in \mathbb{R}^d$ is said to be a *portfolio*. The *Sharpe ratio* of a portfolio v is given by

$$\mathrm{Sh}_v = \frac{v'\mu}{\sqrt{v'\Sigma v}}\,.$$

In this work I will often refer to the *tangential portfolio*

$$w_{\mathrm{T}} = \frac{\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mu}$$

and the *global minimum variance portfolio*

$$w_{\mathrm{MV}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\,.$$

Every d-dimensional random vector $\hat{w} = (\hat{w}_1, \ldots, \hat{w}_d) \in \mathcal{C} \subseteq \mathbb{R}^d$ is said to be an *investment strategy*. Here $\mathcal{C}$ denotes a convex set of portfolio-weight constraints. This means every convex combination of different strategies lies also in $\mathcal{C}$.[4] The weight of the risk-free asset follows implicitly by $\hat{w}_0 = 1 - \hat{w}'\mathbf{1}$. This guarantees that the budget constraint is satisfied for every $\hat{w} \in \mathbb{R}^d$.

A deterministic vector $v \in \mathcal{C}$ is referred to as a *fixed strategy*. Further, a strategy which does not depend on empirical data is said to be *trivial*. For instance, every fixed strategy is trivial. Due to **A1** and **A2** the out-of-sample performance of a buy-and-hold strategy $\hat{w}$ (where the chosen portfolio is liquidated after 1 period) is given by

$$\phi(\hat{w}) = \hat{w}'\mu - \frac{\alpha}{2} \cdot \hat{w}'\Sigma\hat{w}\,,$$

[4] The convexity assumption is weak but very useful for technical reasons. For example, otherwise it could not be even guaranteed that $\mathrm{E}(\hat{w}) \in \mathcal{C}$.

where $\alpha > 0$ is the investor's individual risk-aversion parameter. Hence, under parameter uncertainty, the out-of-sample performance is a stochastic quantity unless the investor has a fixed strategy.

Now suppose that

A4. the investor seeks for a buy-and hold strategy $\hat{w} \in \mathcal{C}$ where the chosen portfolio is liquidated after $T \in \mathbb{N}$ periods.

A5. He aims at maximizing the expected out-of-sample performance of $\hat{w}$ given his investment horizon T, i.e.[5]

$$T\mathrm{E}\Big(\hat{w}'\mu - \frac{\alpha}{2} \cdot \hat{w}'\Sigma\hat{w}\Big)\,.$$

Assumption **A5** guarantees that the investor's optimal decision is not determined by his specific investment horizon and in the following I will suppose that $T = 1$ without loss of generality. An alternative way of quantifying the out-of-sample performance of a strategy would be to calculate

$$\varphi(\hat{w}) = \mathrm{E}(\hat{w}'R) - \frac{\alpha}{2} \cdot \mathrm{Var}(\hat{w}'R)\,,$$

but it can be shown that $\phi(\hat{w}) - \varphi(\hat{w})$ is negligible (Frahm and Memmel, 2010). Hence, in the following I will concentrate on analyzing $\phi(\hat{w})$ (see also Kan and Zhou, 2007).

The *optimal portfolio*, i.e. the portfolio which maximizes $\phi(\cdot)$ without any constraints on the portfolio weights, is denoted by

$$w^* = \frac{\Sigma^{-1}\mu}{\alpha}\,. \tag{1}$$

Assumption **A3** guarantees that the expected return of w^* is positive.[6] The out-of-sample performance of w^* corresponds to

$$\phi(w^*) = \frac{\mu'\Sigma^{-1}\mu}{2\alpha}\,.$$

The quantity $\mu'\Sigma^{-1}\mu$ equals to the squared Sharpe ratio (with respect to the given sampling frequency) of w^* or any other portfolio being proportional to w^* such as the tangential

[5] This is based on a simplifying assumption which is frequently used in the finance literature, namely that the sum of T independent and identically distributed returns corresponds to the return after T periods.

[6] **A3** implies also that the expected return of the global minimum variance portfolio is positive. Roughly speaking, this technical assumption guarantees the existence of the efficient frontier.

portfolio w_{T}. Typical values for annualized ex-post Sharpe ratios of stock-index portfolios which can be found in empirical data are between 0.2 and 0.5 (Cogley and Sargent, 2008; Dimson et al., 2003; Jorion, 1991). This means if μ and Σ refer to monthly data, the squared Sharpe ratio might take values up to 2%. By contrast, typical values for the risk aversion parameter α are between 3 and 5 (Chopra and Ziemba, 1993).

Recall that n corresponds to the number of observations and d is the number of assets.

A6. It is supposed that $n > d + 1$ and $d > 1$.

In the following it is assumed that Σ is estimated by the sample covariance matrix $\widehat{\Sigma}$, i.e.

$$\widehat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}\big(R_t - \hat{\mu}\big)\big(R_t - \hat{\mu}\big)'$$

where $\hat{\mu} = 1/n\sum_{t=1}^{n} R_t$ represents the sample mean vector of $R_1, \ldots, R_n$.

Finally, the *loss* of a strategy $\hat{w}$ is defined as

$$\mathcal{L}(\hat{w}) = \phi(w^*) - \phi(\hat{w})$$

and, according to statistical decision theory, $\mathcal{R}(\hat{w}) = \mathrm{E}\big\{\mathcal{L}(\hat{w})\big\}$ is referred to as the *risk* of the investment strategy $\hat{w}$.

3 The Expected Out-of-Sample Performance

Consider a d-dimensional random vector $X \sim \mathcal{N}_d(\mu, \Sigma)$. A typical instrument in statistical decision theory for examining an estimator $\hat{\theta}$ for μ is given by

$$\mathcal{R}(\hat{\theta}) = \mathrm{E}\big\{\ell(\hat{\theta})\big\} = \mathrm{E}\big\{(\hat{\theta} - \mu)'\Sigma^{-1}(\hat{\theta} - \mu)\big\}\,.$$

This is the so-called *risk function* of the estimator $\hat{\theta}$. The risk function depends on the true but unknown parameter μ.[7] Jorion (1986) suggests to use the quadratic loss function $\ell(\hat{\theta})$ for finding better estimators for expected asset returns in the context of portfolio optimization. His arguments lead to the well-known shrinkage estimators, i.e. the James-Stein and Bayes-Stein estimators, which will be investigated below.

