

Model risk in portfolio optimization

David Stefanovits, Urs Schubiger, Mario V. Wüthrich

February 19, 2014

Abstract

We consider a one-period portfolio optimization problem under model uncertainty. For this purpose, we introduce a measure of model risk. We derive analytical results for this measure of model risk in the mean-variance problem assuming we have observations drawn from a normal variance mixture model. This model allows for heavy tails, tail dependence and leptokurtosis of marginals. The results show that mean-variance optimization is seriously compromised by model uncertainty, in particular, for non-Gaussian data and small sample sizes. To mitigate these shortcomings, we propose a method to adjust the sample covariance matrix in order to reduce model risk.

1 Introduction

Traditional portfolio optimization techniques rely on the true model parameters being known. However, in general, these model parameters are not known and need to be estimated, for example, from historical data. The possible estimation error gives rise to model risk or model uncertainty. Let us illustrate this with a prominent example, the mean-variance optimization introduced by Markowitz [24]. Consider the unconstrained version of the problem

$$\text{maximize } (\boldsymbol{\mu}'\boldsymbol{w} - \kappa\boldsymbol{w}'\Sigma\boldsymbol{w}) \text{ over } \boldsymbol{w} \in \mathbb{R}^n,$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the vector of expected returns, $\Sigma \in \mathbb{R}^{n \times n}$ is the positive definite covariance matrix and $\kappa > 0$ is a risk aversion parameter. This problem has a unique solution $\boldsymbol{w}^* \in \mathbb{R}^n$ given by

$$\boldsymbol{w}^* = \frac{1}{2\kappa}\Sigma^{-1}\boldsymbol{\mu}.$$

This solution crucially depends on the true expected returns $\boldsymbol{\mu}$ and the covariance matrix Σ . The standard approach replaces true parameters by estimates ignoring any model uncertainty. In the literature, it has been shown that this standard approach has serious drawbacks in mean-variance optimization, see e.g. Michaud [28].

The aim of this paper is to study analytically and numerically the consequences of such model uncertainty in the mean-variance problem and derive parameter estimators which take model uncertainty into account. Despite being a standard one-period problem, it becomes mathematically rather involved when considering model uncertainty. We remark that the mean-variance portfolio remains one of the most important benchmarks in the literature on asset allocation and in the asset management industry, see e.g. Litterman,

[20] Chapter 4, and Meucci, [27] Chapter 6. Numerically we will also analyze three other important benchmarks in asset allocation: the minimum variance, the equal contributions to risk and the maximum diversification problems; see e.g. Choueifaty and Coignard [4], Maillard et al. [21] and Clarke et al. [5] for more information about the latter problems. In this paper we consider these four major portfolio optimization problems. There is an extensive literature discussing mean-variance optimization under model uncertainty. We briefly review the main contributions.

A first portfolio selection rule which takes model uncertainty into account has initially been studied in Brown [2, 3], and Klein and Bawa [17]. They use a Bayesian approach under a non-informative prior. In this framework, the optimal portfolio is selected maximizing the expected value of the objective function conditioned on the predictive distribution of the asset returns. This approach mostly outperforms the standard one and has been studied in many subsequent articles, see e.g. Kandel and Stambaugh [16], Barberis [1], Pástor [30], Pástor and Stambaugh [31], Xia [35], and Tu and Zhou [34].

A second approach to deal with model uncertainty has been proposed in Jobson et al. [12], Jorion [14], and Frost and Savarino [9]. This consists in adjusting the estimated parameters, in order to minimize or reduce a certain loss function. The loss function measures the negative effects of model uncertainty. Typically, it is defined by the expected loss in the mean-variance objective function caused by the estimation. In [12], [14] and [9] the mean-variance problem is considered within the Bayesian framework using such a loss function. This approach differs from the one in [3] and subsequent papers. In [12], [14] and [9] informative priors are proposed with the aim of exploiting prior knowledge about the problem and reducing estimation bias and error for the optimal portfolio weights. In [12] Stein-type estimators (also known as shrinkage estimators) for μ and Σ are introduced. In [14] the problem under uncertainty in μ only is studied. In [9] the mean-variance problem with linear constraints is considered under uncertainty in both μ and Σ . Ter Horst et al. [33] analyze the same loss function as in [12], [14] and [9] under the assumption of known covariance matrix Σ and considering only estimation of μ . A uniform rescaling of the optimal portfolio weights is proposed to account for model uncertainty. In Kan and Zhou [15] a closed-form expression for the same loss function as in [12], [14] and [9] is derived in the unconstrained mean-variance problem under uncertainty in μ and Σ . In Jobson and Korkie [11] the asymptotic distributions of the means, variances and optimal portfolio weights are derived in a constrained problem. Mori [29] obtains analytical results on the optimality of certain estimators in the constrained mean-variance problem. General linear equality constraints are considered under uncertainty in μ and Σ .

Ledoit and Wolf [18], partially motivated by portfolio optimization applications, propose a shrinkage estimator for Σ defined by the convex linear combination of the sample covariance matrix and the identity matrix. The optimal weight is determined asymptotically for number of assets and observations going simultaneously to infinity, i.e. they obtain asymptotically optimal covariance estimators.

El Karoui [7, 8] derives asymptotic results which highlight the severe problems arising in high-dimensional optimization problems. The minimum variance problem with linear constraints and uncertainty in μ and Σ is considered under elliptical asset returns.

In this paper we give the following contributions. First, we introduce a measure of model risk based on an alternative loss function to the one initially proposed by Jobson et al.

[12], Jorion [14], and Frost and Savarino [9]. In Section 2 we introduce this measure in a general optimization problem and describe the reason why we believe it is more relevant for applications than the one considered in [12], [14] and [9].

Secondly, most of the articles so far, have quantified model risk assuming i.i.d. multivariate Gaussian observations. This is, of course, not a realistic model for asset returns in the financial market, see e.g. McNeil et al. [25], Chapter 3. In Ter Horst et al. [33] a non-i.i.d. case is considered. In [7, 8] non-i.i.d. elliptical data is considered and the problem is analyzed asymptotically. In this paper we derive analytical results for finite number of assets and observations in a normal variance mixture model. In particular, we study the cases of small and large number of assets relative to the sample size. Note also that some articles in the literature (see e.g. [33]) consider only uncertainty in μ . From the analytical results derived in Kan and Zhou [15], it follows that the uncertainty in the covariance matrix also has a large impact on the optimal portfolio when the number of assets is large compared to the sample size. Therefore, in this paper, we consider uncertainty in both μ and Σ .

Finally, we present an alternative method to adjust the estimated parameters for model uncertainty. We consider modifications of the spectral decomposition of the sample covariance matrix in order to reduce the drawbacks of model risk. In the literature we can find many papers studying the spectrum of the sample covariance matrix statistically, see e.g. Johnstone [13], Ledoit and Wolf [18], and El Karoui [6]. It has been shown that, although the sample covariance matrix is an unbiased estimator, its eigenvalues behave very differently from the true ones, especially when the number of assets is large compared to the sample size. The linear shrinkage procedure of Ledoit and Wolf [18] corresponds to shrinking all the eigenvalues towards the same constant. In Ledoit and Wolf [19] it is argued that this approach works well in the case where the true eigenvalues are close to each other (this is typically the case when the number of assets is small compared to the sample size). However, if the eigenvalues are highly dispersed (typically the case when number of assets is large compared to the sample size) linear shrinkage does not substantially improve the estimation. In their work they propose to apply an individual shrinkage to each sample eigenvalue in order to obtain an improvement in the covariance matrix estimation. This, however, without direct consideration to portfolio optimization. The procedure to estimate the shrinkage intensities for each eigenvalue is based on a large dimensional asymptotic approximation, which relies on results from random matrix theory on the limiting spectral distribution of the sample covariance matrix, see Marchenko and Pastur [22]. In Menchero et al. [26] we find an empirical study on the bias of volatility estimates for mean-variance portfolios and its relations to eigenvalues and eigenvectors of the sample covariance matrix. In this paper we revisit the idea of modifying the eigenvalues individually and apply it to reduce the negative effects of model uncertainty. In particular, we analyze the measure of model uncertainty introduced in Section 2 in the mean-variance problem for the covariance matrix estimator with rescaled eigenvalues.

Organization of the paper. In Section 2 we introduce our measure of model risk for general portfolio optimization problems. Using this measure we prove in Proposition 2.5 the following result: under the assumption of known constraints and unbiased estimators, optimal portfolios are on average negatively affected by model risk. In Section 3 we consider mean-variance optimization assuming we have observations drawn from a normal

variance mixture model. We derive in Theorem 3.4 an analytical formula for our measure of model uncertainty. This provides a realistic description of the effects of model uncertainty. Moreover, we show that this measure of model risk attains its lowest value in the Gaussian case. In Section 4 we introduce an estimator for the covariance matrix with adjusted eigenvalues. In Theorem 4.5 we derive a relationship between our measure of model uncertainty and the adjusted eigenvalues. Using this, we derive in Theorem 4.7 a rule which reduces the effects of model uncertainty under the assumption of known eigenvectors. Even if these are unknown, in practice, we show numerically in Section 5 that the rule provides significant improvements to the portfolio optimization problem.

2 General definitions

We consider the portfolio optimization problem of an investor who allocates his wealth among n risky assets. A portfolio is defined by $\mathbf{w} = (w_1, \dots, w_n)' \in \mathbb{R}^n$, where w_i denotes the proportion of wealth invested in asset i . Let $U : \mathbb{R}^n \rightarrow \mathbb{R}$ be a real-valued function on the set of all portfolios. The function U specifies the investor's preference for portfolios and is called utility function. Let $\mathcal{A} \subset \mathbb{R}^n$ be the set of admissible portfolios which specifies the investment constraints. In this general setting we define the optimal portfolio by

$$\mathbf{w}^*(U, \mathcal{A}) = \operatorname{argmax}_{\mathbf{w} \in \mathcal{A}} U(\mathbf{w}), \quad (1)$$

under the assumption that the maximization problem has a unique solution. We think of $\mathbf{w}^*(U, \mathcal{A}) \in \mathbb{R}^n$ as the optimal portfolio based on utility U and constraints specified by \mathcal{A} . In Example 2.1 we give prominent examples of the general optimization problem (1).

Example 2.1. Assume the investor has complete knowledge of the mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and the covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ of the asset returns. We assume Σ is positive definite and formulate four major portfolio optimization problems. For a detailed description of the first two problems we refer to Meucci [27], Chapter 6.3-6.5. For the latter two see Choueifaty and Coignard [4], Maillard et al. [21], and Clarke et al. [5].

- (i) Mean-variance optimization: the investor is assumed to maximize

$$U^{\text{mv}}(\mathbf{w}) = \boldsymbol{\mu}'\mathbf{w} - \kappa \mathbf{w}'\Sigma\mathbf{w},$$

where $\kappa > 0$ is a given risk aversion parameter. The solution of (1) in the unconstrained case can be calculated using first order conditions and is given by

$$\mathbf{w}^*(U^{\text{mv}}, \mathbb{R}^n) = \frac{1}{2\kappa} \Sigma^{-1} \boldsymbol{\mu}.$$

In the case of linear constraints we set $\mathcal{A} = \{\mathbf{w} \in \mathbb{R}^n \mid A'\mathbf{w} = \mathbf{a}\}$, where $A \in \mathbb{R}^{n \times k}$ is a matrix of rank k and $\mathbf{a} \in \mathbb{R}^k$. The solution of (1) can be calculated using the method of Lagrange multipliers and is given by

$$\mathbf{w}^*(U^{\text{mv}}, \mathcal{A}) = \frac{1}{2\kappa} (\Sigma^{-1} - \Sigma^{-1} A (A' \Sigma^{-1} A)^{-1} A' \Sigma^{-1}) \boldsymbol{\mu} + \Sigma^{-1} A (A' \Sigma^{-1} A)^{-1} \mathbf{a}.$$

(ii) Minimum variance: the investor is assumed to maximize the following utility function

$$U^v(\mathbf{w}) = -\mathbf{w}'\Sigma\mathbf{w}.$$

Consider the following linear equality constraints $\mathcal{A} = \{\mathbf{w} \in \mathbb{R}^n \mid A'\mathbf{w} = \mathbf{a} \text{ and } \mathbf{w}'\boldsymbol{\mu} = \mu_0\}$, where $A \in \mathbb{R}^{n \times k}$ is a matrix of rank k , $\mathbf{a} \in \mathbb{R}^k$ and $\mu_0 \in \mathbb{R}$. In this problem we are looking for the portfolio with minimal variance satisfying linear constraints and having expected return equal to μ_0 . The optimal portfolio can be calculated using the method of Lagrange multipliers and is given by

$$\mathbf{w}^*(U^v, \mathcal{A}) = \Sigma^{-1}\tilde{A}(\tilde{A}'\Sigma^{-1}\tilde{A})^{-1}\tilde{\mathbf{a}},$$

where $\tilde{A} = (A, \boldsymbol{\mu}) \in \mathbb{R}^{n \times (k+1)}$ and $\tilde{\mathbf{a}} = (\mathbf{a}', \mu_0)' \in \mathbb{R}^{k+1}$. In this example note that the set of admissible strategies \mathcal{A} also depends on the distribution of the random returns through the mean vector $\boldsymbol{\mu}$.

(iii) Equal contributions to risk: the investor is assumed to maximize

$$U^{\text{ecr}}(\mathbf{w}) = - \sum_{\substack{i,j=1,\dots,n \\ i < j}} (\text{RC}_i(\mathbf{w}) - \text{RC}_j(\mathbf{w}))^2,$$

where $\text{RC}_i(\mathbf{w}) = w_i(\Sigma\mathbf{w})_i$ denotes the contribution of asset i to the total portfolio variance. This means the investor looks for a portfolio with risk contributions as close to each other as possible with respect to the Euclidean distance. Typically, constraints as in (ii) are chosen.

