UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-15 (10)
Statistics and Econometrics Series
May 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

CALIBRATION OF SHRINKAGE ESTIMATORS FOR PORTFOLIO OPTIMIZATION

Victor DeMiguel, Alberto Martín-Utrera and Franciso J. Nogales*

Abstract

Shrinkage estimators is an area widely studied in statistics. In this paper, we contemplate the role of shrinkage estimators on the construction of the investor's portfolio. We study the performance of shrinking the sample moments to estimate portfolio weights as well as the performance of shrinking the naive sample portfolio weights themselves. We provide a theoretical and empirical analysis of different new methods to calibrate shrinkage estimators within portfolio optimization.

Keywords: Portfolio choice, estimation error, shrinkage estimators, smoothed bootstrap.

* DeMiguel is from London Business School and may be contacted at avmiguel@london.edu. Martin-Utrera and Nogales are both from Universidad Carlos III de Madrid and may be contacted at amutrera@est-econ.uc3m.es and FcoJavier.Nogales@uc3m.es. This work is supported by the Spanish Government through the project MTM2010-16519.

Calibration of shrinkage estimators for portfolio optimization*

Victor DeMiguel Alberto Martín-Utrera Franciso J. Nogales

05 of May, 2011

Abstract

Shrinkage estimators is an area widely studied in statistics. In this paper, we contemplate the role of shrinkage estimators on the construction of the investor's portfolio. We study the performance of shrinking the sample moments to estimate portfolio weights as well as the performance of shrinking the naive sample portfolio weights themselves. We provide a theoretical and empirical analysis of different new methods to calibrate shrinkage estimators within portfolio optimization.

Keywords: Portfolio choice, estimation error, shrinkage estimators, smoothed bootstrap.

*DeMiguel is from London Business School and may be contacted at avmiguel@london.edu. Martin-Utrera and Nogales are both from Universidad Carlos III de Madrid and may be contacted at amutrera@est-econ.uc3m.es and FcoJavier.Nogales@uc3m.es. This work is supported by the Spanish Government through the project MTM2010-16519.

1 Introduction

The classic mean-variance approach to portfolio optimization proposed by Markowitz (1952) constitutes a breakthrough in finance. This is the first mathematical model that formalizes the concept of diversification. This model allows investors to attain lower risk for a pre-specified level of return. But it also requires the knowledge of the distributional moments of stock returns, which are usually unknown. Therefore, they have to be estimated with sample information via maximum likelihood estimation. Usually, maximum likelihood estimators carry a lot of estimation error, making the investor's choice problem even more uncertain. In fact, the classical mean-variance model turns out to offer estimated portfolios below the true efficient frontier (see Jobson and Korkie (1981)). The effects of estimation error is one of the main drawbacks that practitioners have found in portfolio optimization and it has been one of the most active areas of the last decades. To study this issue, many authors have study the effect of estimation error in practice. For instance, Best and Grauer (1991) show that mean-variance portfolios are quite sensitive to small changes on the vector of means. Britten-Jones (1999) proposes a statistical procedure to test mean-variance portfolios. He showed that the sampling error of an international portfolio formed by 11 countries is large. DeMiguel et al. (2009) show that an "inefficient" equally weighted portfolio might outperform estimated optimal portfolios due to estimation error. To address the issue of estimation error, the researcher might attack either the inference about the inputs that form the portfolio weights or the portfolio weights themselves. In this paper we focus on the special case of shrinkage estimators, which arise from the well known literature of James-Stein estimators (see James and Stein (1961)).

As pointed out by Ledoit and Wolf (2003), maximum likelihood estimators are "the most likely parameter values given the data. In other words: let the data speak (and only the data)". Therefore, when the data is not enough to estimate, maximum likelihood might result very cumbersome and erratic. At this point, shrinkage estimators arise as a potential alternative to maximum likelihood estimation. In general, these methods dominate classical estimators in terms of quadratic loss. Moreover, shrinkage estimators have been shown to perform much better than naive estimators in the out-of-sample analysis (see Brandt (2004) for a revision of the literature).

Shrinkage estimators have been widely studied to diminish the estimation error on the estimation of the moments of asset returns. Frost and Savarino (1986) estimate the vector of means with the mean of a posterior density defined by a Normal-Wishart conjugate prior. Jorion (1986) proposes an empirical Bayes-Stein estimator for the vector of means. The author estimates it as the mean of the posterior density function defined by an informative prior which belongs to the class of exponential families. Both works, Frost and Savarino

(1986) and Jorion (1986), hold the underlying concept of inadmissibility of the sample mean (see Stein (1956)). More studies on the estimation of the vector of means based on shrinkage estimators have been made by Jobson et al. (1979) and Jorion (1985). In this paper, we propose a new class of shrinkage estimator for the vector of means constructed as a convex combination between the sample mean and a and a given target. The convex parameter or shrinkage intensity is computed to minimize a quadratic loss function.

Although some work has been devoted to study the estimation error in the vector of means, since Merton (1980) it is well known that variances are more stable along time and therefore, they are easier to be estimated. This issue prompted many investment managers to drop the vector of means out of their models. But even covariance matrices are deeply affected by estimation error, in particular in high-dimensional data sets. To deal with this problem, practitioners have implemented portfolio constraints which help to obtain better out-of-sample performance. Until Jagannathan and Ma (2003), it was not clear why portfolio constraints improved the performance of the minimum-variance portfolios. These authors show that implementing portfolio constraints is equivalent to solve an unconstrained minimum-variance portfolio problem where the covariance matrix is shrunk by the lagrange multipliers of the constrained problem. The literature of shrinkage estimators for the covariance matrix is one of the main practical streams in portfolio optimization. Ledoit and Wolf (2003) propose a shrinkage estimator for the covariance matrix which is a weighted average of the sample covariance matrix and a single-index covariance matrix implied by the market factor model (see Sharpe (1963)). In the same manner, Ledoit and Wolf (2004a) propose a shrinkage estimator for the covariance matrix where the sample covariance matrix is shrunk towards the identity matrix. They show that the resulting matrix is well-conditioned, even if the sample covariance matrix is not. The shrinkage intensity given in Ledoit and Wolf (2004a) is obtained under asymptotic results. On the other hand, DeMiguel et al. (2009) introduce a new class of portfolio constraints in the minimum-variance portfolio problem, where the norm of portfolio weights must be less or equal than a given threshold. They show that these constraints have a shrinkage effect over the sample covariance matrix.

The main motivation of using shrinkage estimators is to reduce the estimation error. When dealing with naive estimators (i.e. maximum likelihood estimators), optimal estimated portfolio weights might be very damaged by the estimation error. The first work where we find analytical expressions defining the estimation error of portfolio weights goes back to Jobson and Korkie (1980). In this work, the authors provide approximations to the distributional properties of the mean-variance portfolio weights. Moreover, they also provide approximations to the portfolio mean and the portfolio variance. The study of estimation error in this fashion is extended in the works of Okhrin and Schmid (2006) and Siegel and Woodgate (2007). Okhrin and Schmid (2006) obtain closed-form expressions for the exact moments of the mean-variance portfolio weights. They also provide explicit expressions for the multivariate density of the global minimum-variance portfolio. Siegel and Woodgate (2007) quantify the "over-optimism" given by mean-variance portfolios. They show that the portfolio mean tends to be biased upwards while the portfolio variance tends to be biased downwards. They provide some adjustments to correct for the "over-optimism" to better understand the poor out-of-sample performance of mean-variance portfolios. Kan and Smith (2008) provide finite-sample expressions for the minimum-variance frontier.

Shrinkage methods for portfolio weights and the impact of estimation error have been studied in the literature as well. Kan and Zhou (2007) studied a three-fund portfolio where the optimal asset allocation is a combination of the mean-variance portfolio, the minimum variance portfolio and a risk-free asset. They show that an optimal combination of the three funds diminishes the effect of estimation error in the investor's utility function. DeMiguel et al. (2009) study another type of three-fund rule where the minimum-variance portfolio is combined with an equally weighted portfolio. Tu and Zhou (2011) study the trade-off between optimal mean-variance portfolios and non-contaminated equally weighted portfolios[1]. The authors show that there exist an optimal trade-off between risky portfolios and an equally weighted portfolio such that the equally weighted portfolio and the considered mean-variance portfolios can be outperformed by this combination.

In this paper, we study shrinkage estimators for the vector of means, the covariance matrix and portfolio weights themselves. Throughout the whole paper, the basic assumption is that returns are independent and identically distributed with unknown vector of means μ and covariance matrix Σ. [2]. The independence assumption is required to calibrate the shrinkage estimators.

We consider a data set R of $N \geq 2$ assets over a sample of T observations, where each observation $R_t \in \mathbb{R}^N$ is independent and identically distributed. The naive estimator of the vector of means is $\widehat{\mu} = (1/T)\sum_{t=1}^{T} R_t$ and the naive estimator of the covariance matrix is $\widehat{\Sigma} = (1/(T-1))\sum_{t=1}^{T}(R_t - \widehat{\mu})(R_t - \widehat{\mu})'$.

In the classical mean-variance formulation, the objective is to maximize a investor's quadratic utility which is defined as the trade-off between portfolio return and portfolio variance. Hence, the estimation of the population moments μ and Σ is quite relevant for

[1]The term non-contaminated stands for the case case where estimation of moments is not required to construct the portfolio, which is the case of the equally weighted portfolio. Due to the absence of estimation error, this portfolio can be seen as a non-contaminated portfolio.

[2]The independence assumption is very strong with high frequency data like weekly, daily or tick by tick data. In the empirical analysis we work with monthly data where the independence assumption is less unrealistic.

the investor's choice. The population optimal mean-variance portfolio is $w = \frac{1}{\gamma}\Sigma^{-1}\mu$ and its naive estimator counterpart is $\widehat{w}_{mv} = \frac{(T-N-2)}{(T-1)}\frac{1}{\gamma}\widehat{\Sigma}^{-1}\widehat{\mu}$[3], where γ is the investor's risk aversion level. In this portfolio, investors can allocate/borrow part of their budget in/from the risk free asset. Throughout the paper, we assume that there is no risk free asset, therefore we normalize the estimated optimal mean-variance portfolio obtaining the tangency portfolio (TP) $\widehat{w}_{TP} = \frac{\widehat{\Sigma}^{-1}\widehat{\mu}}{\iota'\widehat{\Sigma}^{-1}\widehat{\mu}}$.

We also pointed out that many investors were prompted to drop the vector of means out of their models since Merton (1980). In this case, the objective is to minimize the portfolio variance. The population global minimum-variance (GMV) portfolio is defined as $w_{GMV} = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}$, and its naive estimator counterpart is $\widehat{w}_{GMV} = \frac{\widehat{\Sigma}^{-1}\iota}{\iota'\widehat{\Sigma}^{-1}\iota}$, where ι is a proper vector of ones. This portfolio adds up to one because of the denominator $\iota'\Sigma^{-1}\iota$, which normalizes the element $\Sigma^{-1}\iota$. In the following sections, whenever we mention the estimated minimum-variance portfolio, we refer to the non-normalized element, which is $\widehat{w}_{min} = \frac{(T-N-2)}{(T-1)}\widehat{\Sigma}^{-1}\iota$. In general, and for notational simplicity, we set $\widehat{x}$ as the estimator of the random variable x. This statement holds for the following sections.

Our contribution to the literature is threefold: first, we propose new procedures to obtain the shrinkage intensities of the shrinkage estimators for the vector of means and the covariance matrix. We obtain closed form expressions of the shrinkage intensities under the assumption of independent and normally distributed asset returns. We find a more robust calibration method for the shrinkage vector of means, which gives better empirical results than the bayes-stein shrinkage estimator for the vector of means proposed by Jorion (1986). We also propose two different methods to obtain the shrinkage intensities of the shrinkage covariance matrix. In the first one, we obtain a closed form expression which allows us to understand its dependence with the number of assets and observations under consideration. In the second, we come up with a calibration method that accounts explicitly for the condition number of the covariance matrix. As we will show in the empirical analysis, this procedure is very convenient in high-dimensional data sets. Furthermore, we propose a new class of shrinkage estimator over the inverse covariance matrix.

Second, we make an extensive empirical study of four different calibration criteria for shrinking portfolio weights. We apply each calibration method through three different shrinkage portfolios across six different real data sets. Moreover, we account for the impact of transaction costs. To the best of our knowledge, this is the first paper that makes such a empirical analysis in the study of shrinkage portfolios. After adjusting for transaction costs, we find that the expected portfolio variance minimization technique is more robust

[3]We multiply the sample mean-variance portfolio $\overline{w}_{mv} = \frac{1}{\gamma}\widehat{\Sigma}^{-1}\widehat{\mu}$ by a factor $\frac{(T-N-2)}{(T-1)}$ to correct the bias implied by the inverse of $\widehat{\Sigma}$ when returns are assumed to be normally distributed.

across every data set in terms of Sharpe ratio. Moreover, we also observe that when the shrinkage portfolio does not consider the vector of means, a mean-squared loss criteria to calibrate the shrinkage portfolio is very convenient in terms of Sharpe ratio and turnover.