Concerning the statistical loss function $\ell(\hat{\theta})$ Jorion (1986) writes 'The use of this loss function is relatively widespread because it leads to tractable results' and 'also, a quadratic

[7] Here the covariance matrix Σ is considered as fixed.

loss is generally a good local approximation of a more general loss function expanded in a Taylor series.'. However, in the context of portfolio optimization investors are typically interested in finding a strategy which leads to a better expected out-of-sample performance. Hence, it might be questionable whether the presented loss function makes sense from the perspective of optimal asset allocation. In the following I will prove that $\ell(\hat{\theta})$ in fact is equivalent to the actual loss function $\mathcal{L}(\hat{w})$ presented above. This means if somebody is interested in maximizing her expected out-of-sample performance, the choice of $\ell(\hat{\theta})$ is not only ad hoc but even *inevitable*.

Before that I will establish a theorem which is useful for quantifying the expected out-of-sample performance of a strategy.

Theorem 1

*Under the assumptions **A1** to **A5** the risk of an investment strategy $\hat{w} \in \mathcal{C}$ amounts to*

$$\mathcal{R}(\hat{w}) = \frac{\alpha}{2}\,\mathrm{E}\big\{(\hat{w} - w^*)'\Sigma\,(\hat{w} - w^*)\big\} = \frac{\alpha}{2}\,\big\{\mathcal{B}(\hat{w}) + \mathcal{V}(\hat{w})\big\}\,,$$

where

$$\mathcal{B}(\hat{w}) = (\bar{w} - w^*)'\Sigma\,(\bar{w} - w^*) \qquad \text{and} \qquad \mathcal{V}(\hat{w}) = \mathrm{E}\big\{(\hat{w} - \bar{w})'\Sigma\,(\hat{w} - \bar{w})\big\}$$

with $\bar{w} = \mathrm{E}(\hat{w})$ and $w^* = \Sigma^{-1}\mu/\alpha$.

Proof: See the appendix.

This means the risk of a strategy can be separated into a part which quantifies the bias, i.e. the systematic deviation of that strategy from the optimal portfolio, and another part which quantifies the variance of the chosen strategy. Moreover, if $\hat{w}_1$ and $\hat{w}_2$ are such that

$$\mathcal{B}(\hat{w}_1) + \mathcal{V}(\hat{w}_1) < \mathcal{B}(\hat{w}_2) + \mathcal{V}(\hat{w}_2)\,,$$

strategy $\hat{w}_1$ outperforms strategy $\hat{w}_2$ for every investment horizon $T \in \mathbb{N}$ and risk-aversion parameter $\alpha > 0$.

Let v be some fixed strategy, i.e. $\mathcal{V}(v) = 0$ and $\hat{w}$ a non-trivial strategy with $\mathcal{B}(\hat{w}) < \mathcal{B}(v)$ but $\mathcal{V}(\hat{w}) > 0$ due to estimation errors. The main problem of implementing Markowitz' theory is that $\mathcal{V}(\hat{w}) > \mathcal{B}(v) - \mathcal{B}(\hat{w})$ in many practical situations, so that the trivial strategy v outperforms the sophisticated strategy $\hat{w}$.

Now consider the random vector $\hat{\theta} = \alpha\Sigma\hat{w}$. This can be interpreted as an estimator for $\theta = \alpha\Sigma w$, where w is the optimal portfolio under the constraint $\mathcal{C}$. The components of θ will be referred to as *implicit returns* (with respect to the constraint $\mathcal{C}$). This is directly motivated by Eq. 1. For instance, in case $\mathcal{C} = \mathbb{R}^d$ it holds that $\theta = \mu$. The vector of implicit returns is the solution of a reverse mean-variance optimization problem. This means instead of deducing w from θ by calculating $w = \Sigma^{-1}\theta/\alpha$, conversely θ is derived from w by calculating $\theta = \alpha\Sigma w$.[8]

Corollary 1

Under the assumptions ***A1*** *to* ***A5*** *the risk of an investment strategy* $\hat{w} \in \mathcal{C}$ *amounts to*

$$\mathcal{R}(\hat{w}) = \frac{1}{2\alpha}\,\mathcal{R}(\hat{\theta}) = \frac{1}{2\alpha}\left\{\mathcal{B}(\hat{\theta}) + \mathcal{V}(\hat{\theta})\right\},$$

where

$$\mathcal{B}(\hat{\theta}) = (\bar{\theta} - \mu)'\Sigma^{-1}(\bar{\theta} - \mu) \qquad \text{and} \qquad \mathcal{V}(\hat{\theta}) = \mathrm{E}\big\{(\hat{\theta} - \bar{\theta})\Sigma^{-1}(\hat{\theta} - \bar{\theta})\big\}$$

with $\hat{\theta} = \alpha\Sigma\hat{w}$ *and* $\bar{\theta} = \mathrm{E}(\hat{\theta})$.

Proof: See the appendix.

Corollary 1 implies that if $\hat{\theta}_1$ and $\hat{\theta}_2$ are two implicit return estimators such that $\mathcal{R}(\hat{\theta}_1) < \mathcal{R}(\hat{\theta}_2)$, investment strategy $\hat{w}_1 = \Sigma^{-1}\hat{\theta}_1/\alpha$ has a higher expected out-of-sample performance than $\hat{w}_2 = \Sigma^{-1}\hat{\theta}_2/\alpha$ for every investment horizon $T \in \mathbb{N}$ and risk-aversion parameter $\alpha > 0$.[9] This is the answer to the first question which has been posed in the introduction:

(i) It is meaningful to study the risk function in the context of optimal asset allocation, since the estimator for expected asset returns which has the smallest risk can be expected to produce the largest expected out-of-sample performance.