(iv) Maximum diversification: the investor is assumed to maximize

$$U^{\text{md}}(\mathbf{w}) = \frac{\sum_{i=1}^n w_i \sqrt{\Sigma_{ii}}}{\sqrt{\mathbf{w}'\Sigma\mathbf{w}}},$$

where Σ_{ii} is the i -th element on the diagonal of Σ . This means the investor maximizes the benefits of diversification. Typically, constraints as in (ii) are chosen.

Note that in practice $\boldsymbol{\mu}$ and Σ are not observable. Therefore, the investor cannot directly maximize U^{mv} , U^v , U^{ecr} or U^{md} , because neither U nor \mathcal{A} are explicitly known.

Let (Ω, \mathcal{F}, P) be a probability space. Suppose the investor has m observations of the n -dimensional return vectors given by

$$\mathbf{R} = (\mathbf{R}_i)_{i=1,\dots,m} = ((R_{i,1}, \dots, R_{i,n})')_{i=1,\dots,m},$$

i.e. he has m random vectors $\mathbf{R}_1, \dots, \mathbf{R}_m$ on (Ω, \mathcal{F}, P) representing the returns of the n risky assets.

Remark 2.2. In this section we do not make particular assumptions on the distribution or the dependence structure of \mathbf{R} , except for certain integrability conditions specified below.

Based on the observations \mathbf{R} we define the utility estimator by

$$\hat{U} : \mathbf{w} \mapsto u(\mathbf{w}, \mathbf{R}_1, \dots, \mathbf{R}_m),$$

for a function $u : \mathbb{R}^n \times (\mathbb{R}^n)^m \rightarrow \mathbb{R}$. The estimator \hat{U} is called unbiased for U if $\hat{U}(\mathbf{w}) \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and $E[\hat{U}(\mathbf{w})] = U(\mathbf{w})$ for all $\mathbf{w} \in \mathbb{R}^n$. The set of admissible portfolios \mathcal{A} may also be unobservable, see e.g. Example 2.1(ii). Analogously to the utility estimator, we define the set $\hat{\mathcal{A}}$ estimated from the observations, i.e.

$$\hat{\mathcal{A}} = a(\mathbf{R}_1, \dots, \mathbf{R}_m) \subset \mathbb{R}^n,$$

for a suitable function a . The investor then chooses the optimal portfolio (1) by replacing U and \mathcal{A} with the estimates \hat{U} and $\hat{\mathcal{A}}$. This means,

$$\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}}) = \operatorname{argmax}_{\mathbf{w} \in \hat{\mathcal{A}}} \hat{U}(\mathbf{w}) = \operatorname{argmax}_{\mathbf{w} \in a(\mathbf{R}_1, \dots, \mathbf{R}_m)} u(\mathbf{w}, \mathbf{R}_1, \dots, \mathbf{R}_m),$$

where we assume that the maximization problem has a unique solution, almost surely. The fact that the investor can only compute the portfolio $\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}})$ is often neglected in practice, assuming that the true utility of this portfolio is on average similar to its estimated utility.

In order to get a better understanding of this problem we introduce the following loss function

$$L(U, \hat{U}, \hat{\mathcal{A}}) = E[\hat{U}(\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}})) - U(\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}}))], \quad (2)$$

under the assumption $U(\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}})), \hat{U}(\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}})) \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. This loss function measures the difference between the estimated and true expected utility of $\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}})$, which is the investor's portfolio choice. In other words: if $L(U, \hat{U}, \hat{\mathcal{A}}) > 0$ (< 0), then the out-of-sample utility of the portfolio choice is on average lower (higher) than the one measured in-sample. We interpret the loss function (2) as a measure of model risk in the portfolio optimization problem.

Remark 2.3. In the literature the problem of model uncertainty in portfolio optimization is mostly studied under the loss function

$$L^{\text{Jorion}}(U, \mathcal{A}, \hat{U}, \hat{\mathcal{A}}) = U(\mathbf{w}^*(U, \mathcal{A})) - E[U(\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}}))], \quad (3)$$

see e.g. Jobson et al. [12], Jorion [14], Forst and Savarino [9], Ter Horst et al. [33], Mori [29], and Kan and Zhou [15]. According to this definition, the true utility U of the true optimal portfolio $\mathbf{w}^*(U, \mathcal{A})$ is compared to

$$E[U(\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}}))],$$

i.e. the expected true utility of the investor's optimal portfolio choice $\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}})$. In real applications the investor is not able to choose the portfolio $\mathbf{w}^*(U, \mathcal{A})$ because U and \mathcal{A} are unknown. Since it does not reflect the estimated utility, we are not interested in the quantity $U(\mathbf{w}^*(U, \mathcal{A}))$, and the loss function (3) is not particularly informative in practice. We are rather concerned about the true utility U of the actual investor's portfolio choice $\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}})$ relative to

$$E[\hat{U}(\mathbf{w}^*(\hat{U}, \hat{\mathcal{A}}))],$$

which is the expected optimal utility of the investor. For this reason we choose (2) to study the effects of model uncertainty on the portfolio optimization. More recently, El Karoui [8] studies the difference between the variance forecast and the realized variance of the minimum variance portfolio. This difference corresponds to (2) in the case of Example 2.1(ii). A similar measure, called second order risk, has also been studied in Shepard [32].

Remark 2.4. In general (2) and (3) are unobservable since they depend on U .

The measure of model risk (2) satisfies the following property (see Section A.1 in the appendix).

Proposition 2.5. *Let $\hat{\mathcal{A}} = \mathcal{A}$ be known, $U(\mathbf{w}^*(\hat{U}, \mathcal{A})), \hat{U}(\mathbf{w}^*(\hat{U}, \mathcal{A})) \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and \hat{U} be unbiased for U . Then,*

$$L(U, \hat{U}, \mathcal{A}) \geq 0,$$

with strict inequality if $P[\mathbf{w}^(U, \mathcal{A}) \neq \mathbf{w}^*(\hat{U}, \mathcal{A})] > 0$.*

This result states that, assuming known constraints and unbiased utility estimator, the loss function is positive and the portfolio choice $\mathbf{w}^*(\hat{U}, \mathcal{A})$ is on average negatively affected by model uncertainty, i.e. the out-of-sample utility is on average lower than the in-sample one. This result suggests that we might want to introduce bias correction in the utility estimator in order to reduce the loss function. In Section 4 we present a method to accomplish this task.

3 Analysis of the loss function in the mean-variance case

We analyze loss function (2) in the mean-variance case. We focus on this problem since it is one of the most important benchmarks in portfolio optimization and because it is analytically tractable. In Section 5 we consider the three problems presented in Example 2.1(ii)-(iv) from a numerical point of view. In this section, we focus on analytical results. As in Example 2.1, let $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ be the unobservable mean vector and covariance matrix of the asset returns. Assume Σ is symmetric positive definite. The portfolio $\mathbf{w} \in \mathbb{R}^n$ is chosen to maximize the utility function $U^{\text{mv}}(\mathbf{w}) = \boldsymbol{\mu}'\mathbf{w} - \kappa \mathbf{w}'\Sigma\mathbf{w}$, where $\kappa > 0$ is a given risk aversion parameter.

We derive a closed-form expression for the loss function (2) in the unconstrained mean-variance problem under non-Gaussian observations. In Kan and Zhou [15] this has been done for loss function (3) under the assumption of i.i.d. multivariate Gaussian observations. Under the same assumption on the observations, Mori [29] derived analytical results on the optimality of certain estimators with respect to (3) under linear equality constraints. El Karoui [8] studies the asymptotic behaviour of (2) in a minimum-variance problem with linear equality constraints under non-i.i.d. and non-Gaussian observations.

We assume the observations $(\mathbf{R}_i)_{i=1, \dots, m}$ are drawn from a multivariate normal variance mixture distribution

$$\mathbf{R}_i \sim \boldsymbol{\mu} + \sqrt{W}\Sigma^{\frac{1}{2}}\mathbf{Z}_i, \quad (4)$$

where

- (i) $(\mathbf{Z}_i)_{i=1, \dots, m} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_n, I_n)$ for $\mathbf{0}_n$ and I_n being the n -dimensional zero vector and identity matrix, respectively;

(ii) W is an almost surely positive random variable, independent of $(\mathbf{Z}_i)_{i=1,\dots,m}$, satisfying $E[W] = 1$; and

(iii) $\Sigma^{\frac{1}{2}} \in \mathbb{R}^{n \times n}$ is a non-singular matrix such that $\Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})' = \Sigma$.

This assumption encompasses models with heavy tails, tail dependence and leptokurtosis of marginals, see e.g. McNeil et al. [25], Chapter 3. Often, this provides a reasonable description of returns in the financial market.

Remark 3.1. Under the above assumptions we have

(i) $(\mathbf{R}_i|W)_{i=1,\dots,m} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}, W\Sigma)$,

(ii) $E[\mathbf{R}_i] = \boldsymbol{\mu}$, and $\text{Cov}(\mathbf{R}_i) = E[\text{Cov}(\mathbf{R}_i|W)] = E[W]\Sigma = \Sigma$ for all $i = 1, \dots, m$.

Example 3.2. The following special cases are often used in practice.

(i) Multivariate Gauss: set $W = 1$.

(ii) 2-points mixture: let W be discrete and take positive values x_1 and x_2 with probabilities $p \in (0, 1)$ and $1 - p$. The condition $E[W] = 1$ implies $x_1 \in (0, p^{-1})$ and $x_2 = \frac{1-px_1}{1-p}$. This is a regime switching model with 2 variance regimes characterized by x_1 and x_2 .

(iii) Multivariate Student- t : let $W \sim \frac{\nu-2}{\nu} \text{Ig}(\frac{\nu}{2}, \frac{\nu}{2})$, where $\nu > 2$ and $\text{Ig}(\cdot, \cdot)$ denotes the inverse gamma distribution. The scaling is chosen so that $E[W] = 1$. Then, for all $i = 1, \dots, m$, the random vector \mathbf{R}_i has a multivariate Student- t distribution with ν degrees of freedom, mean vector $\boldsymbol{\mu}$ and covariance matrix Σ .

For $\boldsymbol{\mu}$ and Σ we consider the sample estimators

$$\hat{\boldsymbol{\mu}} = \frac{1}{m} \sum_{i=1}^m \mathbf{R}_i \quad \text{and} \quad \hat{\Sigma} = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{R}_i - \hat{\boldsymbol{\mu}})(\mathbf{R}_i - \hat{\boldsymbol{\mu}})'. \quad (5)$$

In this way we estimate the mean-variance utility by $\hat{U}^{\text{mv}}(\mathbf{w}) = \hat{\boldsymbol{\mu}}'\mathbf{w} - \kappa \mathbf{w}'\hat{\Sigma}\mathbf{w}$ for $\mathbf{w} \in \mathbb{R}^n$.

Remark 3.3. Under (4) the sample estimators satisfy the following properties.

(i) Conditional on W we have

$$\hat{\boldsymbol{\mu}}|W \sim \mathcal{N}\left(\boldsymbol{\mu}, \frac{W}{m}\Sigma\right) \quad \text{and} \quad (m-1)\hat{\Sigma}|W \sim \mathcal{W}(\Sigma, m-1),$$

where \mathcal{W} denotes the Wishart distribution. Moreover, $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are conditionally independent given W . See Appendix B for more details.

(ii) For $m > n$ the sample covariance matrix is almost surely positive definite and $\hat{\Sigma}^{-1}$ exists almost surely. See Appendix B for more details.

(iii) For any $\mathbf{w} \in \mathbb{R}^n$ we have

$$\begin{aligned} E[\hat{\boldsymbol{\mu}}] &= E[E[\hat{\boldsymbol{\mu}}|W]] = \boldsymbol{\mu}, \\ E[\hat{\Sigma}] &= \frac{1}{m-1} E[E[(m-1)\hat{\Sigma}|W]] = E[W]\Sigma = \Sigma, \text{ and} \\ E[\hat{U}^{\text{mv}}(\mathbf{w})] &= \mathbf{w}' E[\hat{\boldsymbol{\mu}}] - \kappa \mathbf{w}' E[\hat{\Sigma}] \mathbf{w} = U^{\text{mv}}(\mathbf{w}). \end{aligned}$$

This means $\hat{\boldsymbol{\mu}}$, $\hat{\Sigma}$ and \hat{U}^{mv} are unbiased estimators for $\boldsymbol{\mu}$, Σ and U^{mv} , respectively.

Under (4) we have the following result (see Section A.2 in the appendix).

Theorem 3.4. *Assume $m > n + 4$, then*

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) &= \frac{1}{4\kappa} \alpha(m-1, n) \left\{ \frac{n}{m} (1 + \beta(m-1, n) E[W^{-1}]) \right. \\ &\quad \left. + (\beta(m-1, n) E[W^{-2}] - E[W^{-1}]) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\}, \end{aligned}$$

where $\alpha(m, n) = \frac{m}{m-n-1}$ and $\beta(m, n) = \frac{m(m-1)}{(m-n)(m-n-3)}$.

As we will explain in Remark 3.5(ii) below, this result implies that the drawbacks of model uncertainty cannot completely be eliminated by increasing the sample size m , unless we are in the multivariate Gaussian case with $W = 1$. In practice, we need to find a trade-off between large and recent samples: we need as many observations as possible in order to reduce parameter estimation error; on the other hand we need the most recent ones since, in practice, the i.i.d. assumption is often violated over long time horizons.

