AStA Adv Stat Anal (2012) 96:343–383
DOI 10.1007/s10182-011-0166-1

ORIGINAL PAPER

Multiple tests for the performance of different investment strategies

Gabriel Frahm · Tobias Wickern · Christof Wiechers

Received: 30 November 2010 / Accepted: 22 August 2011 / Published online: 8 October 2011

Abstract In the context of modern portfolio theory, we compare the out-of-sample performance of eight investment strategies which are based on statistical methods with the out-of-sample performance of a family of trivial strategies. A wide range of approaches is considered in this work, including the traditional sample-based approach, several minimum-variance techniques, a shrinkage, and a minimax approach. In contrast to similar studies in the literature, we also consider short-selling constraints and a risk-free asset. We provide a way to extend the concept of minimum-variance strategies in the context of short-selling constraints. A main drawback of most empirical studies on that topic is the use of simple testing procedures which do not account for the effects of multiple testing. For that reason we conduct several hypothesis tests which are proposed in the multiple-testing literature. We test whether it is possible to beat a trivial strategy by at least one of the non-trivial strategies, whether the trivial strategy is better than every non-trivial strategy, and which of the non-trivial strategies is significantly outperformed by naive diversification. The empirical part of our study is conducted using US stock returns from the last four decades, obtained via the CRSP database.

Keywords Asset allocation · Certainty equivalent · Multiple hypothesis tests · Naive diversification · Out-of-sample performance · Portfolio optimization · Sharpe ratio

G. Frahm
Chair for Applied Stochastics, Helmut Schmidt University / University of the Federal Armed Forces Hamburg, Holstenhofweg 85, 22043 Hamburg, Germany
e-mail: frahm@hsu-hh.de

T. Wickern (✉) · C. Wiechers
Research Training Group Risk Management, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany
e-mail: wickern@wiso.uni-koeln.de

C. Wiechers
e-mail: wiechers@wiso.uni-koeln.de

1 Introduction

According to Markowitz (1952), portfolio optimization consists of two stages. The first stage is about forming beliefs about future performances of asset returns. In the second stage the investor calculates an optimal portfolio based on his specific information and tailored to his own risk preferences. In terms of statistical theory, this means that the investor has to estimate some unknown parameters μ and Σ, i.e., the vector of expected asset returns, as well as the variances and covariances of the asset returns. In the last decades people have begun to think about the question of which parameter is more susceptible to estimation errors and which kind of estimation error has a dominant impact on portfolio optimization. Chopra and Ziemba (1993), among many others, clarified that it is the vector of expected asset returns that causes most of the loss due to parameter uncertainty.

Consequently, most of the suggested portfolio strategies involve an improved estimator for μ. To smooth extreme entries, it was proposed to shrink the sample mean toward a certain target. Jorion (1986) built up such a shrinkage estimator based on a quadratic loss function proposed by Stein (1956). This well-known Bayes–Stein estimator can also be viewed from a Bayesian perspective. In general, that means to treat μ and Σ themselves as random variables and to derive the predictive return distributions. First introduced by Mao and Särndal (1966) and further contributed to by Kalymon (1971), Brown (1976) and Klein and Bawa (1976), an expanding part of the literature especially attended to the combination of data with prior knowledge.

Another interesting class of investment strategies is the family of minimum-variance portfolios. It focuses on minimizing the portfolio variance under some constraints on the portfolio weights and is completely independent of the expected returns. For instance, the global minimum-variance portfolio (GMVP) is the optimal portfolio of risky assets if the expected asset returns are equal. While traditionally using the sample covariance matrix as an estimator for the true covariance matrix Σ, Frahm and Memmel (2010) as well as Ledoit and Wolf (2004) propose shrinkage estimators for Σ^{-1} and Σ, respectively. By contrast, DeMiguel et al. (2009a) impose a further constraint on the norm of the portfolio vector to improve the performance of minimum-variance portfolios.

Halldórsson and Tütüncü (2003) deal with the problem of estimation errors by imposing uncertainty regions for μ and Σ during the optimization process itself. Their approach is often denoted as robust portfolio optimization as the investor maximizes the mean-variance objective function under a worst case scenario. Technically speaking, the investor chooses that (μ, Σ)-constellation from the uncertainty sets for which the objective function is minimal. In a second step, the optimization process for the portfolio weights is conducted. Garlappi et al. (2007) show that such a minimax approach corresponds to a convex combination of the sample-based tangential portfolio and the estimated GMVP. Hence, DeMiguel et al. (2009b) argue in Footnote 14 that the Sharpe ratio of a robust portfolio lies between the Sharpe ratios of the sample-based tangential portfolio and the estimated GMVP. Although this is true, it cannot be concluded that the *expected* Sharpe ratio of a robust portfolio lies between the expected marginal Sharpe ratios.[1] Thus, it is worth taking the robust investment strategy explicitly into consideration.

Despite the various approaches to incorporating estimation risk into the optimization process, it is questionable whether the proposed strategies lead to an improvement of out-of-sample performance. For example, Kritzman et al. (2010) argue that a minimum-variance portfolio outperforms the market portfolio as well as the equally-weighted portfolio if large periods of observations are taken into consideration. Nevertheless, in a recent study DeMiguel et al. (2009b) raise the question of whether optimizing a portfolio using time series information is worthwhile to begin with. Their results show that the considered investment strategies do not *significantly* outperform the naive portfolio, where each asset is equally weighted. Their contribution has aroused a heated discussion about the validity of contemporary methods of portfolio optimization. By contrast, almost the same authors (DeMiguel et al. 2009a) claim that they have found an investment strategy which is able to outperform the naive portfolio.

The problem of finding superior investment strategies is highly important for practical issues, because most participants in the mutual funds industry claim that expert knowledge outperforms naive diversification. Hence, it is not surprising that portfolio optimization is still a matter of debate. We want to clarify the question whether the study conducted by DeMiguel et al. (2009b) justifies the conclusion that the naive strategy is preferable. We enter this debate by constructing a *realistic* setting. This includes the opportunity to invest in a risk-free asset and to constrain short selling both of risk-free and risky assets which meets the options and requirements of the mutual funds industry.

Jagannathan and Ma (2003) treat the short-selling constraint as a means of increasing the out-of-sample performance of the minimum-variance strategy. They show that short-selling constraints have the effect of shrinking the largest eigenvalues of the sample covariance matrix and thus reduce estimation risk. However, in our context the short-selling constraint is not a means to an end but an inevitable requirement of the mutual funds industry. Besides, in our setting it is not sufficient to concentrate on the covariance matrix, because due to the risk-free investment opportunity the investor's optimal asset allocation depends on his individual risk preferences and thus on the expected returns. This means the impact of estimation errors can be expected to be substantially larger compared to the minimum-variance strategy in the traditional case.

We will use both the certainty equivalent (CEQ) and the widely accepted Sharpe ratio as a performance measure though only the CEQ is applicable to the investment decision problem if borrowing is constrained. Applying a rolling-window procedure, for each strategy we generate a series of out-of-sample portfolio returns and calculate the corresponding out-of-sample performances. Pairwise differences of the estimated

[1] Let $\mathrm{Sh}_{\mathrm{MV}}$ and Sh_{T} be the Sharpe ratios of the estimated GMVP and the estimated tangential portfolio. Further, consider the Sharpe ratio $\mathrm{Sh}_{\mathrm{rob}} = \omega \mathrm{Sh}_{\mathrm{MV}} + (1-\omega)\mathrm{Sh}_{\mathrm{T}}$ of the robust portfolio, where ω is such that $0 \leq \omega \leq 1$. Note that ω, $\mathrm{Sh}_{\mathrm{MV}}$, and Sh_{T} are random quantities. The problem is that generally $\mathrm{E}(\mathrm{Sh}_{\mathrm{rob}})$ does not correspond to the convex combination $\mathrm{E}(\omega)\mathrm{E}(\mathrm{Sh}_{\mathrm{MV}}) + \mathrm{E}(1-\omega)\mathrm{E}(\mathrm{Sh}_{\mathrm{T}})$. This means the mean value theorem is not applicable and so it cannot be guaranteed that $\mathrm{E}(\mathrm{Sh}_{\mathrm{MV}}) \leq \mathrm{E}(\mathrm{Sh}_{\mathrm{rob}}) \leq \mathrm{E}(\mathrm{Sh}_{\mathrm{T}})$.

performance between the trivial strategy and the respective non-trivial strategy are taken in order to evaluate the investment strategies. It is tested whether these estimated differences are significant in the sense of detecting an outperforming strategy.

When testing for the strategy with the best out-of-sample performance, one is typically faced with a multiple-testing problem, i.e., conducting several hypothesis tests simultaneously. Instead of controlling for the significance level in each single hypothesis test separately, we consider the so-called familywise error rate (FWER). This concept accounts for the probability of rejecting at least one of the true null hypotheses. To the best of our knowledge, this is the first time that state-of-the-art testing procedures have been applied in this context. On the contrary, DeMiguel et al. (2009a, 2009b) simultaneously carry out several pairwise tests on the performance of investment strategies without an explicit adjustment of the significance level. By doing so, they do not consider the principal nature of multiple testing. Hence, their results suffer from the fact that the probability of rejecting a true null hypothesis can be substantially larger than the nominal error rate.

To sum up, this work aims at providing several contributions to the existing literature:

1. We compare the out-of-sample performance of different investment strategies by applying contemporary methods of statistical analysis. More precisely, we take into consideration that comparing different investment strategies is a multiple-testing problem. Empirical studies which are based on simple test procedures (DeMiguel et al. 2009a, 2009b) might lead to wrong conclusions since the probability of rejecting a true null hypothesis can be substantially larger than the nominal error rate of the simple test.
2. In contrast to DeMiguel et al. (2009b) we are not only interested whether there exists an investment strategy which outperforms the trivial strategy but also whether the trivial strategy performs best among all investment strategies which have been taken into consideration in this work. It is worth emphasizing that, despite the fact that these two questions represent two sides of the same coin, they require very different testing procedures.[2] This is because if the trivial strategy is not significantly outperformed by a non-trivial strategy it cannot be concluded that it performs best among all available strategies. Additionally, we would like to know not only if there *exists* an outperforming investment strategy but rather to *identify* all outperforming strategies. The last question is a typical application of multiple-testing procedures (Romano et al. 2008).
3. We aim at constructing a realistic setting with short-selling constraints both to the risk-free and risky assets. Especially, in contrast to other studies (DeMiguel et al. 2009a, 2009b) we take the risk-free asset into consideration. This is extremely important since Tobin's two-fund separation theorem breaks down in the presence of estimation risk (Kan and Zhou 2007).
4. In contrast to DeMiguel et al. (2009b) we also investigate the performance of robust investment strategies (Halldórsson and Tütüncü 2003). We clarify that in

[2] This is due to the typical asymmetry of frequentistic testing procedures, i.e., if H_0 is not rejected nothing can be said about H_1.

the presence of estimation risk the mean value theorem is not applicable and so the argument which has been used by the aforementioned authors is not valid. Additionally, we investigate a method for estimating the covariance matrix of asset returns which has recently been proposed by Frahm and Memmel (2010).
5. We discuss the inadequacy of the portfolio turnover as a performance measure if a static portfolio selection problem is to be investigated. We motivate the use of the certainty equivalent in the presence of estimation risk and short-selling constraints while we do not pass on the Sharpe ratio as a state-of-the-art performance measure abstaining from the investor's individual risk preferences.
6. In contrast to DeMiguel et al. (2009b) our statistical methodology for comparing the different out-of-sample performances does not require that the data are normally distributed and serially independent. Any testing procedure which is based on the normal distribution hypothesis and the assumption of serial independence might lead to wrong conclusions (Frahm 2007; Lo 2002; Ledoit and Wolf 2008). By contrast, we apply a stationary block bootstrap procedure to account for the serial dependence structure of the out-of-sample portfolio returns. As this is a non-parametric method, the normality assumption is not required either.

In Sect. 2 we present the different investment strategies which are taken into consideration. Section 3 contains a detailed explanation of the chosen performance measures and a discussion of the inappropriateness of the portfolio turnover. The multiple-testing procedures are described in Sect. 4. The empirical study can be found in Sect. 5 and Sect. 6 concludes our work.

2 Strategies for asset allocation

In this section, we present the strategies tested in our study. As we explicitly incorporate the money market, the asset universe consists of d risky assets and one risk-free asset. Let μ be the d-dimensional vector of expected *excess* returns of the risky assets,[3] while Σ is the corresponding $d \times d$ positive-definite covariance matrix. The investor aims to allocate his wealth among the assets according to the well-known mean-variance objective function, introduced by Markowitz (1952). In general, the investor's problem is

$$\begin{aligned} &\max_{w}\ w'\mu - \frac{\lambda}{2} w'\Sigma w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1, \end{aligned} \tag{1}$$

where $\lambda > 0$ denotes the investor-specific parameter of risk aversion. Note that $1 - w'\mathbf{1}$ denotes the proportion of wealth which is invested into the risk-free asset. The constraint $w'\mathbf{1} \leq 1$ represents the fact that borrowing is not allowed. The restrictions $w'\mathbf{1} \leq 1$ and $w \geq \mathbf{0}$ meet the situation of a mutual-fund manager with no possibility of issuing bonds or selling risky assets short. Our primary goal is to create a realistic setting even though the optimization problem in (1) does not exhibit a

[3] We will always refer to excess returns, i.e., asset returns minus the corresponding risk-free interest rate. Nevertheless, in the following we will drop the prefix 'excess' for convenience.

closed-form solution. However, this is a standard problem of quadratic optimization and readily tractable with some more computational effort.

If μ and Σ were known to the investor with full precision, the optimal solution w^* of (1) would be unique since Σ is positive definite, and would lead to the maximal CEQ of $w^{*\prime}\mu - (\lambda/2)w^{*\prime}\Sigma w^*$. The problem is that the true moments of the return distribution are not known to the investor and thus have to be estimated. In this paper, the estimation procedure will be purely data-driven, meaning that the only source of information available to an investor is a sample $r_1, \ldots, r_T$ of historical realizations of the d-dimensional random vector r of excess returns. The way μ and Σ are estimated will be referred to as the respective 'philosophy' of the investor.

2.1 Trivial strategies

An investor might be reluctant to place any confidence in historical data. Furthermore, he might not believe in strengths or weaknesses of single assets, but rather in the market of risky assets as a whole. Thus, he completely ignores μ and Σ and allocates his wealth equally into all risky assets. In a world without risk-free assets, this means he simply chooses the allocation $\mathbf{1}/d$, the so-called 'equally-weighted' or 'naive' portfolio.

Since our market is assumed to be endowed with a risk-free asset, we have to extend this trivial rule. In our setting, $\gamma \cdot \mathbf{1}/d$ serves as the allocation rule for naive investors. The proportion of wealth γ with $0 \le \gamma \le 1$ is invested in the risky assets. The question remains of how to determine the parameter γ. At this point it is important to recognize that a naive investor is characterized by a non-optimizing, i.e., trivial, strategy. Hence, γ is not due to an optimization calculus. In our study, we consider five types of naive investor by setting $\gamma \in \{0, 0.25, 0.5, 0.75, 1\}$. Each type characterizes a specific amount of risk aversion. Note that the trivial strategy with $\gamma = 0$ implies that the investor is extremely pessimistic regarding the equity market, whereas the trivial strategy with $\gamma = 1$ is chosen by an optimistic investor who fully believes in the strengths of the equity market.