For this reason I will focus on calculating the risk functions of the 5 expected return estimators mentioned above. Here it is simply assumed that $\mathcal{C} = \mathbb{R}^d$, i.e. the investor aims at finding the optimal portfolio w^* which is the standard case of modern portfolio theory.

[8] As well this is the basic idea behind the Black-Litterman approach for optimal asset allocation (Black and Litterman, 1992).

[9] A similar result under more restrictive assumptions has been derived by Memmel (2004, p. 75).

4 Estimators for Expected Asset Returns

4.1 The Sample Mean Vector

In most practical applications μ is estimated by the sample mean vector $\hat{\mu}$. The following proposition is a well-known result of multivariate analysis.

Proposition 1

*Under the assumptions **A1** to **A5** the risk of $\hat{\mu}$ amounts to*

$$\mathcal{R}(\hat{\mu}) = \frac{d}{n}\,.$$

Proof: See the appendix.

Hence, if the covariance matrix is assumed to be known, the risk of the strategy $\hat{w} = \Sigma^{-1}\hat{\mu}/\alpha$ amounts to

$$\mathcal{R}(\hat{w}) = \frac{1}{2\alpha} \cdot \frac{d}{n}\,.$$

Since in real-world applications the covariance matrix is unknown, the risk even increases due to the estimation errors produced by the sample covariance matrix $\widehat{\Sigma}$. Closed form expressions for the risk under this circumstance have been derived by Kan and Zhou (2007).

It might be helpful to illustrate the latter result by a sample calculation. Suppose that $n = 60$ monthly asset returns have been observed (i.e. the observation period corresponds to 5 years) and the investor has a risk-aversion parameter of $\alpha = 5$. He can calculate the sample covariance matrix for any number $d < n$ of assets. Suppose that the corresponding stock market consists of $d = 30$ assets. This leads to a risk of

$$\mathcal{R}(\hat{w}) = \frac{1}{2 \cdot 5} \cdot \frac{30}{60} = 0.05\,.$$

Even if the squared Sharpe ratio of the tangential portfolio corresponds to 2%,[10] the expected out-of-sample performance of $\hat{w}$ amounts to

$$\mathrm{E}\big\{\phi(\hat{w})\big\} = \frac{0.02}{2 \cdot 5} - 0.05 = -0.048 < 0 = \mathrm{E}\big\{\phi(\mathbf{0})\big\}\,.$$

This means it is much better to put all the money into the risk-free asset instead of using the sample mean vector for portfolio optimization. I will come back to this point later on.

[10] As already mentioned before, this value is at the upper end of the typical interval which can be observed in many empirical studies.

4.2 The Shrinkage Estimators

James-Stein estimators and Bayes-Stein estimators belong to the class of shrinkage estimators which have a long tradition in portfolio optimization (Jobson and Korkie, 1979; Jorion, 1986). James-Stein estimators (Stein, 1956) are of the form

$$\hat{\theta}_{\mathrm{JS}} = \omega_{\mathrm{JS}}\theta + (1-\omega_{\mathrm{JS}})\hat{\mu}\,,$$

where $\theta \in \mathbb{R}^d$ is a fixed vector and

$$\omega_{\mathrm{JS}} = \min\left\{1, \frac{d-2}{n-d+2}\cdot\frac{1}{(\hat{\mu}-\theta)'\widehat{\Sigma}^{-1}(\hat{\mu}-\theta)}\right\}. \tag{2}$$

By contrast, the Bayes-Stein estimator (Jorion, 1985, 1986) can be written as

$$\hat{\theta}_{\mathrm{BS}} = \omega_{\mathrm{BS}}\theta + (1-\omega_{\mathrm{BS}})\hat{\mu}\,,$$

where

$$\omega_{\mathrm{BS}} = \frac{d+2}{(d+2)+(n-d-2)(\hat{\mu}-\theta)'\widehat{\Sigma}^{-1}(\hat{\mu}-\theta)}\,. \tag{3}$$

Jorion (1986) suggests to use the shrinkage target

$$\theta_{\mathrm{MV}} = \frac{\mathbf{1}'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\cdot\mathbf{1}\,, \tag{4}$$

i.e. the vector of expected asset returns under the assumption that the tangential portfolio w_{T} corresponds to the global minimum variance portfolio. In real-world applications μ and Σ are typically replaced by $\hat{\mu}$ and $\widehat{\Sigma}$, i.e. θ_{MV} is replaced by

$$\hat{\theta}_{\mathrm{MV}} = \frac{\mathbf{1}'\widehat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}}\cdot\mathbf{1}\,.$$

Theorem 2

*Under the assumptions **A1** to **A6** the James-Stein estimator*

$$\hat{\theta}_{\mathrm{JS}} = \omega_{\mathrm{JS}}\theta_{\mathrm{MV}} + (1-\omega_{\mathrm{JS}})\hat{\mu}$$

with shrinkage weight (2) and shrinkage target (4) has risk

$$\mathcal{R}(\hat{\theta}_{\mathrm{JS}}) = \frac{d}{n} - 2\mathrm{E}\left\{\omega\left(\frac{\xi'\xi}{n} + \frac{\Delta_{\mathrm{MV}}\xi_1}{\sqrt{n}}\right)\right\} + \mathrm{E}\left(\omega^2\cdot\frac{\chi'\chi}{n}\right)$$

with

$$\omega = \min\left\{1, \frac{d-2}{n-d+2}\cdot\frac{\zeta'\zeta}{\chi'\chi}\right\},$$

where $\xi \sim \mathcal{N}_d(\mathbf{0}, I_d)$ *and* $\zeta \sim \mathcal{N}_{n-d}(\mathbf{0}, I_d)$ *are stochastically independent,* $\chi = (\sqrt{n}\,\Delta_{\text{MV}} + \xi_1, \xi_2, \ldots, \xi_d)$*, and*

$$\Delta_{\text{MV}} = \sqrt{\text{Sh}_{\text{T}}^2 - \text{Sh}_{\text{MV}}^2}\,.$$

Further, Sh_{T} *is the Sharpe ratio of the tangential portfolio and* Sh_{MV} *is the Sharpe ratio of the global minimum variance portfolio.*

Proof: See the appendix.