Remark 3.5. From Theorem 3.4 we observe the following facts.

(i) Jensen's inequality implies

$$E[W^{-2}] = E[(W^{-1})^2] \geq E[W^{-1}] E[W^{-1}] \geq E[W^{-1}] E[W]^{-1} = E[W^{-1}].$$

Hence for $m > n + 4$,

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) &\geq \frac{1}{4\kappa} (\beta(m-1, n) E[W^{-2}] - E[W^{-1}]) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \\ &\geq \frac{1}{4\kappa} (E[W^{-2}] - E[W^{-1}]) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \geq 0, \end{aligned}$$

see also Proposition 2.5. This lower bound is strictly positive in the non-Gaussian case.

(ii) If we fix the number of assets n and let $m \rightarrow \infty$ we obtain

$$L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow \frac{1}{4\kappa} (E[W^{-2}] - E[W^{-1}]) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \geq 0.$$

In particular, in the multivariate Gaussian case we have $L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow 0$ as $m \rightarrow \infty$, see Example 3.2(i).

(iii) For $n, m \rightarrow \infty$ and $n/m \rightarrow c \in (0, 1)$ we get

$$L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow \frac{1}{4\kappa} \frac{1}{(1-c)} \left\{ c + \frac{c}{(1-c)^2} E[W^{-1}] + \left(\frac{1}{(1-c)^2} E[W^{-2}] - E[W^{-1}] \right) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\}.$$

This asymptotic analysis corresponds to large number of risky assets n and comparable sample size m . This limit is considered in El Karoui [7, 8], and Ledoit and Wolf [19]. From this formula we can see that for $c \rightarrow 0^+$ we obtain the same limit as in (ii), and for $c \rightarrow 1^-$ we have $L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow \infty$.

(iv) Note that the loss function depends on the unobservable distribution of W , true mean vector $\boldsymbol{\mu}$ and true covariance matrix Σ .

We consider the special cases for W discussed in Example 3.2 (see Section A.3 in the appendix).

Corollary 3.6. *Let $m > n + 4$. In the three special cases given in Example 3.2 we have for the loss function*

(i) *multivariate Gauss:*

$$L^{\text{Gauss}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) = \frac{1}{4\kappa} \alpha(m-1, n) \left\{ \frac{n}{m} (1 + \beta(m-1, n)) + (\beta(m-1, n) - 1) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\};$$

(ii) *2-points mixture:*

$$L^{\text{2pm}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) = \frac{1}{4\kappa} \alpha(m-1, n) \left\{ \frac{n}{m} - \left(\frac{p}{x_1} + \frac{(1-p)^2}{1-px_1} \right) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} + \frac{n}{m} \beta(m-1, n) \left(\frac{p}{x_1} + \frac{(1-p)^2}{1-px_1} \right) + \beta(m-1, n) \left(\frac{p}{x_1^2} + \frac{(1-p)^3}{(1-px_1)^2} \right) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\},$$

where $p \in (0, 1)$ and $x_1 \in (0, p^{-1})$;

(iii) *multivariate Student-t:*

$$L^{\text{t}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) = \frac{1}{4\kappa} \alpha(m-1, n) \left\{ \frac{n}{m} \left(1 + \beta(m-1, n) \frac{\nu}{\nu-2} \right) + \frac{\nu}{\nu-2} \left(\beta(m-1, n) \frac{2+\nu}{\nu-2} - 1 \right) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\},$$

where $\nu > 2$.

Remark 3.7. From the results of Corollary 3.6 we observe the following facts.

- (i) In the 2-points mixture model we have $L^{2\text{pm}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow \infty$ for $x_1 \rightarrow p^{-1}$ or $x_1 \rightarrow 0$, and $L^{2\text{pm}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow L^{\text{Gauss}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$ for $x_1 \rightarrow 1$. The more we increase the difference between the two variance regimes (i.e. by considering $x_1 \rightarrow p^{-1}$ or $x_1 \rightarrow 0$), the greater the average negative effect of model risk. If we decrease the difference (by considering $x_1 \rightarrow 1$), then the loss function becomes closer to the Gaussian case.
- (ii) In the multivariate Student- t model we have $L^t(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow \infty$ for $\nu \rightarrow 2^+$, and $L^t(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) \rightarrow L^{\text{Gauss}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$ for $\nu \rightarrow \infty$. This means, by making the marginals “less normal” (i.e. by considering $\nu \rightarrow 2^+$), we increase the average negative effect of model risk.

We show that the drawbacks of model uncertainty in the general case are on average bigger than those in the Gaussian case (see Section A.4 in the appendix).

Corollary 3.8. *Let $m > n + 4$. Then, $L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) > L^{\text{Gauss}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$, for any $W \neq 1$.*

In the financial market asset returns show non-Gaussian behaviour, which on average has a negative effect on the optimal portfolio choice. In the next section we consider techniques to reduce the loss function and improve the optimal portfolio.

4 Adjusting for model risk

As Proposition 2.5 and Remark 3.5(i) pointed out, if we solve the mean-variance problem using the unbiased utility estimator \hat{U}^{mv} , model risk deteriorates the investor’s optimal portfolio choice $\mathbf{w}^*(\hat{U}^{\text{mv}}, \mathbb{R}^n)$. Therefore, we might want to introduce a bias correction in \hat{U}^{mv} which allows to reduce the loss function (2). In the literature we can find several papers using this idea to improve the optimal portfolio under loss function (3). See for example Jobson et al. [12], Jorion [14], Frost and Savarino [9], Mori [29], and Kan and Zhou [15]. These approaches typically result in rescaled or shrinkage estimators for the mean or covariance matrix so that loss function (3) is improved. In this work we present an alternative approach based on the classical spectral decomposition of symmetric positive definite matrices.

Remark 4.1. Contrary to Section 3 we do not make specific distributional assumptions on \mathbf{R} in the following results, unless explicitly specified.

Let $\mathbf{R}_1, \dots, \mathbf{R}_m \in \mathcal{L}^2(\Omega, \mathcal{F}, P)^n$, and $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ be unbiased estimators for $\boldsymbol{\mu}$ and Σ . We assume that $\hat{\Sigma}$ is positive definite almost surely. We want to analyze a class of estimators for the covariance matrix Σ obtained by individually rescaling the eigenvalues of $\hat{\Sigma}$. An asymptotic analysis of this approach is proposed in Ledoit and Wolf [19]. In [19] this is done to improve the covariance matrix estimation without specific considerations on mean-variance optimization. Menchero et al. [26] analyze empirically spectral properties of $\hat{\Sigma}$ in the context of mean-variance optimization. The spectral decomposition of the symmetric positive definite covariance matrix Σ is given by

$$\Sigma = T\Lambda T',$$

where

- (i) $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix of eigenvalues (also called eigenvariances in this context) of Σ , which satisfy $0 < \lambda_1 \leq \dots \leq \lambda_n$; and
- (ii) T is an orthogonal matrix with columns $\mathbf{v}_1 = (T_{11}, \dots, T_{n1})', \dots, \mathbf{v}_n = (T_{n1}, \dots, T_{nn})'$ being the eigenvectors (also called eigenportfolios) of Σ corresponding to $\lambda_1, \dots, \lambda_n$, respectively.

In the same way set $\hat{\Sigma} = \hat{T}\hat{\Lambda}\hat{T}'$, where $\hat{T} = (\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_n)$ and $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_n)$ are random matrices with $\hat{T}\hat{T}' = I_n$ and $0 < \hat{\lambda}_1 \leq \dots \leq \hat{\lambda}_n$, almost surely. Note that \hat{T} and $\hat{\Lambda}$ are in general dependent. In this work we study the following class of covariance matrix estimators

$$\hat{\Sigma}_C = \hat{T} \text{diag}(c_1 \hat{\lambda}_1, \dots, c_n \hat{\lambda}_n) \hat{T}',$$

where $c_1, \dots, c_n > 0$ are referred to as *adjustment factors*. For finite n and m , there is only little known about the distribution of $\hat{\Lambda}$ and \hat{T} . Nevertheless, the decomposition of the sample covariance matrix using eigenportfolios and eigenvariances is a useful tool in mean-variance optimization. It is understood that the eigenvalues of $\hat{\Sigma}$ behave very differently from those of Σ . Ledoit and Wolf [18] point out that the largest sample eigenvalues are systematically biased upwards and the smallest ones downwards. This fact has also been verified for market returns data in different empirical studies, see e.g. Menchero et al. [26]. There is statistical literature studying asymptotic spectral properties of $\hat{\Sigma}$ when n and m are large, see e.g. Johnstone [13] and El Karoui [7].

Remark 4.2. Note that $\hat{\Sigma}_C$ is almost surely positive definite and

$$\begin{aligned} \hat{\Sigma}_C &= \hat{T}C\hat{\Lambda}\hat{T}' = \hat{T}C\hat{T}'\hat{T}\hat{\Lambda}\hat{T}' = (\hat{T}C\hat{T}')\hat{\Sigma} = \hat{\Sigma}(\hat{T}C\hat{T}'), \\ \hat{\Sigma}_C^{-1} &= \hat{T}C^{-1}\hat{\Lambda}^{-1}\hat{T}' = \hat{T}C^{-1}\hat{T}'\hat{T}\hat{\Lambda}^{-1}\hat{T}' = (\hat{T}C^{-1}\hat{T}')\hat{\Sigma}^{-1} = \hat{\Sigma}^{-1}(\hat{T}C^{-1}\hat{T}'), \end{aligned}$$

where $C = \text{diag}(c_1, \dots, c_n)$. If $C = I_n$ then $\hat{\Sigma}_C$ is unbiased for Σ .

Using $\hat{\Sigma}_C$ we define the following utility estimator for $\mathbf{w} \in \mathbb{R}^n$

$$\hat{U}_C^{\text{mv}}(\mathbf{w}) = \hat{\boldsymbol{\mu}}'\mathbf{w} - \kappa\mathbf{w}\hat{\Sigma}_C\mathbf{w} = \hat{\boldsymbol{\mu}}'\mathbf{w} - \kappa\mathbf{w}(\hat{T}C\hat{T}')\hat{\Sigma}\mathbf{w}.$$

Our aim is to determine adjustment factors c_1, \dots, c_n so that

$$L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}) \leq L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \hat{\mathcal{A}}), \quad (6)$$

see also Proposition 2.5. In this way we obtain an optimal portfolio $\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}})$ which is more robust to model uncertainty as measured by (2). We can accomplish this in a straightforward way considering the special case of equal adjustment factors (see Section A.5 in the appendix).

Proposition 4.3. Choose $c_1 = \dots = c_n = \bar{c} > 0$. Let $\Sigma_{\bar{c}} = \hat{\Sigma}_C = \bar{c}\hat{\Sigma}$ and $\hat{U}_{\bar{c}}^{\text{mv}}$ be the corresponding utility estimator. Then,

$$L(U^{\text{mv}}, \hat{U}_{\bar{c}}^{\text{mv}}, \mathbb{R}^n) = \frac{1}{4\bar{c}\kappa}E[\hat{\boldsymbol{\mu}}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}] - \frac{1}{2\bar{c}\kappa}\boldsymbol{\mu}'E[\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}] + \frac{1}{4\bar{c}^2\kappa}E[\hat{\boldsymbol{\mu}}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}].$$

In particular, under assumption (4) on the observations, and setting

$$\bar{c} = \beta(m-1, n) \frac{E[W^{-2}]}{E[W^{-1}]}, \quad (7)$$

we have $L(U^{\text{mv}}, \hat{U}_{\bar{c}}^{\text{mv}}, \mathbb{R}^n) \leq L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$.

Remark 4.4. Observe the following about the approach of Proposition 4.3.

- (i) $\hat{\Sigma}_{\bar{c}} = \bar{c}\hat{\Sigma}$ is a simple proportional scaling type estimator. The unconstrained mean-variance portfolio constructed using $\hat{\Sigma}_{\bar{c}}$ is given by

$$\mathbf{w}^*(\hat{U}_{\bar{c}}^{\text{mv}}, \mathbb{R}^n) = \frac{1}{2\kappa} \hat{\Sigma}_{\bar{c}}^{-1} \hat{\boldsymbol{\mu}} = \frac{1}{2\bar{c}\kappa} \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}. \quad (8)$$

- (ii) In general, adjustment factors (7) are unobservable since they depend on the unknown distribution of W . Note that, as observed in Remark 3.5(i), we have $E[W^{-2}] \geq E[W^{-1}]$ and henceforth

$$\bar{c} \geq \beta(m-1, n),$$

with equality in the multivariate Gaussian case.

Scaling $\hat{\Sigma}$ by \bar{c} is useful since it reduces the impact of model risk. However, this proportional scaling is not helpful to improve the relative asset allocation because it does not adjust the correlation structure of $\hat{\Sigma}$. For example, in the unconstrained mean-variance case, all portfolio weights are uniformly reduced by a constant factor $1/\bar{c}$, i.e. the risk aversion is increased by factor \bar{c} , see (7). Since the relative asset allocation remains the same, this cannot be considered as a substantial improvement in the optimal portfolio allocation. Therefore, we consider individual adjustment factors c_1, \dots, c_n for the eigenvariances $\hat{\lambda}_1, \dots, \hat{\lambda}_n$.