2.2 The traditional approach

The traditional sample-based approach to estimate the unknown parameters μ and Σ is to use their sample counterparts

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t \quad \text{and} \quad \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{\mu})(r_t - \hat{\mu})'. \tag{2}$$

This strategy was initially proposed by Markowitz (1952). Note that $\hat{\mu}$ and $\hat{\Sigma}$ are the maximum-likelihood (ML) estimators of the parameters μ and Σ when $r_1, \ldots, r_T$ are assumed to be independent realizations of a normally distributed random vector r. The traditional strategy is characterized by

$$\begin{aligned} \hat{w}_{\mathrm{T}} &= \arg\max_{w}\; w'\hat{\mu} - \frac{\lambda}{2} w'\hat{\Sigma} w \\ &\text{subject to} \quad w \ge \mathbf{0}, \qquad w'\mathbf{1} \le 1. \end{aligned} \tag{3}$$

2.3 Minimum-variance strategies

The pursuit of a minimum-variance strategy can be seen in the light of estimation problems. It is widely accepted that even small estimation errors in the vector of expected returns heavily distort the solution of the optimization problem (1). The portfolio which is suggested by the traditional approach can differ substantially from the true, optimal portfolio. Put another way, its *realized* (but not the suggested) Sharpe ratio can be very small since the expected asset returns are unknown. Hence, it might be better to search for a minimum-variance portfolio. In the following we would like to clarify this issue.

The GMVP has been advocated by many authors (Jagannathan and Ma 2003; Kempf and Memmel 2006; Ledoit and Wolf 2003). On the one hand choosing the GMVP is closely related to the basic idea of Markowitz (1952), i.e., searching for an efficient portfolio by diversification. On the other hand there are no expected asset returns which have to be estimated and so the impact of estimation errors can be substantially reduced. However, one might ask why it should be appropriate to search for a minimum-variance portfolio if the investor is interested in maximizing a mean-variance utility function or the Sharpe ratio according to Tobin's two-fund separation theorem (Tobin 1958). Our hypothesis is that the estimated GMVP in general is substantially closer to w^* than most portfolios which are suggested as a solution of (1) where μ and Σ are substituted by some estimates.

In the simple framework without risk-free asset and without short-selling constraint the optimization problem of the global minimum-variance strategy is

$$\min_{w} \; w'\Sigma w \quad \text{s.t.} \quad w'\mathbf{1} = 1 \tag{4}$$

which leads to the (true) global minimum-variance portfolio $w_{\mathrm{MV}}^{(s)} = \Sigma^{-1}\mathbf{1}/\mathbf{1}'\Sigma^{-1}\mathbf{1}$ and its (true) variance $1/\mathbf{1}'\Sigma^{-1}\mathbf{1} > 0$. Note that the superscript 's' in $w_{\mathrm{MV}}^{(s)}$ refers to the simple framework without risk-free asset and without short-selling constraint whereas the subscript 'MV' refers to the minimum-variance strategy. If a risk-free asset were added a trivial solution to the minimum-variance problem would be to invest solely in the risk-free asset resulting in a variance of 0. A more sophisticated way to expand minimum-variance strategies can be conducted when realizing that the problem in (4) is equivalent to

$$\max_{w} \; w'\theta\mathbf{1} - \frac{\lambda}{2} w'\Sigma w \quad \text{s.t.} \quad w'\mathbf{1} = 1, \tag{5}$$

for any $\theta > 0$. We adapt the minimum-variance strategy to our framework by changing the constraint in (5) to $w'\mathbf{1} \leq 1$ (possible investment in the risk-free asset) and adding the short-selling constraint $w \geq \mathbf{0}$. It remains to estimate θ. To this end we assume that returns follow a joint normal distribution with mean $\theta\mathbf{1}$ and covariance matrix Σ. Estimating the mean on behalf of the ML method complies with

$$\tilde{\theta} := \frac{\mathbf{1}'\tilde{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'\tilde{\Sigma}^{-1}\mathbf{1}}, \tag{6}$$

where $\tilde{\Sigma}$ is an estimator for Σ. Interestingly, $\tilde{\theta}$ is nothing else than the expected return of the minimum-variance strategy in the conventional setting. In our framework, the hereinafter presented minimum-variance strategies will mainly differ in the way Σ is estimated.

Traditionally, an investor substitutes Σ with the sample covariance matrix $\hat{\Sigma}$. This again implies

$$\hat{\theta}_{\mathrm{MV}} := \frac{\mathbf{1}'\hat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}} \quad \text{and} \quad \hat{\mu}_{\mathrm{MV}} := \hat{\theta}_{\mathrm{MV}}\mathbf{1}, \tag{7}$$

Thus, in our setting the traditional variance-minimizing investor obtains his optimal portfolio $\hat{w}_{\mathrm{MV}}$ by solving

$$\begin{aligned} &\max_{w} \ w'\hat{\mu}_{\mathrm{MV}} - \frac{\lambda}{2}w'\hat{\Sigma}w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1. \end{aligned} \tag{8}$$

Frahm and Memmel (2010) derive a shrinkage estimator for the global minimum-variance portfolio in the conventional setting without risk-free asset and without short-selling constraint. More precisely, they suppose to approximate the solution of the optimization problem in (4) by the shrinkage estimator

$$\hat{w}_{\mathrm{FM}}^{(s)} = \hat{\phi}_{\mathrm{FM}} w_R + (1 - \hat{\phi}_{\mathrm{FM}})\hat{w}_{\mathrm{MV}}^{(s)}, \tag{9}$$

where w_R is an arbitrary reference portfolio satisfying $w_R'\mathbf{1} = 1$ and $\hat{w}_{\mathrm{MV}}^{(s)} = \hat{\Sigma}^{-1}\mathbf{1}/(\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1})$ is the traditional estimator for the GMVP in the setting of (4). According to the proposal of Frahm and Memmel (2010) we set $w_R := \mathbf{1}/d$. The shrinkage intensity is defined by

$$\hat{\phi}_{\mathrm{FM}} := \min\left\{\frac{d-3}{T-d+2}\frac{\hat{\sigma}_{\mathrm{MV}}^2}{\hat{\sigma}_{\mathrm{R}}^2 - \hat{\sigma}_{\mathrm{MV}}^2}, 1\right\}, \tag{10}$$

where $\hat{\sigma}_{\mathrm{R}}^2$ and $\hat{\sigma}_{\mathrm{MV}}^2$ are the estimated variances of the return of the portfolios w_R and $\hat{w}_{\mathrm{MV}}^{(s)}$, respectively, viz.

$$\hat{\sigma}_{\mathrm{R}}^2 := w_R'\hat{\Sigma}w_R = \frac{1}{d^2}\mathbf{1}'\hat{\Sigma}\mathbf{1}, \quad \text{and} \quad \hat{\sigma}_{\mathrm{MV}}^2 := \hat{w}_{\mathrm{MV}}^{(s)\prime}\hat{\Sigma}\hat{w}_{\mathrm{MV}}^{(s)} = \frac{1}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}. \tag{11}$$

In Theorem 8 of Frahm and Memmel (2010), it is shown that $\hat{w}_{\mathrm{FM}}^{(s)}$ may also be obtained by calculating

$$\hat{\Sigma}_{\mathrm{FM}}^{-1} := \hat{\phi}_{\mathrm{FM}}\frac{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}{d}I_d + (1 - \hat{\phi}_{\mathrm{FM}})\hat{\Sigma}^{-1}, \tag{12}$$

and using $\hat{\Sigma}_{\mathrm{FM}}$ as a shrinkage estimator for Σ in the formula for $w_{\mathrm{MV}}^{(s)}$, i.e., $\hat{w}_{\mathrm{FM}}^{(s)} = \hat{\Sigma}_{\mathrm{FM}}^{-1}\mathbf{1}/(\mathbf{1}'\hat{\Sigma}_{\mathrm{FM}}^{-1}\mathbf{1})$. However, we use $\hat{\Sigma}_{\mathrm{FM}}$ in our setting to calculate

$$\hat{\theta}_{\mathrm{FM}} := \frac{\mathbf{1}'\hat{\Sigma}_{\mathrm{FM}}^{-1}\hat{\mu}}{\mathbf{1}'\hat{\Sigma}_{\mathrm{FM}}^{-1}\mathbf{1}} \quad \text{and} \quad \hat{\mu}_{\mathrm{FM}} := \hat{\theta}_{\mathrm{FM}}\mathbf{1}. \tag{13}$$

The overall optimization framework for the 'Frahm–Memmel-type' investor is

$$\begin{aligned} &\max_{w}\ w'\hat{\mu}_{\mathrm{FM}} - \frac{\lambda}{2} w'\hat{\Sigma}_{\mathrm{FM}} w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1. \end{aligned} \tag{14}$$

The maximizer of (14) will be referred to as $\hat{w}_{\mathrm{FM}}$. Frahm and Memmel (2010) showed that the shrinkage estimator $\hat{w}^{(s)}_{\mathrm{FM}}$ dominates the traditional estimator $\hat{w}^{(s)}_{\mathrm{MV}}$ for the global minimum-variance portfolio with respect to the out-of-sample variance of the portfolio return. More precisely, we have

$$\mathrm{E}\big\{\big(\hat{w}^{(s)}_{\mathrm{FM}} - w^{(s)}_{\mathrm{MV}}\big)'\Sigma\big(\hat{w}^{(s)}_{\mathrm{FM}} - w^{(s)}_{\mathrm{MV}}\big)\big\} < \mathrm{E}\big\{\big(\hat{w}^{(s)}_{\mathrm{MV}} - w^{(s)}_{\mathrm{MV}}\big)'\Sigma\big(\hat{w}^{(s)}_{\mathrm{MV}} - w^{(s)}_{\mathrm{MV}}\big)\big\}. \tag{15}$$

The dominance result remain valid when 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. In our setting, we are concerned with the linear *inequality* constraints $w \geq \mathbf{0}$ and $w'\mathbf{1} \leq 1$ so that we cannot expect the 'Frahm–Memmel-type' strategy to dominate the traditional minimum-variance strategy in general. Nevertheless, it may be assumed that the 'Frahm–Memmel-type' strategy also works well in our setting since the shrinkage approach in (12) is expected to reduce the estimation error substantially.

Another way to handle the distortion when estimating Σ is to presume a special structure in the variances and covariances of the returns. In particular, assume that all returns are equicorrelated and have the same variance, i.e.,

$$\Sigma = \bar{\sigma}^2\big\{\rho\mathbf{1}\mathbf{1}' + (1-\rho)I_d\big\}. \tag{16}$$

Since the risky assets do not differ from each other in mean and variance, by construction, it ends up in equal portfolio weights. But the amount of wealth invested in the risk-free asset depends on the parameters $\bar{\sigma}^2$ and ρ. We again conduct an ML estimation under the assumption of multivariate normally distributed returns and the special structure in (16). First, it can be shown that the ML estimator for θ coincides in this case with the grand mean $\tilde{\mu} = \mathbf{1}'\hat{\mu}/d$. Following Frahm (2009b), consider the quantities

$$\tilde{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \tilde{\mu})^2, \tag{17}$$

$$\tilde{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \tilde{\mu})(r_{jt} - \tilde{\mu}), \quad i, j = 1, \ldots, d,\ i \neq j, \tag{18}$$

so that we obtain the ML estimates for the equicorrelation structure, i.e.,

$$\tilde{\sigma}^2 = \frac{1}{d}\sum_{i=1}^{d}\tilde{\sigma}_i^2, \tag{19}$$

$$\tilde{\rho} = \frac{2}{d(d-1)}\sum_{i=1}^{d-1}\sum_{j=i+1}^{d}\frac{\tilde{\sigma}_{ij}}{\tilde{\sigma}^2}. \tag{20}$$

An investor believing in the assumption of equicorrelation obtains his optimal portfolio $\hat{w}_{\text{EC}}$ as the solution of

$$\begin{aligned} &\max_{w}\ w'\tilde{\mu}\mathbf{1} - \frac{\lambda}{2}\tilde{\sigma}^2 w'\{\tilde{\rho}\mathbf{1}\mathbf{1}' + (1-\tilde{\rho})I_d\}w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1. \end{aligned} \tag{21}$$

Ledoit and Wolf (2004) derive a shrinkage estimator for Σ rather than Σ^{-1} as in Frahm and Memmel (2010). Their estimator reads

$$\hat{\Sigma}_{\text{LW}} := \hat{\phi}_{\text{LW}}\hat{G} + (1-\hat{\phi}_{\text{LW}})\hat{\Sigma}, \tag{22}$$

with $\hat{\phi}_{\text{LW}}$ being the shrinkage intensity and $\hat{G} = [\hat{g}_{ij}]$ a $d \times d$ matrix which is specified below. An equicorrelation structure similar to the assumption in (16) underlies the shrinkage target. More precisely,

$$\hat{\rho}_{\text{LW}} := \frac{2}{d(d-1)} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j}, \tag{23}$$

where $\hat{\Sigma} = [\hat{\sigma}_{ij}]$ with $\hat{\sigma}_{ij}$ being the ijth component of the sample covariance matrix and $\hat{\sigma}_i = \sqrt{\hat{\sigma}_{ii}}$ $(i = 1, \ldots, d)$. Now the components of the matrix $\hat{G}$ are given by

$$\hat{g}_{ii} = \hat{\sigma}_i^2 \quad \text{and} \quad \hat{g}_{ij} = \hat{\rho}_{\text{LW}}\hat{\sigma}_i\hat{\sigma}_j. \tag{24}$$

The (theoretically) optimal shrinkage intensity ϕ_{LW} is chosen to be the minimal distance between the matrix $\Sigma_{\text{LW}} := \phi_{\text{LW}}\hat{G} + (1-\phi_{\text{LW}})\hat{\Sigma}$ and Σ in terms of the Frobenius norm. The problem is that ϕ_{LW} is unknown and thus Σ_{LW} is not feasible. Hence, Ledoit and Wolf (2004) derive a consistent estimator $\hat{\phi}_{\text{LW}}$ which is used as a substitute for ϕ_{LW} in (22). The resulting estimator for Σ is denoted by $\hat{\Sigma}_{\text{LW}}$.

It is worth emphasizing that Ledoit and Wolf (2004) only seek to improve the estimation of covariance matrices without taking into account a particular portfolio optimization problem. However, they conclude: "All types of portfolio optimization procedures [...] would benefit from shrinking the sample covariance matrix." Thus, we are convinced that it is meaningful to apply the Ledoit–Wolf estimator for the covariance matrix to our specific setting though the estimation of the vector of expected asset returns is an independent problem which has not been considered neither by Frahm and Memmel (2010) nor by Ledoit and Wolf (2004). The corresponding investment strategy will be referred to as the 'Ledoit–Wolf-type' strategy and its optimization problem is

$$\begin{aligned} &\max_{w}\ w'\hat{\mu}_{\text{LW}} - \frac{\lambda}{2} w'\hat{\Sigma}_{\text{LW}} w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1, \end{aligned} \tag{25}$$

with

$$\hat{\mu}_{\text{LW}} := \frac{\mathbf{1}'\hat{\Sigma}_{\text{LW}}^{-1}\hat{\mu}}{\mathbf{1}'\hat{\Sigma}_{\text{LW}}^{-1}\mathbf{1}} \cdot \mathbf{1}. \tag{26}$$

The optimal solution is referred to as $\hat{w}_{\text{LW}}$.