Third, we propose two different methods to calibrate shrinkage portfolios. In the first one, we assume that asset returns are independent and normally distributed (i.e. $N(\mu, \Sigma)$). This assumption allows us the obtain closed form expressions of the optimal shrinkage intensities in general. These closed form expressions allows us to study the dependence of the shrinkage intensities with respect to the number of assets and observations under consideration. The second method to calibrate shrinkage portfolios does not make any assumption about the distribution of asset returns. In this case, the optimal shrinkage intensities are obtained via smoothed bootstrap. We show that this calibration method gives good results in high-dimensional data sets.

The paper is organized as follows. Section 2 study and extend the analysis of shrinkage estimators of the moments of stock returns. Section 3 study the mixture of portfolios, understood as an alternative way of shrinking portfolio weights, and their calibration. In Section 4, we make an extended empirical analysis of shrinkage estimators. Section 5 concludes.

2 Shrinkage estimators of moments

We first address the problem of estimation error within the estimation of moments. This is a natural approach to deal with the poor performance of classical methods in portfolio optimization. Because of estimation error within naive sample moments[4], portfolio weights are highly contaminated and this is why classical methods perform badly in real life. Throughout this section, we study shrinkage estimators which try to alleviate the effect of estimation error of the naive sample moments. We closely follow the shrinkage framework described in Ledoit and Wolf (2004a), where the naive estimator is shrunk toward a target element. The shrinkage intensity is chosen to be the optimal shrinkage that minimizes the mean-squared loss of the shrinkage estimator with respect to the true value. Hence, if we want to estimate some population property μ_x of a random variable x, such that $\widehat{\mu}_x$ is the naive estimator, we define its mean-squared loss as follows:

$$MSE\left(\widehat{\mu}_x\right) = E\left(\left(\widehat{\mu}_x - \mu_x\right)^2\right) = \underbrace{E\left(\left(\widehat{\mu}_x - E\left(\widehat{\mu}_x\right)\right)^2\right)}_{\text{Variance}} + \underbrace{\left(\mu_x - E\left(\widehat{\mu}_x\right)\right)^2}_{\text{Squared Bias}}. \quad (1)$$

[4]For the rest of the paper, naive estimator are understood as maximum likelihood estimators.

Therefore, our criteria to select the shrinkage intensity is the expected mean-squared loss. This section deals with the estimation of the moments of stock returns. To diminish the estimation error, we study shrinkage estimators for the vector of means and the covariance matrix. In Section 2.1, we derive a shrinkage estimator for the vector of means. In Section 2.2 we derive a shrinkage estimator for the covariance matrix. We also derive a shrinkage estimator for the inverse covariance matrix. Throughout this section and Section 3, we understand shrinkage estimators as a convex combination of some naive sample estimator and a target element:

$$\text{Shrinkage Estimator} = (1-\alpha) \times \text{Sample Estimator} + \alpha \times \text{Target}, \tag{2}$$

where α is the shrinkage intensity. This parameter determines the "strength" under which the naive sample estimator is shrunk towards the target element. This value is between zero and one. Therefore, the maximum "strength" is one, being the shrinkage estimator equal to the target element. If α is zero, the "strength" is zero and therefore the shrinkage estimator corresponds with the naive sample estimator. The optimal shrinkage intensity α is defined by a mean-squared loss criteria.

2.1 Estimation of the vector of means

Shrinkage rules to diminish the estimation error of the vector of means have been a matter of study in the literature. Frost and Savarino (1986) assume an informative prior where all stocks have the same expected values, variance and covariances, such that the predictive mean turns out to be a weighted average of the sample mean and a grand mean equal to the historical average return for all stocks. Jorion (1986) estimates the vector of means by integrating a predictive density function defined by an exponential prior which is only specified for the vector of means. DeMiguel et al. (2009) study the performance of James-Stein shrinkage estimators for the mean, defined as a weighted average of the naive sample mean $\widehat{\mu}$ and the portfolio mean of the minimum-variance portfolio. The James-Stein shrinkage intensity parameter is in general expressed as follows:

$$\alpha_{JS} = \frac{N+2}{(N+2) + T\left(\mu - \mu^{\text{target}}\right)\Sigma^{-1}\left(\mu - \mu^{\text{target}}\right)'}, \tag{3}$$

where μ^{target} is a target vector. Expression (3) is estimated with sample information to construct the shrinkage estimator $\widehat{\mu}_{JS} = (1-\widehat{\alpha}_{JS})\widehat{\mu} + \widehat{\alpha}_{JS}\mu^{\text{target}}$. This estimator belongs to the class of estimators with lower expected quadratic loss than the usual sample estimator $\widehat{\mu}$ (see James and Stein (1961)).

We propose another type of shrinkage estimator where the mean is also a weighted average of the sample vector of means and the target $\iota = \mathbb{1} \in \mathbb{R}^N$, scaled by a scalar factor ν. The scale factor ν is introduced to approach the target to the true vector of means. Since the considered target has zero variance, its squared bias defines the mean-squared loss. Therefore, to control for the bias we include the scale factor ν. The optimal shrinkage intensity α and scale factor ν, are obtained by solving the following problem:

$$\min_{\alpha,\nu} \quad E\left[\|\widetilde{\mu} - \mu\|_2^2\right] \tag{4}$$

$$\text{s.t.} \quad \widetilde{\mu} = (1-\alpha)\widehat{\mu} + \alpha\nu\iota, \tag{5}$$

where $\|x\|_2^2 = \sum_{i=1}^N x_i^2$, α is the shrinkage intensity and ν is the scale factor. Problem (4) minimizes the expected squared loss of the estimated vector of means with respect to the true vector of means. The variables of the problem are the shrinkage intensity α and the scale factor ν. The constraint (5) shows that the estimated vector of means $\widetilde{\mu}$ is a convex combination of the sample vector of means with the scaled target ι. We choose the target $\iota = \mathbb{1} \in \mathbb{R}^N$ because when the shrinkage intensity is equal to one, the solution of the estimated mean-variance portfolio with our shrinkage estimator for the vector of means would be the minimum-variance portfolio, which is known to perform better than the mean-variance portfolio due to the estimation error (see Merton (1980)). On the other hand, introducing constraint (5) in problem (4), we have that:

$$\min_{\alpha,\nu} \quad E\left[\|\widetilde{\mu} - \mu\|_2^2\right] = (1-\alpha)^2 E\left[\|\widehat{\mu} - \mu\|_2^2\right] + \alpha^2 \left\|\nu\iota - \mu\right\|^2. \tag{6}$$

We observe that the optimal scale parameter ν is $\nu^* = \text{argmin}_\nu \left\|\nu\iota - \mu\right\|^2$, such that the squared bias of the scaled target $\nu\iota$ is minimized. The following proposition gives the optimal values for α and ν.

Proposition 2.1. *Assuming R_t are independent and normally distributed, the optimal finite shrinkage parameter α and the optimal scale factor ν of problem (4) are:*

$$\nu_\mu^* = \overline{\mu}, \tag{7}$$

and

$$\alpha_\mu^* = \frac{(1/T)\overline{\sigma^2}}{(1/T)\overline{\sigma^2} + \overline{\mu}_2 - \overline{\mu}^2}, \tag{8}$$

where $\overline{\mu}$ is the scalar average of the vector of means, $\overline{\mu}_2 = (1/N)\mu'\mu$, and $\overline{\sigma^2}$ is the average of the population variances.

Proof. The proof of Proposition 2.1 is given in the appendix. □

The optimal shrinkage intensity α_μ^* for the vector of means shows some interesting properties. First, we see that when $T \to \infty$, the optimal shrinkage intensity converges to zero. This confirms that the shrinkage estimator holds with the Glivenko-Cantelli theorem[5]. Second, from formula (8), we observe that the larger the average population variance, the closer the shrinkage intensity is to one. Third, it is important to point out that the optimal shrinkage intensity does not depend on the number of assets. The advantage of our estimator is twofold: first we scale the target such that the bias of the target is reduced. Second, the computation of the shrinkage intensity does not involve the inverse covariance matrix of asset returns Σ^{-1}. This is an advantage because when estimating the shrinkage intensity, Σ^{-1} might be extremely large and very erratic if the condition number[6] of Σ tends to infinite. It might happen when N is large and T is relatively small. When this happens, we usually refer to that matrix as an ill-conditioned matrix. To control for this issue, we may shrink the covariance matrix as well. In the following subsection, we deal with shrinkage estimators for the covariance matrix.

2.2 Estimation of the covariance matrix

In this section, we study shrinkage estimators for the covariance matrix. There is a vast literature of shrinkage estimator for the covariance matrix. In particular, we focus on the shrinkage estimators which define the shrinkage intensity by a mean-squared loss criteria. The general framework is based on a convex combination between a naive estimator of the covariance matrix and a target matrix that imposes some structure on the naive estimator. Ledoit and Wolf (2003) propose as a target matrix the single-index covariance matrix implied by a market factor model. On the other hand, Ledoit and Wolf (2004b) propose a target defined by a "constant correlation matrix", where all the correlations are equal to the average of all the sample correlations. Ledoit and Wolf (2004a) shrink the naive sample covariance matrix toward a scaled identity matrix. Due to the high bias of the identity matrix, it is scaled by a factor ν, imposing the structure that all the variances are equal to ν and the covariances are zero. Likewise, we estimate the covariance matrix as an optimal convex combination between the sample covariance matrix and a scaled identity matrix. We focus on this shrinkage method because the target is deterministic, i.e. its variance is zero.

[5]The Glivenko-Cantelli theorem states that the sample probability function converges to the population probability function when the number of sample observations tends to infinite (see Shorack and Wellner (1986)).

[6]The condition number is the ratio between the largest eigenvalue and the smallest eigenvalue. Big condition numbers are usually referred to define unstable systems of equations in linear algebra.

Ledoit and Wolf (2004a) obtain an asymptotic optimal shrinkage intensity α assuming that observations are independent and identically distributed. We instead assume normality to obtain finite closed-form expressions for the optimal shrinkage intensity and the scale factor. Under these assumptions, we can observe how the shrinkage intensity evolves with the number of observations. Ledoit and Wolf (2004a) propose the following minimization problem:

$$\min_{\alpha,\nu} \quad E\left[\left\|\widetilde{\Sigma} - \Sigma\right\|_F^2\right] \tag{9}$$

$$\text{s.t.} \quad \widetilde{\Sigma} = (1-\alpha)\widehat{\Sigma} + \alpha\nu I, \tag{10}$$

where $\|X\|_F^2 = \text{tr}(X'X)$, $\text{tr}(\cdot)$ is the trace operator, α is the shrinkage intensity and ν is the scale factor. Note in (9) that we are minimizing the expected quadratic loss of the estimator with respect to the true moment like it was made in (4).

Ledoit and Wolf (2004a) assume that the observations are independent and identically distributed, but they do not assume any particular distribution. They show that the optimal shrinkage intensity can be asymptotically estimated as follows:

$$\alpha_a = \frac{\frac{1}{T^2}\sum_{t=1}^{T}\left\|R_t'R_t - \widehat{\Sigma}\right\|_F^2}{\left\|\widehat{\Sigma} - mI\right\|_F^2}, \tag{11}$$

where $m = \text{tr}\left(\widehat{\Sigma}\right)/N$.

The challenge of shrinkage estimators is to obtain an optimal shrinkage intensity with statistical meaning. We contribute to the literature defining a finite closed-form expression for the optimal shrinkage intensity when asset returns are independent and normally distributed. This expression allows us to understand how the optimal shrinkage intensity is affected by the sample window length. Moreover, we can also observe how the optimal shrinkage intensity might be affected by the number of assets under consideration. Therefore, our contribution is to give a closed form expression which help us to understand better the optimality of the shrinkage intensity. Thus, the finite optimal values that minimize problem (9) are defined in the following proposition.

Proposition 2.2. *When returns are independent and normally distributed, the optimal target parameter and the optimal finite shrinkage intensity parameter under a quadratic loss criterion for the covariance matrix are:*

$$\nu_\Sigma^* = \overline{\sigma^2}, \tag{12}$$

and

$$\alpha_{\Sigma}^{*} = \frac{\frac{N}{T-1}\left(\frac{tr\left(\Sigma^{2}\right)}{N} + N\left(\overline{\sigma^{2}}\right)^{2}\right)}{\frac{N}{T-1}\left(\frac{T}{N}tr\left(\Sigma^{2}\right) - (T-N-1)\left(\overline{\sigma^{2}}\right)^{2}\right)}. \tag{13}$$

Proof. The proof of Proposition 2.2 is given in the appendix. □

From formula (13), we observe that if $T \rightarrow \infty$, the optimal shrinkage intensity converges to zero. It confirms that the shrinkage estimator for the covariance matrix is sample-dependent. Thus, when T converges to infinite, according with the Glivenko-Cantelli theorem, the sample covariance matrix becomes the population covariance matrix. Moreover, from formula (13), we also observe that the larger the number of assets, the closer the shrinkage intensity is to one. Unlike the vector of means, it is very important the number of assets on the estimation of the covariance matrix. This is because for the covariance matrix, the number of parameters to be estimated grows quadratically with the number of assets[7].