The key observation is that the risk of the James-Stein estimator depends on the parameters μ and Σ only through the quantity Δ_{MV}. This can be interpreted as a measure for the distance between the Sharpe ratios Sh_{T} and Sh_{MV}. In contrast to the unknown parameters μ and Σ, the quantities Sh_{T}^2 and Sh_{MV}^2 are much more easy to grasp. As already mentioned, several empirical studies indicate the possible interval for Sh_{T} and note that $\text{Sh}_{\text{MV}}^2 \leq \text{Sh}_{\text{T}}^2$, i.e. $0 \leq \Delta_{\text{MV}} \leq \text{Sh}_{\text{T}}$.

The next theorem implies that the same property holds for the Bayes-Stein estimator, too.

Theorem 3

Under the assumptions ***A1*** *to* ***A6*** *the Bayes-Stein estimator*

$$\hat{\theta}_{\text{BS}} = \omega_{\text{BS}}\theta_{\text{MV}} + (1 - \omega_{\text{BS}})\hat{\mu}$$

with shrinkage weight (3) and shrinkage target (4) has risk

$$\mathcal{R}(\hat{\theta}_{\text{BS}}) = \frac{d}{n} - 2\text{E}\left\{\omega\left(\frac{\xi'\xi}{n} + \frac{\Delta_{\text{MV}}\xi_1}{\sqrt{n}}\right)\right\} + \text{E}\left(\omega^2 \cdot \frac{\chi'\chi}{n}\right)$$

with

$$\omega = \frac{d+2}{d+2+(n-d-2)\chi'\chi/\zeta'\zeta}\,,$$

where $\xi \sim \mathcal{N}_d(\mathbf{0}, I_d)$ *and* $\zeta \sim \mathcal{N}_{n-d}(\mathbf{0}, I_d)$ *are stochastically independent,* $\chi = (\sqrt{n}\,\Delta_{\text{MV}} + \xi_1, \xi_2, \ldots, \xi_d)$*, and*

$$\Delta_{\text{MV}} = \sqrt{\text{Sh}_{\text{T}}^2 - \text{Sh}_{\text{MV}}^2}\,.$$

Further, Sh_{T} *is the Sharpe ratio of the tangential portfolio and* Sh_{MV} *is the Sharpe ratio of the global minimum variance portfolio.*

Proof: See the appendix.

4.3 The Minimum-Variance Estimator

The minimum-variance estimator corresponds to

$$\hat{\theta}_{\mathrm{MV}} = \frac{\mathbf{1}'\widehat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}} \cdot \mathbf{1} \,.$$

According to the next theorem, the risk of the minimum-variance estimator can be split into a bias part $\mathcal{B}(\hat{\theta}_{\mathrm{MV}})$ and a variance part $\mathcal{V}(\hat{\theta}_{\mathrm{MV}}) = \mathcal{V}_{\hat{\mu}}(\hat{\theta}_{\mathrm{MV}}) + \mathcal{V}_{\widehat{\Sigma}}(\hat{\theta}_{\mathrm{MV}})$. Here $\mathcal{V}_{\hat{\mu}}(\hat{\theta}_{\mathrm{MV}})$ represents the impact of $\hat{\mu}$ and $\mathcal{V}_{\widehat{\Sigma}}(\hat{\theta}_{\mathrm{MV}})$ the impact of $\widehat{\Sigma}$.

Theorem 4

*Under the assumptions **A1** to **A6** the minimum-variance estimator*

$$\hat{\theta}_{\mathrm{MV}} = \frac{\mathbf{1}'\widehat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}} \cdot \mathbf{1}$$

has risk

$$\mathcal{R}(\hat{\theta}_{\mathrm{MV}}) = \mathcal{B}(\hat{\theta}_{\mathrm{MV}}) + \mathcal{V}_{\hat{\mu}}(\hat{\theta}_{\mathrm{MV}}) + \mathcal{V}_{\widehat{\Sigma}}(\hat{\theta}_{\mathrm{MV}})$$

with

$$\mathcal{B}(\hat{\theta}_{\mathrm{MV}}) = \Delta^2_{\mathrm{MV}} \,, \quad \mathcal{V}_{\hat{\mu}}(\hat{\theta}_{\mathrm{MV}}) = \frac{1}{n} \,, \quad \textit{and} \quad \mathcal{V}_{\widehat{\Sigma}}(\hat{\theta}_{\mathrm{MV}}) = \frac{(d-1)/n + \Delta^2_{\mathrm{MV}}}{n-d-1} \,,$$

where

$$\Delta_{\mathrm{MV}} = \sqrt{\mathrm{Sh}^2_{\mathrm{T}} - \mathrm{Sh}^2_{\mathrm{MV}}} \,.$$

Further, Sh_{T} *is the Sharpe ratio of the tangential portfolio and* $\mathrm{Sh}_{\mathrm{MV}}$ *is the Sharpe ratio of the global minimum variance portfolio.*

Proof: See the appendix.

4.4 The CAPM Estimator

Let $w \in \mathbb{R}^d$ ($w \neq \mathbf{0}$) be some fixed strategy, e.g. the equally-weighted portfolio $w = \mathbf{1}/d$ or some vector of market capitalizations. Consider the linear regression model

$$R_i = \alpha_i + \beta_i R_w + \varepsilon_i \,, \qquad i = 1, \ldots, d \,,$$

where $R_w = w'R$ is the return of the chosen portfolio w. Further, α_i and β_i are such that R_w and ε_i are uncorrelated as well as $\mathrm{E}(\varepsilon_i) = 0$.[11]

[11] It can be shown that $\alpha'\Sigma^{-1}\alpha = \mathrm{Sh}^2_{\mathrm{T}} - \mathrm{Sh}^2_w$, i.e. if $w = w_{\mathrm{T}}$ it holds that $\alpha = (\alpha_1, \ldots, \alpha_d) = \mathbf{0}$.