Another strategy belonging to the family of minimal variance strategies is the 2-norm-constrained approach. Originally, and extending the results of Jagannathan and Ma (2003), DeMiguel et al. (2009a) bring this strategy into discussion when introducing a whole new family of so-called norm-constrained strategies. Following their study, and because the 2-norm-constrained strategy has been found to be among the best of a whole range of strategies (see DeMiguel et al. 2009a, Table 4), we have decided to include this strategy.

The basic idea is to improve the performance of the traditional minimum-variance strategy by constraining the Euclidean norm of the portfolio weights. Applying this idea to our setting, the optimization problem of the 2-norm-constrained investor is

$$\begin{aligned} &\max_{w}\ w'\hat{\mu}_{\mathrm{MV}} - \frac{\lambda}{2} w'\hat{\Sigma} w \\ &\text{subject to} \quad \|w\|_2 \le \delta, \qquad w \ge \mathbf{0}, \qquad w'\mathbf{1} \le 1. \end{aligned} \tag{27}$$

As DeMiguel et al. (2009a) remark (see their Proposition 2), there is a connection between an additional constraint on the norm of the portfolio weight vector and replacing the sample covariance matrix $\hat{\Sigma}$ by a shrunk version, where the shrinkage target and intensity is determined by the type of the norm and the bound on the norm of the portfolio weights, δ.[4]

2.4 A Bayesian strategy

A Bayesian approach to counter estimation error is to view μ and Σ as random variables rather than deterministic quantities. The posterior distribution of the unknown parameters reflects both information from the historical data and some prior knowledge. The predictive return distribution accounts for the estimation risk by incorporating the posterior distribution of the asset returns (Bade et al. 2008).

Jorion (1986) introduces an informative conjugate prior for μ . The resulting estimator is a linear combination of the sample mean $\hat{\mu}$ and the mean of the traditional minimum-variance strategy $\hat{\mu}_{\mathrm{MV}}$, viz.

$$\hat{\mu}_{\mathrm{BS}} = \frac{T}{T + \hat{\phi}_{\mathrm{BS}}} \hat{\mu} + \frac{\hat{\phi}_{\mathrm{BS}}}{T + \hat{\phi}_{\mathrm{BS}}} \hat{\mu}_{\mathrm{MV}} \tag{28}$$

with

$$\hat{\phi}_{\mathrm{BS}} = \frac{T}{T - d - 2} \frac{(d+2)}{(\hat{\mu} - \hat{\mu}_{\mathrm{MV}})'\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_{\mathrm{MV}})}. \tag{29}$$

The relation in (28) may also be viewed as a 'Stein-type' shrinkage approach with $\hat{\mu}_{\mathrm{MV}}$ as the shrinkage target (Stein 1956).

Note that the sample covariance matrix in (29) is adjusted according to the original suggestion of Jorion (1986). Even though this approach focuses on the estimation of

[4]In our empirical section, the parameter δ is calculated according to a cross-validation procedure described by DeMiguel et al. (2009a).

expected returns, it also affects the estimation of the covariance matrix. The covariance structure is now assessed by

$$\hat{\Sigma}_{\mathrm{BS}} = \frac{T + \hat{\phi}_{\mathrm{BS}} + 1}{T + \hat{\phi}_{\mathrm{BS}}} \frac{T}{T - d - 2} \hat{\Sigma} + \frac{\hat{\phi}_{\mathrm{BS}}}{T(T + \hat{\phi}_{\mathrm{BS}} + 1)} \frac{T}{T - d - 2} \frac{\mathbf{1}\mathbf{1}'}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}. \tag{30}$$

Consequently, the Jorion optimal strategy $\hat{w}_{\mathrm{BS}}$ is the solution to the investment problem

$$\begin{aligned} &\max_{w} \; w'\hat{\mu}_{\mathrm{BS}} - \frac{\lambda}{2} w'\hat{\Sigma}_{\mathrm{BS}} w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1. \end{aligned} \tag{31}$$

The shrinkage of the vector of expected returns can also be interpreted economically. Assets with a relatively high historical mean will be adjusted downwards, while assets with a low historical return are adjusted upwards. Numerically, the shrunk vector $\hat{\mu}_{\mathrm{BS}}$ leads to less extreme portfolio weights than the traditional approach (3), which then are less vulnerable to unexpected realized returns.

2.5 A minimax strategy

A relatively new set of strategies, first introduced by Halldórsson and Tütüncü (2003) and in a more general framework by Goldfarb and Iyengar (2003), deals with the problem of estimation error during the optimization process itself. The parameters μ and Σ are not substituted by point estimates and not considered as random variables. By contrast, certain confidence regions $\Theta_\mu \subset \mathbb{R}^d$ and $\Theta_\Sigma \subset \mathbb{R}^{d \times d}$ are established, in which the true values of these parameters are assumed to be. Then, unlike the Bayesian procedure, the optimal value of the investor's problem for any of the parameter constellations $(\mu, \Sigma) \in \Theta_\mu \times \Theta_\Sigma$ is found. As in this paper only uncertainty regions for μ are under consideration, so the optimization problem reads

$$\begin{aligned} &\max_{w} \left(\min_{\mu \in \Theta_\mu} w'\mu \right) - \frac{\lambda}{2} w'\hat{\Sigma} w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1. \end{aligned} \tag{32}$$

Thus, the input for the expected excess returns is taken as uncertain, but for the covariance structure, the sample covariance matrix $\hat{\Sigma}$ will be employed. This can be justified by the findings of Chopra and Ziemba (1993) and others, stating that errors in the estimated means are much more disturbing than errors in the covariance structure. A main advantage of this approach—as well as of the Bayesian approach—is that investors can incorporate expert knowledge, such as analysts' forecasts, into their estimates of future returns.

One way to construct confidence regions is based on the observation that the quantity $(T - d)/d \cdot (\mu - \hat{\mu})'\hat{\Sigma}^{-1}(\mu - \hat{\mu})$ is F-distributed with d and $T - d$ degrees of freedom if the underlying returns are multivariate normally distributed (Press 1972, p. 132). That gives rise to a theoretical choice for the uncertainty set

$$\Theta_\mu := \left\{ \mu \in \mathbb{R}^d : (\mu - \hat{\mu})'\hat{\Sigma}^{-1}(\mu - \hat{\mu}) \leq \frac{d}{T - d}\kappa^2 \right\}, \tag{33}$$

where κ^2 is chosen as a $(1-\alpha)$ quantile of the $F_{d,T-d}$-distribution. This guarantees that the uncertainty region Θ_μ will contain the true parameter μ with a probability of at least $(1-\alpha)$.

The minimax approach suffers from over-conservatism in the estimation of the means. The implicit adjustment of the means used for calculating the optimal weights will always be downwards. But as Ceria and Stubbs (2006) observed, this does not meet with reality, as the realized returns will sometimes be smaller than the sample mean, but sometimes they may be greater. Thus, they proposed the use of what they call 'zero-net-alpha adjustment', that is, to add the constraint

$$\mathbf{1}' D(\mu - \hat{\mu}) = 0 \tag{34}$$

with a prespecified, symmetric matrix D to the set of constraints in problem (32). For $D = I_d$, this additional constraint will ensure that in absolute terms, the sum of negative deviations of the mean return of assets will be offset by the same amount of positive deviations. In our study, D is chosen to equal $\sqrt{T}(\hat{\Sigma}^{\frac{1}{2}})^{-1}$ with $\hat{\Sigma} = \hat{\Sigma}^{\frac{1}{2}}\hat{\Sigma}^{\frac{1}{2}}$.

Solving the inner minimization problem in (32) first, we can rewrite the optimization problem subject to the additional constraint (34) in a computationally tractable form:

$$\begin{aligned}&\max_w \; w'\hat{\mu} - \sqrt{\frac{d}{T-d}}\kappa \left\| \left(\hat{\Sigma} - \frac{\hat{\Sigma}^{\frac{1}{2}}\mathbf{1}\mathbf{1}'\hat{\Sigma}^{\frac{1}{2}}}{d}\right)^{\frac{1}{2}} w \right\|_2 - \frac{\lambda}{2} w'\hat{\Sigma} w \\ &\text{subject to} \quad w \geq \mathbf{0}, \qquad w'\mathbf{1} \leq 1,\end{aligned} \tag{35}$$

which—even though not a quadratic program anymore—can be solved numerically by applying second order cone programming methods. The penalizing term in (35) can be interpreted as a correction for estimation error. Even though such a penalizing term is already in place in the Bayes–Stein-framework, it is of a different type. This is because in (35), the size of the particular weights plays a role in the penalizing term, while this is not the case in (31).

3 Performance measurement

In the following we want to consider the requirements of a performance measure that is suitable in our setting. At this point, one should be distinctly aware that we are considering a static portfolio selection problem. This means it is assumed that the investor searches for a buy-and-hold portfolio which is liquidated after one period. In Sect. 5 we will focus on out-of-sample portfolio returns which arise from a rolling-window procedure and include different vectors of portfolio weights. This is by no means real portfolio rebalancing but rather an overlapping repetition of a static optimization process. The investor's problem remains in allocating wealth only once though he is free to regard the potential outcome of his decision rule in the past.

In comparable empirical studies, the portfolio turnover is often used as a performance measure (see, e.g., DeMiguel et al. 2009a, 2009b; Behr et al. 2010). It consists basically of differences in the portfolio weights of a given asset over time. It aims to quantify the amount of trading required by an investment strategy and results in a comparison of portfolio return and transaction costs. As mentioned above, static portfolio selection does not require any intermediate trading. Consequently, we must not take the portfolio turnover into consideration as a performance measure. In our case, performance is only influenced by the expected portfolio return and its variance. To this end, two alternatives appear in the literature: the first and predominant measure is the Sharpe ratio. For a portfolio strategy m, it is defined as the expected portfolio return $\mu_{P,m}$ divided by its standard deviation $\sigma_{P,m}$, viz.

$$\mathrm{Sh}_m := \frac{\mu_{P,m}}{\sigma_{P,m}}. \tag{36}$$

It can be interpreted as a 'reward-to-liability' ratio, and was initially introduced by Sharpe (1966). Its main advantage is that it is a simple and intuitive measure for the performance of an investment strategy. So the Sharpe ratio has become one of the most prominently used performance measures regarding the comparison and ranking of mutual funds.

By contrast, the CEQ appears to be a valuable alternative. For portfolio strategy m, it is defined by

$$\mathrm{CEQ}_m := \mu_{P,m} - \frac{\lambda}{2}\sigma^2_{P,m}. \tag{37}$$

The CEQ can be thought of as the maximum risk-free rate of excess return an investor would be willing to give up in order to invest according to strategy m. Thus it depends on the individual risk preference of the investor, which might be seen as a drawback compared to the Sharpe ratio. However, we are essentially interested in quantifying the specific impact of the investor's risk aversion on the performance of the different strategies.

We now explain why the CEQ is a reasonable performance measure in our setting. Assume for a moment the idealized situation in which the parameters μ and Σ are known to the investor with full precision. It is clear that this is an unrealistic assumption but our goal is to demonstrate that even in this idealized setting it is inappropriate to make an investment decision based on the Sharpe ratio.[5] Consider portfolios 1 and 2 which are both composed of risky assets only. Due to the known parameters of the joint return distribution, we are able to display the corresponding points of the two portfolios in a μ–σ coordinate system, as in Fig. 1.

Our setting contains a risk-free asset and prevents any asset from being sold short. Thus, the investor is free to combine portfolio 1 and 2 with the risk-free asset as long as the weight of the risk-free investment takes values between 0 and 1. Mathematically speaking, our setting admits any *convex* combination of a portfolio of risky assets with the risk-free asset. In Fig. 1, those convex combinations are depicted by the solid lines which connect zero with portfolios 1 and 2, respectively. Note that the risk-free asset is located at zero in the μ–σ space whenever one considers excess returns. By contrast, the dashed lines represent linear combinations of the two

[5] It is even more inappropriate if the estimation risk is taken into account. The reason is that under parameter uncertainty the two-fund separation theorem breaks down (Kan and Zhou 2007).

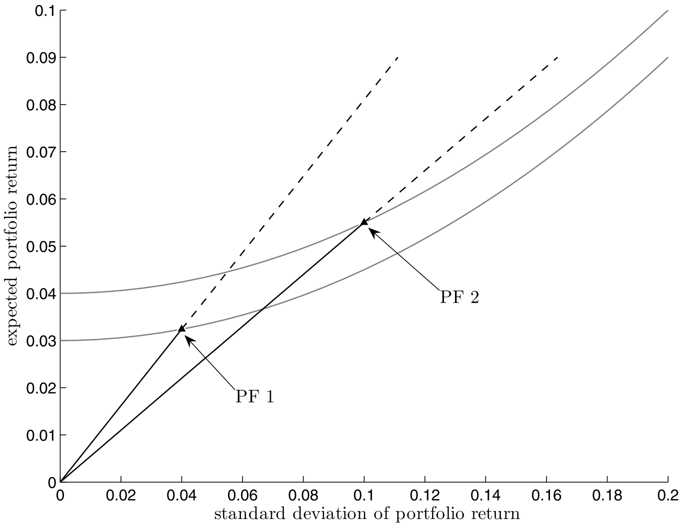

Fig. 1 Depicted are portfolios 1 and 2, respectively, in the μ–σ space. The *solid lines* represent admissible combinations of the portfolios with the risk-free asset. Inaccessible combinations are displayed by *dashed lines*

portfolios with the risk-free asset which are not attainable. This is because the weight of the risk-free asset would have to take a negative value, or—in other words—the investor would have to sell the risk-free asset short in order to reach these μ–σ combinations.

The slopes of the straight lines which pass through zero and portfolios 1 and 2, respectively, correspond to the Sharpe ratios of the two portfolios. It is easy to see that portfolio 1 exhibits a higher Sharpe ratio than portfolio 2. The investor's risk preferences are reflected by the curvature of the indifference curves. The intersection point of these curves with the ordinate represents the CEQ of the two portfolios. Portfolio 1 exhibits a CEQ of 0.03, whereas portfolio 2 has a greater CEQ of 0.04, but the Sharpe ratio of portfolio 1 is greater than the Sharpe ratio of portfolio 2. Thus, the valuation of the two portfolios varies depending on whether the investor relies on the Sharpe ratio or on the CEQ as a performance measure. The CEQ incorporates the whole optimization process, i.e., choosing a portfolio of risky assets *and* combining it with the risk-free asset according to the investor's specific risk preferences. As one can see in Fig. 1, relying on the Sharpe ratio as a *decision criterion* can lead to false portfolio choices when the weight of the risk-free asset is restricted. Nevertheless, we will take both the Sharpe ratio and the certainty equivalent into consideration when it comes to *performance measurement*. This makes our empirical results comparable to other studies and enables us to evaluate the performance of the strategies independently of the investor's risk preferences.