A heuristic approach to select the adjustment factors is to set for $k = 1, \dots, n$

$$c_k = \frac{\lambda_k}{E[\hat{\lambda}_k]}. \quad (9)$$

This rule ensures that the eigenvariances of $\hat{\Sigma}_C$ are unbiased estimators for the ones of Σ . In the next result we calculate $L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n)$ for general c_1, \dots, c_n in the unconstrained mean-variance case (see Section A.6 in the appendix). We first introduce some additional notation to simplify the representation. Since the family $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of eigenportfolios builds an orthonormal basis for \mathbb{R}^n we can write the optimal portfolio as

$$\mathbf{w}^*(\hat{U}^{\text{mv}}, \hat{\mathcal{A}}) = \sum_{k=1}^n \pi_k(\mathbf{w}^*(\hat{U}^{\text{mv}}, \hat{\mathcal{A}}), \Sigma) \mathbf{v}_k,$$

where $\pi_k(\mathbf{w}, \Sigma) = (T'\mathbf{w})_k$ denotes the projection of $\mathbf{w} \in \mathbb{R}^n$ on the eigenportfolio \mathbf{v}_k . In the same way we define $\pi_k(\mathbf{w}, \hat{\Sigma})$ and $\pi_k(\mathbf{w}, \hat{\Sigma}_C)$. Note that, in general, these two random variables depend on both \hat{T} and $\hat{\Lambda}$.

Theorem 4.5. *The loss function for the unconstrained mean-variance problem and the estimator $\hat{\Sigma}_C$ is given by*

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n) &= \frac{1}{2\kappa} \sum_{k=1}^n E[(\pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma})^2 - \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma})\pi_k(\boldsymbol{\mu}, \hat{\Sigma}))c_k^{-1}\hat{\lambda}_k^{-1}] \\ &\quad + \frac{1}{4\kappa} \sum_{k=1}^n E[\pi_k(\hat{\Sigma}_C^{-1}\hat{\boldsymbol{\mu}}, \Sigma)^2\lambda_k - \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma})^2c_k^{-1}\hat{\lambda}_k^{-1}]. \end{aligned}$$

From the general formula of Theorem 4.5 it is difficult to derive a simple rule to select the adjustment factors c_1, \dots, c_n . However, this can be done under the following simplifying assumptions. For the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ we consider the following assumption

$$\hat{\Sigma} = T\hat{\Lambda}T', \quad (10)$$

where

- (i) $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_n)$ independent of $\hat{\boldsymbol{\mu}}$,
- (ii) $0 < \hat{\lambda}_1 \leq \dots \leq \hat{\lambda}_n$ almost surely and integrable, and
- (iii) $E[\hat{\boldsymbol{\mu}}] = \boldsymbol{\mu}$.

This means that we consider only estimation uncertainty in the means and eigenvariances, and assume independence between eigenvariances and means.

Remark 4.6. The estimator $\hat{\Sigma}$ is almost surely positive definite. The estimator $\hat{\Sigma}$ is unbiased for Σ if and only if $\hat{\Lambda}$ is unbiased for Λ .

We derive the following result for the loss function (see Section A.7 in the appendix).

Theorem 4.7. *Under (10) we have*

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n) &= \frac{1}{2\kappa} \sum_{k=1}^n c_k^{-1} \text{Var}[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)] E[\hat{\lambda}_k^{-1}] \\ &\quad + \frac{1}{4\kappa} \sum_{k=1}^n E[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)^2] (c_k^{-2} E[\hat{\lambda}_k^{-2}] \lambda_k - c_k^{-1} E[\hat{\lambda}_k^{-1}]). \end{aligned}$$

Assume $\hat{\Sigma}$ is unbiased for Σ and set

$$c_k = \frac{\lambda_k E[\hat{\lambda}_k^{-2}]}{E[\hat{\lambda}_k^{-1}]}. \quad (11)$$

Then, $c_k \geq 1$ for all $k = 1, \dots, n$ and $L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n) \leq L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$.

In the simplified situation (10) we obtain a rule for selecting the adjustment factors c_1, \dots, c_n which reduce the effect of estimation uncertainty in $\hat{\Lambda}$.

In Section 5 we show that adjustment factors (7), (9) and (11) considerably reduce the loss function (2) in the mean-variance problem. This reduction is particularly large in the

multivariate Student- t case with small values of ν , or in the case n is close to m . The numerical results show that the adjustment factors (11) reduce the drawbacks of model uncertainty also when we use the sample estimator for the covariance matrix. We interpret this fact as the property that estimation errors in the eigenportfolios and eigenvariances of the sample covariance matrix may have low correlation with each other. Therefore, selecting the adjustment factors based on the stand-alone effect of estimation uncertainty in the eigenvariances is helpful. Note that the adjustment factors (7), (9) and (11) are, in general, not observable and need to be estimated. This issue is also addressed in Section 5.

5 Case studies

5.1 Simulated observations

In the first place we compute loss function (2) using return observations simulated from known $\boldsymbol{\mu}$ and Σ . We consider the following four problems, see Example 2.1.

- (i) Unconstrained mean-variance: $U = U^{\text{mv}}$ and $\mathcal{A} = \mathbb{R}^n$. We set $\kappa = 1$ for the risk aversion parameter.
- (ii) Long only fully invested minimum variance: $U = U^{\text{v}}$ and $\mathcal{A} = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{1}'\mathbf{w} = 1 \text{ and } w_i \geq 0 \text{ for all } i = 1, \dots, n\}$.
- (iii) Long only fully invested equal contributions to risk: $U = U^{\text{ecr}}$ and \mathcal{A} as in (ii).
- (iv) Long only fully invested maximum diversification: $U = U^{\text{md}}$ and \mathcal{A} as in (ii).

We set $n = 50$ assets and consider different values for the number of observations m . We set $\boldsymbol{\mu}$ and Σ equal to the sample estimates computed from the dataset of Section 5.2. We simulate independently N samples of m return observations from the three distributional models of Example 3.2. In the 2-points mixture case we set $x_1 = 5$ and $p = 0.1$; in the multivariate Student- t case we set $\nu = 3$.

In Figures 5.1 and 5.2 we see the values of loss function (2) in the unconstrained mean-variance problem (i). As N becomes larger we have convergence to the values given by Theorem 3.4. Note also that, as expected by Corollary 3.8, the drawbacks of model risk are considerably more severe in the 2-points mixture and the multivariate Student- t models compared to the multivariate Gaussian case. We observe that the loss function is decreasing in the sample size m , steep for m close to n and flat for $m \gg n$. From this analysis we conclude that the unconstrained mean-variance portfolio is severely compromised by model risk and particularly so in the case of non-Gaussian returns and m close to n . An adjustment in the sense of Section 4 is therefore necessary in this problem.

In Figures 5.3, 5.4 and 5.5 we provide loss function (2) in problem (i) using the covariance matrix estimator $\hat{\Sigma}_C$, where c_1, \dots, c_n are chosen according to (7), (9) and (11). Note that rule (7) simply increases the risk aversion, rule (9) is a heuristic approach and rule (11) is derived under simplifying assumptions. These estimators provide a significant reduction in the loss function, in particular for small m , compare Figures 5.3, 5.4 and 5.5 to Figure 5.2. We conclude that using the adjustment factors of Section 4 we obtain optimal portfolios

which are more robust to model risk. In practice, the adjustment factors (7), (9) and (11) need to be estimated, and therefore the loss function improvement is less pronounced. Observe that the loss function curves in Figures 5.3, 5.4 and 5.5 are less steep for m close to n compared to Figure 5.2. In Figures 5.6, 5.7 and 5.8 we show the adjustment factors computed for each eigenvariance. For (7) we have constant scaling (see Figure 5.6), and for (9) and (11) each eigenvariance is scaled individually (see Figures 5.7 and 5.8). In Figure 5.7 we also observe the stylized facts of Ledoit and Wolf [18], and Menchero et al. [26], i.e. small eigenvariances are substantially biased downwards and large eigenvariances are slightly biased.

Note that problem (i) requires estimation of $\boldsymbol{\mu}$ and Σ , whereas problems (ii), (iii) and (iv) require only estimation of Σ . We compare the behaviour of the loss functions in (ii), (iii) and (iv) for different values of m , since they are based on estimation of the same parameters. In Figures 5.9, 5.10 and 5.11 we present these loss functions. We observe that these problems are not severely deteriorated by model risk even in the case of n close to m or non-Gaussian observations because they only require estimation of Σ . The equal contributions to risk and maximum diversification portfolios are less affected by the distribution of the observations compared to the minimum variance portfolio.

5.2 Large-cap U.S. equity portfolios

We consider an investment universe of 50 large-cap U.S. stocks from different industry sectors. The list of companies is given in Appendix C. We set $n = 50$ and enumerate the companies according to the list. For these stocks we consider daily total returns (i.e. dividends are included) for the period from January 2, 1995 to October 14, 2011 (4380 daily observations). We study the mean-variance optimization problem under long only fully invested constraints, i.e. $\mathcal{A} = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{1}'\mathbf{w} = 1 \text{ and } w_i \geq 0 \text{ for all } i = 1, \dots, n\}$, see Example 2.1(i). In this numerical example we set $\kappa = 100$ for the risk aversion parameter. In this way the utility function is mainly determined by the portfolio variance and the problem is comparable to a constrained minimum variance problem, see Example 2.1(ii). For parameter estimation we use $m = 100$, i.e. for each trading day we consider the previous 100 daily observations to estimate $\boldsymbol{\mu}$ and Σ . Note that the estimation of $\boldsymbol{\mu}$ is required in this problem, unlike in problems (ii)-(iv) of Section 5.1. As shown in Section 5.1, the estimator $\hat{\Sigma}_C$, where c_1, \dots, c_n are selected according to (7) and (11), considerably reduces the negative effects of model risk in the unconstrained mean-variance problem. In this section we apply (7) and (11) to the problem of long only fully invested constraints. Observe that for non-Gaussian observations the adjustment factors (7) depend on the unobservable distribution of W . To estimate these values we calibrate a specific normal variance mixture model to the data. The adjustment factors (11) are also unobservable since they depend on the true covariance matrix. In order to estimate them we rely on the following simulation procedure. For each trading day, we estimate $\hat{\Sigma}$ and consider it as the true covariance matrix. We then simulate $N = 1000$ samples of 100 observations using a specific normal variance mixture model calibrated to the data. These samples are used, on the other hand, to estimate sample covariance matrices $\Sigma^{(1)}, \dots, \Sigma^{(N)}$. Diagonalization of these matrices provide the eigenvariances

$$(\hat{\lambda}_i^{(j)})_{i=1, \dots, n, j=1, \dots, N},$$

which are ordered such that $\hat{\lambda}_1^{(j)} \leq \dots \leq \hat{\lambda}_n^{(j)}$ for all $j = 1, \dots, N$. Then, we set for $i = 1, \dots, n$

$$c_i = \hat{\lambda}_i \frac{\sum_{k=1}^N (\hat{\lambda}_i^{(k)})^{-2}}{\sum_{k=1}^N (\hat{\lambda}_i^{(k)})^{-1}}. \quad (12)$$

We perform out-of-sample backtesting of the following five investment strategies, where we set $\kappa = 100$ and $\mathcal{A} = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{1}'\mathbf{w} = 1 \text{ and } w_i \geq 0 \text{ for all } i = 1, \dots, n\}$.

- (i) The classical mean-variance portfolio $\mathbf{w}^*(\hat{U}^{\text{mv}}, \mathcal{A})$ based on the sample estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$.
- (ii) Portfolio $\mathbf{w}^*(\hat{U}_{\bar{c}}^{\text{mv}}, \mathcal{A})$ based on the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}_{\bar{c}}$, where \bar{c} is determined by (7) under a multivariate Gaussian model.
- (iii) As in (ii), where \bar{c} is determined by (7) under a multivariate Student- t model.
- (iv) Portfolio $\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \mathcal{A})$ based on the estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}_C$, where c_1, \dots, c_n are determined numerically by (12) under a multivariate Gaussian model.
- (v) As in (iv), where c_1, \dots, c_n are determined numerically by (12) under a multivariate Student- t model.

In Tables 5.1 and 5.2 we present the out-of-sample backtesting results of investment strategies (i)-(v).

Remark 5.1. In Tables 5.1 and 5.2 we report maximum drawdown and average turnover values of the optimal portfolios. We briefly recall how these numbers are calculated. Let $\mathbf{r}_1, \dots, \mathbf{r}_T \in \mathbb{R}^n$ be a sample of T holding period returns and $\mathbf{w}_1, \dots, \mathbf{w}_T$ be T vectors of portfolio weights. The cumulative profit and loss $(p_t)_{t=1, \dots, T}$ is defined by $p_1 = 1$ and $p_t = p_{t-1} + \mathbf{w}'_{t-1} \mathbf{r}_t$ for $t = 2, \dots, T$. The drawdown curve $(\text{DD}_t)_{t=1, \dots, T}$ is computed from the cumulative profit and loss as

$$\text{DD}_t = \max_{1 \leq k \leq t} p_k - p_t.$$

This quantity measures the decline in value from a historical peak. This is a very important quantity in the asset management industry since it measures the potential loss of an investor entering the strategy in an unfavourable point in time. The maximum drawdown is defined by $\text{MDD} = \max_{1 \leq t \leq T} \text{DD}_t$. The turnover $(\text{TO}_t)_{1 \leq t \leq T}$ is defined by $\text{TO}_0 = 0$ and for $t = 2, \dots, T$

$$\text{TO}_t = \sum_{i=1}^n |w_{t,i} - w_{t-1,i}|.$$

This measures trading activity: higher turnover means higher transaction costs and more difficulties in the practical realization of the strategy. Note that this definition does not consider the trading activity necessary to rebalance the portfolio for changes in prices. The average turnover is defined by $\text{ATO} = T^{-1} \sum_{t=1}^T \text{TO}_t$.