2.3 Estimation of the inverse covariance matrix

We also propose an shrinkage estimator for the estimated inverse covariance matrix. In portfolio optimization, the inverse of the covariance matrix is rather important. This issue has a bigger effect when we are dealing with many assets, where inverting a covariance matrix might be very cumbersome due to the condition number. In fact, when the number of assets N is approximately equal to the number of observations T, the covariance matrix tends to be singular. In this situation, the estimation error of the naive estimator for the inverse covariance matrix explodes to infinity. Likewise, we develop another shrinkage estimator for the inverse of the covariance matrix. The motivation of this part is to understand the properties of the shrinkage intensity and how it is affected by the number of assets N and the sample window length T. This insight is important to understand better the risk implied by the naive estimator of the inverse covariance matrix. We keep the mean-squared loss framework to calibrate the shrinkage intensity in the estimation of the

[7]In fact, for N assets, the number of parameters to estimate the covariance matrix is $\frac{N\times(N+1)}{2}$.

inverse covariance matrix[8]:

$$\min_{\alpha,\nu} \quad E\left(\left\|\widetilde{\Sigma}^{-1} - \Sigma^{-1}\right\|_F^2\right) \tag{14}$$

$$\text{s.t.} \quad \widetilde{\Sigma}^{-1} = (1-\alpha)\widehat{\Sigma}_u^{-1} + \alpha\nu I. \tag{15}$$

Problem (14) minimizes the mean-squared loss of the estimated shrinkage inverse covariance matrix with respect to the true inverse covariance matrix. The shrinkage estimator for the inverse covariance matrix is formed as a convex combination between the unbiased estimator of the inverse covariance matrix and a scaled identity matrix.

The following proposition gives the optimal values for the shrinkage intensity and the scale factor.

Proposition 2.3. *Assuming asset returns are independent and identically distributed, the optimal scale factor and the optimal finite shrinkage intensity under a quadratic loss criterion for the inverse covariance matrix are:*

$$\nu^*_{\Sigma^{-1}} = \overline{\sigma^{-2}}, \tag{16}$$

and

$$\alpha^*_{\Sigma^{-1}} = \frac{tr\,(\Omega) - tr\,(\Sigma^{-1}\Sigma^{-1})}{tr\,(\Omega) - \left(\overline{\sigma^{-2}}\right)^2 N}, \tag{17}$$

where $\Omega = \frac{(T-N-2)}{(T-N-1)(T-N-4)}\left(tr\,(\Sigma^{-1})\,\Sigma^{-1} + (T-N-2)\Sigma^{-2}\right)$*, and* $\overline{\sigma^{-2}}$ *is the average of the inverse population variances.*

Proof. The proof of Proposition 2.3 is given in the appendix. □

Proposition 2.3 gives the optimal values for the shrinkage intensity α and the scale factor ν. We observe that the number of assets N has a deep impact on the expression of the optimal shrinkage intensity. In fact, the larger the number of assets, the smaller the denominator of Ω, and hence the larger the value of the shrinkage intensity.

[8]To simplify the computations, the sample estimator of the covariance matrix is multiplied by the factor $(T-1)/(T-N-2)$, so that the estimated inverse covariance matrix is unbiased $\widehat{\Sigma}_u = \frac{T-1}{T-N-2}\widehat{\Sigma}$.

2.4 Estimation of the covariance matrix considering the condition number

In this section, we propose an alternative procedure to calculate the shrinkage intensity. With this method we pretend to account for the mean-squared loss of the estimated covariance matrix and its condition number. The way in which we account for both features is by two measures. First, we consider the *relative improvement in average loss* (RIAL) (see Ledoit and Wolf (2004a)) of the shrinkage covariance matrix $\widetilde{\Sigma} = (1-\alpha)\widehat{\Sigma} + \alpha\nu I$:

$$RIAL\left(\widetilde{\Sigma}\right) = \frac{E\left(\left\|\widehat{\Sigma} - \Sigma\right\|_F^2\right) - E\left(\left\|\widetilde{\Sigma} - \Sigma\right\|_F^2\right)}{E\left(\left\|\widehat{\Sigma} - \Sigma\right\|_F^2\right)}. \tag{18}$$

RIAL is a performance measure of the shrinkage covariance matrix $\widetilde{\Sigma}$. This measure is bounded above, being one its maximum value. This is attained when the mean-squared loss is negligible relative to the mean-squared loss of the naive sample estimator of the covariance matrix $\widehat{\Sigma}$. Therefore, the ideal shrinkage estimator would give a RIAL close to one. Assuming normality, we have closed form expressions for $RIAL\left(\widetilde{\Sigma}\right)$. The closed form expressions that form the $RIAL\left(\widetilde{\Sigma}\right)$ are given in the Appendix.

On the other hand, the second element that we consider to compute the shrinkage intensity is the condition number of the shrinkage covariance matrix:

$$\delta_{\widetilde{\Sigma}} = \frac{(1-\alpha)\lambda_{\text{max}} + \alpha 1}{(1-\alpha)\lambda_{\text{min}} + \alpha 1}, \tag{19}$$

where $\delta_{\widetilde{\Sigma}}$ is the condition number of the shrinkage covariance matrix, λ_{max} is the maximum eigenvalue of the naive sample covariance matrix $\widehat{\Sigma}$ and λ_{min} is the minimum eigenvalue of the naive sample covariance matrix $\widehat{\Sigma}$. Since we are shrinking the naive sample covariance matrix to the identity matrix, all the eigenvalues of the identity are one and that is why α is multiplied by one. Ideally, the smallest condition number is one, which is attained when α is zero. In that case our shrinkage covariance matrix would be the identity.

Therefore, we propose the following problem to find an optimal shrinkage intensity that accounts both for the mean-squared loss and the condition number of the shrinkage covariance matrix:

$$\alpha = \text{argmin}\left\{\delta_{\widetilde{\Sigma}} - RIAL\left(\widetilde{\Sigma}\right)\right\}. \tag{20}$$

The above formulation is ideally solved when the true covariance matrix is the identity matrix (i.e. $\Sigma = I$). In that case, the shrinkage intensity would be one and therefore, the value of the objective function in (20) would be zero. The RIAL is introduced in negative because the real objective is to maximize the RIAL as much as possible, while the objective of the condition number is to minimize it as much as possible. Since problem (20) is a very nonlinear optimization problem, it is difficult to obtain a closed form solution. Instead, we solve the problem with numerical methods.

2.5 Discussion of the shrinkage estimators

In this section, we illustrate the theoretical findings of the previous sections with a simulation study. Our population probability distribution is defined by a data set formed with 48 industry portfolios. Then, the sample moments of this data set constitute the population moments of the multivariate normal distribution that governs asset returns, i.e. $N(\mu, \Sigma)$. We use the population moments to compute the true optimal shrinkage intensities and the theoretical values of the mean-squared losses of the shrinkage moments. The objective of the experiment is to measure the dependence of the shrinkage intensity and the mean squared loss of shrinkage estimators with the sample window length T.

Figure 1 shows the evolution of the shrinkage intensities for the estimation of the mean vector of returns, the covariance matrix and the inverse covariance matrix. The sample moments of the data set formed by 48 industry portfolios define the multivariate normal distribution required to compute the theoretical values. Figure 1a shows the evolution of the shrinkage intensities for the estimator of the vector of means. In general, we observe that our proposed shrinkage estimator is more conservative than the classical James-Stein estimator. Our shrinkage estimator contains a higher shrinkage intensity than the James-Stein estimator for every window's length T. Then, we can say that it is more conservative since the target is more weighted under our new shrinkage estimator. This might be due to the fact that the James-Stein estimator utilizes the inverse of the covariance matrix in the computation of the shrinkage intensity. It might give erratic values if the condition number of the covariance matrix is large.

Figure 1b shows the shrinkage intensities of the asymptotic shrinkage estimator for the covariance matrix (solid line) the finite-shrinkage estimator for the covariance matrix (dashed line), the finite-shrinkage estimator for the covariance matrix that accounts for the condition number (dotted line) and the finite-shrinkage estimator for the inverse covariance matrix (dot-dashed line). This plot shows that the finite-shrinkage estimator for the covariance matrix shrinks less than the finite-shrinkage estimator for the inverse covariance matrix. This fact suggests that the inverse of the covariance matrix has an additional source of estimation error and therefore, it has to be shrunk to the scaled identity matrix by a larger quantity. The main problem of shrinking the inverse matrix is that for medium-large data sets it gives too much weight to the scaled identity matrix, loosing some information about the covariances among assets.

Figure 2 shows the results of the experiment made to compute the theoretical mean-squared loss[9] of the shrinkage moments $(\widetilde{\mu}, \widetilde{\Sigma}, \widetilde{\Sigma}^{-1})$. Figure 2a depicts the mean-squared loss for the vector of means when it is computed by the James-Stein estimator (dashed line) given in (3) and the finite shrinkage estimator (dot-dashed line) given by problem (4)-(5). Since the shrinkage estimator minimizes the mean-squared loss, it is obvious that the mean-squared loss of this estimator will be lower than the mean-squared loss of the James-Stein estimator. Figure 2b shows the mean-squared loss of the finite shrinkage estimator for the covariance matrix (dashed line), finite shrinkage estimator for the covariance matrix that accounts for the condition number (dotted line) and the asymptotic shrinkage estimator for the covariance matrix (dot-dashed line). We can observe that the finite and asymptotic shrinkage estimators offer very similar losses and both of them are very small. Finally, Figure 2c depicts the mean-squared loss of the finite-shrinkage estimator for the inverse covariance matrix (dot-dashed line) given by problem (14)-(15). We have also computed the mean-squared loss of the inverse for the shrinkage covariance matrices obtained by the asymptotic shrinkage estimator (solid line), the finite shrinkage estimator (dashed line) and the finite shrinkage estimator that accounts for the condition number (dotted line). The mean-squared loss of the last three estimators cannot be computed theoretically[10], thus we make a Monte Carlo simulation to compute the mean-squared loss of the estimated inverse covariance matrix. Then, for each window's length T, we generate 1000 samples $R \in \mathbb{R}^{T\times N}$ from a multivariate normal distribution defined with the population moments obtained from the 48 industry portfolios. Then, we compute the sample mean-squared loss of the estimated inverse covariance matrix as follows:

$$\text{MSE}\left(\Sigma^{-1}\right) = \frac{\sum_{i=1}^{1000} \left\| \widetilde{\Sigma}_i^{-1} - \Sigma^{-1} \right\|_F^2}{1000}, \tag{21}$$

where $\widetilde{\Sigma}$ is the estimated shrinkage covariance matrix. Figure 2c shows that the mean-squared loss of the shrinkage inverse covariance matrix considered in problem (14)-(15) is lower than the mean-squared loss of the inverse matrix of the shrinkage estimators for

[9]In the case of the asymptotic shrinkage estimator, we compute the shrinkage intensity with the optimal asymptotic shrinkage intensity, but the mean-squared loss is computed assuming normality and a finite sample.

[10]It is because it is not clear which is the distribution of the estimator $\left((1-\alpha)\widehat{\Sigma} + \alpha\nu I\right)^{-1}$.

the covariance matrix (both the asymptotic and the finite shrinkage estimators). It means that shrinking first the covariance matrix and then invert it will give higher mean-squared loss than making the shrinkage directly over the inverse covariance matrix. The problem of shrinking the inverse covariance matrix is that the shrinkage intensity is high, as we can see from Figure 1b. It implies too much structure over the naive sample inverse covariance matrix, and it makes us loose information about the covariances among assets.

Figure 1b shows that the shrinkage intensity is very small for the finite shrinkage estimator. If the data set is big (say N>50), the condition number of the covariance matrix might be very extreme, so the shrinkage intensity might be insufficient to reduce the condition number. It generates that the inverse covariance matrix used to calculate optimal portfolios might give very unstable weights[11]. This is why we introduce the finite-shrinkage estimator that accounts for the condition number. In fact, this estimator gives lower mean-squared loss of the inverse covariance matrix for small samples as we observe in Figure 2c. It is because in small samples it is more likely that the condition number of the sample covariance matrix will be high. In that case it is important to shrink the covariance matrix accounting for the condition number.

To conclude, we summarize this section about shrinkage over the moments of stock returns in three points: first, we have obtained an alternative shrinkage estimator for the vector of means which offers a finite shrinkage intensity higher than the one given by the James-Stein estimator. Moreover, our shrinkage intensity does not depends on the inverse covariance matrix, as the James-Stein shrinkage intensity, which is specially good when the condition number of the covariance matrix of returns is large. Second, we observe from the simulation study that the shrinkage intensity for the shrinkage covariance matrix is small for medium-large data sets. This issue might be cumbersome when the condition number of the sample covariance matrix is large, fact that is likely with real data. We solve this problem by introducing a new shrinkage intensity which is optimally calculated to minimize the relative mean-squared error of the shrinkage covariance matrix and the condition number. Third, we observe that shrinking over the inverse covariance matrix gives too much weight to the scaled identity matrix. Too much structure is imposed over the naive sample inverse covariance matrix and all the information about the covariances might be lost.

[11]This means that little changes on the covariance matrix would generate very different results.