The CAPM estimator corresponds to

$$\hat{\theta}_{\mathrm{CAPM}} = \hat{\eta}\hat{\beta}\,,$$

where $\hat{\beta}$ is the vector of OLS estimates for $\beta = (\beta_1, \ldots, \beta_d)$ and $\hat{\eta}$ is the sample mean of the portfolio returns $R_{w1}, \ldots, R_{wn}$, i.e.

$$\hat{\theta}_{\mathrm{CAPM}} = w'\hat{\mu} \cdot \frac{\widehat{\Sigma} w}{w'\widehat{\Sigma} w} = (w'\hat{\mu})\hat{\beta}\,.$$

Theorem 5

Under the assumptions ***A1*** *to* ***A6*** *the CAPM estimator*

$$\hat{\theta}_{\mathrm{CAPM}} = (w'\hat{\mu})\hat{\beta}$$

with $\hat{\beta} = \widehat{\Sigma} w / (w'\widehat{\Sigma} w)$ *has risk*

$$\mathcal{R}(\hat{\theta}_{\mathrm{CAPM}}) = \mathcal{B}(\hat{\theta}_{\mathrm{CAPM}}) + \mathcal{V}_{\hat{\mu}}(\hat{\theta}_{\mathrm{CAPM}}) + \mathcal{V}_{\widehat{\Sigma}}(\hat{\theta}_{\mathrm{CAPM}})$$

with

$$\mathcal{B}(\hat{\theta}_{\mathrm{CAPM}}) = \Delta_w^2\,, \quad \mathcal{V}_{\hat{\mu}}(\hat{\theta}_{\mathrm{CAPM}}) = \frac{1}{n}\,, \quad \textit{and} \quad \mathcal{V}_{\widehat{\Sigma}}(\hat{\theta}_{\mathrm{CAPM}}) = \left(\mathrm{Sh}_w^2 + \frac{1}{n}\right)\frac{d-1}{n-3}\,,$$

where

$$\Delta_w = \sqrt{\mathrm{Sh}_{\mathrm{T}}^2 - \mathrm{Sh}_w^2}$$

with Sh_{T} *being the Sharpe ratio of the tangential portfolio and* Sh_w *the Sharpe ratio of the portfolio* w *.*

Proof: See the appendix.

The preceding theorems provide the answer to the second question which has been posed in the introduction:

(ii) The risks of the different expected return estimators depend only on the number of observations, the number of assets, the Sharpe ratio of the tangential portfolio, and the Sharpe ratio of the reference portfolio of the respective estimator.

5 Numerical Results

Hitherto I derived analytically the risk functions of the different estimators for expected asset returns. Now I will present some numerical approximations for typical values of the number of observations n, the number of assets d, and the Sharpe ratios Sh_{T}, $\mathrm{Sh}_{\mathrm{MV}}$, and Sh_w. Suppose that n is the number of monthly asset returns. As already mentioned, the typical range for the annualized ex-post Sharpe ratios of stock-index portfolios is between 0.2 and 0.5. This means the upper bound for the monthly Sharpe ratio Sh_{T} is about $0.5/\sqrt{12} \approx 0.15$. Since $\mathrm{Sh}_{\mathrm{MV}}, \mathrm{Sh}_w \leq \mathrm{Sh}_{\mathrm{T}}$, it holds that $0 \leq \Delta_{\mathrm{MV}}, \Delta_w \leq \mathrm{Sh}_{\mathrm{T}} \leq 0.15$.

The most simple estimator, i.e. the trivial estimator $\hat{\theta}_{\mathrm{tr}} = \mathbf{0}$ has not been discussed, yet. Its risk is easy to calculate, viz

$$\mathcal{R}(\mathbf{0}) = \mu' \Sigma^{-1} \mu = \mathrm{Sh}_{\mathrm{T}}^2 .$$

The question whether it is possible to beat the standard estimators by the trivial estimator will be investigated later on. First of all I will compare the standard estimators only.

5.1 The Standard Estimators

Table 1 provides the risks of the standard estimators for different numbers of observations and numbers of assets. It is worth pointing out that in real-world applications the risks of the James-Stein estimator will be larger than those which are suggested by Table 1. This is because the values which are contained in the panels are calculated on the basis of Theorem 2, where it is assumed that θ_{MV} is known. However, in practice the vector θ_{MV} is replaced by the minimum-variance estimator $\hat{\theta}_{\mathrm{MV}}$. This means the reported values for the James-Stein estimator represent lower risk bounds, whereas the true risks can be considerably larger in real-world applications. The same argument holds also for the Bayes-Stein estimator.[12]

For comparing the minimum-variance estimator with the CAPM estimator it can be assumed that $\mathrm{Sh}_{\mathrm{MV}} = \mathrm{Sh}_w$ for the sake of simplicity. The values reported in Table 1 are calculated under the assumption that $\mathrm{Sh}_{\mathrm{T}} = 0.15$.[13] The left hand side of the table represents the case where $\Delta_{\mathrm{MV}} = \Delta_w = 0$, i.e. where the standard estimators attain their lower

[12]However, Monte Carlo simulations indicate that the values which are calculated on the basis of Theorem 3 are quite good approximations for the case where θ_{MV} is replaced by $\hat{\theta}_{\mathrm{MV}}$.

[13]Other values of Sh_{T} between 0 and 0.15 produce similar outcomes.