In Tables 5.1 and 5.2 we see the following: first, we observe that the volatilities of the five investment strategies are similar over the entire period and in most single years (lower for investment strategies (ii)-(v) compared to (i)). In terms of annualized return, Sharpe ratio and maximum drawdown investment strategies (ii)-(v) outperform (i) over the entire period and in most single years. In particular, investment strategy (v) shows the best annualized return, Sharpe ratio and maximum drawdown. The average turnover of investment strategies (ii)-(v) is lower than the one of (i) over the entire period and in most single years. In this example we conclude the following. Investment strategies (ii) and (iv) slightly improve the mean-variance optimization strategy. Investment strategies (iii) and (v) provide significant improvement, particularly so for (v).

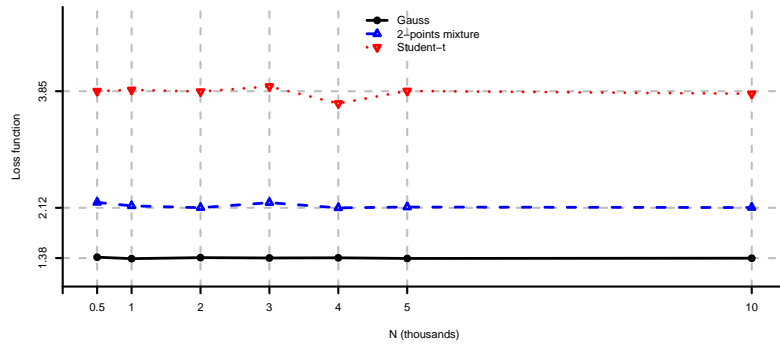


Figure 5.1: $L^{\text{Gauss}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$, $L^{2\text{pm}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$ and $L^t(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$ computed by simulation for $n = 50$, $m = 100$ and different values of N . The values given by Theorem 3.4 are approximately 1.3814, 2.1236 and 3.8453, respectively.

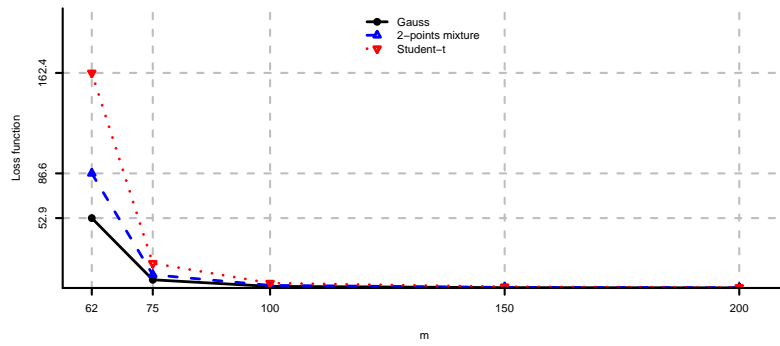


Figure 5.2: $L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$ computed using Theorem 3.4 for $n = 50$ and different values of m .

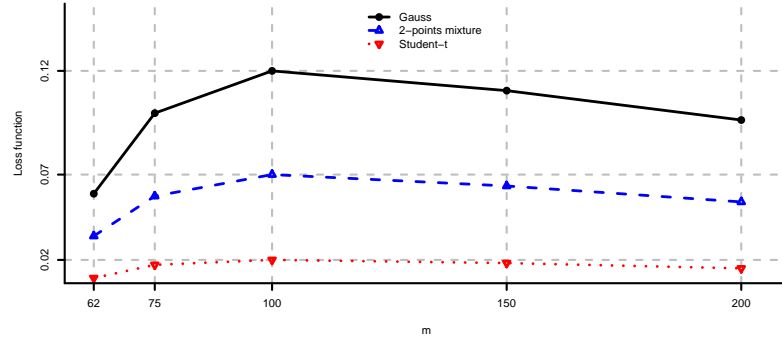


Figure 5.3: $L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n)$ computed by simulation ($N = 1000$) for $n = 50$ and different values of m . The adjustment factors c_1, \dots, c_n are selected according to (7).

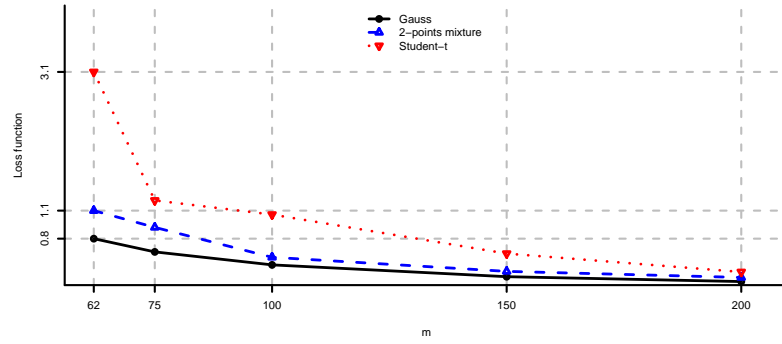


Figure 5.4: As in Figure 5.3 where c_1, \dots, c_n are selected according to (9).

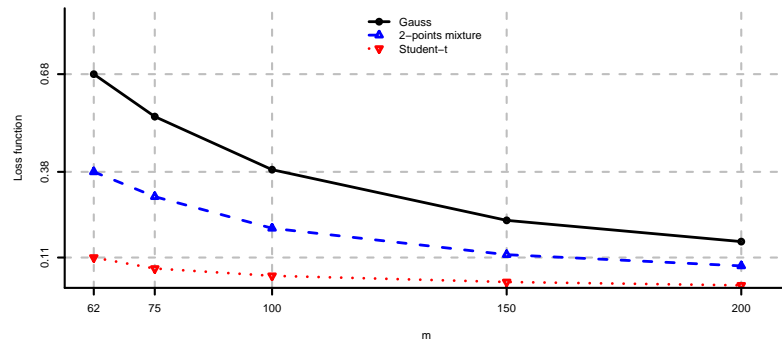


Figure 5.5: As in Figure 5.3 where c_1, \dots, c_n are selected according to (11).

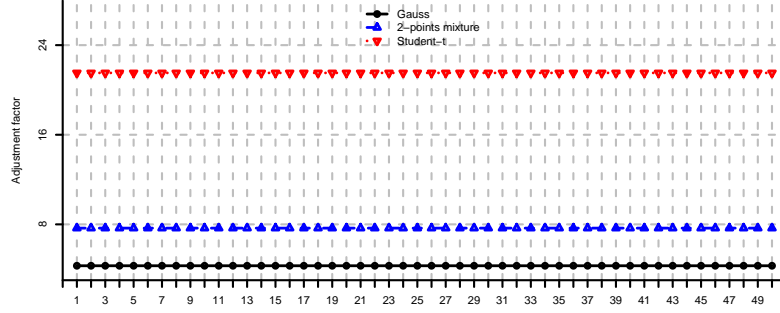


Figure 5.6: Adjustment factors (7) computed by simulation for $N = 1000$ and $m = 100$, for each eigenvariance.

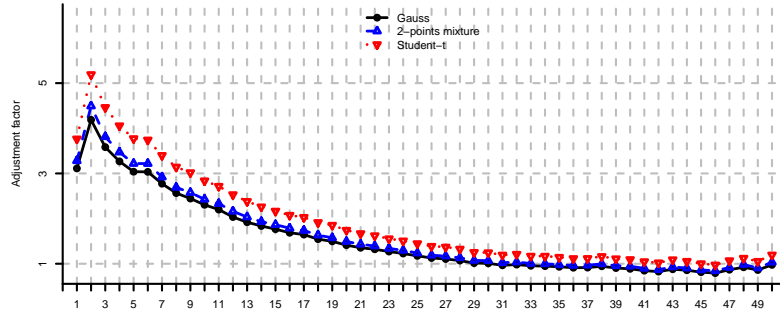


Figure 5.7: Adjustment factors (9) computed by simulation for $N = 1000$ and $m = 100$, for each eigenvariance.

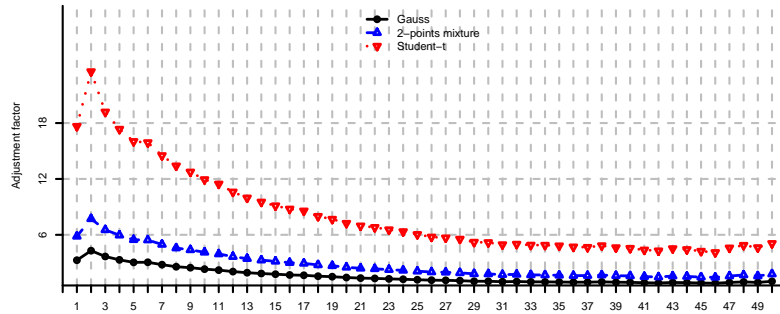


Figure 5.8: Adjustment factors (11) computed by simulation for $N = 1000$ and $m = 100$, for each eigenvariance.

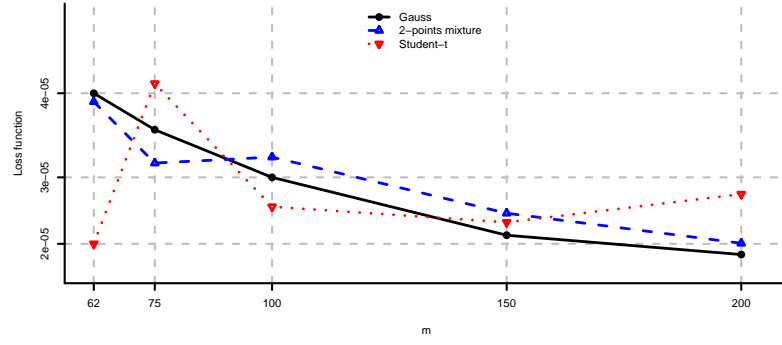


Figure 5.9: $L(U^v, U^v, \mathcal{A})$ for $N = 1000$ and different values of m .

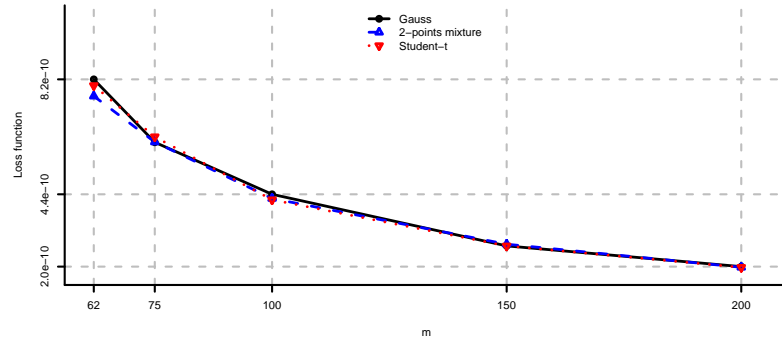


Figure 5.10: $L(U^{\text{ecr}}, U^{\text{ecr}}, \mathcal{A})$ for $N = 1000$ and different values of m .

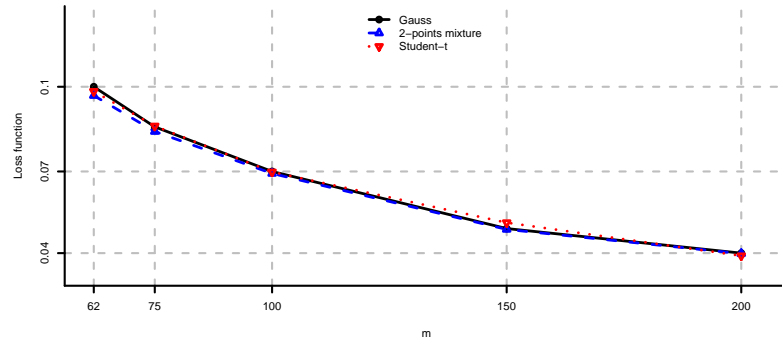


Figure 5.11: $L(U^{\text{md}}, U^{\text{md}}, \mathcal{A})$ for $N = 1000$ and different values of m .