We are interested in the mean and the standard deviation of the *out-of-sample* portfolio returns.[6] In the following we denote the out-of-sample return of the mth in-

[6] A series of out-of-sample portfolio returns can be generated by means of a rolling-windows procedure. The sample of asset returns is divided into an in-sample part and an out-of-sample part. The parameters μ

vestment strategy in the nth month by $R_{m,n}$ $(n = 1, \ldots, N)$. We estimate the certainty equivalent of the mth investment strategy as

$$\widehat{\mathrm{CEQ}}_m = \hat{\mu}_{P,m} - \frac{\lambda}{2}\hat{\sigma}^2_{P,m}, \quad m = 1, \ldots, M, \tag{38}$$

where $\hat{\mu}_{P,m} = N^{-1}\sum_{n=1}^{N} R_{m,n}$ and $\hat{\sigma}^2_{P,m} = N^{-1}\sum_{n=1}^{N}(R_{m,n} - \hat{\mu}_{P,m})^2$ are the sample versions of $\mu_{P,m}$ and $\sigma^2_{P,m}$ and N is the sample size of the portfolio returns. The corresponding Sharpe ratio-estimate of the mth strategy is

$$\widehat{\mathrm{Sh}}_m = \frac{\hat{\mu}_{P,m}}{\hat{\sigma}_{P,m}}, \quad m = 1, \ldots, M. \tag{39}$$

4 Some multiple hypothesis tests

In this section, we present the theoretical foundations of the hypothesis tests which are carried out later on. We go into details because the test procedures of comparable empirical analyses have led to some irritations. In particular, we refer to the studies of DeMiguel et al. (2009a, 2009b). Their analyses involve testing several hypotheses simultaneously. Thus, it is necessary to control the probability of rejecting one or more correct null hypotheses. Otherwise, test decisions do not remain valid. Those inaccuracies occur despite the fact that a broad literature is concerned with the derivation of appropriate multiple-testing procedures. For a nice overview of existing methods, see Romano et al. (2008).

At the present time it is widely discussed in the literature whether portfolio optimization should be passed over in favor of spreading wealth uniformly over all assets (see, e.g., DeMiguel et al. 2009a, 2009b). We face this discussion by asking whether the trivial strategy can significantly outperform some or all of the other investment strategies in terms of the CEQ or the Sharpe ratio. In addition, and as a matter of consequence, we want to test for the superiority of the non-trivial strategies. First, consider the hypotheses

$$H_{0\wedge}: \bigwedge_{m=1}^{M-1} \Delta_m \geq 0 \quad \text{vs.} \quad H_{1\wedge}: \neg H_{0\wedge}, \tag{40}$$

where Δ_m is the true performance difference between the trivial strategy and the mth non-trivial strategy, i.e., $\Delta_m = \mathrm{CEQ}_M - \mathrm{CEQ}_m$ in case of the certainty equivalent and $\Delta_m = \mathrm{Sh}_M - \mathrm{Sh}_m$ in case of the Sharpe ratio. Note that we are not interested in the significance of the underlying single hypotheses in (40) but rather in the significance of the hypothesis $H_{0\wedge}$. Strictly speaking, this is not a multiple-testing problem

and Σ are estimated in-sample and lead to the vectors of portfolio weights characterizing the investment strategies. Each of these vectors is multiplied by the vector of asset returns in the first month of the out-of-sample part forming an out-of-sample portfolio return. Then, the in-sample part of asset returns is shifted forward in steps of one month leading to a series of out-of-sample portfolio returns.

because we do not test several hypotheses simultaneously. Instead, this is a *joint* hypothesis test which will be referred to as the intersection test.

If it is possible to reject $H_{0\wedge}$ we know that there is at least one non-trivial strategy performing significantly better than the trivial one. However, we are not able to identify any *particular* outperforming strategy by means of this test. Remember that we are not able to deduce *any* statistical decisions if $H_{0\wedge}$ cannot be rejected. In particular, we must not treat naive diversification as an outperforming strategy in this case. For this purpose, we have to conduct another test, viz.

$$H_{0\vee} : \bigvee_{m=1}^{M-1} \Delta_m \le 0 \quad \text{vs.} \quad H_{1\vee} : \neg H_{0\vee}. \tag{41}$$

If we can reject $H_{0\vee}$, the trivial strategy turns out to be significantly the best among all non-trivial strategies. Note that this is again a joint hypothesis test which will be referred to as the union test. Although this test is formed only by interchanging the null and the alternative in (40) and turning the inequalities, the possible test decisions as well as the test procedures are completely different.

With regard to the findings of DeMiguel et al. (2009b) and others, there might be little chance that the trivial strategy turns out to be the best among *all* the other strategies. Consequently, our third test strives for detecting *as many* strategies *as possible* which are outperformed by the trivial strategy. Consider the single hypotheses

$$H_{0,m} : \Delta_m \le 0 \quad \text{vs.} \quad H_{1,m} : \Delta_m > 0, \quad m = 1, \dots, M-1. \tag{42}$$

This is a typical multiple-testing problem because we test $M-1$ hypotheses simultaneously. The test procedure differs notably from that of a joint test. The concept of controlling the error of the first kind is widened to the control of the familywise error rate (FWER). A multiple-testing procedure can be used in principle for testing (40). This is simply done by rejecting $H_{0\wedge}$ if the multiple test rejects at least one $H_{0,m}$. However, the goal of a multiple test is to find *every* false single null hypotheses. By contrast, a joint test is only designed to find *any* false single null hypothesis. So it is possible that $H_{0\wedge}$ is rejected by a joint test whereas no $H_{0,m}$ can be rejected by a multiple test.

In the following, we describe the test procedures of the three test problems in (40), (41), and (42). The t-statistic

$$t_m = \frac{\hat{\Delta}_m}{\text{se}(\hat{\Delta}_m)}, \quad m = 1, \dots, M-1 \tag{43}$$

is fundamental for the hereinafter presented test statistics. Here, $\hat{\Delta}_m$ is the *estimated* performance difference between the trivial strategy and the mth competing strategy and $\text{se}(\hat{\Delta}_m)$ denotes the (estimated) standard error of $\hat{\Delta}_m$. Note that $\hat{\Delta}_m = \widehat{\text{CEQ}}_M - \widehat{\text{CEQ}}_m$ or $\hat{\Delta}_m = \widehat{\text{Sh}}_M - \widehat{\text{Sh}}_m$ depending on the context. We begin with the description of the test procedure for the intersection test based on the hypotheses defined in (40). Hansen (2005) proposes a test for comparing the performance of several models in terms of expected loss and calls it a 'test for superior predictive ability' (SPA). While originally constructed to employ sample means, Theorem 1 in Hansen (2005, p. 369) offers a straightforward extension to functionals of sample means when using an appropriate test statistic. In our case,

$$T_{\mathrm{SPA}} = \min\left\{ \min_{m=1,\ldots,M-1} t_m, 0 \right\} \tag{44}$$

complies with the required properties. The decision criterion improves upon the suggestion of White (2000) concerning the ability to detect false hypotheses while asymptotically satisfying a significance level of α. In concrete terms, the distribution of the test statistic T_{SPA} is not estimated under the worst case scenario $\Delta_m = 0$ for all $M-1$ differences of certainty equivalents. On the contrary, it is assumed $\Delta_m = \hat{\Delta}_m$ if $t_m > \sqrt{2 \log\log N}$. As a result, the influence of poorly performing strategies is reduced and the power of the test increases. The null distribution of the test statistic T_{SPA} is approximated via bootstrapping. In particular, let

$$\left(R^*_{m,n,k}\right), \quad m = 1, \ldots, M, \quad n = 1, \ldots, N, \quad k = 1, \ldots, K \tag{45}$$

be the kth bootstrap sample created from the series of out-of-sample portfolio returns $(R_{m,n})$. From $(R^*_{m,n,k})$ it is possible to compute the bootstrap sample estimators $\hat{\Delta}^*_{m,k}$ and $\mathrm{se}(\hat{\Delta}^*_{m,k})$ as well as the bootstrap approximation for the null distribution of the test statistic T_{SPA}, viz.

$$\hat{F}^*_{(1),K}(x) := \begin{cases} \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}(\min_m \frac{\hat{\Delta}^*_{m,k} - \hat{\Delta}_m \mathbb{1}(t_m \le \sqrt{2 \log\log N})}{\mathrm{se}(\hat{\Delta}^*_{m,k})} \le x), & x < 0; \\ 1, & x \ge 0. \end{cases} \tag{46}$$

The p-value inferred by $\hat{F}^*_{(1),K}$ is a consistent estimate of the correct p-value (see Hansen 2005). We reject $H_{0\wedge}$ on a significance level of α if

$$T_{\mathrm{SPA}} < \hat{F}^{*-1}_{(1),K}(\alpha). \tag{47}$$

Next, we describe the procedure of the union test from (41). Remember that this test investigates the superiority of the trivial strategy over *all* non-trivial strategies under consideration. Assume that $H_{0,\vee}$ is rejected whenever *each* single hypothesis $H_{0,m}$ is rejected and let A_m denote the event where $H_{0,m}$ is rejected ($m = 1, \ldots, M-1$). Note that $\mathbb{P}(\bigcap_m A_m) \le \min_m \mathbb{P}(A_m)$ and under $H_{0,\vee}$ at least for one single hypothesis we have $\mathbb{P}(A_m) \le \alpha_m$ where the significance level α_m is assigned to the mth single hypothesis $H_{0,m}$. That means the union test works on the significance level α whenever $\max_m \alpha_m \le \alpha$. The least conservative choice for the single significance levels is given by $\alpha_1, \ldots, \alpha_{M-1} = \alpha$ and so $H_{0,\vee}$ can be rejected if and only if $\max_m p_m \le \alpha$. Here, p_m is the p-value of the mth underlying single hypothesis test. Consequently, our test statistic is

$$T_{\mathrm{LRT}} = \max\left\{ \min_{m=1,\ldots,M-1} t_m, 0 \right\}. \tag{48}$$

Now let z_1 be the strategy with the smallest t-value. The null distribution of T_{LRT} is again approximated via bootstrapping, viz.

$$\hat{F}^*_{z_1,K}(x) := \begin{cases} 0, & x < 0; \\ \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\left(\frac{\hat{\Delta}^*_{z_1,k} - \hat{\Delta}_{z_1}}{\text{se}(\hat{\Delta}^*_{z_1,k})} \le x\right), & x \ge 0. \end{cases} \tag{49}$$

Then, $H_{0\vee}$ is rejected if

$$T_{\text{LRT}} > \hat{F}^{*-1}_{z_1,K}(1-\alpha). \tag{50}$$

At a first glance the presented union test might seem to suffer from a lack of power as would be the case with the Bonferroni correction if applied to the intersection test in (40). The question is whether the test can be improved by taking the correlations between single tests properly into consideration. Frahm (2009a) shows that the proposed test for $H_{0,\vee}$ in fact represents a *likelihood-ratio* test.[7] In particular, it is shown that this likelihood-ratio test is neither determined by the number of single null hypotheses nor by the correlations between the single hypothesis tests.

Finally, we explain the procedure of the multiple-testing problem in (42). Now, we are faced with several hypotheses which are simultaneously tested. Let $I_0 \subset \{1, \ldots, M-1\}$ be the indices of the set of true null hypotheses. Then, the FWER is the probability that any null hypothesis $H_{0,m}$ with $m \in I_0$ is rejected. We seek to asymptotically control the FWER at level α. The stepwise multiple (StepM) test proposed by Romano and Wolf (2005) yields this property and exhibits a good average power. As it starts by examining only the most significant hypothesis, the StepM procedure is of a stepdown nature. If a hypothesis has been rejected in a previous step, it is not considered any more in subsequent steps. If no (further) null hypotheses are rejected in a step, the procedure stops.

The following description of the stepwise procedure corresponds to Algorithms 4.1 and 4.2 in Romano and Wolf (2005). Let z_s ($s = 1, \ldots, M-1$) be the strategy with the sth smallest t-value, i.e.

$$t_{z_1} \le t_{z_2} \le \cdots \le t_{z_s} \le \cdots \le t_{z_{M-1}}. \tag{51}$$

Let B_y denote the number of rejections in the first y steps (with $B_0 = 0$). Further, consider the empirical bootstrap distribution of the maximum of the $M-s$ smallest t-values,

$$\hat{F}^*_{(M-s),K}(x) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\left(\max_{z_1,\ldots,z_{M-s}} \frac{\hat{\Delta}^*_{m,k} - \hat{\Delta}_m}{\text{se}(\hat{\Delta}^*_{m,k})} \le x\right), \quad s = 1, \ldots, M-1. \tag{52}$$

In the first step, our test statistics are simply $t_1, \ldots, t_{M-1}$ and we reject $H_{0,m}$ on a significance level of α if $t_m > \hat{F}^{*-1}_{(M-1),K}(1-\alpha)$. If $B_1 = 0$, then the procedure stops.

[7] This explains the subscript of the test statistic T_{LRT}.

Otherwise, we move on to the next step. In step y, we consider the $M - B_{y-1} - 1$ smallest t-statistics

$$\{t_{z_s} : 1 \le s \le M - B_{y-1} - 1\}. \tag{53}$$

We reject the null hypothesis H_{0,z_s} in step y if

$$t_{z_s} > \hat{F}^{*-1}_{(M-B_{y-1}-1),K}(1-\alpha). \tag{54}$$

We want to add a few remarks on the robust estimation of the standard errors $\text{se}(\hat{\Delta}_m)$ and $\text{se}(\hat{\Delta}^*_{m,k})$. The simplest case occurs when the underlying portfolio returns of the strategies are assumed to be multivariate normally distributed and serially independent. Jobson and Korkie (1981) derived an analytical expression for the standard error of the estimated difference of two Sharpe ratios under this assumption. Their formula has frequently been used in the literature, see, e.g., the test procedure of DeMiguel et al. (2009b). Lo (2002) as well as Ledoit and Wolf (2008) point out that the Jobson–Korkie-test is inappropriate in the case of serially correlated returns or a heavy-tailed return distribution. According to the stylized facts of financial time series, asset returns are heavy-tailed and squared returns are serially correlated (see, e.g., McNeil et al. 2005, pp. 117ff). Any testing procedure which is based on the normal distribution hypothesis and the assumption of serial independence might lead to wrong conclusions.