$n=30$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.167	0.833	—	—
JS	0.054	0.223	—	—
BS	**0.038**	0.172	—	—
MV	0.039	0.233	—	—
CAPM	0.042	**0.083**	—	—
$n=60$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.083	0.417	0.833	—
JS	0.025	0.035	0.119	—
BS	0.019	0.101	0.193	—
MV	**0.018**	**0.028**	0.107	—
CAPM	0.019	0.0332	**0.0503**	—
$n=120$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.042	0.208	0.417	0.833
JS	0.012	0.013	0.017	0.060
BS	0.010	0.051	0.103	0.201
MV	**0.009**	**0.011**	**0.014**	0.052
CAPM	**0.009**	0.015	0.021	**0.034**
$n=240$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.021	0.104	0.208	0.417
JS	0.006	**0.005**	0.006	0.008
BS	0.005	0.026	0.051	0.103
MV	**0.004**	**0.005**	**0.005**	**0.007**
CAPM	0.005	0.007	0.010	0.015

$n=30$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.167	0.833	—	—
JS	0.068	0.240	—	—
BS	**0.050**	0.184	—	—
MV	0.062	0.262	—	—
CAPM	0.061	**0.086**	—	—
$n=60$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.083	0.417	0.833	—
JS	0.039	0.054	0.138	—
BS	**0.030**	0.112	0.205	—
MV	0.041	0.052	0.132	—
CAPM	0.040	**0.046**	**0.054**	—
$n=120$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.042	0.208	0.417	0.833
JS	0.024	**0.032**	0.037	0.081
BS	**0.019**	0.062	0.113	0.213
MV	0.031	0.033	0.037	0.075
CAPM	0.031	0.033	**0.034**	**0.038**
$n=240$	$d=5$	$d=25$	$d=50$	$d=100$
SM	0.021	0.104	0.208	0.417
JS	0.015	**0.024**	**0.026**	0.029
BS	**0.013**	0.036	0.062	0.114
MV	0.027	0.027	0.028	0.030
CAPM	0.027	0.027	0.028	**0.028**

Table 1: Risks of the sample mean estimator (SM), James-Stein estimator (JS), Bayes-Stein estimator (BS), minimum-variance estimator (MV), and CAPM estimator (CAPM) for different numbers of asset returns (n) and numbers of assets (d). Further, it is assumed that $\mathrm{Sh}_{\mathrm{T}} = 0.15$ and $\Delta_{\mathrm{MV}} = \Delta_w = 0$ (left hand side) as well as $\Delta_{\mathrm{MV}} = \Delta_w = 0.15$ (right hand side). Bold entries indicate the best estimators within the corresponding panel.

risk bounds. Table 1 shows that the CAPM estimator clearly outperforms the minimum-variance estimator if the number of assets is close to the number of observations. Otherwise the risks of the minimum-variance and the CAPM estimator are not essentially different.

The values on the right hand side of Table 1 are obtained under the worst case scenario $\Delta_{\mathrm{MV}} = \Delta_w = 0.15$. In that case the shrinkage estimators become comparably good as the minimum-variance and the CAPM estimator. This is because the shrinkage estimators are convex combinations of $\hat{\theta}_{\mathrm{MV}}$ and $\hat{\mu}$ and so they can take a benefit from shrinking back to $\hat{\mu}$ if $\hat{\theta}_{\mathrm{MV}}$ is a bad target. Nevertheless, the left hand side of Table 1 represents a more realistic scenario. It can be concluded that the minimum-variance and CAPM estimator dominate the other estimators in most practical situations.[14]

Hence, the answer to the third question which has been posed in the introduction is:

(iii) The minimum-variance estimator and the CAPM estimator exhibit approximately the same risks and are preferable among the standard estimators in most practical situations. The CAPM estimator dominates the minimum-variance estimator only if the number of assets is close to the number of observations.

This means in most practical situations it is best to choose the global minimum variance portfolio or some benchmark portfolio among all non-trivial strategies which have been considered in this work.[15] These results have been obtained *analytically* and not by Monte Carlo simulation, re-sampling or empirical investigation and so they do not suffer from the typical drawbacks. Jorion (1991) compares different investment strategies which correspond to the standard estimators discussed above (except for the James-Stein estimator) by their ex-post Sharpe ratios using long-term financial data from 1931 to 1987. In fact, he finds that the minimum-variance and the CAPM estimator produce the best strategies but has to defend his findings against contrary results which have been reported in the literature (Grauer and Hakansson, 1995). Interestingly, Jorion's empirical evidence perfectly agree with the analytical results presented in this work.

[14] Only if the number of assets is small ($d = 5$), the Bayes-Stein estimator becomes slightly preferable.

[15] More precisely, these risky portfolios are combined with the risk-free asset according to the investor's individual risk-aversion.

5.2 The Trivial Estimator

The risk of any trivial estimator $\hat{\theta}_{\mathrm{tr}} = \theta \in \mathbb{R}^d$ corresponds to

$$\mathcal{R}(\theta) = (\theta - \mu)'\Sigma^{-1}(\theta - \mu)\,,$$

which can be estimated by

$$\widehat{\mathcal{R}}(\theta) = (\theta - \hat{\mu})'\widehat{\Sigma}^{-1}(\theta - \hat{\mu})\,. \tag{5}$$

Theorem 6

*Let $\hat{\theta}_{\mathrm{tr}} = \theta \in \mathbb{R}^d$ be some trivial estimator for $\mu \in \mathbb{R}^d$ and consider the estimator $\widehat{\mathcal{R}}(\theta)$ given by Eq. 5 for its risk $\mathcal{R}(\theta)$. Under the assumptions **A1** to **A6** it holds that*

$$\widehat{\mathcal{R}}(\theta) \sim \frac{\chi^2_d\big\{n\mathcal{R}(\theta)\big\}}{\chi^2_{n-d}}\,,$$

where $\chi^2_d\big\{n\mathcal{R}(\theta)\big\}$ and χ^2_{n-d} are stochastically independent.

Proof: See the appendix.

The lower risk bounds presented on the left hand side of Table 1 allow for clarifying the circumstances under which it is better to choose some trivial estimator rather than a standard estimator. More precisely, if $\mathcal{R}(\theta)$ is smaller than each of the lower risk bounds, it is guaranteed that the trivial estimator outperforms the standard estimators. In the following I will concentrate on the trivial estimator $\hat{\theta}_{\mathrm{tr}} = \mathbf{0} \in \mathbb{R}^d$, which will be compared with the minimum-variance and the CAPM estimator only, due to the arguments given in Section 5.1.