| | Return annualized | | | Volatility annualized | | | Sharpe ratio | | | Maximum drawdown | | | Average turnover | | |
|------|-------------------|--------|--------|-----------------------|-------|-------|--------------|------|-------|------------------|-------|-------|------------------|-------|-------|
| | (i) | (ii) | (iii) | (i) | (ii) | (iii) | (i) | (ii) | (iii) | (i) | (ii) | (iii) | (i) | (ii) | (iii) |
| All | 7.63 | 8.03 | 8.66 | 15.06 | 15.01 | 15.00 | 0.51 | 0.53 | 0.58 | 47.80 | 47.16 | 45.31 | 13.38 | 12.78 | 12.25 |
| 2011 | 10.85 | 11.85 | 12.01 | 14.09 | 13.97 | 13.88 | 0.77 | 0.85 | 0.87 | 10.04 | 9.58 | 9.75 | 13.75 | 12.95 | 12.31 |
| 2010 | 9.13 | 9.26 | 8.90 | 11.12 | 10.97 | 10.86 | 0.82 | 0.84 | 0.82 | 7.63 | 7.35 | 7.33 | 10.39 | 9.94 | 9.80 |
| 2009 | 22.10 | 21.35 | 24.27 | 17.94 | 18.05 | 18.19 | 1.23 | 1.18 | 1.33 | 19.56 | 20.06 | 18.83 | 10.84 | 10.69 | 10.28 |
| 2008 | -27.67 | -26.57 | -25.93 | 28.82 | 28.74 | 28.91 | neg | neg | neg | 39.68 | 38.59 | 38.05 | 12.95 | 12.31 | 11.60 |
| 2007 | 10.29 | 10.06 | 10.72 | 11.33 | 11.23 | 11.09 | 0.91 | 0.90 | 0.97 | 7.05 | 6.87 | 6.70 | 14.27 | 13.27 | 12.80 |
| 2006 | 9.30 | 10.04 | 10.91 | 8.05 | 7.96 | 7.89 | 1.16 | 1.26 | 1.38 | 5.93 | 5.37 | 4.78 | 15.06 | 13.62 | 12.40 |
| 2005 | 1.66 | 2.02 | 1.94 | 9.59 | 9.51 | 9.41 | 0.17 | 0.21 | 0.21 | 7.29 | 6.79 | 6.62 | 15.75 | 15.16 | 14.48 |
| 2004 | 9.17 | 9.76 | 10.38 | 9.45 | 9.35 | 9.26 | 0.97 | 1.04 | 1.12 | 7.68 | 7.18 | 6.79 | 14.28 | 13.50 | 12.59 |
| 2003 | 20.49 | 21.05 | 21.44 | 12.16 | 12.10 | 12.07 | 1.69 | 1.74 | 1.78 | 8.57 | 8.59 | 8.80 | 13.03 | 12.64 | 12.05 |
| 2002 | -13.33 | -13.29 | -12.20 | 18.12 | 18.25 | 18.26 | neg | neg | neg | 27.35 | 27.56 | 27.19 | 12.33 | 12.05 | 11.74 |
| 2001 | -7.23 | -6.52 | -5.41 | 13.97 | 13.89 | 13.94 | neg | neg | neg | 16.49 | 16.20 | 16.00 | 13.93 | 13.45 | 13.14 |
| 2000 | 6.64 | 6.29 | 6.83 | 17.99 | 17.99 | 18.01 | 0.37 | 0.35 | 0.38 | 22.37 | 22.42 | 22.34 | 14.70 | 14.49 | 13.95 |
| 1999 | -2.76 | -1.62 | -0.41 | 15.32 | 15.30 | 15.25 | neg | neg | neg | 15.12 | 14.52 | 14.01 | 14.45 | 13.77 | 13.47 |
| 1998 | 18.57 | 19.22 | 19.09 | 17.63 | 17.57 | 17.61 | 1.05 | 1.09 | 1.08 | 17.41 | 17.57 | 17.88 | 13.82 | 13.53 | 13.35 |
| 1997 | 26.78 | 26.58 | 26.09 | 15.91 | 15.76 | 15.59 | 1.68 | 1.69 | 1.67 | 8.05 | 8.07 | 8.13 | 12.54 | 11.82 | 11.12 |
| 1996 | 18.62 | 19.32 | 20.01 | 12.01 | 11.94 | 11.88 | 1.55 | 1.62 | 1.68 | 7.33 | 7.13 | 7.17 | 14.93 | 14.22 | 13.70 |
| 1995 | 17.80 | 18.60 | 19.35 | 7.76 | 7.60 | 7.47 | 2.29 | 2.45 | 2.59 | 3.77 | 3.72 | 3.78 | 9.72 | 9.26 | 8.83 |

Table 5.1: Out-of-sample backtesting results of the strategies (i), (ii) and (iii). All numbers are expressed in percent except for Sharpe ratios. The return and volatility values are annualized assuming 250 trading days per year.

| | Return annualized | | | Volatility annualized | | | Sharpe ratio | | | Maximum drawdown | | | Average turnover | | |
|------|-------------------|--------|--------|-----------------------|-------|-------|--------------|------|------|------------------|-------|-------|------------------|-------|-------|
| | (i) | (iv) | (v) | (i) | (iv) | (v) | (i) | (iv) | (v) | (i) | (iv) | (v) | (i) | (iv) | (v) |
| All | 7.63 | 7.64 | 9.34 | 15.06 | 14.74 | 14.55 | 0.51 | 0.52 | 0.64 | 47.80 | 45.43 | 41.17 | 13.38 | 11.40 | 9.55 |
| 2011 | 10.85 | 11.46 | 13.43 | 14.09 | 13.95 | 13.35 | 0.77 | 0.82 | 1.01 | 10.04 | 10.10 | 9.65 | 13.75 | 11.63 | 8.65 |
| 2010 | 9.13 | 9.42 | 7.85 | 11.12 | 10.83 | 10.43 | 0.82 | 0.87 | 0.75 | 7.63 | 7.09 | 6.83 | 10.39 | 9.34 | 8.34 |
| 2009 | 22.10 | 17.63 | 18.27 | 17.94 | 16.48 | 16.00 | 1.23 | 1.07 | 1.14 | 19.56 | 19.49 | 19.72 | 10.84 | 8.77 | 7.83 |
| 2008 | -27.67 | -25.08 | -20.81 | 28.82 | 28.09 | 28.06 | neg | neg | neg | 39.68 | 37.17 | 33.29 | 12.95 | 11.06 | 10.74 |
| 2007 | 10.29 | 11.40 | 13.34 | 11.33 | 11.23 | 10.77 | 0.91 | 1.02 | 1.24 | 7.05 | 6.67 | 5.60 | 14.27 | 12.58 | 9.21 |
| 2006 | 9.30 | 9.41 | 12.84 | 8.05 | 7.98 | 7.80 | 1.16 | 1.18 | 1.65 | 5.93 | 6.12 | 4.26 | 15.06 | 12.41 | 9.22 |
| 2005 | 1.66 | 2.60 | 0.69 | 9.59 | 9.55 | 9.28 | 0.17 | 0.27 | 0.07 | 7.29 | 6.80 | 7.42 | 15.75 | 13.71 | 11.14 |
| 2004 | 9.17 | 9.63 | 10.66 | 9.45 | 9.39 | 9.10 | 0.97 | 1.03 | 1.17 | 7.68 | 7.27 | 6.63 | 14.28 | 12.32 | 9.32 |
| 2003 | 20.49 | 20.85 | 22.23 | 12.16 | 11.98 | 11.90 | 1.69 | 1.74 | 1.87 | 8.57 | 7.95 | 8.03 | 13.03 | 11.12 | 9.81 |
| 2002 | -13.33 | -15.84 | -15.64 | 18.12 | 17.72 | 18.16 | neg | neg | neg | 27.35 | 28.48 | 29.53 | 12.33 | 10.59 | 9.98 |
| 2001 | -7.23 | -10.55 | -5.53 | 13.97 | 13.75 | 13.67 | neg | neg | neg | 16.49 | 17.88 | 15.71 | 13.93 | 10.92 | 9.44 |
| 2000 | 6.64 | 8.22 | 11.63 | 17.99 | 17.71 | 17.74 | 0.37 | 0.46 | 0.66 | 22.37 | 21.88 | 21.99 | 14.70 | 12.09 | 10.29 |
| 1999 | -2.76 | -0.29 | 4.73 | 15.32 | 14.97 | 14.79 | neg | neg | 0.32 | 15.12 | 13.74 | 10.51 | 14.45 | 11.69 | 9.65 |
| 1998 | 18.57 | 16.90 | 15.64 | 17.63 | 17.57 | 17.45 | 1.05 | 0.96 | 0.90 | 17.41 | 18.08 | 18.41 | 13.82 | 12.30 | 11.17 |
| 1997 | 26.78 | 26.78 | 27.04 | 15.91 | 15.78 | 15.10 | 1.68 | 1.70 | 1.79 | 8.05 | 7.86 | 7.87 | 12.54 | 10.87 | 8.82 |
| 1996 | 18.62 | 18.54 | 22.43 | 12.01 | 12.03 | 11.77 | 1.55 | 1.54 | 1.91 | 7.33 | 6.94 | 7.78 | 14.93 | 13.33 | 11.13 |
| 1995 | 17.80 | 19.73 | 20.95 | 7.76 | 7.73 | 7.10 | 2.29 | 2.55 | 2.95 | 3.77 | 3.62 | 4.05 | 9.72 | 8.49 | 6.94 |

Table 5.2: Out-of-sample backtesting results of the strategies (i), (iv) and (v). All numbers are expressed in percent except for Sharpe ratios. The return and volatility values are annualized assuming 250 trading days per year.

A Proofs

A.1 Proof of Proposition 2.5

Since $\mathbf{w}^*(U, \mathcal{A}) \in \mathcal{A}$ is the maximal point of U in \mathcal{A} , we have

$$U(\mathbf{w}^*(U, \mathcal{A})) \geq U(\mathbf{w}^*(\hat{U}, \mathcal{A})), \text{ almost surely.} \quad (13)$$

Since $\mathbf{w}^*(\hat{U}, \mathcal{A}) \in \mathcal{A}$ is the maximal point of \hat{U} in \mathcal{A} , we also have

$$\hat{U}(\mathbf{w}^*(\hat{U}, \mathcal{A})) \geq \hat{U}(\mathbf{w}^*(U, \mathcal{A})), \text{ almost surely.} \quad (14)$$

Taking the expectation of (14) and using the assumption that \hat{U} is unbiased for U , we obtain

$$E[\hat{U}(\mathbf{w}^*(\hat{U}, \mathcal{A}))] \geq E[\hat{U}(\mathbf{w}^*(U, \mathcal{A}))] = U(\mathbf{w}^*(U, \mathcal{A})). \quad (15)$$

Combining (13) and (15) we get the desired inequality

$$L(U, \hat{U}, \mathcal{A}) = E[\hat{U}(\mathbf{w}^*(\hat{U}, \mathcal{A}))] - U[(\mathbf{w}^*(\hat{U}, \mathcal{A}))] \geq U(\mathbf{w}^*(U, \mathcal{A})) - U(\mathbf{w}^*(U, \mathcal{A})) = 0.$$

Regarding the statement on strict inequality: because of the uniqueness of the maximum of U , the assumption $P[\mathbf{w}^*(U, \mathcal{A}) \neq \mathbf{w}^*(\hat{U}, \mathcal{A})] > 0$ implies that there is a set in Ω with non-zero probability such that the inequalities (13) and (14) hold in strict sense. Therefore, the inequalities obtained taking the expected values must hold strictly as well. This proves the claim. \square

A.2 Proof of Theorem 3.4

From Example 2.1(i) we have

$$\mathbf{w}^*(\hat{U}^{\text{mv}}, \mathbb{R}^n) = \frac{1}{2\kappa} \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}.$$

Thus, for the loss function we obtain

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) &= E \left[\hat{U}^{\text{mv}} \left(\frac{1}{2\kappa} \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \right) - U^{\text{mv}} \left(\frac{1}{2\kappa} \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \right) \right] \\ &= \frac{1}{4\kappa} E \left[2\hat{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \hat{\Sigma} \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}} - 2\boldsymbol{\mu}' \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \right] \quad (16) \\ &= \frac{1}{4\kappa} E [\hat{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}] - \frac{1}{2\kappa} \boldsymbol{\mu}' E [\hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}] + \frac{1}{4\kappa} E [\hat{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}]. \end{aligned}$$

From assumption (4) on the observations we have $(\mathbf{R}_i|W)_{i=1,\dots,m} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}, W\Sigma)$. Thus, Proposition B.2 in the Appendix implies that the sample estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are conditionally independent, given W , with conditional distributions given by

$$\hat{\boldsymbol{\mu}}|W \sim \mathcal{N}(\boldsymbol{\mu}, m^{-1}W\Sigma) \quad \text{and} \quad (m-1)\hat{\Sigma}|W \sim \mathcal{W}(W\Sigma, m-1).$$

For the inverse sample covariance matrix we have

$$E[\hat{\Sigma}^{-1}|W] = (m-1)E[(m-1)\hat{\Sigma}]^{-1}|W = \alpha(m-1, n)W^{-1}\Sigma^{-1}, \quad (17)$$

and

$$\begin{aligned} E[\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}|W] &= (m-1)^2 E[(m-1)\hat{\Sigma}]^{-1}\Sigma[(m-1)\hat{\Sigma}]^{-1}|W \\ &= \frac{\alpha(m-1, n)\beta(m-1, n)}{m-2} \text{tr}(W^{-1}\Sigma^{-1}\Sigma)W^{-1}\Sigma^{-1} \\ &\quad + \frac{\beta(m-1, n)(m-1)}{m-2} W^{-1}\Sigma^{-1}\Sigma W^{-1}\Sigma^{-1} \\ &= \alpha(m-1, n)\beta(m-1, n)W^{-2}\Sigma^{-1}, \end{aligned} \quad (18)$$

where we have used Proposition B.3 in the Appendix, $\text{tr}(W^{-1}\Sigma^{-1}\Sigma) = nW^{-1}$ and

$$(m-2)^{-1}(\alpha(m-1, n)n + m - 1) = \alpha(m-1, n).$$

Using (17), (18) and the conditional independence between $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ we compute the three expectations in (16)

$$\begin{aligned} E[\hat{\boldsymbol{\mu}}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}] &= E[E[\hat{\boldsymbol{\mu}}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}|W]] = \sum_{i,j=1}^n E[E[\hat{\Sigma}_{ij}^{-1}\hat{\mu}_i\hat{\mu}_j|W]] \\ &= \sum_{i,j=1}^n E[E[\hat{\Sigma}_{ij}^{-1}|W]E[\hat{\mu}_i\hat{\mu}_j|W]] \\ &= \alpha(m-1, n) \sum_{i,j=1}^n E\left[W^{-1}\Sigma_{ij}^{-1}\left(\frac{W}{m}\Sigma_{ij} + \mu_i\mu_j\right)\right] \\ &= \alpha(m-1, n)\left(\frac{n}{m} + E[W^{-1}]\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\right), \end{aligned} \tag{19}$$

$$\begin{aligned} \boldsymbol{\mu}'E[\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}] &= \boldsymbol{\mu}'E[E[\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}|W]] = \boldsymbol{\mu}'E[E[\hat{\Sigma}^{-1}|W]E[\hat{\boldsymbol{\mu}}|W]] \\ &= \alpha(m-1, n)\boldsymbol{\mu}'E[W^{-1}\Sigma^{-1}\boldsymbol{\mu}] = \alpha(m-1, n)E[W^{-1}]\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}, \end{aligned} \tag{20}$$

and

$$\begin{aligned} E[\hat{\boldsymbol{\mu}}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}] &= E[E[\hat{\boldsymbol{\mu}}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}|W]] = \sum_{i,j=1}^n E[E[(\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1})_{ij}\hat{\mu}_i\hat{\mu}_j|W]] \\ &= \sum_{i,j=1}^n E[E[(\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1})_{ij}|W]E[\hat{\mu}_i\hat{\mu}_j|W]] \\ &= \alpha(m-1, n)\beta(m-1, n) \sum_{i,j=1}^n E\left[W^{-2}\Sigma_{ij}^{-1}\left(\frac{W}{m}\Sigma_{ij} + \mu_i\mu_j\right)\right] \\ &= \alpha(m-1, n)\beta(m-1, n)\left(\frac{n}{m}E[W^{-1}] + E[W^{-2}]\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\right). \end{aligned} \tag{21}$$

Finally, substituting the last three expectations into (16), proves the claim. \square

A.3 Proof of Corollary 3.6

We obtain the following expressions for the expectations $E[W^{-1}]$ and $E[W^{-2}]$.