Frahm (2007) extends the concept of Jobson and Korkie (1981) by assuming only that the underlying return process is strongly stationary and ergodic. Then, the standard error of the difference in Sharpe ratios involves not only the covariance structure, but also the autocovariance structure of the underlying returns. However, it was not the primary concern of Frahm (2007) to derive a particular estimator for the standard error. For this purpose, we will rely on a procedure of Ledoit and Wolf (2008), based on the assumption

$$\sqrt{N}(\hat{\nu}_m - \nu_m) \xrightarrow{d} N(0, \Psi_m), \tag{55}$$

where $\nu_m = [\,\mu_{P,m}\ \mu_{P,M}\ \gamma_{P,m}\ \gamma_{P,M}\,]'$ and $\hat{\nu}_m = [\,\hat{\mu}_{P,m}\ \hat{\mu}_{P,M}\ \hat{\gamma}_{P,m}\ \hat{\gamma}_{P,M}\,]'$. Here, $\gamma_{P,m} = \text{E}(R^2_{m,n})$ and $\mu_{P,M}$ $(\gamma_{P,M})$ is the mean of the (squared) return of the trivial strategy. Ψ_m/N is the unknown asymptotic covariance matrix of the vector $\hat{\nu}_m$. Assume the vector $\mathfrak{R}_{m,n} = [\,R_{m,n}\ R_{M,n}\ R^2_{m,n}\ R^2_{M,n}\,]'$. Then we have

$$\Psi_m = \sum_{q=-\infty}^{\infty} \Gamma_m(q), \tag{56}$$

where $\Gamma_m(q) = \text{Cov}(\mathfrak{R}_{m,n+q}, \mathfrak{R}_{m,n})$ is the autocovariance function of $(\mathfrak{R}_{m,n})$. By contrast, we only have the empirical autocovariance matrices

$$\hat{\Gamma}_m(q) = \frac{1}{N}\sum_{n=1}^{N-q}(\mathfrak{R}_{m,n+q} - \hat{\nu}_m)(\mathfrak{R}_{m,n} - \hat{\nu}_m)', \quad q = 0, 1, \ldots, N-1, \tag{57}$$

at hand. Note that Δ_m can be viewed as a function $f(\nu_m)$, regardless whether we use the certainty equivalent or the Sharpe ratio as our performance measure. Applying the Delta method leads to

$$\sqrt{N}(\hat{\Delta}_m - \Delta_m) \xrightarrow{d} N\big(0, \nabla' f(\nu_m)\Psi_m \nabla f(\nu_m)\big). \tag{58}$$

Note that the gradient of the function $f(\nu_m)$ equals

$$\nabla f(\nu_m) = \begin{cases} [-(1+\lambda\mu_{P,m}) \;\; 1+\lambda\mu_{P,M} \;\; \frac{\lambda}{2} \;\; -\frac{\lambda}{2}]' \quad \text{(CEQ)}, \\ [-\frac{\gamma_{P,m}}{(\gamma_{P,m}-\mu_{P,m}^2)^{3/2}} \;\; \frac{\gamma_{P,M}}{(\gamma_{P,M}-\mu_{P,M}^2)^{3/2}} \;\; \frac{\mu_{P,m}}{2(\gamma_{P,m}-\mu_{P,m}^2)^{3/2}} \;\; -\frac{\gamma_{P,M}}{2(\gamma_{P,M}-\mu_{P,M}^2)^{3/2}}]' \\ \quad \text{(Sharpe ratio)}. \end{cases} \tag{59}$$

The standard error for the difference Δ_m can be estimated as

$$\text{se}(\hat{\Delta}_m) = \sqrt{\frac{\nabla' f(\hat{\nu}_m)\hat{\Psi}_m \nabla f(\hat{\nu}_m)}{N}}. \tag{60}$$

We are in need of a consistent estimator for Ψ_m to obtain a robust estimate of $\text{se}(\hat{\Delta}_m)$. For this purpose we use a HAC estimation procedure developed by Andrews and Monahan (1992) which accounts for the time series characteristics of the portfolio return sample like heteroscedasticity and autocorrelation. In particular, the following steps are conducted:

1. Center the data $(\mathfrak{R}_{m,n})$ and prewhiten them by use of a VAR(1) model.
2. Calculate the empirical autocovariance function of the residuals of the VAR(1) model.
3. Compute the bandwidth parameter for use with the quadratic spectral kernel function (cf. Andrews 1991 for more details).
4. Combine kernel and autocovariance function to estimate the asymptotic covariance matrix of the VAR(1) residuals.
5. Recolor the estimated covariance matrix with the aid of the estimated VAR(1) parameters to obtain $\hat{\Psi}_m$.

Now, we concentrate on the bootstrap sample. To face the time series character of our sample $(R_{m,n})$ we run a stationary block bootstrap with an average block length $b = 5$ and $K = 10\,000$ bootstrap iterations. Politis and Romano (1994) introduce this procedure and demonstrate its robustness regarding the choice of a block length. In contrast to the circular block bootstrap (cf. Politis and Romano 1992), blocks with random lengths are generated according to the geometric distribution with parameter $p = 1/b$. This leads to a less variable estimate of variance since the stationary bootstrap estimate can be viewed as a weighted average over b of the variance estimate based on the circular block bootstrap. Here,

$$\text{se}\big(\hat{\Delta}^*_{m,k}\big) = \sqrt{\frac{\nabla' f(\hat{\nu}^*_{m,k})\hat{\Psi}^*_{m,k} \nabla f(\hat{\nu}^*_{m,k})}{N}}, \tag{61}$$

and $\hat{\Psi}^*_{m,k}$ is the (estimated) asymptotic covariance matrix of $\hat{\nu}^*_{m,k}$.

Assume the vector $\mathfrak{R}^*_{m,n,k} = [\, R^*_{m,n,k} \; R^*_{M,n,k} \; (R^*_{m,n,k})^2 \; (R^*_{M,n,k})^2 \,]'$. Then, $\hat{\Psi}^*_{m,k}$ is the sum of the bootstrap sample autocovariance matrices $\hat{\Gamma}^*_{m,k}(q)$ up to a lag length of $b-1$:

$$\hat{\Psi}^*_{m,k} = \hat{\Gamma}^*_{m,k}(0) + \sum_{q=1}^{b-1} \bigl(\hat{\Gamma}^*_{m,k}(q) + \hat{\Gamma}^*_{m,k}(q)'\bigr), \tag{62}$$

where

$$\hat{\Gamma}^*_{m,k}(q) = \frac{1}{N} \sum_{n=1}^{N-q} \bigl(\mathfrak{R}^*_{m,n+q,k} - \hat{\nu}^*_{m,k}\bigr)\bigl(\mathfrak{R}^*_{m,n,k} - \hat{\nu}^*_{m,k}\bigr)', \quad q = 0, 1, 2, \ldots, N-1. \tag{63}$$

5 Empirical study

5.1 Data

In this section we apply the presented investment strategies to real data. For this purpose we use a sample of stock returns arising from the CRSP data set which covers price information of common stocks traded on the NYSE, AMEX and NASDAQ. Our sample incorporates monthly returns between January 1969 and December 2008.[8]

Besides the stock market with its risky assets it is possible to invest at a risk-free interest rate. This is reflected by use of excess returns. To adjust the CRSP return sample appropriately we use 3-month treasury bills provided online by the Federal Reserve System (2009).

The following procedure is based on Jagannathan and Ma (2003). For each year, beginning in 1979, we consider a set of assets consisting of all stocks which exhibit return data for the last 10 years as well as the subsequent year. The amount of stocks in each set of assets ranges from 1239 assets in the time period 1969–1979 to 2992 assets in the period 1998–2008.[9]

From each set of assets we constitute an asset universe by drawing 100 stocks randomly and without replacement. In each month the investors make use of the last 120 monthly excess returns to estimate the parameters of the joint return distribution. In our study we calculate M vectors of portfolio weights corresponding to M investment strategies. Then the estimation window is switched to the next month in order to keep up an estimation window of 120 observations. This is done from January to December for the given asset universe. For the following estimation window a new asset universe is considered. This procedure is repeated until the end of 2008 is reached. In

[8] We also analyzed weekly return data. However, the findings of our monthly study were endorsed, confer also Sect. 5.2.

[9] For the weekly analysis, the set of assets rises from 1399 stocks in the period 1969–1979 to 2977 assets in the period 1998–2008.

Table 1 Monthly out-of-sample certainty equivalents of the investment strategies for various risk-aversion constellations. In the first part of the Table, we report the results for the five trivial strategies under consideration. Remember that the trivial strategies differ in the proportion of wealth γ invested in risky assets. For example, the Trivial-75-Strategy allocates three quarters of the wealth equally to the risky assets whereas only one quarter is given to the risk-free asset

Strategy	$\lambda = 1$	$\lambda = 3$	$\lambda = 5$	$\lambda = 7$	$\lambda = 10$
Naive strategies					
Trivial-0	0.000000	0.000000	0.000000	0.000000	0.000000
Trivial-25	0.001823	0.001677	0.001530	0.001384	0.001165
Trivial-50	0.003499	0.002914	0.002329	0.001744	0.000866
Trivial-75	0.005030	0.003713	0.002396	0.001080	−0.000895
Trivial-100	0.006413	0.004073	0.001732	−0.000609	−0.004120
Optimizing strategies					
Traditional	0.000747	−0.001129	−0.002936	−0.003759	−0.005575
Jorion	0.003242	0.001486	0.000974	0.000695	0.000486
GMVP (trad.)	0.005085	0.003904	0.002784	0.001876	0.000740
Frahm–Memmel	0.005413	0.004439	0.003356	0.002397	0.001024
Ledoit–Wolf	0.005214	0.004335	0.003558	0.002769	0.001885
Equicorrelation	0.006121	0.003122	0.000934	−0.000258	−0.000364
2-Norm	0.004800	0.003695	0.002576	0.002091	0.001318
Minimax	0.000646	0.000412	0.000280	0.000210	0.000147

that way we try to avoid any kind of survival bias in our data. Altogether, we obtain a series of $N = 360$ out-of-sample portfolio returns for each investment strategy.[10]

5.2 Results

The estimated certainty equivalents for the different investment strategies are reported in Table 1.[11] Our study distinguishes five risk-aversion constellations. As we expected, the performance of all naive strategies—except for the Trivial-0 strategy—diminishes for increasing risk aversion. This is easy to explain as the respective portfolio weights are constant across all λ-constellations implying the same estimators for the expected portfolio return and the return's variance. A higher risk aversion leads to a stronger penalization of variance in the CEQ. Thus, the performance necessarily decreases. Note that the corresponding estimator for the Sharpe ratio is constant over all naive strategies and all risk-aversion parameters, confer also Table 2.

[10] The procedure remains the same for weekly data. Here, the estimation window covers the last 574 weekly excess returns and is weekly moved forward. Consequently, 52 out-of-sample portfolio returns are computed per asset universe, resulting in a total number of 1560 out-of-sample portfolio returns for each of the M strategies.

[11] Throughout this section, we present the results of our study based on monthly return data. For weekly returns, our findings generally correspond to what we discuss hereafter and can be made available upon request.

Table 2 Monthly out-of-sample Sharpe ratios of the investment strategies for various risk-aversion constellations. Note that all trivial strategies exhibit the same Sharpe ratio value independent of the risk-aversion parameter λ. Hence, we list the results for the various naive allocation rules in aggregated form

Strategy	$\lambda = 1$	$\lambda = 3$	$\lambda = 5$	$\lambda = 7$	$\lambda = 10$
Trivial	0.1568	0.1568	0.1568	0.1568	0.1568
Traditional	0.0553	0.0910	0.1034	0.1167	0.1127
Jorion	0.1056	0.1040	0.1068	0.1066	0.1066
GMVP (trad.)	0.1945	0.1890	0.1762	0.1621	0.1406
Frahm–Memmel	0.1851	0.1871	0.1853	0.1856	0.1833
Ledoit–Wolf	0.2008	0.2075	0.2060	0.2015	0.1942
Equicorrelation	0.1517	0.1370	0.1228	0.1074	0.1028
2-Norm	0.1638	0.1754	0.1661	0.1719	0.1637
Minimax	0.0996	0.0861	0.0848	0.0871	0.0871

Interestingly, the ranking of the naive strategies varies. Less risk-averse investors ($\lambda = 1$ and $\lambda = 3$) prefer to invest purely in risky assets. For $\lambda = 5$, the investor is willing to give a quarter of his wealth to the risk-free asset, for $\lambda = 7$ he likes to allocate equal proportions of his wealth to risky and risk-free positions and finally, a very risk-averse investor ($\lambda = 10$) wants to save three quarters of wealth in the risk-free asset. These results are rather intuitive and place once again special emphasis on the use of the CEQ as a suitable performance measure.

The results for the group of non-trivial strategies stress the point that acting according to the traditional sample-based approach is a poor decision rule. The consideration of uncertainty sets for the parameters of interest leads to a slight improvement of performance over the sample-based strategy (cf. the results of the minimax strategy). The celebrated Jorion strategy performs better but cannot compete with the minimum-variance strategies.

In fact, the extended minimum-variance concept leads to better results. The Frahm–Memmel-type strategy performs best when the investor is less risk averse, whereas the Ledoit–Wolf-type strategy is amazingly robust against higher risk aversion. Comparing the results for the 2-norm constrained strategy and the traditional minimum-variance strategy shows how additional constraints can smooth the performance across different λ-constellations.

We now come back to the question of whether portfolio optimization should be passed over in favor of following naive allocation rules. A first understanding is that in almost all constellations we are able to identify an optimizing strategy which exhibits a higher out-of-sample CEQ than every single one of the five naive strategies. An exception is the Trivial-100 strategy which performs best for $\lambda = 1$. In the more realistic case of $\lambda = 3$, the Frahm–Memmel-type strategy turns out to be the best in terms of the estimated CEQ. For higher risk aversion, every single one of the trivial strategies is beaten by the Ledoit–Wolf-type strategy. The absolute value of the CEQ is also of interest. For example, an investor with a risk aversion of $\lambda = 3$ values the risky portfolio returns of the Frahm–Memmel-type strategy as a risk-free monthly extra premium of 0.44%.

In Table 2 we report the estimated Sharpe ratios of the investment strategies. Note that the Sharpe ratio is constant per definition for all naive allocation rules with $\gamma \neq 0$ regardless of how risk-averse the investor is. This is because the proportionality of the portfolio weights leads to proportional (sample) means and standard deviations of the portfolio returns. When calculating the Sharpe ratio, the proportionality factor is simply canceled out. Thus, we report the various trivial strategies in aggregated form. Furthermore, we have to exclude the Trivial-0-Strategy from our analysis since a Sharpe ratio value is not defined.

Comparing the performance of the non-trivial strategies confirms the impression that the group of minimum-variance strategies make themselves stand out from the remaining strategies, even in terms of the Sharpe ratio. Among these, the Ledoit–Wolf-type strategy ranks best for all investigated risk-aversion parameters. The Frahm–Memmel-type as well as the 2-norm constrained strategy show a remarkable robustness against increasing risk aversion while the performance of the traditional minimum-variance strategy is competitive only for a weaker risk aversion.

In contrast to the strategies which follow the extended minimum-variance concept, the traditional sample-based approach as well as the Jorion strategy, the equicorrelation and the minimax strategy are outperformed by the trivial strategy. On top of that, the Sharpe ratios of the sample-based and the equicorrelation strategy vary considerably with increasing risk aversion. Note that in general, the Sharpe ratio of an investment strategy is not constant for different values of the risk-aversion parameter though λ is not explicitly included in the calculation of the Sharpe ratio. However, the optimization problems of all non-trivial strategies include λ as a parameter (confer Sect. 2), leading to different portfolio return samples for different risk-aversion parameters.

We proceed with testing the null that the CEQ of the trivial strategy is at least as high as the CEQ of all competing strategies. As we proposed five versions of naive investments in our setting and considered five values of risk aversion, each test is carried out 25 times separately. Table 3 reports the results of the intersection test. Remember that we seek rejection of the superiority of the naive allocation rule. Taking on a significance level of $\alpha = 5\%$, three of the 25 tests can be rejected.

Interestingly, rejection involves the extreme constellations of being very risk-averse and investing all wealth in risky assets (lower right of Table 3) or being prepared to take more risks and giving the money only to the risk-free asset (upper left of Table 3). Considering both the out-of-sample values of the CEQ in Table 1 and the values of the minimum test statistic in Table 3 we may derive optimal trivial strategies in agreement to the investor's risk behavior. Higher risk aversion clearly prefers the Trivial-25 strategy, whereas a medium risk aversion corresponds to the Trivial-50 strategy ($\lambda = 7$) and the Trivial-75 strategy ($\lambda = 5$). A less risk-averse investor ($\lambda = 1$ and $\lambda = 3$) will make use of the Trivial-100 strategy. Thus, the traditionally used naive strategy is optimal only for a relatively high risk taker. To be more realistic, an investor prefers to mix risky and risk-free positions.