According to Theorem 4, the lower risk bound of the minimum-variance estimator is

$$\inf_{\Delta_{\mathrm{MV}}} \mathcal{R}(\hat{\theta}_{\mathrm{MV}}) = \frac{1}{n} + \frac{(d-1)/n}{n-d-1}\,.$$

Further, from Theorem 5 it can be concluded that the CAPM estimator exhibits minimum risk in case $\Delta_w = 0$, i.e. $\mathrm{Sh}_w = \mathrm{Sh}_{\mathrm{T}}$. Hence, its lower risk bound amounts to

$$\inf_{\Delta_w} \mathcal{R}(\hat{\theta}_{\mathrm{CAPM}}) = \frac{1}{n} + \left(\mathrm{Sh}_{\mathrm{T}}^2 + \frac{1}{n}\right)\frac{d-1}{n-3}\,.$$

This leads to the following theorem.

Theorem 7

*Suppose that the assumptions **A1** to **A6** are satisfied. In case*

$$\mathcal{R}(\mathbf{0}) = \mathrm{Sh}_{\mathrm{T}}^2 < \frac{(n-2)/n}{n-d-1}$$

the risk of the trivial estimator $\hat{\theta}_{\mathrm{tr}} = \mathbf{0} \in \mathbb{R}^d$ *is smaller than the risk of the minimum-variance estimator and the CAPM estimator.*

Proof: See the appendix.

For instance, in case $n = 60$ and $d = 25$ the lower risk bound given by Theorem 7 corresponds to 0.028. Since the upper bound for $\mathrm{Sh}_{\mathrm{T}}^2$ is only $0.15^2 = 0.023 < 0.028$, the trivial estimator clearly outperforms both the minimum-variance and the CAPM estimator under these circumstances. The same argument holds for many other constellations of n and d. Especially, if the number of assets is large compared to the number of observations, the trivial estimator outperforms every standard estimator.

Hence, the answer to the fourth question which has been posed in the introduction is:

(iv) Instead of applying the non-trivial strategies which have been considered in this work, in many practical situations it is better to renounce parameter estimation altogether and put the money straight away into the risk-free investment. This holds in particular if the number of assets is large compared to the number of observations.

6 Empirical Study

In this section I will present an empirical study which shall clarify the question whether it is possible to beat the standard estimators by the trivial estimator. By applying Theorem 6 in combination with Theorem 7 it is possible to conduct an exact hypothesis test for

$$H_0\colon\ \mathcal{R}(\mathbf{0}) \geq \frac{(n-2)/n}{n-d-1} \qquad \text{vs.} \qquad H_1\colon\ \mathcal{R}(\mathbf{0}) < \frac{(n-2)/n}{n-d-1}$$

on a significance level of α $(0 < \alpha < \frac{1}{2})$. This is done by comparing

$$\widehat{\mathcal{R}}(\mathbf{0}) = \hat{\mu}'\widehat{\Sigma}^{-1}\hat{\mu} = \widehat{\mathrm{Sh}}_{\mathrm{T}}^2$$

with the α-quantile of

$$\frac{\chi_d^2\left(\frac{n-2}{n-d-1}\right)}{\chi_{n-d}^2}\,.$$

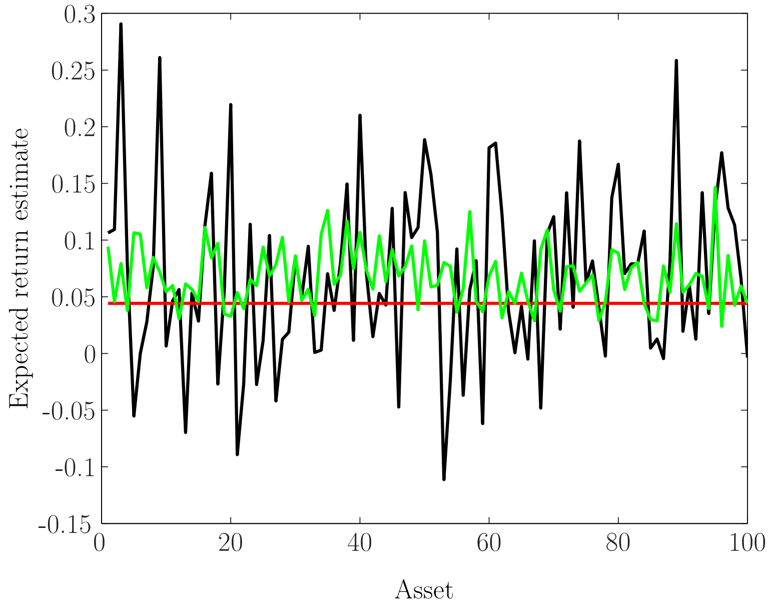

Figure 1: Estimates for annualized expected returns based on the sample mean vector (black), minimum-variance estimator (red), and CAPM estimator (green) given $n = 120$ observations and $d = 100$ assets.

The latter can be simply computed by Monte Carlo simulation and the null hypothesis is rejected if $\widehat{\mathrm{Sh}}_{\mathrm{T}}^2$ falls below the α-quantile. In that case the trivial estimator is significantly better than every standard estimator. Intuitively speaking, if the ex-ante Sharpe ratio of the tangential portfolio is sufficiently small, the asset returns cannot be expected to be essentially different from zero and so it is better to use the trivial estimator rather than some standard estimator which suffers from estimation errors.

For the empirical study I use a sample of monthly stock returns which have been obtained from the CRSP data set. This data set consists of stock prices which are observed on the NYSE, AMEX, and NASDAQ. The sample incorporates monthly returns between January 1969 and December 2008. The risk-free interest rate is calculated on the basis of 3-month treasury bill secondary market rates, which are provided online by the Federal Reserve System. For each year beginning in 1979 I constitute a separate asset universe consisting of all assets which exhibit return data for the last $n = 120$ months.[16] Finally, from each asset universe I randomly draw $d = 100$ stocks without replacement. By this way 30 asset

[16]The amount of stocks in each asset universe ranges from 1 239 assets in the time period 1969–1979 to 2 992 assets in the period 1998–2008.