- (i) Multivariate Gauss: the statement follows directly setting $W = 1$.
- (ii) 2-points mixture: set $W \sim x_2 + (x_1 - x_2)\text{Bernoulli}(p)$, where $x_1 \in (0, p^{-1})$ and $x_2 = \frac{1-px_1}{1-p}$. Then,

$$E[W^{-1}] = \frac{p}{x_1} + \frac{(1-p)^2}{1-px_1} \quad \text{and} \quad E[W^{-2}] = \frac{p}{x_1^2} + \frac{(1-p)^3}{(1-px_1)^2}.$$

(iii) Multivariate Student- t : set $W \sim \frac{\nu-2}{\nu} \text{Ig}(\frac{\nu}{2}, \frac{\nu}{2})$. Then, $W^{-1} \sim \frac{\nu}{\nu-2} \text{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$,

$$E[W^{-1}] = \frac{\nu}{\nu-2} \text{ and } E[W^{-2}] = \frac{\nu^2}{(\nu-2)^2} \left(\frac{2}{\nu} + 1 \right).$$

Substituting these expressions into Theorem 3.4 proves the statements. \square

A.4 Proof of Corollary 3.8

By Jensen's inequality we have for non-deterministic W

$$E[W^{-1}] > E[W]^{-1} = 1,$$

$$E[W^{-2}] > E[W]^{-2} = 1,$$

and

$$E[W^{-1}] = E[(W^{-2})^{\frac{1}{2}}] < E[W^{-2}]^{\frac{1}{2}} < E[W^{-2}].$$

In the last inequality we use that $E[W^{-2}] > 1$. Using these inequalities in combination with the result of Theorem 3.4 we have

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) &= \frac{1}{4\kappa} \alpha(m-1, n) \left\{ \frac{n}{m} (1 + \beta(m-1, n) E[W^{-1}]) \right. \\ &\quad \left. + (\beta(m-1, n) E[W^{-2}] - E[W^{-1}]) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\} \\ &> \frac{1}{4\kappa} \alpha(m-1, n) \left\{ \frac{n}{m} (1 + \beta(m-1, n)) \right. \\ &\quad \left. + E[W^{-2}] (\beta(m-1, n) - 1) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\} \\ &> L^{\text{Gauss}}(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n). \end{aligned}$$

This proves the claim. \square

A.5 Proof of Proposition 4.3

We introduce the function $l(\bar{c}) = L(U^{\text{mv}}, \hat{U}_{\bar{c}}^{\text{mv}}, \mathbb{R}^n)$. Using equation (16) we have

$$\begin{aligned} l(\bar{c}) &= \frac{1}{4\kappa} E[\hat{\boldsymbol{\mu}}' \hat{\Sigma}_{\bar{c}}^{-1} \hat{\boldsymbol{\mu}}] - \frac{1}{2\kappa} E[\boldsymbol{\mu}' \hat{\Sigma}_{\bar{c}}^{-1} \hat{\boldsymbol{\mu}}] + \frac{1}{4\kappa} E[\hat{\boldsymbol{\mu}}' \hat{\Sigma}_{\bar{c}}^{-1} \Sigma \hat{\Sigma}_{\bar{c}}^{-1} \hat{\boldsymbol{\mu}}] \\ &= \frac{1}{4\bar{c}\kappa} E[\hat{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}] - \frac{1}{2\bar{c}\kappa} \boldsymbol{\mu}' E[\hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}] + \frac{1}{4\bar{c}^2\kappa} E[\hat{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\boldsymbol{\mu}}]. \end{aligned}$$

Observe that $l(1) = L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n)$. Using (19), (20) and (21) we have

$$\begin{aligned} l(\bar{c}) &= \alpha(m-1, n) \left\{ \frac{n}{m} \left(\frac{1}{\bar{c}} + \frac{1}{\bar{c}^2} \beta(m-1, n) E[W^{-1}] \right) \right. \\ &\quad \left. + \left(\frac{1}{\bar{c}^2} \beta(m-1, n) E[W^{-2}] - \frac{1}{\bar{c}} E[W^{-1}] \right) \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} \right\}. \end{aligned}$$

From Remark 3.5(i) we have

$$\beta(m-1, n)E[W^{-2}] - E[W^{-1}] \geq E[W^{-2}] - E[W^{-1}] \geq 0.$$

Set

$$\bar{c} = \beta(m-1, n) \frac{E[W^{-2}]}{E[W^{-1}]} \geq 1.$$

Then, we obtain

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n) &= l(1) \geq \frac{1}{4\kappa} \alpha(m-1, n) \frac{n}{m} \left(1 + \beta(m-1, n) E[W^{-1}] \right) \\ &\geq \frac{1}{4\kappa} \alpha(m-1, n) \frac{n}{m} \left(\frac{1}{\bar{c}} + \frac{1}{\bar{c}^2} \beta(m-1, n) E[W^{-1}] \right) \\ &= l(\bar{c}) = L(U^{\text{mv}}, \hat{U}_{\bar{c}}^{\text{mv}}, \mathbb{R}^n). \end{aligned}$$

This proves the claim. □

A.6 Proof of Theorem 4.5

First observe that $\hat{\Sigma}_C$ and $\hat{\Sigma}$ have the same eigenportfolios $(\hat{\mathbf{v}}_k)_{k=1, \dots, n}$ with eigenvariances $(c_k \hat{\lambda}_k)_{k=1, \dots, n}$ and $(\hat{\lambda}_k)_{k=1, \dots, n}$, respectively. Since the set of eigenportfolios $(\mathbf{v}_k)_{k=1, \dots, n}$ is an orthonormal basis of \mathbb{R}^n we have

$$\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}) = \sum_{k=1}^n \pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \Sigma) \mathbf{v}_k,$$

and for $k, l = 1, \dots, n$

$$\mathbf{v}_k' \Sigma \mathbf{v}_l = \delta_{kl} \lambda_k.$$

In the same way for the set of eigenportfolios $(\hat{\mathbf{v}}_k)_{k=1, \dots, n}$ we have

$$\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}) = \sum_{k=1}^n \pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \hat{\Sigma}) \hat{\mathbf{v}}_k,$$

and for $k, l = 1, \dots, n$

$$\hat{\mathbf{v}}_k' \hat{\Sigma}_C \hat{\mathbf{v}}_l = \delta_{kl} c_k \hat{\lambda}_k.$$

The loss function $L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \hat{\mathcal{A}})$ for any $\hat{\mathcal{A}}$ can be written as

$$\begin{aligned} L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}) &= E \left[\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}})' (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) - \kappa \mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}})' \hat{\Sigma}_C \mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}) \right. \\ &\quad \left. + \kappa \mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}})' \Sigma \mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}) \right]. \end{aligned}$$

Using the orthonormal basis $(\mathbf{v}_k)_{k=1,\dots,n}$ and $(\hat{\mathbf{v}}_k)_{k=1,\dots,n}$ we obtain

$$\begin{aligned}
L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}) &= E[\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}})'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})] \\
&\quad - \kappa \sum_{k,l=1}^n E\left[\pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \hat{\Sigma}) \pi_l(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \hat{\Sigma}) \hat{\mathbf{v}}_k' \hat{\Sigma}_C \hat{\mathbf{v}}_l\right] \\
&\quad + \kappa \sum_{k,l=1}^n E\left[\pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \Sigma) \pi_l(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \Sigma) \mathbf{v}_k' \Sigma \mathbf{v}_l\right] \quad (22) \\
&= E[\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}})'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})] \\
&\quad + \kappa \sum_{k=1}^n E\left[\pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \Sigma)^2 \lambda_k - \pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \hat{\mathcal{A}}), \hat{\Sigma})^2 c_k \hat{\lambda}_k\right].
\end{aligned}$$

In the unconstrained problem we have

$$\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \mathbb{R}^n) = \frac{1}{2\kappa} \hat{\Sigma}_C^{-1} \hat{\boldsymbol{\mu}},$$

and therefore we obtain

$$\begin{aligned}
\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \mathbb{R}^n)' \hat{\boldsymbol{\mu}} &= \frac{1}{2\kappa} \hat{\boldsymbol{\mu}}' \hat{\Sigma}_C^{-1} \hat{\boldsymbol{\mu}} = \sum_{k=1}^n \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma}_C^{-1})^2 c_k^{-1} \hat{\lambda}_k^{-1} = \sum_{k=1}^n \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma})^2 c_k^{-1} \hat{\lambda}_k^{-1}, \\
\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \mathbb{R}^n)' \boldsymbol{\mu} &= \frac{1}{2\kappa} \hat{\boldsymbol{\mu}}' \hat{\Sigma}_C^{-1} \boldsymbol{\mu} = \sum_{k=1}^n \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma}_C^{-1}) \pi_k(\boldsymbol{\mu}, \hat{\Sigma}_C^{-1}) c_k^{-1} \hat{\lambda}_k^{-1} \\
&= \sum_{k=1}^n \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma}) \pi_k(\boldsymbol{\mu}, \hat{\Sigma}) c_k^{-1} \hat{\lambda}_k^{-1}, \\
\pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \mathbb{R}^n), \Sigma) &= \frac{1}{2\kappa} \pi_k(\hat{\Sigma}_C^{-1} \hat{\boldsymbol{\mu}}, \Sigma), \quad \text{and} \\
\pi_k(\mathbf{w}^*(\hat{U}_C^{\text{mv}}, \mathbb{R}^n), \hat{\Sigma}) &= (\hat{T}' \mathbf{w}^*(\hat{U}_C^{\text{mv}}, \mathbb{R}^n))_k = \frac{1}{2\kappa} (\hat{T}' \hat{\Sigma}_C^{-1} \hat{\boldsymbol{\mu}})_k = \frac{1}{2\kappa} (\hat{T}' \hat{T} C^{-1} \hat{\Lambda}^{-1} \hat{T}' \hat{\boldsymbol{\mu}})_k \\
&= \frac{1}{2\kappa} c_k^{-1} \hat{\lambda}_k^{-1} (\hat{T}' \hat{\boldsymbol{\mu}})_k = \frac{1}{2\kappa} c_k^{-1} \hat{\lambda}_k^{-1} \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma}).
\end{aligned}$$

Plugging these expressions into (22) we obtain

$$\begin{aligned}
L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n) &= \frac{1}{2\kappa} \sum_{k=1}^n E\left[(\pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma}) - \pi_k(\boldsymbol{\mu}, \hat{\Sigma})) c_k^{-1} \hat{\lambda}_k^{-1} \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma})\right] \\
&\quad + \frac{1}{4\kappa} \sum_{k=1}^n E\left[\pi_k(\hat{\Sigma}_C^{-1} \hat{\boldsymbol{\mu}}, \Sigma)^2 \lambda_k - \pi_k(\hat{\boldsymbol{\mu}}, \hat{\Sigma})^2 c_k^{-1} \hat{\lambda}_k^{-1}\right].
\end{aligned}$$

This proves the statement. □

A.7 Proof of Theorem 4.7

Under assumption (10) we have $\pi_k(\cdot, \Sigma) = \pi_k(\cdot, \hat{\Sigma}) = \pi_k(\cdot, \hat{\Sigma}_C)$ almost surely. Then, using Theorem 4.5, $E[\hat{\boldsymbol{\mu}}] = \boldsymbol{\mu}$ and the independence between $\hat{\boldsymbol{\mu}}$ and $\hat{\Lambda}$ we obtain

$$\begin{aligned}
L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n) &= \frac{1}{2\kappa} \sum_{k=1}^n E \left[(\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)^2 - \pi_k(\hat{\boldsymbol{\mu}}, \Sigma) \pi_k(\boldsymbol{\mu}, \Sigma)) c_k^{-1} \hat{\lambda}_k^{-1} \right] \\
&\quad + \frac{1}{4\kappa} \sum_{k=1}^n E [\pi_k(\hat{\Sigma}_C^{-1} \hat{\boldsymbol{\mu}}, \Sigma)^2 \lambda_k - \pi_k(\hat{\boldsymbol{\mu}}, \Sigma)^2 c_k^{-1} \hat{\lambda}_k^{-1}] \\
&= \frac{1}{2\kappa} \sum_{k=1}^n c_k^{-1} \text{Var}[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)] E[\hat{\lambda}_k^{-1}] \\
&\quad + \frac{1}{4\kappa} \sum_{k=1}^n E[(T' T C^{-1} \hat{\Lambda}^{-1} T' \hat{\boldsymbol{\mu}})_k^2 \lambda_k - \pi_k(\hat{\boldsymbol{\mu}}, \Sigma)^2 c_k^{-1} \hat{\lambda}_k^{-1}] \\
&= \frac{1}{2\kappa} \sum_{k=1}^n c_k^{-1} \text{Var}[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)] E[\hat{\lambda}_k^{-1}] \\
&\quad + \frac{1}{4\kappa} \sum_{k=1}^n E[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)^2] (c_k^{-2} E[\hat{\lambda}_k^{-2}] \lambda_k - c_k^{-1} E[\hat{\lambda}_k^{-1}]).
\end{aligned} \tag{23}$$