In Table 4, we present the results of the intersection test if the performance is measured by the (monthly) Sharpe ratio. The null hypothesis that the trivial strategy performs best is far away from being rejected. Note that the test statistic T_{SPA} exhibits in all cases a negative value indicating that there is always at least one non-trivial

Table 3 Results of the intersection test for the difference in monthly certainty equivalents. Listed are the values of the test statistic T_{SPA}. The p-values (*in brackets*) are obtained from the empirical bootstrap distribution $\hat{F}^*_{(1),K}$

Benchmark	$\lambda=1$	$\lambda=3$	$\lambda=5$	$\lambda=7$	$\lambda=10$
Trivial-0	−2.9975	−2.6659	−2.2428	−1.8633	−1.4592
	(0.0167)	(0.0377)	(0.0835)	(0.1414)	(0.2693)
Trivial-25	−2.5024	−1.9876	−1.5415	−1.1209	−0.6517
	(0.0590)	(0.1555)	(0.3168)	(0.5169)	(0.7228)
Trivial-50	−1.7130	−1.3433	−0.8895	−0.6709	−0.6606
	(0.2652)	(0.4159)	(0.6407)	(0.6786)	(0.6296)
Trivial-75	−1.3350	−0.5699	−0.7433	−1.0229	−1.4459
	(0.3968)	(0.7948)	(0.5675)	(0.3721)	(0.1663)
Trivial-100	0.0000	−0.1940	−0.8291	−1.5117	−2.3909
	(1.0000)	(0.7116)	(0.3959)	(0.1911)	(0.0293)

Table 4 Results of the intersection test for the difference in monthly Sharpe ratios. Listed are the values of the test statistic T_{SPA}. Note that we only perform one test per λ-constellation since the Sharpe ratio-differences and their standard errors are constant, no matter which of the trivial strategies is considered as benchmark. The p-values (*in brackets*) are obtained from the empirical bootstrap distribution $\hat{F}^*_{(1),K}$

Benchmark	$\lambda=1$	$\lambda=3$	$\lambda=5$	$\lambda=7$	$\lambda=10$
Trivial	−0.7967	−0.8749	−0.8107	−0.7715	−0.6853
	(0.6861)	(0.6182)	(0.6111)	(0.5553)	(0.5861)

strategy with a higher Sharpe ratio than the trivial strategy, confer also Table 2. Thus, we should not think about treating these test results as a proof for the superiority of the naive diversification rule. Note that it suffices to perform only one intersection test per λ-constellation. This is because the differences of the Sharpe ratio as well as the standard errors of these differences are constant irrespective of which trivial strategy is the benchmark.

To sum up, we find in all but one constellation a non-trivial strategy that beats the trivial strategy in terms of the out-of-sample CEQ. Moreover, the naive diversification rule is outperformed by all of the minimum-variance strategies in terms of the estimated Sharpe ratio. Despite that, it is difficult to reject the null hypothesis of the intersection test except for extreme constellations. It is straightforward to conclude that in most cases the standard error of the minimum test statistic T_{SPA} is too large to deduce significant test decisions. A possible way out would be to enlarge the sample size N. DeMiguel et al. (2009b) calculate some critical values for N yielding a rejection of the null that the trivial strategy performs better than the traditional sample-based approach. Their analysis is based on some particular Sharpe ratio-constellations. The insight of these theoretical calculations coincides with our finding that financial data usually do not provide a sample size from which one can

Table 5 Results of the union test for the difference in monthly certainty equivalents. Listed are the values of the test statistic T_{LRT}. The p-values (*in brackets*) are obtained from the empirical bootstrap distribution $\hat{F}^*_{z_1,K}$ where z_1 is the index of the minimum value of all t-statistics

Benchmark	$\lambda=1$	$\lambda=3$	$\lambda=5$	$\lambda=7$	$\lambda=10$
Trivial-0	0	0	0	0	0
	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)
Trivial-25	0	0	0	0	0
	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)
Trivial-50	0	0	0	0	0
	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)
Trivial-75	0	0	0	0	0
	(1.0000)	(1.0000)	(1.0000)	(1.0000)	(1.0000)
Trivial-100	0.4721	0	0	0	0
	(0.3248)	(1.0000)	(1.0000)	(1.0000)	(1.0000)

deduce significant decisions. This result holds true even if we consider weekly return data by which the sample size N is more than quadrupled.

Nevertheless, we should be wary of promoting naive decision rules as a consequence. In Table 5, the results of the union test are given. This test tries to detect some significance in the fact that the trivial strategy is better than every single one of the optimizing strategies. In only one of the 25 cases does the test statistic show a positive value. It would have been possible to deduce the results either from the comparison of certainty equivalents in Table 1 or from the results of the intersection test in Table 3. We report Table 5 anyway as we want to give a notion of how far naive allocation rules are from being significantly the best. Note that the test results based on the Sharpe ratio coincide with the ones of Table 5 except for the case $\lambda = 1$. Here, test statistic T_{LRT} takes a value of zero, too, if the Sharpe ratio is considered as a performance measure.

A weaker claim might be that the trivial strategy performs better than at least one of the optimizing strategies. Having a look back at Table 1 we notice that the estimated CEQ of the traditional sample-based approach is almost always below the estimated CEQ of the naive strategies. There is even more to say: The traditional sample-based approach exhibits a negative expected CEQ for $\lambda \geq 3$ indicating that the investor should put his wealth completely into the risk-free asset rather than to buy stocks. On the other hand, the certainty equivalents of the naive decision rules mixing risk-free and risky assets are obviously positive. Despite that it is almost never possible to find a trivial strategy which is significantly better than any of the optimizing strategies. Table 9 in the appendix reports the results of the StepM test if the certainty equivalent is used as performance measure.

Recall that this test is a stepdown procedure. If on the first step a strategy was found to be significantly outperformed by the benchmark, a further test step is conducted for the remaining strategies. The p-value belonging to a particular strategy typically decreases on subsequent steps. Thus, it is more likely to reject a particular hypothesis on a subsequent step due to a reduced total number of hypotheses under consideration. In our case, the test merely consists of one step. In other words, we cannot reject any of the null hypotheses in 24 of the 25 cases for the differences in certainty equivalents. The p-values are in most cases far away from the critical threshold of 5%. Only one time, when examining the Trivial-50 strategy for a risk aversion of $\lambda = 10$, it is possible to reject the superiority of the traditional sample-based approach.[12]

The StepM test is also conducted for the difference in Sharpe ratios, see Table 10 in the appendix for the results. As has already been the case with the intersection and the union test, we do not distinguish between single trivial strategies since the value of the t-statistic does not vary for different γ-constellations. Hence, we perform the test only five times for five risk-aversion constellations. We cannot reject the null hypothesis that the respective non-trivial strategy is better than the benchmark in any case such that the StepM test consists in all cases of only one step.

5.3 Analyzing the dependence structure of the certainty equivalents

In Sect. 5.2 we have seen that it is difficult to reject any of the test hypotheses due to large standard errors. This is even though we find some empirical evidence against the family of trivial strategies. Closely related to the computation of standard errors is the analysis of the dependence structure of the performance measure for different investment strategies. In the following, we concentrate on the certainty equivalent but the Sharpe ratio might be analyzed in the same way. Table 6 reports the average correlations between the certainty equivalents of different strategies induced by the stationary block bootstrap procedure for $\lambda = 5$.[13] It is obvious that most of the certainty equivalents are highly positive correlated except for the CEQ of the minimax strategy which exhibits rather moderate correlations to the CEQ of the other strategies.

It is questionable whether the stationary block bootstrap—or bootstrap procedures in general—can successfully detect the dependence structure on such short samples, if it is strong. To this end, we want to test this ability by means of a simulation study. In order to create samples as close as possible to the real sample of portfolio returns, we proceed as follows. Extending the notation of Sect. 4, assume the vector $\mathfrak{R}_n = [\,R_{1,n} \ldots R_{M,n}\ (R_{1,n})^2 \ldots (R_{M,n})^2\,]'$ and the vector $\mathfrak{R}_n^* = [\,R_{1,n}^* \ldots R_{M,n}^*\ (R_{1,n}^*)^2 \ldots (R_{M,n}^*)^2\,]'$. Here, $(R_{m,n}^*)$ is a random bootstrap sample created from the original sample of portfolio returns $(R_{m,n})$ using the stationary bootstrap procedure. Furthermore, let

$$\hat{\Gamma}(q) = \frac{1}{N} \sum_{n=1}^{N-q} (\mathfrak{R}_{n+q} - \hat{\nu})(\mathfrak{R}_n - \hat{\nu})', \quad q = 0, 1, \ldots, N-1 \tag{64}$$

[12] In the case of weekly data, the traditional sample-based strategy is more often found to be significantly outperformed by the trivial strategy. However, it is not a new fact that simply treating the sample estimators as the true parameters μ and Σ is a poor decision rule in the light of estimation error.

[13] The estimated correlation matrices for the other four risk-aversion constellations are very similar to the one in Table 6.

Table 6 Average correlations between the monthly certainty equivalents of different strategies, calculated on the bootstrap samples for $\lambda = 5$. We used 10 000 bootstrap replications of the underlying sample of portfolio returns. Note that the pairwise correlation of the Trivial-0-Strategy with any other strategy is zero per definition

	T-25	T-50	T-75	T-100	trad	BS	MV	FM	LW	EC	2-n	mm
T-25[a]	1.0000	0.9982	0.9932	0.9857	0.7675	0.6930	0.6002	0.8196	0.5697	0.9144	0.6765	0.1322
T-50[a]	0.9982	1.0000	0.9984	0.9940	0.7788	0.6992	0.6120	0.8344	0.5782	0.9292	0.6881	0.1301
T-75[a]	0.9932	0.9984	1.0000	0.9986	0.7869	0.7028	0.6211	0.8457	0.5843	0.9401	0.6968	0.1277
T-100[a]	0.9857	0.9940	0.9986	1.0000	0.7923	0.7042	0.6279	0.8538	0.5883	0.9476	0.7030	0.1252
trad[a]	0.7675	0.7788	0.7869	0.7923	1.0000	0.7824	0.6496	0.7352	0.6074	0.7587	0.6967	0.1557
BS[a]	0.6930	0.6992	0.7028	0.7042	0.7824	1.0000	0.8725	0.8119	0.7861	0.7584	0.8930	0.2912
MV[a]	0.6002	0.6120	0.6211	0.6279	0.6496	0.8725	1.0000	0.8655	0.9137	0.6870	0.9380	0.1959
FM[a]	0.8196	0.8344	0.8457	0.8538	0.7352	0.8119	0.8655	1.0000	0.8046	0.8660	0.8654	0.1785
LW[a]	0.5697	0.5782	0.5843	0.5883	0.6074	0.7861	0.9137	0.8046	1.0000	0.6110	0.8290	0.1823
EC[a]	0.9144	0.9292	0.9401	0.9476	0.7587	0.7584	0.6870	0.8660	0.6110	1.0000	0.7737	0.1534
2-n[a]	0.6765	0.6881	0.6968	0.7030	0.6967	0.8930	0.9380	0.8654	0.8290	0.7737	1.0000	0.2014
mm[a]	0.1322	0.1301	0.1277	0.1252	0.1557	0.2912	0.1959	0.1785	0.1823	0.1534	0.2014	1.0000

[a]The names of the investment strategies are abbreviated as follows: T-25—Trivial-25-Strategy, T-50—Trivial-50-Strategy, T-75—Trivial-75-Strategy, T-100—Trivial-100-Strategy, trad—traditional sample-based strategy, BS—Bayesian 'Jorion' Strategy, MV—traditional minimum-variance strategy, FM—Frahm–Memmel-type strategy, LW—Ledoit–Wolf-type strategy, EC—Equicorrelation Strategy, 2-n—2-norm constrained strategy, mm—minimax strategy

be the empirical autocovariance function of $(\mathfrak{R}_n)$, where $\hat{\nu} = [\hat{\mu}_{P,1} \ \ldots \ \hat{\mu}_{P,M} \ \hat{\gamma}_{P,1} \ \ldots \ \hat{\gamma}_{P,M}]'$. In the following, we synonymously refer to $(\mathfrak{R}_n^*)$ as the stochastic process *and* the corresponding time series. Let the true (asymptotic) covariance matrix of $(\mathfrak{R}_n^*)$ be denoted by Ψ^* and remember that Ψ^* is also the true covariance matrix of $\sqrt{N}\hat{\nu}^*$ where $\hat{\nu}^* = [\hat{\mu}_{P,1}^* \cdots \hat{\mu}_{P,M}^* \ \hat{\gamma}_{P,1}^* \cdots \hat{\gamma}_{P,M}^*]'$ is the vector of the sample means of $(\mathfrak{R}_n^*)$. According to Proposition 3.2 in Lahiri (2003) we are able to calculate

$$\Psi^* = \hat{\Gamma}(0) + \sum_{q=1}^{N-1} \left(\frac{N-q}{N}(1-p)^q + \frac{q}{N}(1-p)^{N-q} \right) \big(\hat{\Gamma}(q) + \hat{\Gamma}(q)' \big). \tag{65}$$

Note that the estimated certainty equivalents $\widehat{\mathrm{CEQ}} = \big[\widehat{\mathrm{CEQ}}_1 \ldots \widehat{\mathrm{CEQ}}_M\big]'$ can be seen as an M-variate function $g(\hat{\nu})$ of the sample means of $(\mathfrak{R}_n)$—just the way we have handled the differences in certainty equivalents in Sect. 4. Applying once again the Delta method leads to

$$\mathrm{Cov}(\widehat{\mathrm{CEQ}}^*) = \nabla' g(\hat{\nu}) \Psi^* \nabla g(\hat{\nu}), \tag{66}$$

where $\widehat{\mathrm{CEQ}}^* = \big[\widehat{\mathrm{CEQ}}_1^* \ldots \widehat{\mathrm{CEQ}}_M^*\big]'$ is the vector of the certainty equivalents of the M investment strategies, estimated on a random bootstrap sample, and

$$\nabla g(\hat{\nu}) = \begin{bmatrix} 1+\lambda\hat{\mu}_{P,1} & 0 & \ldots & 0 & -\lambda/2 & 0 & \ldots & 0 \\ 0 & 1+\lambda\hat{\mu}_{P,2} & \ldots & 0 & 0 & -\lambda/2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1+\lambda\hat{\mu}_{P,M} & 0 & 0 & \ldots & -\lambda/2 \end{bmatrix}' \tag{67}$$

is the gradient of the function $g(\hat{\nu})$. Now, we apply the stationary bootstrap procedure with average block length $b = 1/p = 5$ to create a time series $(\mathfrak{R}_n^*)$ and use the original sample $(\mathfrak{R}_n)$ in combination with (66) to compute the theoretical covariance matrix of the vector $\widehat{\mathrm{CEQ}}^*$, see Table 7.