	1	2	3	4	5	6	7	8	9	10
$\widehat{\mathrm{Sh}}^2_{\mathrm{T}}$	**1.584**	**1.907**	**2.989**	**3.113**	**1.898**	**3.078**	4.446	**3.101**	**1.876**	3.903
	(0)	(0)	(0.034)	(0.045)	(0)	(0.041)	(0.275)	(0.043)	(0)	(0.160)

	11	12	13	14	15	16	17	18	19	20
$\widehat{\mathrm{Sh}}^2_{\mathrm{T}}$	**2.738**	**2.338**	9.086	**2.613**	3.287	4.305	**1.698**	**2.142**	3.375	6.251
	(0.017)	(0.004)	(0.915)	(0.011)	(0.063)	(0.244)	(0)	(0.002)	(0.074)	(0.650)

	21	22	23	24	25	26	27	28	29	30
$\widehat{\mathrm{Sh}}^2_{\mathrm{T}}$	7.029	**3.551**	4.986	4.187	4.651	**2.141**	3.430	**2.895**	**2.432**	3.539
	(0.759)	(0.010)	(0.397)	(0.218)	(0.321)	(0.002)	(0.082)	(0.027)	(0.006)	(0.097)

Table 2: Results of the hypothesis tests for 30 asset universes with $n = 120$ observations and $d = 100$ assets. Bold entries indicate asset universes where the trivial estimator had outperformed the standard estimators on the significance level $\alpha = 0.05$.

universes are produced and the hypothesis test described above is applied separately on each asset universe.

Figure 1 illustrates a typical realization of the sample mean vector, the minimum-variance estimator and the CAPM estimator for such an asset universe.[17] Further, Table 2 provides the results of the empirical study. In 17 out of 30 asset universes it was better (on the significance level of $\alpha = 0.05$) to use the trivial estimator instead of any standard estimator. This means in most cases the excess returns which can be expected on the stock markets are not essentially different from zero or, in other words, the expected original asset returns are not essentially different from the risk-free interest rate.

7 Conclusion

It has been shown that the risk function of an estimator for expected or implicit asset returns is proportional to the expected out-of-sample performance of the investment strategy based on that estimator. For that reason I investigated the risk function of 5 estimators for expected asset returns which are frequently advocated in the literature. It turns out that the risks of the different estimators depend only on the number of observations, the number of assets, and the Sharpe ratios of the tangential portfolio as well as the reference portfolio of the respective estimator. The minimum-variance and the CAPM estimator exhibit approximately the same risks and are preferable among the standard estimators

[17]The chosen reference portfolio of the CAPM estimator corresponds to $w = \mathbf{1}/d$.

in most practical situations. The CAPM estimator performs substantially better only if the number of assets is close to the number of observations. This means among all non-trivial strategies which have been considered in this work, in most practical situations it is best to choose a portfolio which is proportional to the global minimum variance portfolio or some benchmark portfolio such as the equally-weighted portfolio. Moreover, it has been shown that if the squared Sharpe ratio of the tangential portfolio is smaller than $\frac{1}{n}\,(n-2)/(n-d-1)$ (where n is the number of observations and d the number of assets) the asset returns cannot be expected to be essentially different from the risk-free interest rate and so it is better to renounce parameter estimation altogether and put the money straight away into the risk-free investment. An exact hypothesis test has been derived for deciding whether the squared Sharpe ratio undershoots the critical threshold. Finally, this hypothesis test has been applied to 30 asset universes, each one containing 120 empirical observations of 100 assets from the NYSE, AMEX, and NASDAQ. In 17 out of the 30 asset universes it was significantly better to choose the risk-free investment.

References

F.M. Aparicio and J. Estrada (2001), 'Empirical distributions of stock returns: European securities markets, 1990–95', *European Journal of Finance* **7**, pp. 1–21.

A. Bade, G. Frahm, and U. Jaekel (2008), 'A general approach to Bayesian portfolio optimization', *Mathematical Methods of Operations Research* **70**, pp. 337–356.

F. Black and R. Litterman (1992), 'Global portfolio optimization', *Financial Analysts Journal* **48**, pp. 28–43.

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

T. Cogley and T.J. Sargent (2008), 'The market price of risk and the equity premium: a legacy of the Great Depression', *Journal of Monetary Economics* **55**, pp. 454–476.

V. DeMiguel, L. Garlappi, and R. Uppal (2009), 'Optimal versus naive diversification: how inefficient is the $1/N$ portfolio strategy?', *Review of Financial Studies* **22**, pp. 1915–1953.

E. Dimson, P. Marsh, and M. Staunton (2003), 'Global evidence on the equity risk premium', *Journal of Applied Corporate Finance* **15**, pp. 27–38.

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

G. Frahm (2009), 'Asymptotic distributions of robust shape matrices and scales', *Journal of Multivariate Analysis* p. DOI: 10.1016/j.jmva.2008.11.007.

G. Frahm and C. Memmel (2010), 'Dominating estimators for minimum variance portfolios', Discussion paper, University of Cologne, Department of Economic and Social Statistics, Germany.

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

P.A. Frost and J.E. Savarino (1988), 'For better performance: constrain portfolio weights', *Journal of Portfolio Management* **14**, pp. 29–34.

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

R.R. Grauer and N.H. Hakansson (1995), 'Stein and CAPM estimators of the means in asset allocation', *International Review of Financial Analysis* **4**, pp. 35–66.

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

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

P. Jorion (1985), 'International portfolio diversification with estimation risk', *Journal of Business* **58**, pp. 259–278.

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

P. Jorion (1991), 'Bayesian and CAPM estimators of the means: Implications for portfolio selection', *Journal of Banking and Finance* **15**, pp. 717–727.

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

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

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

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

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

C. Memmel (2004), *Schätzrisiken in der Portfoliotheorie*, Ph.D. thesis, University of Cologne, Department of Economic and Social Statistics, Germany.

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

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

M. Wolf (2007), ‘Resampling vs. shrinkage for benchmarked managers’, *Wilmott Magazine* **January 2007**, pp. 76–81.