This proves the first part of the theorem. Set for $k = 1, \dots, n$

$$c_k = \frac{\lambda_k E[\hat{\lambda}_k^{-2}]}{E[\hat{\lambda}_k^{-1}]}.$$

By Jensen's inequality we have for $k = 1, \dots, n$

$$c_k \geq \frac{\lambda_k E[\hat{\lambda}_k^{-1}]^2}{E[\hat{\lambda}_k^{-1}]} = \lambda_k E[\hat{\lambda}_k^{-1}] \geq \lambda_k E[\hat{\lambda}_k]^{-1} = 1,$$

where we have used the assumption of unbiased $\hat{\Sigma}$. Then, using (23) we have for these c_k 's

$$\begin{aligned}
L(U^{\text{mv}}, \hat{U}_C^{\text{mv}}, \mathbb{R}^n) &= \frac{1}{2\kappa} \sum_{k=1}^n c_k^{-1} \text{Var}[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)] E[\hat{\lambda}_k^{-1}] \leq \frac{1}{2\kappa} \sum_{k=1}^n \text{Var}[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)] E[\hat{\lambda}_k^{-1}] \\
&\leq \frac{1}{2\kappa} \sum_{k=1}^n \left(\text{Var}[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)] E[\hat{\lambda}_k^{-1}] + \frac{1}{2} E[\pi_k(\hat{\boldsymbol{\mu}}, \Sigma)^2] (E[\hat{\lambda}_k^{-2}] \lambda_k - E[\hat{\lambda}_k^{-1}]) \right) \\
&= L(U^{\text{mv}}, \hat{U}^{\text{mv}}, \mathbb{R}^n),
\end{aligned}$$

where in the last step, we have used (23) for $c_1 = \dots = c_n = 1$. This proves the claim. \square

B Wishart distribution

We recall the definition of the Wishart distribution and some properties we use in this paper. For more details on the Wishart distribution we refer to Mardia et al. [23]. Let X

be a $m \times n$ random matrix whose rows are i.i.d multivariate Gaussian distributed random vectors with mean vector $\mathbf{0}_n$ and positive definite covariance matrix Σ . Then, $M = X'X$ has a Wishart distribution with scale matrix Σ and m degrees of freedom denoted by $\mathcal{W}(\Sigma, m)$. In particular, for $n = 1$, we have $X = (X_1, \dots, X_m)'$ with i.i.d. $X_k \sim \mathcal{N}(0, \sigma^2)$. Hence, $M = \sum_{k=1}^m X_k^2 \sim \sigma^2 \chi_m^2$, i.e. chi-squared distributed. The mean and the variance of a Wishart matrix M are given by,

$$E[M] = m\Sigma, \quad \text{Var}[M_{ij}] = m(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}), \quad (24)$$

for $i, j = 1, \dots, n$. The first important fact of the Wishart distribution is the following.

Proposition B.1. *Let $A \in \mathbb{R}^{p \times n}$ and $M \sim \mathcal{W}(\Sigma, m)$. Then $AMA' \sim \mathcal{W}(A\Sigma A', m)$.*

In particular, this proposition implies that for $m \geq n$ the random matrix M is positive definite with probability 1 if Σ is positive definite. To verify this let $\mathbf{a} \in \mathbb{R}^n$ be a non-zero vector. Using the above proposition, we have that

$$\mathbf{a}'M\mathbf{a} \sim \mathcal{W}(\mathbf{a}'\Sigma\mathbf{a}, m) \sim \mathbf{a}'\Sigma\mathbf{a}\chi_m^2,$$

which implies $\mathbf{a}'M\mathbf{a} > 0$ almost surely since $\mathbf{a}'\Sigma\mathbf{a} > 0$. We also rely on the following key property.

Proposition B.2. *Let $\mathbf{R}_1, \dots, \mathbf{R}_m$ be n -dimensional i.i.d multivariate Gaussian random vectors with mean $\boldsymbol{\mu}$ and covariance matrix Σ . Let $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ be the sample mean and sample covariance matrix given by*

$$\hat{\boldsymbol{\mu}} = \frac{1}{m} \sum_{i=1}^m \mathbf{R}_i \quad \text{and} \quad \hat{\Sigma} = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{R}_i - \hat{\boldsymbol{\mu}})(\mathbf{R}_i - \hat{\boldsymbol{\mu}})'$$

Then, $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are independent, $\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \frac{1}{m}\Sigma)$ and $(m-1)\hat{\Sigma} \sim \mathcal{W}(\Sigma, m-1)$.

Note that the sample covariance matrix estimator is a Wishart matrix with only $m-1$ degrees of freedom. This is because in the sample covariance matrix estimator we subtract the sample mean from the observations and not the true mean. Finally, important for this paper are the first two moments of the inverse of a Wishart matrix. In Haff [10] the following result is proven.

Proposition B.3. *Let $M \sim \mathcal{W}(\Sigma, m)$, $A \in \mathbb{R}^{n \times n}$ positive semi definite and $m > n + 3$. Then,*

$$(i) \quad E[M^{-1}] = \frac{\alpha(m, n)}{m} \Sigma^{-1}; \text{ and}$$

$$(ii) \quad E[M^{-1}AM^{-1}] = \frac{\alpha(m, n)\beta(m, n)}{m^2(m-1)} \text{tr}(\Sigma^{-1}A)\Sigma^{-1} + \frac{\beta(m, n)}{m(m-1)} \Sigma^{-1}A\Sigma^{-1};$$

where $\alpha(m, n) = \frac{m}{m-n-1}$ and $\beta(m, n) = \frac{m(m-1)}{(m-n)(m-n-3)}$.

C Equity universe for case studies

| Nr. | Company | Bloomberg ticker | Industry sector |
|-----|---------------------------------------|------------------|------------------------|
| 1 | Wells Fargo & Company | WFC US Equity | financials |
| 2 | JP Morgan Chase & Co. | JPM US Equity | financials |
| 3 | Citigroup, Inc. | C US Equity | financials |
| 4 | Bank of America Corporation | BAC US Equity | financials |
| 5 | American Express Company | AXP US Equity | financials |
| 6 | American International Group | AIG US Equity | financials |
| 7 | PNC Financial Services Group | PNC US Equity | financials |
| 8 | General Electric Company | GE US Equity | industrials |
| 9 | United Technologies Corporation | UTX US Equity | industrials |
| 10 | 3M Company | MMM US Equity | industrials |
| 11 | Caterpillar, Inc. | CAT US Equity | industrials |
| 12 | Boeing Company | BA US Equity | industrials |
| 13 | Union Pacific Corporation | UNP US Equity | industrials |
| 14 | Honeywell International | HON US Equity | industrials |
| 15 | Wal-Mart Stores, Inc. | WMT US Equity | consumer staples |
| 16 | McDonald's Corporation | MCD US Equity | consumer discretionary |
| 17 | Comcast Corporation | CMCSA US Equity | consumer discretionary |
| 18 | Walt Disney Company | DIS US Equity | consumer discretionary |
| 19 | Home Depot, Inc. | HD US Equity | consumer discretionary |
| 20 | CVS Caremark Corporation | CVS US Equity | consumer staples |
| 21 | Costco Wholesale Corporation | COST US Equity | consumer staples |
| 22 | Apple Inc. | AAPL US Equity | information technology |
| 23 | Microsoft Corporation | MSFT US Equity | information technology |
| 24 | International Business Machines (IBM) | IBM US Equity | information technology |
| 25 | Oracle Corporation | ORCL US Equity | information technology |
| 26 | Intel Corporation | INTC US Equity | information technology |
| 27 | Hewlett-Packard Company | HPQ US Equity | information technology |
| 28 | EMC Corporation Common | EMC US Equity | information technology |
| 29 | Exxon Mobil Corporation | XOM US Equity | energy |
| 30 | Chevron Corporation Common | CVX US Equity | energy |
| 31 | Schlumberger N.V. | SLB US Equity | energy |
| 32 | ConocoPhillips | COP US Equity | energy |
| 33 | Occidental Petroleum | OXY US Equity | energy |
| 34 | Anadarko Petroleum | APC US Equity | energy |
| 35 | Apache Corporation | APA US Equity | energy |
| 36 | Procter & Gamble Company | PG US Equity | consumer staples |
| 37 | Coca-Cola Company | KO US Equity | consumer staples |
| 38 | Pepsico, Inc. | PEP US Equity | consumer staples |
| 39 | Altria Group, Inc. | MO US Equity | consumer staples |
| 40 | Colgate-Palmolive Company | CL US Equity | consumer staples |
| 41 | Ford Motor Company | F US Equity | consumer discretionary |
| 42 | Nike, Inc. | NKE US Equity | consumer discretionary |
| 43 | Kimberly-Clark Corporation | KMB US Equity | consumer staples |
| 44 | Johnson & Johnson | JNJ US Equity | health care |
| 45 | Pfizer, Inc. Common Stock | PFE US Equity | health care |
| 46 | Merck & Company, Inc. | MRK US Equity | health care |
| 47 | Abbott Laboratories | ABT US Equity | health care |
| 48 | Bristol-Myers Squibb Company | BMJ US Equity | health care |
| 49 | Amgen Inc. | AMGN US Equity | health care |
| 50 | UnitedHealth Group | UNH US Equity | health care |

References

- [1] Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance* 55(1): 225-264.
- [2] Brown, S.J. (1976). Optimal portfolio choice under uncertainty: a Bayesian approach. Ph.D. thesis, University of Chicago.
- [3] Brown, S.J. (1978). The portfolio choice problem: comparison of certainty equivalence and optimal Bayes portfolios. *Communications in Statistics - Simulation and Computation* 7(4): 321-334.
- [4] Choueifaty, Y., and Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management* 35(1): 4051.
- [5] Clarke, R., de Silva, H., and Thorley, S. (2013). Risk parity, maximum diversification, and minimum variance: an analytic perspective. *Journal of Portfolio Management* 39(3): 39-53.
- [6] El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics* 36(6): 2757-2790.
- [7] El Karoui, N. (2009). High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: risk underestimation. *Annals of Statistics* 38(6): 3487-3566.
- [8] El Karoui, N. (2010). On the realized risk of high-dimensional Markowitz portfolios. Preprint.
- [9] Frost, P.A., and Savarino, J.E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis* 21(3): 293-305.
- [10] Haff, L.R. (1979). An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis* 9(4): 531-544.
- [11] Jobson, J.D., and Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association* 75(371): 544-554.
- [12] Jobson, J.D., Korkie, B., and Ratti, V. (1979). Improved estimation for Markowitz portfolios using James-Stein type estimators. *Proceedings of the American Statistical Association, Business and Economics Statistics Section* 41: 279-284.
- [13] Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics* 29(2): 295-327.
- [14] Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21(3): 279-292.
- [15] Kan, R., and Zhou, G. (2007). Portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42(3): 621-656.

- [16] Kandel, S., and Stambaugh, R.F. (1996). On the predictability of stock returns: an asset-allocation perspective. *Journal of Finance* 51(2): 385-424.
- [17] Klein, R.W., and Bawa V.S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics* 3(3): 215-231.
- [18] Ledoit, O., and Wolf, M. (2004). A well conditioned estimator for large dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2): 365-411.
- [19] Ledoit, O., and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics* 40(2): 1024-1060.
- [20] Litterman, B. (2003). *Modern Investment Management: An Equilibrium Approach*. Wiley.
- [21] Maillard, S., Roncalli, T., and Teiletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management* 36(4): 60-70.
- [22] Marchenko, V.A., and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik* 72(114): 507-536.
- [23] Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press.
- [24] Markowitz, H.M. (1952). Mean-variance analysis in portfolio choice and capital markets. *Journal of Finance* 7(1): 77-91.
- [25] McNeil, A.J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press.
- [26] Menchero, J., Wang, J., and Orr, D.J. (2011). Optimization bias adjustment: improving risk forecasts using eigen-adjusted covariance matrices. MSCI Barra Research Insight.
- [27] Meucci, A. (2007). *Risk and Asset Allocation*. Springer.
- [28] Michaud, R. (1989). The Markowitz optimization enigma: is “optimized” optimal? *Financial Analysts Journal* 45(1): 31-42.
- [29] Mori, H. (2004). Finite sample properties of estimators for the optimal portfolio weights. *Japan Statistical Society* 34(1): 27-46.
- [30] Pástor, L. (2000). Portfolio selection and asset pricing models. *Journal of Finance* 55(1): 179-223.
- [31] Pástor, L., and Stambaugh, R.F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics* 56(3): 335-381.
- [32] Shepard, P.G. (2009). Second order risk. Preprint on arXiv.
- [33] Ter Horst, J.R., De Roon, F.A., and Werker B.J.M. (2004). An alternative approach to estimation risk. Preprint available on <http://citeseerx.ist.psu.edu>.

- [34] Tu, J., and Zhou, G. (2004). Data-generating process uncertainty: what difference does it make in portfolio decisions? *Journal of Financial Economics* 72(2): 385-421.
- [35] Xia, L. (2001). Learning about predictability: the effect of parameter uncertainty on dynamic asset allocation. *Journal of Finance* 56(1): 205-246.