We consider $(\mathfrak{R}_n)$ as our underlying population and the time series $(\mathfrak{R}_n^*)$ as our sample. Thus, by resampling from $(\mathfrak{R}_n^*)$ we generate a bootstrap sample $(\mathfrak{R}_n^{**})$. For this purpose, we apply once again the stationary block bootstrap procedure with average block length $b = 5$. We estimate the covariance matrix of $(\mathfrak{R}_{n,k}^{**})$ on the kth bootstrap sample by

$$\hat{\Psi}_k^{**} = \hat{\Gamma}_k^{**}(0) + \sum_{q=1}^{b-1} \big(\hat{\Gamma}_k^{**}(q) + \hat{\Gamma}_k^{**}(q)' \big), \tag{68}$$

where $\hat{\Gamma}_k^{**}(q)$ is the empirical autocovariance function of $(\mathfrak{R}_{n,k}^{**})$ for a lag length $0 \leq q \leq N-1$. Finally, we get an estimate of the covariance matrix of $\widehat{\mathrm{CEQ}}_k^{**}$ (which

Table 7 Theoretical covariance matrix of the vector $\widehat{\mathrm{CEQ}}^*$. The values for the Trivial-0-Strategy are zero by definition and thus they are not reported. Note that the original sample of portfolio returns for $\lambda = 10$ served as a basis and Equation 66 was applied. We scaled all reported values by a factor of 10^5

	T-25	T-50	T-75	T-100	trad	BS	MV	FM	LW	EC	2-n	mm
T-25	0.0499	0.1025	0.1579	0.2162	0.1702	0.0194	0.0608	0.1112	0.0456	0.0919	0.0542	0.0009
T-50	0.1025	0.2115	0.3271	0.4492	0.3565	0.0403	0.1276	0.2335	0.0948	0.1922	0.1127	0.0018
T-75	0.1579	0.3271	0.5075	0.6991	0.5588	0.0629	0.2004	0.3672	0.1476	0.3007	0.1754	0.0027
T-100	0.2162	0.4492	0.6991	0.9659	0.7772	0.0870	0.2792	0.5120	0.2038	0.4176	0.2423	0.0037
trad	0.1702	0.3565	0.5588	0.7772	0.9587	0.1051	0.3370	0.4971	0.2556	0.3683	0.2822	0.0035
BS	0.0194	0.0403	0.0629	0.0870	0.1051	0.0157	0.0517	0.0645	0.0389	0.0459	0.0450	0.0011
MV	0.0608	0.1276	0.2004	0.2792	0.3370	0.0517	0.2059	0.2366	0.1506	0.1492	0.1649	0.0027
FM	0.1112	0.2335	0.3672	0.5120	0.4971	0.0645	0.2366	0.3776	0.1769	0.2407	0.1933	0.0033
LW	0.0456	0.0948	0.1476	0.2038	0.2556	0.0389	0.1506	0.1769	0.1509	0.1036	0.1236	0.0023
EC	0.0919	0.1922	0.3007	0.4176	0.3683	0.0459	0.1492	0.2407	0.1036	0.2122	0.1313	0.0028
2-n	0.0542	0.1127	0.1754	0.2423	0.2822	0.0450	0.1649	0.1933	0.1236	0.1313	0.1527	0.0035
mm	0.0009	0.0018	0.0027	0.0037	0.0035	0.0011	0.0027	0.0033	0.0023	0.0028	0.0035	0.0011

is nothing else than the vector of certainty equivalents estimated on the kth bootstrap sample) by applying the Delta method, viz.

$$\widehat{\mathrm{Cov}}(\widehat{\mathrm{CEQ}}_k^{**}) = \nabla' g\big(\hat{\nu}_k^{**}\big)\hat{\Psi}_k^{**}\nabla g\big(\hat{\nu}_k^{**}\big), \tag{69}$$

where $\hat{\nu}_k^{**}$ is the vector of the sample means of $(\mathfrak{R}_{n,k}^{**})$. A bootstrap estimate for the covariance matrix of $\widehat{\mathrm{CEQ}}^*$ follows by averaging over all K bootstrap replications, see Table 8 for the results:

$$\widehat{\mathrm{Cov}}(\widehat{\mathrm{CEQ}}^*) = \frac{1}{K}\sum_{k=1}^{K}\widehat{\mathrm{Cov}}(\widehat{\mathrm{CEQ}}_k^{**}). \tag{70}$$

The comparison of the Tables 7 and 8 shows that the entries of the theoretical covariance matrix exceed their estimated counterparts in almost all cases. Thus, the stationary block bootstrap procedure leads to a slight underestimation of the dependence structure of the certainty equivalents. However, the specific outcome in Table 8 depends heavily on the realized path of $(\mathfrak{R}^*)$. We repeatedly carried out the random experiment and find for some of the simulated time series that $\mathrm{Cov}(\widehat{\mathrm{CEQ}}^*)$ is partly overestimated. By averaging over the estimates $\widehat{\mathrm{Cov}}(\widehat{\mathrm{CEQ}}^*)$ of 100 simulated time series $(\mathfrak{R}^*)$, the theoretical covariance matrix in Table 7 is well approximated. Apart from that, the theoretical *correlation* matrix of $\widehat{\mathrm{CEQ}}^*$ is well reproduced by the bootstrap estimate, confer Tables 11 and 12 in the Appendix. This is despite the fact that the corresponding covariance matrix is underestimated. Hence, we are concerned with more or less even-underestimated variances and covariances.

6 Conclusion

We compare the performance of eight non-trivial investment strategies with that of the trivial strategies. We include a wide range of approaches to find the optimal asset allocation, considering the traditional sample-based approach, several minimum-variance techniques, a shrinkage approach and a minimax procedure. We focus on establishing a realistic setting including short-selling constraints and a risk-free asset. We propose a possible way to widen the concept of minimum-variance strategies to this setting. In general, we are convinced that the comparison of performance should always be based on a consistent setting. Thus, we suggest a family of naive strategies to compete with the optimizing strategies.

The use of a suitable performance measure is essential for our study. We demonstrate the advantages of the CEQ over the Sharpe ratio in our setting. The CEQ incorporates both the relative portfolio weights of the risky assets and the fraction invested in the risk-free asset. On the other hand, the Sharpe ratio allows a comparison independently of the individual risk preferences of the investor. We base our study on out-of-sample portfolio returns of the strategies. Our empirical study indicates that both the average CEQ and the average Sharpe ratio of the Frahm–Memmel-type and Ledoit–Wolf-type strategy are higher than those of all naive strategies in almost all constellations. Especially for a medium-sized risk aversion, these minimum-variance

Table 8 Estimated covariance matrix of $\widehat{\mathrm{CEQ}}^*$. The table reports the average values induced by the stationary block bootstrap procedure which was applied to the time series $(\mathfrak{R}_n^*)$. We created $K = 10\,000$ bootstrap replications. All reported values are scaled by a factor of 10^5

	T-25	T-50	T-75	T-100	trad	BS	MV	FM	LW	EC	2-n	mm
T-25	0.0470	0.0969	0.1497	0.2054	0.1641	0.0169	0.0538	0.1112	0.0407	0.0810	0.0439	0.0007
T-50	0.0969	0.2004	0.3106	0.4275	0.3439	0.0356	0.1145	0.2348	0.0856	0.1698	0.0928	0.0014
T-75	0.1497	0.3106	0.4829	0.6665	0.5394	0.0561	0.1821	0.3709	0.1349	0.2665	0.1468	0.0022
T-100	0.2054	0.4275	0.6665	0.9222	0.7506	0.0785	0.2567	0.5194	0.1885	0.3711	0.2059	0.0030
trad	0.1641	0.3439	0.5394	0.7506	0.9325	0.0993	0.3115	0.4872	0.2400	0.3515	0.2457	0.0038
BS	0.0169	0.0356	0.0561	0.0785	0.0993	0.0140	0.0439	0.0582	0.0354	0.0414	0.0364	0.0011
MV	0.0538	0.1145	0.1821	0.2567	0.3115	0.0439	0.1701	0.2103	0.1324	0.1306	0.1287	0.0029
FM	0.1112	0.2348	0.3709	0.5194	0.4872	0.0582	0.2103	0.3802	0.1596	0.2285	0.1661	0.0030
LW	0.0407	0.0856	0.1349	0.1885	0.2400	0.0354	0.1324	0.1596	0.1420	0.0977	0.1061	0.0039
EC	0.0810	0.1698	0.2665	0.3711	0.3515	0.0414	0.1306	0.2285	0.0977	0.1839	0.1044	0.0022
2-n	0.0439	0.0928	0.1468	0.2059	0.2457	0.0364	0.1287	0.1661	0.1061	0.1044	0.1108	0.0028
mm	0.0007	0.0014	0.0022	0.0030	0.0038	0.0011	0.0029	0.0030	0.0039	0.0022	0.0028	0.0005

strategies seem to outperform the benchmark. However, these results are not significant.

To be more precise, the null hypothesis that trivial strategies perform at least as well as all non-trivial strategies cannot be rejected except for extreme constellations. This finding coincides with that of DeMiguel et al. (2009b) but it cannot be rejected either that at least one of the non-trivial strategies performs better than the trivial ones. The multiple tests reveal only that for a very high risk aversion of $\lambda = 10$ the superiority of the traditional sample-based strategy can be rejected in the case of the CEQ. The tests with which the performance is examined based on the Sharpe ratio show even less significant results.

All in all, our results show by considering historical data and applying contemporary methods of multiple testing it is hardly possible to promote any specific investment strategy. These statistical results indicate that it is hard to find any strategy which is significantly the best because the sample size is too small. We carry out an additional analysis based on weekly return data in order to investigate the impact of an increased sample size. Our finding is that a quadruplication of the sample size still does not lead to significant results while the ranking of the investment strategies principally persists favoring the extended minimum-variance concept.

Acknowledgements Many thanks belong to Karl Mosler, Michael Wolf and two anonymous referees for their very important comments on the manuscript. Wickern and Wiechers gratefully acknowledge financial support provided by the German Research Foundation (DFG).

Appendix

Table 9 Results of the StepM test for the difference in monthly certainty equivalents. Listed are the values of the t-statistic The p-values are given *in brackets*

	$\lambda = 1$ Step 1	$\lambda = 3$ Step 1	$\lambda = 5$ Step 1	$\lambda = 7$ Step 1	$\lambda = 10$ Step 1	 Step 2
Benchmark: Trivial-0 strategy						
Traditional	−0.1476	0.2528	0.6853	0.9337	1.4134	
	(0.9167)	(0.7811)	(0.6053)	(0.4921)	(0.2945)	
Jorion	−1.2525	−1.0440	−1.0544	−1.0522	−1.0522	
	(0.9986)	(0.9926)	(0.9929)	(0.9914)	(0.9926)	
GMVP	−2.9399	−2.3335	−1.7302	−1.1792	−0.4589	
	(1.0000)	(1.0000)	(0.9999)	(0.9956)	(0.9446)	
Frahm–Memmel	−2.6013	−2.0529	−1.4945	−1.0354	−0.4236	
	(1.0000)	(1.0000)	(0.9992)	(0.9907)	(0.9393)	
Ledoit–Wolf	−2.9975	−2.6659	−2.2428	−1.8633	−1.4592	
	(1.0000)	(1.0000)	(1.0000)	(0.9999)	(0.9993)	
Equicorrelation	−1.7953	−0.9433	−0.2951	0.0959	0.1878	
	(1.0000)	(0.9896)	(0.9228)	(0.8273)	(0.7885)	
2-Norm	−2.1350	−1.8756	−1.3935	−1.3398	−1.0492	
	(1.0000)	(1.0000)	(0.9984)	(0.9976)	(0.9925)	
Minimax	−1.6511	−1.3542	−1.2684	−1.2935	−1.2940	
	(0.9998)	(0.9987)	(0.9974)	(0.9974)	(0.9975)	

Table 9 (*Continued*)

	$\lambda = 1$ Step 1	$\lambda = 3$ Step 1	$\lambda = 5$ Step 1	$\lambda = 7$ Step 1	$\lambda = 10$ Step 1	Step 2
Benchmark: Trivial-25 strategy						
Traditional	0.2347	0.7691	1.2680	1.6121	2.1864	
	(0.9408)	(0.6833)	(0.3924)	(0.2160)	(0.0612)	
Jorion	−0.7600	0.2010	0.8508	1.2241	1.2001	
	(1.0000)	(0.9302)	(0.6238)	(0.3936)	(0.3950)	
GMVP	−2.4116	−1.7223	−1.0069	−0.4011	0.3429	
	(1.0000)	(1.0000)	(1.0000)	(0.9968)	(0.8540)	
Frahm–Memmel	−2.5024	−1.8600	−1.1882	−0.6380	0.0830	
	(1.0000)	(1.0000)	(1.0000)	(0.9995)	(0.9332)	
Ledoit–Wolf	−2.4270	−1.9876	−1.5415	−1.1209	−0.6517	
	(1.0000)	(1.0000)	(1.0000)	(0.9999)	(0.9977)	
Equicorrelation	−1.8868	−0.6519	0.2796	0.9521	1.4231	
	(1.0000)	(0.9998)	(0.8991)	(0.5515)	(0.2811)	
2-Norm	−1.8956	−1.4152	−0.7722	−0.6113	−0.1617	
	(1.0000)	(1.0000)	(0.9995)	(0.9994)	(0.9727)	
Minimax	1.3664	1.4723	1.4523	1.3448	1.1416	
	(0.3897)	(0.2968)	(0.3005)	(0.3303)	(0.4249)	
Benchmark: Trivial-50 strategy						
Traditional	0.6749	1.2582	1.7854	2.0909	2.5626	
	(0.7881)	(0.4260)	(0.1881)	(0.1000)	(0.0387)	
Jorion	0.1637	1.2363	1.1644	0.8398	0.2712	
	(0.9730)	(0.4381)	(0.4684)	(0.6212)	(0.8482)	(0.7957)
GMVP	−1.1645	−0.7414	−0.3396	−0.0963	0.0880	
	(1.0000)	(0.9999)	(0.9917)	(0.9600)	(0.9024)	(0.8569)
Frahm–Memmel	−1.7130	−1.3433	−0.8895	−0.5604	−0.1302	
	(1.0000)	(1.0000)	(0.9994)	(0.9946)	(0.9479)	(0.9125)
Ledoit–Wolf	−1.1895	−0.9914	−0.8324	−0.6709	−0.6606	
	(1.0000)	(1.0000)	(0.9990)	(0.9971)	(0.9931)	(0.9822)
Equicorrelation	−1.6861	−0.1359	0.9359	1.7699	1.7696	
	(1.0000)	(0.9888)	(0.5969)	(0.1814)	(0.1799)	(0.1624)
2-norm	−1.0341	−0.6175	−0.1920	−0.2701	−0.3236	
	(1.0000)	(0.9999)	(0.9830)	(0.9788)	(0.9738)	(0.9480)
Minimax	1.7855	1.4879	1.1506	0.8421	0.3850	
	(0.2170)	(0.3090)	(0.4760)	(0.6202)	(0.8053)	(0.7516)

Table 9 (*Continued*)