3 Shrinkage estimation of portfolio weights

In this section, we introduce an analytical study of shrinkage estimators of portfolio weights. From a statistical point of view, shrinkage portfolios can be understood as a linear combination of portfolios which contain smaller estimation error than naive portfolios. In the literature, combination of portfolios and its performance have been studied. Kan and Zhou (2007) propose a three-fund portfolio which combines the risk-free asset with the mean-variance portfolio and the minimum-variance portfolio. DeMiguel et al. (2009) propose a mixture of minimum-variance portfolio and equally weighted portfolio. Tu and Zhou (2011) propose a bunch of portfolio models constructed as an optimal convex combination of different mean-variance portfolios and the equally weighted portfolio. Our view of shrinkage estimators holds with the definition of shrinkage used in the previous section and it is related to the concept of combination of portfolios described in Tu and Zhou (2011). We consider a shrinkage portfolio as a weighted average or convex combination of some naive portfolio and a target portfolio. The target may be any portfolio, but in general it must satisfy two properties: first, it has to be a portfolio with low volatility of its weights (or even null volatility, like the 1/N rule). Second, it has to perform "well" in practice[12]. The convex combination of the naive portfolio with the target portfolio offers insightful information. The shrinkage intensity (i.e. the parameter which defines the convex combination) specifies the trade-off between the naive portfolio[13] and the target portfolio. This trade-off is optimal under some calibration criteria, which defines the value of the shrinkage intensity. Hence, the importance of the calibration criteria. Therefore, the amount of shrinkage is defined by the calibration method and it is crucial for the performance of the portfolio. Then, according with the framework described above, our definition of shrinkage portfolios is as follows:

$$\widehat{w}_{sp} = (1-\alpha)\widehat{w}_n + \alpha\nu w_{\text{target}}, \tag{22}$$

where $\widehat{w}_n$ is the naive portfolio estimator of the true portfolio w^*, w_{target} is the target portfolio, α is the shrinkage intensity and ν is a scale parameter. The scale parameter is the main difference between our methodology and the methodology of Tu and Zhou (2011). We introduce this parameter to correct for the bias of the target portfolio. Later, we go through this topic in more detail.

[12]The term "well" stands for portfolios which are well known to obtain good out-of-sample results, like the 1/N portfolio or the minimum-variance portfolio.

[13]Unlike DeMiguel et al. (2009), the meaning of naive stands for the sample portfolio estimator. Relying on maximum likelihood estimators might be very "naive" for finite samples.

We have pointed out that to fully understand the meaning of the shrinkage intensity, we have to understand the calibration method. In this paper, we calibrate shrinkage intensities from four different perspectives: expected out-of-sample utility maximization, mean-squared loss minimization, expected out-of-sample portfolio variance minimization and Sharpe ratio. The out-of-sample utility, or portfolio variance, is the utility, or portfolio variance, of the estimated portfolio $\widehat{w}$, given the true moments μ and Σ, that is $U(\widehat{w}) = \widehat{w}'\mu - (\gamma/2)\,\widehat{w}'\Sigma\widehat{w}$, or $\widehat{w}'\Sigma\widehat{w}$. From now on, and for the sake of simplicity, the expected out-of-sample utility maximization and the expected out-of-sample portfolio variance minimization will be named as expected utility maximization and expected variance minimization, respectively.

Thus, the following expressions define each calibration method:

$$\max_{\alpha} \quad E\left(f_{\text{sl}}\left(\widehat{w}_{sp}\right)\right) = \max_{\alpha} \quad E\left(\|\widehat{w}_{sp} - w^*\|_2^2\right), \tag{23}$$

$$\max_{\alpha} \quad E\left(f_{\text{ut}}\left(\widehat{w}_{sp}\right)\right) = \max_{\alpha} \quad E\left(\widehat{w}_{sp}'\mu - (\gamma/2)\,\widehat{w}_{sp}'\Sigma\widehat{w}_{sp}\right), \tag{24}$$

$$\max_{\alpha} \quad E\left(f_{\text{var}}\left(\widehat{w}_{sp}\right)\right) = \max_{\alpha} \quad E\left(\widehat{w}_{sp}'\Sigma\widehat{w}_{sp}\right), \tag{25}$$

$$\max_{\alpha} \quad E\left(f_{\text{SR}}\left(\widehat{w}_{sp}\right)\right) = \max_{\alpha} \quad \frac{E\left(\widehat{w}_{sp}'\mu\right)}{E\left(\widehat{w}_{sp}'\Sigma\widehat{w}_{sp}\right)}. \tag{26}$$

Equation (23) solves for the optimal shrinkage intensity α that minimizes the mean-squared loss. For this method, the shrinkage intensity will give us the optimal trade-off between the naive portfolio and the scaled target portfolio which minimizes the mean-squared loss. On the other hand, equation (24) solves for the optimal shrinkage intensity α that maximizes the expected utility. The shrinkage intensity obtained with this method gives the optimal trade-off between the naive portfolio and the scaled target portfolio in terms of expected utility, that is, it is the optimal trade-off that maximizes the expected utility. Equation (25) solves for the optimal shrinkage intensity α that minimizes the expected portfolio variance. Hence, the shrinkage intensity gives the trade-off between the naive portfolio and the scaled target portfolio which defines the combination that minimizes the portfolio variance of the combined portfolio. The last calibration method considered in equation (26) solves for the optimal trade-off between the naive portfolio and the scaled target portfolio that maximizes the Sharpe ratio.

In any good calibration method there exist an underlying concept about investor's preferences. Therefore, the most reasonable calibration methods are the expected utility maximization, the expected variance minimization and the Sharpe ratio maximization. The mean-squared loss minimization method does not represent any clear desire of investor's preferences. We decide to take into account this calibration method for two reasons. First, to be consistent with the mean-squared loss framework showed in Section 2 for the moments of asset returns. Second, and more important, because there exist an underlying investor's desire within this method. In fact, minimizing the mean-squared loss of any portfolio model is the same to minimizing the expected squared error. Therefore, assuming that returns are independent and identically distributed, the true optimal portfolio w^* is the same throughout time because the distribution of returns does not change. Therefore, each time that we calibrate our model, we are minimize the expected squared error. In this way, this calibration method penalizes big errors over small errors. Hence, penalizing big errors over small ones is a mechanism to control for the turnover of our portfolios. But as we have said, this reasoning is feasible when the true optimal portfolio w^* does not change throughout time. In general it is not true, but this might be the case when the vector of means is not considered such that the combination of portfolios is constructed only with the covariance matrix, which is well known to be more stable than the vector of means (see Merton (1980)). Therefore, this calibration method is convenient to control for the turnover only with portfolio models which do not consider the vector of means.

The calibration method is as important as the choice of portfolios that form the shrinkage portfolio. We have pointed out that some calibration criteria might only succeed under some specifications. This is why we require some specific structure in our shrinkage portfolios. The following sections give a detailed description of the shrinkage portfolios considered in this paper.

3.1 Shrinkage from mean-variance to minimum-variance portfolio

The combination of the mean-variance portfolio with the minimum-variance portfolio is a well known strategy pioneered by Kan and Zhou (2007). The authors define this strategy as a three-fund portfolio because some wealth goes to the risk free asset, and the rest of the wealth is divided between the mean-variance portfolio and the minimum-variance portfolio. According with our definition of shrinkage portfolios, the mixture of the mean-variance and minimum-variance portfolios is as follows:

$$\widehat{w}_{\text{mv-min}} = (1-\alpha)\widehat{w}_{\text{mv}} + \alpha\nu\widehat{w}_{\text{min}}, \tag{27}$$

where parameter ν is a scale parameter that corrects for the bias of the naive minimum-variance portfolio. The bias correction is not with respect to the true minimum-variance portfolio, because it is our target, but rather we correct for the bias with respect to the true mean-variance portfolio, this is:

$$\nu = \arg\min_{\nu} \left\| \nu E\left(\widehat{w}_{min}\right) - w^*_{mv} \right\|_2^2, \tag{28}$$

where w^*_{mv} is the true mean-variance portfolio. Deriving the first order condition of problem (28) we obtain that the optimal scale parameter is $\nu^* = \frac{E(\widehat{w}_{min})' w^*_{mv}}{E(\widehat{w}_{min})' E(\widehat{w}_{min})}$. Now, we can define the optimal value of the shrinkage intensity α given the optimal scale parameter ν^*. The following proposition gives the expressions for the optimal shrinkage intensities for calibration methods given in (24)-(25):

Proposition 3.1. *If asset returns are independent and identically distributed, then the shrinkage intensities of calibration methods given in (23)-(26) for the optimal combination between mean-variance and minimum variance portfolios are:*

$$\alpha_{msl} = \frac{E\left(\left\|\widehat{w}_{mv} - w^*_{mv}\right\|_2^2\right) - \tau_{mv-min}}{E\left(\left\|\widehat{w}_{mv} - w^*_{mv}\right\|_2^2\right) + E\left(\left\|\nu^*\widehat{w}_{min} - w^*_{mv}\right\|_2^2\right) - 2\tau}, \tag{29}$$

$$\alpha_{ut} = \frac{E\left(\widehat{\sigma}^2_{mv}\right) - \nu^* E\left(\widehat{\rho}_{mv,min}\right)}{E\left(\widehat{\sigma}^2_{mv}\right) + \nu^{*2} E\left(\widehat{\sigma}^2_{min}\right) - 2\nu^* E\left(\widehat{\rho}_{mv,min}\right)} - \\ - \frac{E(\widehat{\mu}_{mv}) - E(\nu^*\widehat{\mu}_{min})}{E\left(\widehat{\sigma}^2_{mv}\right) + \nu^{*2} E\left(\widehat{\sigma}^2_{min}\right) - 2\nu^* E\left(\widehat{\rho}_{mv,min}\right)} \frac{1}{\gamma}, \tag{30}$$

$$\alpha_{var} = \frac{E\left(\widehat{\sigma}^2_{mv}\right) - \nu^* E\left(\widehat{\rho}_{mv,min}\right)}{E\left(\widehat{\sigma}^2_{mv}\right) + \nu^{*2} E\left(\widehat{\sigma}^2_{min}\right) - 2\nu^* E\left(\widehat{\rho}_{mv,min}\right)}, \tag{31}$$

$$\alpha_{SR} = arg\max_{\alpha} \frac{(1-\alpha)E(\widehat{\mu}_{mv}) + \alpha\nu^* E(\widehat{\mu}_{min})}{\sqrt{(1-\alpha)^2 E\left(\widehat{\sigma}^2_{mv}\right) + \alpha^2\nu^{*2} E\left(\widehat{\sigma}^2_{min}\right) + 2(1-\alpha)\alpha\nu^* E\left(\widehat{\rho}_{mv,min}\right)}}, \tag{32}$$

where $\tau_{mv-min} = E\left(\left(\widehat{w}_{mv} - w^*_{mv}\right)'\left(\nu^*\widehat{w}_{min} - w^*_{mv}\right)\right)$, $E\left(\widehat{\sigma}^2_{mv}\right) = E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{mv}\right)$, $E\left(\widehat{\sigma}^2_{min}\right) = E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right)$, $E\left(\widehat{\rho}_{mv,min}\right) = E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{min}\right)$, $E\left(\widehat{\mu}_{mv}\right) = E\left(\widehat{w}'_{mv}\mu\right)$ *and* $E\left(\widehat{\mu}_{min}\right) = E\left(\widehat{w}'_{min}\mu\right)$.

Proof. The proof of Proposition 3.1 is given in the Appendix. □

Proposition 3.1 gives the optimal shrinkage intensities of the shrinkage portfolio formed by the scaled naive mean-variance portfolio and the naive minimum-variance portfolio. The shrinkage intensities of every calibration method is in general a fraction of the error/risk of the naive mean-variance portfolio with respect to the total error/risk of the naive mean-variance portfolio and the scaled minimum-variance portfolio. This fraction specifies how much the naive mean-variance portfolio is shrunk to the scaled naive minimum-variance portfolio.

Equation (29) gives the optimal shrinkage intensity that minimizes the mean-squared loss of the considered shrinkage portfolio. For this calibration method, the shrinkage intensity is defined as the ratio of the error of the naive mean-variance portfolio with respect to the total error formed by error of the naive mean-variance portfolio and the error of naive scaled minimum-variance portfolio. The definition of error for the naive mean-variance portfolio is given by the mean-squared loss and a corrector element τ which is a measure of the dependence between the naive mean-variance portfolio and the scaled naive minimum-variance portfolio. The definition of error for the naive scaled minimum-variance portfolio is referred as the mean-squared loss of the naive scaled minimum-variance portfolio with respect to the true mean-variance portfolio minus the corrector parameter τ.

Equation (30) gives the optimal shrinkage intensity that maximizes the expected utility. In this case, the shrinkage intensity is defined as the ratio of the risk of the naive mean-variance portfolio with respect to the total risk composed by the risk of the naive mean-variance portfolio and the risk of the scaled naive minimum-variance portfolio. This ratio is smoothed by the ratio of the expected return of the naive mean-variance portfolio, in excess of the expected return of the scaled naive minimum-variance portfolio, divided by the total risk. The risk of the naive mean-variance portfolio is defined as its variance minus a corrector term defined by the covariance between the naive mean-variance portfolio and the scaled minimum-variance portfolio. The risk of the scaled naive minimum-variance portfolio is defined as the portfolio variance of the scaled naive minimum-variance portfolio minus a corrector term defined by the covariance between the naive mean-variance portfolio and the scaled minimum-variance portfolio. Therefore, the shrinkage intensity under this calibration method is defined by the relative risk of the naive mean-variance portfolio smoothed by the excess profitability of the naive mean-variance portfolio.

Equation (31) gives the optimal shrinkage intensity that minimizes the expected portfolio variance. In this case, the shrinkage intensity is defined as the ratio of the risk of the naive mean-variance portfolio with respect to the total risk. The definition of risk is the same as in the case of the expected utility calibration method.

Problem (32) offers the optimal shrinkage intensity that maximizes the Sharpe ratio. It has no explicit formula, but we solve problem (32) with numerical methods.

3.2 Shrinkage from mean-variance to equally weighted portfolio

In this section we introduce the shrinkage portfolio formed by the mean-variance and equally weighted portfolios. The mixture of this portfolio has been already studied in the literature. Tu and Zhou (2011) propose combinations of sophisticated risky portfolios with the 1/N rule. In particular, they study four different risky portfolios to combine with the 1/N rule:the classical mean-variance portfolio of Markowitz (1952), the Kan and Zhou (2007) three-fund portfolio, the bayesian mean-variance portfolio of Jorion (1986) and the implied mean-variance portfolio derived from a factor model of MacKinlay and Pastor (2000). In this paper we focus on the simplest one, the combination of the classical mean-variance portfolio with the equally weighted portfolio. As we have mentioned, our main motivation is the calibration of portfolios. Therefore, any good calibration method could be extended to more sophisticated combination of portfolios. Thus, the shrinkage portfolio formed by the mean-variance and equally weighted portfolios is defined as follows:

$$\widehat{w}_{\text{mv-ew}} = (1-\alpha)\widehat{w}_{\text{mv}} + \alpha\nu\frac{\iota}{N}, \tag{33}$$

where the scale parameter ν is a bias correction term which we define by:

$$\nu^* = \arg\min_{\nu} \left\| \nu\frac{\iota}{N} - w^*_{mv} \right\|_2^2. \tag{34}$$

Deriving the first order conditions of problem (34, we obtain that the optimal scale factor is $\nu^* = \iota' w^*_{mv}$. The following proposition gives the optimal shrinkage intensities of the calibration methods (23)-(26):

Proposition 3.2. *Assuming that returns are independent and identically distributed, the shrinkage intensities for the optimal combination of the mean-variance portfolio with the equally weighted portfolio are:*

$$\alpha_{msl} = \frac{E\left(\|\widehat{w}_{mv} - w^*_{mv}\|_2^2\right) - \tau_{mv-ew}}{E\left(\|\widehat{w}_{mv} - w^*_{mv}\|_2^2\right) + \|\nu^* w_{ew} - w^*_{mv}\|_2^2}, \tag{35}$$

$$\alpha_{ut} = \frac{E\left(\widehat{\sigma}^2_{mv}\right) - \nu^* E\left(\widehat{\rho}_{mv,ew}\right)}{E\left(\widehat{\sigma}^2_{mv}\right) + \nu^{*2}\sigma^2_{ew} - 2\nu^* E\left(\widehat{\rho}_{mv,ew}\right)} - \frac{E(\widehat{\mu}_{mv}) - \nu^*\mu_{ew}}{E\left(\widehat{\sigma}^2_{mv}\right) + \nu^{*2}\sigma^2_{ew} - 2\nu^* E\left(\widehat{\rho}_{mv,ew}\right)}\frac{1}{\gamma}, \tag{36}$$

$$\alpha_{var} = \frac{E\left(\widehat{\sigma}^2_{mv}\right) - \nu^* E\left(\widehat{\rho}_{mv,ew}\right)}{E\left(\widehat{\sigma}^2_{mv}\right) + \nu^{*2}\sigma^2_{ew} - 2\nu^* E\left(\widehat{\rho}_{mv,ew}\right)}, \tag{37}$$

$$\alpha_{SR} = arg\max_{\alpha} \frac{(1-\alpha)E(\widehat{\mu}_{mv}) + \alpha\nu^*\mu_{ew}}{\sqrt{(1-\alpha)^2 E\left(\widehat{\sigma}^2_{mv}\right) + \alpha^2\nu^{*2}\sigma^2_{ew} + 2(1-\alpha)\alpha\nu^* E\left(\widehat{\rho}_{mv,ew}\right)}}, \tag{38}$$

where $\tau_{mv-ew} = E\left((\widehat{w}_{mv} - w^*_{mv})\right)'\left(\nu^* w_{ew} - w^*_{mv}\right)$ $\sigma^2_{ew} = \frac{1}{N^2}\iota'\Sigma\iota$ *is the equally weighted portfolio variance,* $E\left(\widehat{\sigma}_{mv}\right) = E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{mv}\right)$ *is the expected mean-variance portfolio variance,* $E\left(\widehat{\rho}_{mv,ew}\right) = E\left(\widehat{w}'_{mv}\Sigma w_{ew}\right)$ *is the covariance between mean-variance and equally weighted portfolio,* $\mu_{ew} = w'_{ew}\mu = \overline{\mu}$ *is the equally weighted expected portfolio return and* $E\left(\widehat{\mu}_{mv}\right) = E\left(\widehat{w}'_{mv}\mu\right)$ *is the mean-variance expected portfolio return.*

Proof. The proof of Proposition 3.2 is given in the Appendix. □

Proposition 3.2 gives the optimal shrinkage intensities of the shrinkage portfolio formed by the naive mean-variance portfolio and the scaled equally weighted portfolio. For this shrinkage portfolio, we also observe that the optimal shrinkage intensities are defined as the relative error/risk of the naive mean-variance portfolio with respect to the total error/risk of the naive mean-variance portfolio and the scaled minimum-variance portfolio.

Equation (35) gives the optimal shrinkage intensity that minimizes the mean-squared loss. The shrinkage intensity is a ratio defined as the error of the naive mean-variance portfolio relative to the total error formed by the error of the naive mean-variance portfolio and the error of the scaled equally weighted portfolio. In this shrinkage portfolio, the error of the naive mean-variance portfolio is defined by its mean-squared loss. The error of the scaled equally weighted portfolio is defined by the mean-squared loss with respect to the true mean-variance portfolio w_{mv}^*. In this case, there is no corrector term because the equally weighted portfolio and the naive mean-variance portfolio are independent with each other. Since the scaled equally weighted is deterministic (i.e. zero variance) and the naive mean-variance portfolio is unbiased, the optimal shrinkage intensity gives the optimal trade-off between bias and variance.

Equation (36) gives the optimal shrinkage intensity that maximizes the expected utility. The shrinkage intensity is defined as the ratio of the risk of the naive mean-variance portfolio with respect to the total risk of the naive mean-variance portfolio and the scaled equally weighted portfolio. This ratio is smoothed by the ratio of the expected return of the naive mean-variance portfolio, in excess of the expected return of the scaled equally weighted portfolio, divided by the total risk. The risk of the naive mean-variance portfolio is defined as its variance minus a corrector term defined by the covariance between the naive mean-variance portfolio and the scaled equally weighted portfolio. The risk of the scaled equally portfolio is defined as the portfolio variance of the scaled equally weighted portfolio minus the portfolio covariance between the naive mean-variance portfolio and the scaled equally weighted portfolio. Therefore, the shrinkage intensity under this calibration method is defined by the relative risk of the naive mean-variance portfolio smoothed by the excess profitability of the naive mean-variance portfolio.

Equation (37) gives the optimal shrinkage intensity that minimizes the expected portfolio variance. The shrinkage intensity is defined as the ratio of the risk of the naive mean-variance portfolio with respect to the total risk of the naive mean-variance portfolio and the scaled equally weighted portfolio.

Problem (38) offers the optimal shrinkage intensity that maximizes the Sharpe ratio. Like for the previous shrinkage portfolio, it has no explicit formula, but we again solve problem (38) with numerical methods.

3.3 Shrinkage from minimum-variance to equally weighted portfolio

In this section, we study the shrinkage portfolio formed by the minimum-variance and equally weighted portfolio. Likewise to the previous shrinkage portfolios, this mixture of portfolios has been studied in DeMiguel et al. (2009). The authors provide a linear combination of portfolios formed by the minimum-variance portfolio and the 1/N rule or equally weighted portfolio. We use the same portfolios to study the third shrinkage portfolio of the paper. It takes the following expression:

$$\widehat{w}_{\text{min-ew}} = (1-\alpha)\widehat{w}_{\text{min}} + \alpha\nu\frac{\iota}{N}, \tag{39}$$

where the scale parameter ν is a bias correction term which is optimally defined by:

$$\nu^* = \text{arg}\min_{\nu} \left\|\nu\frac{\iota}{N} - w^*_{min}\right\|_2^2. \tag{40}$$

Developing the first order conditions in (40), we obtain that the optimal scale parameter is $\nu^* = \iota' w^*_{min}$. Once again, we obtain the optimal shrinkage intensities for the calibration methods illustrated in (23)-(26). The following proposition gives the expressions.

Proposition 3.3. *Assuming that returns are independent and identically distributed, the shrinkage intensities for the optimal combination of the minimum-variance portfolio with the equally weighted portfolio are:*

$$\alpha_{msl} = \frac{E\left(\|\widehat{w}_{min} - w^*_{min}\|_2^2\right) - \tau_{min-ew}}{E\left(\|\widehat{w}_{min} - w^*_{min}\|_2^2\right) + \|\nu^* w_{ew} - w^*_{min}\|_2^2}, \tag{41}$$

$$\alpha_{ut} = \frac{E\left(\widehat{\sigma}^2_{min}\right) - \nu^* E\left(\widehat{\rho}_{min,ew}\right)}{E\left(\widehat{\sigma}^2_{min}\right) + \nu^{*2}\sigma^2_{ew} - 2\nu^* E\left(\widehat{\rho}_{min,ew}\right)} - \frac{E(\widehat{\mu}_{min}) - \nu^*\mu_{ew}}{E\left(\widehat{\sigma}^2_{min}\right) + \nu^{*2}\sigma^2_{ew} - 2\nu^* E\left(\widehat{\rho}_{min,ew}\right)}\frac{1}{\gamma}, \tag{42}$$

$$\alpha_{var} = \frac{E\left(\widehat{\sigma}^2_{min}\right) - \nu^* E\left(\widehat{\rho}_{min,ew}\right)}{E\left(\widehat{\sigma}^2_{min}\right) + \nu^{*2}\sigma^2_{ew} - 2\nu^* E\left(\widehat{\rho}_{min,ew}\right)}, \tag{43}$$

$$\alpha_{SR} = arg\max_{\alpha} \frac{(1-\alpha)E(\widehat{\mu}_{min}) + \alpha\nu^*\mu_{ew}}{\sqrt{(1-\alpha)^2 E\left(\widehat{\sigma}^2_{min}\right) + \alpha^2\nu^{*2}\sigma^2_{ew} + 2(1-\alpha)\alpha\nu^* E\left(\widehat{\rho}_{min,ew}\right)}}, \tag{44}$$

where $\tau_{min-ew} = E\left((\widehat{w}_{min} - w^*_{min})\right)'\left(\nu^* w_{ew} - w^*_{min}\right)$, $\sigma^2_{ew} = \frac{1}{N^2}\iota'\Sigma\iota$ *is the equally weighted portfolio variance,* $E\left(\widehat{\sigma}_{min}\right) = E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right)$ *is the expected variance of the minimum-variance portfolio,* $E\left(\widehat{\rho}_{min,ew}\right) = E\left(\widehat{w}'_{min}\Sigma w_{ew}\right)$ *is the covariance between minimum-variance and equally weighted portfolio,* $\mu_{ew} = w'_{ew}\mu = \overline{\mu}$ *is the equally weighted expected portfolio return and* $E\left(\widehat{\mu}_{min}\right) = E\left(\widehat{w}'_{min}\mu\right)$ *is the minimum-variance expected portfolio return.*

Proof. The proof of Proposition 3.3 is given in the Appendix. □

Equation (41) gives the optimal shrinkage intensity that minimizes the mean-squared loss. The shrinkage intensity is a ratio defined as the error of the naive minimum-variance portfolio relative to the total error formed by the error of the naive minimum-variance portfolio and the error of the scaled equally weighted portfolio. In this shrinkage portfolio, the error of the naive minimum-variance portfolio is defined by its mean-squared loss. The error of the scaled equally weighted portfolio is defined by the mean-squared loss with respect to the true minimum-variance portfolio w_{min}. In this case, we again do not have corrector term because the equally weighted portfolio and the naive minimum-variance portfolio are independent with each other. Since the scaled equally weighted is deterministic (i.e. zero variance) and the naive minimum-variance portfolio is unbiased, the optimal shrinkage intensity gives the optimal trade-off between bias and variance.