	$\lambda = 1$ Step 1	$\lambda = 3$ Step 1	$\lambda = 5$ Step 1	$\lambda = 7$ Step 1	$\lambda = 10$ Step 1	Step 2
Benchmark: Trivial-75 strategy						
Traditional	1.1714	1.7410	2.1807	2.2136	2.2153	
	(0.4912)	(0.1855)	(0.0755)	(0.0734)	(0.0742)	
Jorion	1.1019	1.2226	0.7361	0.1797	−0.5712	
	(0.5326)	(0.4225)	(0.6573)	(0.8456)	(0.9593)	
GMVP	−0.0308	−0.1040	−0.2008	−0.3833	−0.7295	
	(0.9902)	(0.9740)	(0.9611)	(0.9577)	(0.9731)	
Frahm–Memmel	−0.3021	−0.5699	−0.7433	−1.0229	−1.4459	
	(0.9997)	(0.9982)	(0.9946)	(0.9945)	(0.9984)	
Ledoit–Wolf	−0.0976	−0.3132	−0.5569	−0.7426	−1.1782	
	(0.9949)	(0.9906)	(0.9904)	(0.9861)	(0.9943)	
Equicorrelation	−1.3350	0.6842	1.5051	1.4157	−0.4213	
	(1.0000)	(0.7130)	(0.2642)	(0.2930)	(0.9410)	
2-norm	0.1615	0.0114	−0.1068	−0.5392	−0.9459	
	(0.9629)	(0.9563)	(0.9496)	(0.9735)	(0.9868)	
Minimax	1.8897	1.3284	0.7740	0.3090	−0.3576	
	(0.1630)	(0.3664)	(0.6377)	(0.8011)	(0.9295)	
Benchmark: Trivial-100 strategy						
Traditional	1.6979	2.1318	2.1188	1.4854	0.6440	
	(0.1640)	(0.0669)	(0.0721)	(0.2156)	(0.4967)	
Jorion	1.5750	0.9720	0.2685	−0.4133	−1.2781	
	(0.2068)	(0.4656)	(0.7592)	(0.9043)	(0.9827)	
GMVP	0.5320	0.0661	−0.3842	−0.8109	−1.4355	
	(0.7377)	(0.8650)	(0.9365)	(0.9613)	(0.9887)	
Frahm–Memmel	0.5490	−0.1940	−0.8291	−1.5117	−2.3909	
	(0.7299)	(0.9237)	(0.9823)	(0.9957)	(0.9998)	
Ledoit–Wolf	0.4721	−0.0954	−0.6271	−1.0434	−1.7524	
	(0.7646)	(0.9038)	(0.9682)	(0.9794)	(0.9962)	
Equicorrelation	1.3177	1.4233	0.7340	−0.2273	−1.6151	
	(0.3148)	(0.2563)	(0.5505)	(0.8613)	(0.9943)	
2-norm	0.8175	0.1772	−0.3551	−0.9926	−1.5420	
	(0.5862)	(0.8302)	(0.9321)	(0.9762)	(0.9927)	
Minimax	1.8837	1.1128	0.3901	−0.2120	−1.0519	
	(0.1128)	(0.3933)	(0.7057)	(0.8575)	(0.9683)	

Table 10 Results of the StepM test for the difference in monthly Sharpe ratios. Listed are the values of the t-statistic. The p-values are given *in brackets*

	$\lambda = 1$ Step 1	$\lambda = 3$ Step 1	$\lambda = 5$ Step 1	$\lambda = 7$ Step 1	$\lambda = 10$ Step 1
Benchmark: Trivial strategy					
Traditional	2.2072	1.6105	1.3820	1.0129	1.0803
	(0.0872)	(0.2532)	(0.3403)	(0.4988)	(0.4437)
Jorion	1.1535	1.0253	0.9555	0.9658	0.9558
	(0.4609)	(0.5084)	(0.5348)	(0.5198)	(0.5013)
GMVP	−0.7301	−0.6002	−0.3454	−0.0918	0.2700
	(0.9939)	(0.9844)	(0.9566)	(0.9111)	(0.8014)
Frahm–Memmel	−0.7900	−0.8270	−0.7678	−0.7715	−0.6853
	(0.9950)	(0.9937)	(0.9859)	(0.9857)	(0.9773)
Ledoit–Wolf	−0.7967	−0.8749	−0.8107	−0.6930	−0.5575
	(0.9952)	(0.9944)	(0.9880)	(0.9827)	(0.9672)
Equicorrelation	1.0064	1.3686	1.4839	1.9149	2.0145
	(0.5382)	(0.3496)	(0.2949)	(0.1505)	(0.1246)
2-Norm	−0.1741	−0.3824	−0.1756	−0.2736	−0.1192
	(0.9615)	(0.9679)	(0.9326)	(0.9398)	(0.9057)
Minimax	0.6939	0.7850	0.7569	0.6870	0.6141
	(0.7036)	(0.6300)	(0.6317)	(0.6475)	(0.6591)

Table 11 Theoretical correlation matrix of the vector $\widehat{\text{CEQ}}^*$, denoted by $\text{Corr}(\widehat{\text{CEQ}}^*)$

	T-25	T-50	T-75	T-100	trad	BS	MV	FM	LW	EC	2-n	mm
T-25	1.0000	0.9981	0.9929	0.9850	0.7785	0.6922	0.6003	0.8101	0.5260	0.8939	0.6216	0.1154
T-50	0.9981	1.0000	0.9983	0.9937	0.7916	0.6997	0.6115	0.8263	0.5307	0.9071	0.6271	0.1151
T-75	0.9929	0.9983	1.0000	0.9985	0.8011	0.7044	0.6200	0.8387	0.5332	0.9164	0.6301	0.1144
T-100	0.9850	0.9937	0.9985	1.0000	0.8076	0.7066	0.6260	0.8477	0.5339	0.9224	0.6310	0.1135
trad	0.7785	0.7916	0.8011	0.8076	1.0000	0.8563	0.7585	0.8261	0.6720	0.8167	0.7376	0.1076
BS	0.6922	0.6997	0.7044	0.7066	0.8563	1.0000	0.9090	0.8372	0.7994	0.7942	0.9192	0.2730
MV	0.6003	0.6115	0.6200	0.6260	0.7585	0.9090	1.0000	0.8485	0.8542	0.7139	0.9298	0.1785
FM	0.8101	0.8263	0.8387	0.8477	0.8261	0.8372	0.8485	1.0000	0.7409	0.8502	0.8051	0.1600
LW	0.5260	0.5307	0.5332	0.5339	0.6720	0.7994	0.8542	0.7409	1.0000	0.5789	0.8145	0.1753
EC	0.8939	0.9071	0.9164	0.9224	0.8167	0.7942	0.7139	0.8502	0.5789	1.0000	0.7295	0.1790
2-n	0.6216	0.6271	0.6301	0.6310	0.7376	0.9192	0.9298	0.8051	0.8145	0.7295	1.0000	0.2705
mm	0.1154	0.1151	0.1144	0.1135	0.1076	0.2730	0.1785	0.1600	0.1753	0.1790	0.2705	1.0000

Table 12 Estimated correlation matrix of the vector $\widehat{\mathrm{CEQ}}^*$ using the stationary block bootstrap procedure. We calculate $\widehat{\mathrm{Corr}}(\widehat{\mathrm{CEQ}}_k^{**}) = \mathrm{diag}(\widehat{\mathrm{Cov}}(\widehat{\mathrm{CEQ}}_k^{**}))^{-1/2}\widehat{\mathrm{Cov}}(\widehat{\mathrm{CEQ}}_k^{**})\mathrm{diag}(\widehat{\mathrm{Cov}}(\widehat{\mathrm{CEQ}}_k^{**}))^{-1/2}$ in the kth bootstrap sample. The table reports the average values $\widehat{\mathrm{Corr}}(\widehat{\mathrm{CEQ}}^*) = K^{-1}\sum_{k=1}^{K}\widehat{\mathrm{Corr}}(\widehat{\mathrm{CEQ}}_k^{**})$. We created $K = 10\,000$ bootstrap replications

	T-25	T-50	T-75	T-100	trad	BS	MV	FM	LW	EC	2-n	mm
T-25	1.0000	0.9983	0.9937	0.9868	0.7778	0.6456	0.5828	0.8248	0.4920	0.8679	0.5985	0.1415
T-50	0.9983	1.0000	0.9985	0.9945	0.7889	0.6591	0.6003	0.8430	0.5018	0.8807	0.6128	0.1404
T-75	0.9937	0.9985	1.0000	0.9987	0.7969	0.6698	0.6149	0.8574	0.5095	0.8900	0.6244	0.1391
T-100	0.9868	0.9945	0.9987	1.0000	0.8024	0.6781	0.6270	0.8686	0.5153	0.8964	0.6337	0.1374
trad	0.7778	0.7889	0.7969	0.8024	1.0000	0.8598	0.7646	0.8112	0.6520	0.8381	0.7533	0.1803
BS	0.6456	0.6591	0.6698	0.6781	0.8598	1.0000	0.8932	0.7917	0.7914	0.8043	0.9208	0.4099
MV	0.5828	0.6003	0.6149	0.6270	0.7646	0.8932	1.0000	0.8206	0.8588	0.7130	0.9386	0.2967
FM	0.8248	0.8430	0.8574	0.8686	0.8112	0.7917	0.8206	1.0000	0.6859	0.8547	0.8075	0.2034
LW	0.4920	0.5018	0.5095	0.5153	0.6520	0.7914	0.8588	0.6859	1.0000	0.5983	0.8421	0.4282
EC	0.8679	0.8807	0.8900	0.8964	0.8381	0.8043	0.7130	0.8547	0.5983	1.0000	0.7198	0.2329
2-n	0.5985	0.6128	0.6244	0.6337	0.7533	0.9208	0.9386	0.8075	0.8421	0.7198	1.0000	0.3516
mm	0.1415	0.1404	0.1391	0.1374	0.1803	0.4099	0.2967	0.2034	0.4282	0.2329	0.3516	1.000

References

Andrews, D.W.K.: Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica **59**, 817–858 (1991)

Andrews, D.W.K., Monahan, J.C.: An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica **60**, 953–966 (1992)

Bade, A., Frahm, G., Jaekel, U.: A general approach to Bayesian portfolio optimization. Math. Methods Oper. Res. **70**, 337–356 (2008)

Behr, P., Güttler, A., Miebs, F.: On portfolio optimization: Imposing the right constraints matters. Discussion Paper, University of Frankfurt (2010)

Brown, S.J.: Optimal portfolio choice under uncertainty: a Bayesian approach. PhD thesis, University of Chicago (1976)

Ceria, S., Stubbs, R.A.: Incorporating estimation errors into portfolio selection: robust portfolio construction. J. Asset Manag. **7**, 109–127 (2006)

Chopra, V.K., Ziemba, W.T.: The effect of errors in means, variances, and covariances on optimal portfolio choice. J. Portf. Manag. **19**, 6–11 (1993)

DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R.: A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manag. Sci. **55**, 798–812 (2009a)

DeMiguel, V., Garlappi, L., Uppal, R.: Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? Rev. Financ. Stud. **22**, 1915–1953 (2009b)

Frahm, G.: Testing for the best alternative with an application to performance measurement. Discussion Paper, University of Cologne (2007)

Frahm, G.: The likelihood-ratio test for $\wedge$-hypotheses. Working Paper, University of Cologne (2009a)

Frahm, G.: Maximum-likelihood estimator for equicorrelated random vectors. Working Paper, University of Cologne (2009b)

Frahm, G., Memmel, C.: Dominating estimators for the global minimum variance portfolio. J. Econom. **159**, 289–302 (2010)

Garlappi, L., Uppal, R., Wang, T.: Portfolio selection with parameter and model uncertainty: a multi-prior approach. Rev. Financ. Stud. **20**, 41–81 (2007)

Goldfarb, D., Iyengar, G.: Robust portfolio selection problems. Math. Oper. Res. **28**, 1–38 (2003)

Halldórsson, B.V., Tütüncü, R.H.: An interior-point method for a class of saddle-point problems. J. Optim. Theory Appl. **116**, 559–590 (2003)

Hansen, P.R.: A test for superior predictive ability. J. Bus. Econ. Stat. **23**, 365–380 (2005)

Jagannathan, R., Ma, T.: Risk reduction in large portfolios: Why imposing the wrong constraints helps. J. Finance **58**, 1651–1684 (2003)

Jobson, J.D., Korkie, B.M.: Performance hypothesis testing with the Sharpe and Treynor measures. J. Finance **26**, 889–908 (1981)

Jorion, P.: Bayes–Stein estimation for portfolio analysis. J. Financ. Quant. Anal. **21**, 279–292 (1986)

Kalymon, B.A.: Estimation risk in the portfolio selection model. J. Financ. Quant. Anal. **6**, 559–582 (1971)

Kan, R., Zhou, G.: Optimal portfolio choice with parameter uncertainty. J. Financ. Quant. Anal. **42**, 621–656 (2007)

Kempf, A., Memmel, C.: Estimating the global minimum variance portfolio. Schmalenbach Bus. Rev. **58**, 332–348 (2006)

Klein, R.W., Bawa, V.S.: The effect of estimation risk on optimal portfolio choice. J. Financ. Econ. **3**, 215–231 (1976)

Kritzman, M., Page, S., Turkington, D.: In defense of optimization: the fallacy of $1/N$. Financ. Anal. J. **66**, 31–39 (2010)

Lahiri, S.N.: Resampling Methods for Dependent Data. Springer, New York (2003)

Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Finance **10**, 603–621 (2003)

Ledoit, O., Wolf, M.: Honey, I shrunk the sample covariance matrix. J. Portf. Manag. **30**, 110–119 (2004)

Ledoit, O., Wolf, M.: Robust performance hypothesis testing with the Sharpe ratio. J. Empir. Finance **15**, 850–859 (2008)

Lo, A.W.: The statistics of Sharpe ratios. Financ. Anal. J. **58**, 36–52 (2002)

Mao, J.C.T., Särndal, C.E.: A decision theory approach to portfolio selection. Manag. Sci. **12**, B-323–334 (1966)

Markowitz, H.: Portfolio selection. J. Finance **7**, 77–91 (1952)

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

Politis, D.N., Romano, J.P.: A circular block-resampling procedure for stationary data. In: LePage, R., Billard, L. (eds.) Exploring the Limits of Bootstrap, pp. 263–270. Wiley, New York (1992)

Politis, D.N., Romano, J.P.: The stationary bootstrap. J. Am. Stat. Assoc. **89**, 1303–1313 (1994)

Press, J.S.: Applied Multivariate Analysis. Holt, Rinehart and Winston, New York (1972)

Romano, J.P., Shaikh, A.M., Wolf, M.: Formalized data snooping based on generalized error rates. Econom. Theory **24**, 404–447 (2008)

Romano, J.P., Wolf, M.: Stepwise multiple testing as formalized data snooping. Econometrica **73**, 1237–1282 (2005)

Sharpe, W.F.: Mutual fund performance. J. Bus. **39**, 119–138 (1966)

Stein, C.: Inadmissibility 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. University of California Press, Berkeley (1956)

The Federal Reserve System: 3-month treasury bill secondary market rate on a discount basis. http://www.federalreserve.gov/releases/H15/data/Monthly/H15_TB_M3.txt (2009). Accessed 14 October 2009

Tobin, J.: Liquidity preference as behavior towards risk. Rev. Econ. Stud. **25**, 65–86 (1958)

White, H.: A reality check for data snooping. Econometrica **68**, 1097–1126 (2000)