Equation (42) gives the optimal shrinkage intensity that maximizes the expected utility. The shrinkage intensity is defined as the ratio of the risk of the naive minimum-variance portfolio with respect to the total risk of the naive minimum-variance portfolio and the scaled equally weighted portfolio. This ratio is smoothed by the ratio of the expected return of the naive minimum-variance portfolio, in excess of the expected return of the scaled equally weighted portfolio, divided by the total risk. The risk of the naive minimum-variance portfolio is defined as its variance minus a corrector term defined by the covariance between the naive minimum-variance portfolio and the scaled equally weighted portfolio. The risk of the scaled equally portfolio is defined as the portfolio variance of the scaled equally weighted portfolio minus the covariance between the naive minimum-variance portfolio and the scaled equally weighted portfolio. Therefore, the shrinkage intensity under this calibration method is defined by the relative risk of the naive mean-variance portfolio smoothed by the excess profitability of the naive mean-variance portfolio.

Equation (43) gives the optimal shrinkage intensity that minimizes the expected portfolio variance. The shrinkage intensity is defined as the ratio of the risk of the naive minimum-variance portfolio with respect to the total risk of the naive minimum-variance portfolio and the scaled equally weighted portfolio.

Problem (44) offers the optimal shrinkage intensity that maximizes the Sharpe ratio. It has no explicit formula, but we solve problem (44) with numerical methods.

So far, we have assumed that returns are independent and identically distributed. In the next section, we give closed form expressions for the defined optimal shrinkage intensities. To obtain closed form expressions of the optimal shrinkage intensities we have to assume that returns are normally distributed. Finally, we drop the normality assumption to propose an alternative method for calibrating shrinkage portfolios.

3.4 Parametric calibration

In this section, we give closed form expressions to the expected values that define all the optimal shrinkage parameters. To do that, we assume that asset returns are independent and normally distributed (i.e. $N(\mu, \Sigma)$). The objective of this section is to give the implementations details of the shrinkage intensities and discuss how they evolve when the estimation window T increases. Then, we first define the expected values that we need to compute the optimal shrinkage intensities.

Proposition 3.4. *Assuming that returns are independent and normally distributed, i.e. $N(\mu, \Sigma)$, we can obtain closed-form expressions for the following elements:*

The mean-squared loss of the naive mean-variance portfolio:

$$E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right) = \frac{a}{\gamma^2}\left[tr\left(\Sigma^{-1}\right)\left(\frac{(T-2)}{T} + \mu'\Sigma^{-1}\mu\right) + \right. \\ \left. +(T-N-2)\mu'\Sigma^{-2}\mu\right] - \frac{1}{\gamma^2}\mu'\Sigma^{-2}\mu. \tag{45}$$

The mean-squared loss of the naive minimum-variance portfolio with respect to the true mean-variance portfolio:

$$E\left(\|\nu\widehat{w}_{min} - w_{mv}\|_2^2\right) = \nu^2 a\left[tr\left(\Sigma^{-1}\right)\iota'\Sigma^{-1}\iota + (T-N-2)\iota'\Sigma^{-2}\iota\right] + \\ + \frac{1}{\gamma^2}\mu'\Sigma^{-2}\mu - 2\frac{\nu}{\gamma}\iota'\Sigma^{-2}\mu. \tag{46}$$

The mean-squared loss of the naive minimum-variance portfolio:

$$E\left(\|\widehat{w}_{min} - w_{min}\|_2^2\right) = a\left[tr\left(\Sigma^{-1}\right)\iota'\Sigma^{-1}\iota + \right. \\ \left. +(T-N-2)\iota'\Sigma^{-2}\iota\right] - \iota'\Sigma^{-2}\iota. \tag{47}$$

The expected value of the naive mean-variance portfolio variance:

$$E\left(\widehat{\sigma}_{mv}\right) = E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{mv}\right) = \frac{1}{\gamma^2}\left(a(T-2)\left(\frac{N}{T} + \mu'\Sigma^{-1}\mu\right)\right). \tag{48}$$

The expected value of the naive minimum-variance portfolio variance:

$$E\left(\widehat{\sigma}_{min}\right) = E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right) = a(T-2)\iota'\Sigma^{-1}\iota. \tag{49}$$

The expected value of the covariance between the naive mean-variance and naive minimum-variance portfolio:

$$E\left(\widehat{\rho}_{mv,min}\right) = E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{min}\right) = a(T-2)\frac{1}{\gamma}\mu'\Sigma^{-1}\iota. \tag{50}$$

The corrector term τ_{mv-min}:

$$\tau_{mv-min} = \nu^{*}\left(\frac{a}{\gamma}\left[tr\left(\Sigma^{-1}\right)\mu'\Sigma^{-1}\iota + (T-N-2)\mu'\Sigma^{-2}\iota\right] - \frac{1}{\gamma}\mu'\Sigma^{-2}\iota\right), \tag{51}$$

where $a = \frac{(T-N-2)}{(T-N-1)(T-N-4)}$.

Proof. The proof of Proposition 3.4 is given in the appendix. □

Proposition 3.4 introduces some of the elements which form the optimal shrinkage parameters. As we observe, all of them are affected by the window's length T and the number of assets N. All the expressions show that when $N \geq 4$ and the ratio N/T tends to zero, the expressions showed in Proposition 3.4 tends to the population counterpart[14]. In the case of $E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right)$ and $E\left(\|\widehat{w}_{min} - w_{min}\|_2^2\right)$, the population counterparts will be zero in both cases.

3.4.1 Discussion of optimal shrinkage intensities

In this section, we make an experiment to study how the optimal shrinkage intensities and the theoretical mean-squared losses of shrinkage portfolios evolve with different window's length T. In the experiment we take again the sample formed with 48 industry portfolio as the population. We use the sample moments as the population elements that define the multivariate normal distribution, $N(\mu, \Sigma)$, and governs the dynamics of the stock returns. Figure 3 shows the results of the experiment.

Figure 3a shows the shrinkage intensities given for the shrinkage portfolio formed by the mean-variance portfolio and the minimum-variance portfolio. In this case, we observe that the shrinkage intensities obtained by the mean-squared loss minimization, the expected utility maximization and the Sharpe ratio maximization methods offer similar values. Moreover, from Figure 3a we observe that they seem to converge at the same rate. On the other hand, we observe that the shrinkage intensity obtained via expected portfolio

[14] Here, what we mean by population counterpart is that the expected values converge to the expressions where the estimated quantity is replaced by the population one. For instance, in $E\left(\|\nu\widehat{w}_{min} - w_{mv}\|_2^2\right)$, the population counterpart will be $\|\nu w_{min} - w_{mv}\|_2^2$.

variance minimization is always near one, which means that the minimum-variance portfolio has always a much lower variance than the mean-variance portfolio, independently of the window's length T.

Figure 3b shows the shrinkage intensities given for the shrinkage portfolio formed by the mean-variance portfolio and the equally weighted portfolio. In this shrinkage portfolio, we observe that the portfolio variance minimization is again the most conservative method. This calibration method gives a higher weight to the equally weighted portfolio than the other two calibration methods. DeMiguel et al. (2009) show that an equally weighted portfolio has a good out-of-sample performance in terms of Sharpe Ratio. One of the main reasons why it happens is that the equally weighted portfolio has a small portfolio variance. Therefore, it is reasonable to find that the shrinkage intensity obtained via portfolio variance minimization is higher. Moreover, we observe that the shrinkage intensity given by the Sharpe ratio maximization method gives the lowest value among all the considered calibration methods.

Figure 3c shows the shrinkage intensities given for the shrinkage portfolio formed with the minimum-variance portfolio and the equally weighted portfolio. In this case, we observe that the most conservative calibration method is the mean-squared loss minimization method up to $T \approx 550$. From that point onwards, the most conservative calibration method is the Sharpe ratio maximization. Thus, up to $T \approx 550$, the mean-squared loss minimization method offers a higher shrinkage intensity than the three other methods. This is because the equally weighted is a portfolio with a very low estimation error. Since the equally weighted portfolio is deterministic, the bias of this portfolio is the element that defines its squared-loss. Moreover, with the scale parameter ν^*, it is expected to reduce the bias and this is why the shrinkage intensity of this method is higher. On the other hand, we observe that the shrinkage intensities obtained via utility maximization and portfolio variance minimization are the same. It is because the smoothing element of expression (42) given by the ratio proportional to the profitability of the minimum-variance portfolio μ_{min} in excess to the profitability of the equally weighted portfolio μ_{ew} is negligible. Therefore, both portfolios must offer a similar expected return in this example.

Figures 3d, 3e and 3f obviously show that the mean-squared loss minimization offers the best shrinkage intensities to minimize the mean-squared loss. By definition, α_{msl} is obtained to minimize the mean-squared loss. In Figure 3d, we observe that the mean-squared loss of the expected utility maximization, the mean-squared loss minimization and the Sharpe ratio maximization converge at the same rate. Actually, we can see that their values are quite similar. In Figure 3e, we observe that the mean-squared loss of the Sharpe ratio maximization method increases while the sample window's length T is small. This is because in small samples, the shrinkage intensity of this method decreases faster than it should be decreasing. Then, it gives more weight to the naive mean-variance portfolio, which is well know to be highly contaminated by estimation risk, and this is why the mean-squared loss increases. Finally, in Figure 3f, we observe that the mean-squared losses of the four calibration methods converge with a similar rate.

3.4.2 Performance of calibrated portfolios

In this section, we study two performance measures which are interesting for an investor. The first performance measure that we study is the Sharpe ratio, defined by the expected value of the portfolio return divided by the square root of the expected value of the portfolio variance. This is a classical performance measure which represents the expected profit per unit of risk. Therefore, we define the expected Sharpe ratio as follows:

$$SR_t = \frac{E\left(r_p^t\right)}{\sqrt{Var\left(r_p^t\right)}}, \tag{52}$$

where $r_p^t = \widehat{w}_{t-1}' R_t$ is the portfolio return at time t that corresponds with the estimated portfolio $\widehat{w}_{t-1}$ at time $t-1$.

The second performance measure that we study is the *relative improvement in average loss* (RIAL, see Ledoit and Wolf (2004a)) of each calibrated shrinkage portfolio. This performance measure is defined as:

$$\text{RIAL}(\widehat{w}_{sp}) = \frac{E\left(\|\widehat{w}_n - w^*\|_2^2\right) - E\left(\|\widehat{w}_{sp} - w^*\|_2^2\right)}{E\left(\|\widehat{w}_n - w^*\|_2^2\right)}, \tag{53}$$

where $\widehat{w}_{sp}$ is the shrinkage portfolio composed by a naive estimator of the true portfolio w^* and an scaled target portfolio w_{target}. This measure gives an idea of the relative improvement of the shrinkage portfolio with respect to the estimated benchmark portfolio. If RIAL is near 1, it means that the mean-squared loss of the shrinkage portfolio is negligible compared with the mean-squared loss of the estimated reference portfolio $\widehat{w}_n$. If RIAL is negative, it means that the estimated reference portfolio $\widehat{w}_n$ has a smaller mean-squared loss than the shrinkage portfolio. This performance measure gives an idea of the statistical error that contains the shrinkage portfolio relative to the statistical error of the naive estimator, which is assumed to be high in small samples.

The experiment consists on computing the Sharpe ratios and the RIAL's, given by expressions (52)-(53), of the studied shrinkage portfolios under every different calibration method and for different data sets. We consider four data sets: 5IndP, 10IndP, 38IndP, 48IndP and 100FF. All the data sets are defined in Table 1. The objective of including data sets with different dimensions is to observe whether calibration methods are robust or not along different data sets with different characteristics. To make the experiment closer to reality, we consider an estimation window's length T=150, which is quite reasonable when we deal with monthly data. In the experiment, we consider the sample moments obtained across every the data set as the true population moments of a multivariate normal distribution. Therefore we are computing the theoretical values of the Sharpe ratio and the RIAL. In Table 3, we have the values obtained for the Sharpe ratio. Table 3 gives important insights about the shrinkage portfolio considered. First, we can obviously see that the Sharpe ratio maximization method gives gives the best Sharpe ratio across every data set. After this calibration method, the expected utility maximization method gives the best Sharpe ratio across every data sets in all shrinkge portfolios except for the min-ew portfolio in the 5IndP data set, where the mean-squared loss minimization gives a slightly better value.

Table 4 gives the RIAL's of the considered shrinkage portfolios calibrated with the methods presented in this section. For this performance measure, the upper bound is established by the values given by the mean-squared loss minimization method. It is easy to observe that minimizing the mean-squared loss is equivalent to maximize the RIAL. In Table 4, we observe three important issues. First, for the shrinkage portfolio formed by the mean-variance and minimum-variance portfolios, the RIAL obtained for the expected utility calibration method offers similar values with respect to the RIAL obtained for the same portfolio calibrated via mean-squared loss minimization, except for the data set 5IndP. Moreover, we also observe that this portfolio obtains slightly worse RIAL's when it is estimated via portfolio variance minimization. Second, the shrinkage portfolio formed by the mean-variance and equally weighted portfolio obtains slightly worse RIAL's when it is calibrated via Sharpe ratio maximization, except for the data set 5IndP. Moreover, we also observe that the values obtained with the portfolio variance calibration method are very similar to the values obtained by the mean-squared loss minimization method. Third, the shrinkage portfolio formed by the minimum-variance and equally weighted portfolio obtain very similar RIAL's when the portfolio is calibrated either by the expected utility maximization method or the portfolio variance minimization method. Both calibration methods give better RIAL's throughout every data set than the values obtained for the shrinkage portfolio calibrated via Sharpe ratio maximization except for the data sets 38IndP and 48IndP.

Finally, to complete the analysis, we make the same experiment but this time, instead of changing the number of assets under consideration, we fix the number of assets and change the window's length T. Figure 4 shows the results of the evolution of the expected Sharpe ratio and the RIAL's when the estimation window's length T changes. In this experiment, we take the data set formed by 48 industry portfolios (48IndP data set of Table 1) and take the whole sample as the population distribution. The sample moments are considered the population moments that define the multivariate normal distribution required to compute the theoretical values of the expected Sharpe Ratio and the RIAL. Figures 4a, 4b and 4c confirms the evidences found in the previous experiment. Obviously, the upper bound in all the shrinkage portfolios is established by the Sharpe ratio obtained with the calibration method that maximizes the Sharpe ratio. Like in the previous experiment, we observe that for the mv-min portfolio, the Sharpe ratio attained by the mean-squared loss minimization method and the expected utility maximization method are very similar and slightly higher than the results obtained by the expected variance minimization method. Moreover, we observe that the expected utility maximization is the second best calibration method for the mv-ew shrinkage portfolio. For the min-ew shrinkage portfolio, we can see that the mean-squared loss minimization method is slightly worse that the other calibration methods in terms of Sharpe ratio.

On the other hand, Figures 4d, 4e and 4f depicts the RIAL's of the considered shrinkage portfolios under every different calibration method. Figure 4d shows the RIAL's of the shrinkage portfolio formed by the mean-variance and the minimum-variance portfolio. We observe that the RIAL's of the shrinkage portfolio calibrated via mean-squared loss minimization, expected utility maximization and Sharpe ratio maximization evolve in a similar manner. It holds with the results obtained in Figure 3a, where the shrinkage intensities corresponding to these calibration methods are similar. On the other hand, we observe that the RIAL of the shrinkage portfolio calibrated via portfolio variance minimization is negative when the window's length is larger than 600 observations. It is because this calibration method gives a shrinkage intensity of near one independently of the window's length T, so that the portfolio is practically the minimum-variance portfolio. It means that with 600 observations we can estimate fairly well the mean-variance portfolio and therefore shrinking too much to the minimum-variance portfolio gives bad results in terms of RIAL.

Figure 4e shows the RIAL's of the shrinkage portfolio formed by the mean-variance and the equally weighted portfolio. We observe that the RIAL's of this shrinkage portfolio evolves in a similar manner for the mean-squared loss minimization method and the expected utility maximization method. On the other hand, we observe that the Sharpe ratio maximization method to calibrate the mv-ew portfolio gives lower RIAL that the other three calibration methods up to $T \approx 650$. For that window's length, the expected variance minimization method gives worse results and it even attains negative values when $T \approx 1355$. It means that under that calibration method, the mean-squared error of the sample mean-variance portfolio is better than the combination of the mean-variance portfolio with an equally weighted portfolio. But it does not mean that there is no a better combination of the mean-variance portfolio and the equally weighted portfolio. In fact, we observe that the shrinkage portfolios calibrated via expected utility maximization or mean-squared loss minimization offer a lower mean-squared error than the sample mean-variance portfolio.

Figure 4f shows the RIAL's of the shrinkage portfolio formed by the minimum-variance and the equally weighted portfolio. We observe that the mean-squared error offers a lower RIAL than the other two calibration methods. Moreover, we observe that the RIAL's of the three calibration method calibrated via expected utility maximization is the same as the RIAL of the shrinkage portfolio calibrated via portfolio variance minimization. It holds with the resutls obtained in Figure 3c. We also observe that the RIAL of every calibration method converge faster to zero than any of the shrinkage portfolios considered in the study. It is because this shrinkage portfolio does not consider the vector of means, which is usually more difficult to estimate. Therefore, the benchmark portfolio is fairly well estimated for small window's length and therefore, the shrinkage intensity must go faster to zero as well. This is exactly what we observe in Figure 3c.

We observe that results do not differ too much in terms of Sharpe ratio when the sample size T is small compared with the number of assets. Moreover, we have also observed that in general, a mixture of portfolios formed by the mean-variance and minimum variance portfolios give bad results in terms of Sharpe ratio for a moderate samples size of $T = 150$. Therefore, we can assume that, in general, it is better to consider a mixture of some risky portfolio (either the mean-variance portfolio or the minimum-variance portfolio) and the equally weighted portfolio.

3.5 Non-Parametric calibration of portfolios

In this section, we propose an alternative procedure to calibrate shrinkage portfolios. This framework only requires that stock returns must be independent and identically distributed[15], but we do not specify any particular distribution. This method is based on the well-known non-parametric technique of bootstrap (Efron (1979)). This technique was originally proposed to study the statistical properties of some statistic of interest. Then, let us assume that we want to study some characteristic θ of some probability distribution F, such that this characteristic of interest is expressed as $\theta(F)$. Then, let us assume that we have a sample $X \in \mathbb{R}^{T\times N}$ which has been randomly drawn from F. Therefore, we estimate $\theta(F)$ from the sample X, which constitutes what is called the empirical distribution $\widehat{F}$. Therefore, the value of our estimator is completely specified by the empirical distribution $\widehat{F}$, hence we express it as $\theta\left(\widehat{F}\right)$. Of course, this statistic, $\theta\left(\widehat{F}\right)$, is a random variable and its randomness is defined by the true probability distribution F. It means that F defines

[15]Again, the independence assumption is very unreal for high frequency data. This is why in the empirical study we only focus on monthly data.

the probability distribution of $\theta\left(\widehat{F}\right)$ as well, this is $F \to \theta\left(\widehat{F}\right)$. Therefore, the main goal of bootstrap is to approximate that relation as much as possible. To do that, B (bootstrap) samples $X_b \in \mathbb{R}^{T\times N}$, where $b = 1, \dots, B$, are drawn from the empirical distribution $\widehat{F}$. Each (bootstrap) sample X_b is constructed by drawing random observations with replacement from $\widehat{F}$. For each (bootstrap) sample X_b, we calculate the statistic of interest, which is now defined as $\theta\left(\widehat{F}_b\right)$, where $\widehat{F}_b$ is the empirical distribution of sample X_b. The B statistics obtained from each (bootstrap) sample should define the randomness of the statistic, which is now defined by the original empirical distribution $\widehat{F}$. Therefore, with the bootstrap procedure, we are defining $\widehat{F} \to \theta\left(\widehat{F}_b\right)$, and it must approach $F \to \theta\left(\widehat{F}\right)$ as much as possible.

Then, we use bootstrap to obtain the optimal shrinkage intensities established in sections , and . As we have seen, the optimal shrinkage intensities are defined by a bunch of some expected values. Up to that point, there is no assumption about the distribution of stock returns and they are only assumed to be independent and identically distributed. But to obtain closed form expressions of those expected values, we need to assume some distribution. In this section, we do not make any specification about that distribution. Instead, we use bootstrap to estimate the expected values of the optimal shrinkage intensities given in the previous section. In essence, each expected value correspond to the mean of some random statistic $\theta\left(\widehat{F}\right)$. Thus, we construct B (bootstrap) samples and for each sample we compute the desired statistic $\theta\left(\widehat{F}_b\right)$. Once the B statistics have been computed, we take the average of those statistics as an approximation of the expected value:

$$\overline{\mu}_{\theta(\widehat{F})} = \frac{\sum_{b=1}^{B} \theta\left(\widehat{F}_b\right)}{B}, \tag{54}$$

where $\overline{\mu}_{\theta(\widehat{F})}$ is the bootstrap estimator of the expected value of the statistic of interest. One of the main problems with bootstrap is that the empirical distribution $\widehat{F}$ is a discrete probability function. Therefore, randomly draws from $\widehat{F}$ might gives many repeated observations. This issue turn out to be quite cumbersome when we are trying to estimate the inverse covariance matrix because too many repeated observations might turn out in a sample with singular covariance matrix. Thus, to correct this problem we add an error term for each extracted observation. This is what is called smoothed bootstrap. We apply the multivariate version of the smoothed bootstrap described in Efron (1979) (page 7). Then, now each extracted observation is defined as follows:

$$X_i^* = \widehat{\mu} + (I + \Sigma_Z)^{-1/2}\left[X_i - \widehat{\mu} + \widehat{\Sigma}^{1/2} Z_i\right], \tag{55}$$

where I is the identity matrix, X_i is the i-th row observation from $X \in \mathbb{R}^{T\times N}$, $\widehat{\mu}$ is the sample vector of means of X, $\widehat{\Sigma}$ is the sample covariance matrix of X, and Z_i is a multivariate random variable having zero vector of means and covariance matrix Σ_Z. The peculiarity of this technique is that X_i^* is a random variable which has mean $\widehat{\mu}$ and covariance matrix $\widehat{\Sigma}$ under the empirical distribution $\widehat{F}$.

Furthermore, we consider another non-parametric method to estimate the expected values that form the optimal shrinkage intensities. In this case, we estimate the expectations of the shrinkage intensities obtained from the calibration criteria given in (24)-(26) via cross-validation. Cross-validation consists on splitting the sample data into two parts. In the first part, which is called the "training data", we compute the naive portfolio. This portfolio is then evaluated in the remaining data, which is usually defined as the "validating data". In our experiments, the training data is formed by the whole sample minus one given observation. The portfolio that is computed from the training data is evaluated in the remaining observation, which give us an out-of-sample portfolio return defined as follows:

$$r_p^i = R_i' \widehat{w}^i, \tag{56}$$

where R_i is the i-th sample return and $\widehat{w}^i$ is the estimated portfolio computed without return R_i. Then, we do the same for $i = 1, \ldots, T$, obtaining a time series of T out-of-sample portfolio returns. We use that time series to estimate the expected values that form the optimal shrinkage intensities. We do not report the results of this method because they are similar or worse than the results obtained with the smoothed bootstrap calibration.

In this section we do not consider to make a simulation experiment as we did with the parametric calibration to discuss the evolution of the shrinkage intensities and the performance of the shrinkage portfolios. It is mainly because we have closed form expressions for the parametric shrinkage intensities, which makes the discussion much cheaper in computational terms. Thus, to make the comparison between the performance of the parametric calibration and the non-parametric calibration, we make an empirical analysis across six different real data sets.

4 Empirical Results

In this section, we study the out-of sample performance of portfolios formed with shrinkage estimators considered in section 2. We also study the out-of-sample performance of the shrinkage portfolios considered in section 3. Table 2 lists the portfolios considered in the analysis. Panel A list the existing benchmark portfolios which constitute our benchmark portfolios in the empirical analysis. The first portfolio is the classical mean-variance portfolio of Markowitz (1952). The second portfolio is the classical mean-variance portfolio with shrinkage moments proposed by Jorion (1986). The next portfolios are the combination of portfolios proposed in the literature. The first one is the mixture of the mean-variance and minimum-variance portfolio of Kan and Zhou (2007). The second is the mixture of the mean-variance and equally weighted portfolios of Tu and Zhou (2011). The third is the mixture of the minimum-variance and equally weighted portfolio of DeMiguel et al. (2009). The sixth portfolio is the global minimum-variance portfolio. The seventh portfolio is the global minimum-variance portfolio formed with the shrinkage covariance matrix of Ledoit and Wolf (2004a). The eighth portfolio is the naive equally weighted portfolio and the ninth portfolio is the value weighted or market portfolio. Panel B list the portfolios constructed with the new shrinkage estimators of moments proposed in section 2. Thus, the tenth portfolio is the global minimum-variance portfolio formed with the shrinkage covariance matrix of Ledoit and Wolf (2004a) using the finite optimal shrinkage intensity proposed in section 2. Portfolio eleven is the global minimum-variance portfolio formed with the shrinkage covariance matrix of Ledoit and Wolf (2004a) using the finite optimal shrinkage intensity that account for the matrix condition number. Portfolio twelve is a mean-variance portfolio where the vector of means is estimated with the shrinkage estimator proposed in section 2. The last three portfolios that we study are the shrinkage portfolios proposed in section 3. In the empirical analysis, we calculate the shrinkage intensities of these portfolio using the three calibration methods defined in section 3. The empirical analysis is made across six different data sets. Table 1 list the various data sets that we consider in the analysis. There are three types of data sets: small data set (less than 25 assets), medium data sets (between 25 and 50 assets) and large data sets (more than 50 assets). Small and medium data sets consider all the stocks from NYSE, AMEX and NASDAQ, and they are pooled in industry portfolios to form each of the data sets. On the other hand, 100FF is the one hundred Fama and French (1992) portfolios of firms sorted by size and book-to-market. The last portfolio, SP100 is formed by 100 random assets taken from the S&P500.

4.1 Out-of-sample performance evaluation

We compare the out-of-sample performance of the shrinkage portfolios to the portfolios in the literature using three different criteria: (i) out-of-sample portfolio Sharpe ratio accounting for transaction costs, (ii) portfolio turnover (trading volume), and (iii) out-of-sample portfolio variance. We use the "rolling-horizon" procedure to compute the out-of-sample performance measure. The "rolling-horizon" is defined as follows: first, we choose a window over which to estimate the portfolio. The length of the window is $M < T$, where T

is the total number of observations of the data set. In the empirical analysis, our estimation window has a length of $M = 150$, which corresponds with 15 years of data (with monthly frequency). Second, we compute the various portfolios using the return data over the estimation window. Third, we repeat the "rolling-window" procedure for the next month by including the next data point and dropping the first data point of the estimation window. We continue doing this until the end of the data set. Therefore, at the end we have a time series of $T - M$ portfolio weight vectors for each of the portfolios considered in the analysis; that is w_t^i for $t = M, \dots, T-1$ and portfolio i.

The out-of-sample returns are computed by holding the portfolio weights for one month w_t^i evaluated with the asset-return vector of the next month: $r_{t+1}^i = R_{t+1}' w_t^i$, where R_{t+1} denotes the asset-return vector at time $t+1$ and r_{t+1}^i is the out-of-sample portfolio return at time $t+1$ of portfolio i. We use the times series of portfolio returns and portfolio weights of each strategy to compute the out-of-sample variance, Sharpe ratio and turnover:

$$(\widehat{\sigma}^i)^2 = \frac{1}{T-M-1} \sum_{t=M}^{T-1} \left(w_t^{i\prime} R_{t+1} - \overline{r}^i \right), \tag{57}$$

$$\text{with } \overline{r}^i = \frac{1}{T-M} \sum_{t=M}^{T-1} \left(w_t^{i\prime} R_{t+1} \right), \tag{58}$$

$$\widehat{SR}^i = \frac{\overline{r}^i}{\widehat{\sigma}^i}, \tag{59}$$

$$\text{Turnover} = \frac{1}{T-M-1} \sum_{t=M}^{T-1} \sum_{j=1}^{N} \left(\left| w_{j,t+1}^i - w_{j,t^+}^i \right| \right), \tag{60}$$

where $w_{j,t}^i$ denotes the estimated portfolio weight of asset j at time t under policy i and w_{j,t^+}^i is the estimated portfolio weight of asset j accumulated at time $t+1$, which implies that the turnover is equal to the sum of the absolute value of the rebalancing trades across the N available assets over the $T - M - 1$ trading dates, normalized by the total number of trading dates.

To account for transaction costs in the empirical analysis, the definition of portfolio return is slightly corrected by the implied cost of rebalancing the portfolio. Then, the definition of portfolio return net of transaction costs is:

$$\underline{r}_{t+1}^i = (1 + R_{t+1}' w_t^i) \left(1 - \kappa \sum_{j=1}^{N} \left| w_{j,t+1}^i - w_{j,t^+}^i \right| \right) - 1, \tag{61}$$

where κ is the chargeable fee for rebalancing the portfolio. In the empirical analysis, expressions (57)-(60) are computed using portfolio returns discounted by transaction costs.

4.2 Discussion of the out-of-sample performance

Table 5 reports the Sharpe ratio adjusted by transaction costs of the benchmark portfolios and the portfolios estimated with shrinkage moments. In the analysis we assume 50 basis points of transaction costs, which are high enough to approach to reality. Panel A shows the out-of-sample annualized Sharpe ratio of the benchmark portfolios adjusted with transaction costs. Panel B reports portfolios formed with shrinkage moments, using the shrinkage intensities proposed in this paper. We first compare the portfolios which consider the vector of means. We observe that the mean-variance portfolio with the new proposed shrinkage vector of means (f-mv) has higher adjusted Sharpe ratio than the classical mean-variance portfolio (mv) and the bayes-stein (bs) mean-variance portfolio across every data set except for the SP100 data set. Moreover, portfolio f-mv has higher adjusted Sharpe ratio than the equally weighted portfolio in small data sets.

On the other hand, the new portfolios that do not consider the vector of means and shrink the covariance matrix or the inverse covariance matrix, f-lw, i-lw and c-lw, have a good performance across every data set. Although the minimum-variance portfolio with shrinkage matrix calibrated parametrically (f-lw) does not perform better than the minimum-variance portfolio with the Ledoit and Wolf (2004a) shrinkage covariance matrix (lw) in any data set, portfolio f-lw has a good performance, having larger adjusted Sharpe ratio across every data set than the portfolios which consider the vector of means and also better than the equally weighted portfolio for small and large data sets. The minimum-variance portfolio that shrinks the inverse covariance matrix (i-lw) is very conservative. This portfolio is constructed using the shrinkage inverse covariance matrix proposed in Section 2.3. For large data sets, the shrinkage intensity approaches to one, which makes it tend to the 1/N portfolio. Finally, the new proposed minimum-variance portfolio that shrinks the covariance matrix accounting for the condition number (c-lw) has an excellent performance for medium and large data sets, beating the minimum-variance portfolio lw formed with the shrinkage covariance matrix of Ledoit and Wolf (2004a).

Table 6 reports the annualized Sharpe ratio adjusted by transaction costs of the shrinkage portfolios. Panel A reports the adjusted Sharpe ratios of the shrinkage portfolios calibrated via parametric assumptions (see Section 3.4). Panel B reports the adjusted Sharpe ratios of the shrinkage portfolios calibrated via bootstrap (see Section 3.5). First, we compare the results shown in Panel A. We observe that across every data set, the expected variance minimization is the best calibration method for the mv-min and mv-ew shrinkage portfolios. We also observe that the mean-squared loss minimization method offers the best adjusted Sharpe ratios for the min-ew shrinkage portfolio in medium and large data sets. This issue confirms our intuition of this calibration method. Since the min-ew shrinkage portfolio does not consider the vector of means, the mean-squared loss minimization method will offer a calibration method stable and controlling for transaction costs. This makes the adjusted Sharpe ratio greater than the adjusted Sharpe ratio for the other calibration methods.

The results given by the bootstrap calibration in Panel B are slightly different. We observe that for the mv-min portfolio the best calibration method is the expected variance minimization in small and medium data sets. But for large data sets, the Sharpe ratio maximization method gives slightly better results than the expected variance minimization method. The same occurs for the mv-ew shrinkage portfolio. For the min-ew shrinkage portfolio, we observe that the results obtained by the mean-squared loss minimization, the expected utility maximization and the expected variance minimization are quite similar. We can also observe that, compared with the parametric calibration, the bootstrap procedure to calibrate mv-ew and min-ew shrinkage portfolio is better for medium and large data sets.

Table 7 reports the turnover of the benchmark portfolios and the portfolios estimated with shrinkage moments. Panel A shows the turnover of the benchmark portfolios. Panel B shows the turnover of portfolios with shrinkage estimators using the calibration methods proposed in Section 2. We observe that the mean-variance portfolios estimated with the shrinkage estimator proposed in Section 2.1 offers lower turnover than the benchmark portfolios that consider the vector of means (mv and bs) across every data set except for the SP100 data set, where the bayes-stein mean-variance portfolio (bs) attains a lower turnover. On the other hand, we observe that the minimum-variance portfolio with shrinkage covariance matrix calibrated parametrically (f-lw) does not improve the minimum-variance portfolio with the shrinkage covariance matrix of Ledoit and Wolf (2004a) (lw) in any of the data sets, although it does outperform the results obtained by the minimum-variance portfolio with naive estimators (min). The minimum-variance portfolio formed with the shrinkage inverse covariance matrix proposed in Section 2.3 (i-lw) and the minimum-variance portfolio with a shrinkage covariance matrix that accounts for the condition number (c-lw) offers smaller turnover than portfolio lw across every data set.

Table 8 reports the turnover of shrinkage portfolios. Panel A shows the turnover of the shrinkage portfolios with parametric calibration. Panel B reports the turnover of the shrinkage portfolios with bootstrap calibration. First, we compare portfolios reported in Panel A. We observe that the expected variance minimization method offers the lowest turnover to calibrate portfolios mv-min and mv-ew. The mean-squared loss minimization method offers the lowest turnover to calibrate min-ew portfolio across every data set except for the 5IndP portfolio. It confirms our interpretation of this calibration technique pointed out in Section 3.

From Panel B of Table 8, we observe that the bootstrap procedure to calibrate shrinkage portfolios mv-ew and min-ew offers the lowest turnover in large data sets. It is mainly because the non-parametric calibration shrinks these portfolios to the equally weighted portfolio, which is the portfolio that reaches the lowest turnover among all the considered portfolios. In Panel B, the conclusions about the calibration methods in the parametric case hold for the bootstrap calibration method as well. Again, we observe that the mv-ew and min-ew shrinkage portfolios attain better results with bootstrap calibration than parametric calibration for medium and large data sets.

Table 9 reports the standard deviation of the benchmark portfolios and the portfolios estimated with shrinkage moments. Panel A shows the standard deviation of the benchmark portfolios and Panel B shows the standard deviation of portfolios composed by shrinkage estimators calibrated with the new proposed methods. We observe that among the portfolios that consider the vector of means, the portfolio formed with the estimator proposed in Section 2.1 (f-mv) offers smaller standard deviation across every data set except for the SP100 where the bayes-stein mean-variance portfolio (bs) attains lower standard deviation. On the other hand, we observe that the minimum-variance portfolio formed with the shrinkage estimator for the covariance matrix that accounts for the condition number (c-lw) gives the smallest standard deviation across every data set.

Table 10 reports the standard deviation of the shrinkage portfolios. Panel A shows the standard deviation of the shrinkage portfolios calibrated via parametric assumptions. Panel B reports the standard deviation of the shrinkage portfolios calibrated via bootstrap. We observe that for the mv-min and mv-ew shrinkage portfolios, the expected variance minimization offers the smallest standard deviation across every data set in Panel A. Moreover, we also observe that for the min-ew shrinkage portfolio, the expected utility maximization calibration technique and the expected variance minimization technique gives the same results across every data set. Furthermore, we observe that the min-ew shrinkage portfolio calibrated via mean-squared loss minimization attains the lowest standard deviation for medium and large data sets.

In panel B, we observe that the calibration of mv-min and mv-ew shrinkage portfolios via bootstrap gives very similar results among the mean-squared loss minimization, the expected utility maximization and the expected variance minimization methods.

Throughout this section, we have pointed out several key aspects. We observed the benefits of controlling for the condition number on shrinking the covariance matrix. This matrix property makes the resulting portfolio weights more stable. As a result, we obtain larger Sharpe ratios with smaller transaction costs for medium and large data sets. We have also observed that for shrinkage portfolios considering the vector of means, the expected portfolio variance minimization criteria is the most robust criteria. It gives better Sharpe ratios and turnover in general. We have also observed that the mean-squared loss criteria is very convenient for shrinkage portfolios that do not consider the vector of means.

5 Conclusions

Shrinkage estimators is a powerful tool in statistics which provide the researcher an alternative way of making inference less contaminated by estimation error. Apart from many other sources of risk, estimation error is one of the main problems within portfolio optimization. In this paper, we provide an extensive study of shrinkage techniques applied to portfolio optimization. We first study and extend the shrinkage methods to estimate the moments of a multivariate normal distribution, where we came up with a new class of shrinkage estimator for the vector of means. We also came up with an alternative way of calculating the shrinkage intensity of the shrinkage covariance matrix that accounts for its condition number, which gives very good results in terms of Sharpe ratio for medium and large data sets. We have also studied four different calibration criteria to shrink portfolio weights. Working with real data, the expected variance minimization criteria to calibrate shrinkage portfolios is the most robust method across every considered data set. Finally, we have shown that the smoothed bootstrap is a practical and simple procedure to calibrate shrinkage portfolios. Moreover, this non-parametric calibration procedure works very well in medium and large data sets.

A Proof of propositions

In this part, we proof all the propositions. Before going throughout all the propositions, we state two lemmas that will be used along the proofs:

Lemma A.1. *Let us x be a random vector in $\mathbb{R}^N$ with mean μ and covariance matrix Σ, and being A a definite positive matrix in $\mathbb{R}^{N\times N}$, the expected value of the quadratic form $x'Ax$ is:*

$$E(x'Ax) = tr\left(A\Sigma\right) + \mu' A\mu. \tag{62}$$

Proof. Trivial. □

Lemma A.2. *Given a sample $R \in \mathbb{R}^{T\times N}$ of independent and normally distributed observations, this is $R_t \sim N\left(\mu, \Sigma\right)$, the unbiased sample covariance matrix $\widehat{\Sigma} = \frac{\sum_{t=1}^{T}\left(R_t - \overline{R}\right)^2}{T-1}$, where $\overline{R} = \frac{\sum_{t=1}^{T} R_t}{T}$, has a Wishart distribution $\widehat{\Sigma} \sim \mathcal{W}\left(\frac{\Sigma}{T-1}, T-1\right)$. On the other hand, the unbiased estimator of the inverse covariance matrix $\widehat{\Sigma}_u^{-1} = \frac{T-N-2}{T-1}\widehat{\Sigma}^{-1}$ has an inverse-Wishart distribution $\widehat{\Sigma}_u^{-1} \sim \mathcal{W}^{-1}\left((T-N-2)\Sigma^{-1}, T-1\right)$. Then, the expected values of $\widehat{\Sigma}\widehat{\Sigma}$, $\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}$ and $\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}$ are:*

$$E\left(\widehat{\Sigma}\widehat{\Sigma}\right) = \frac{T}{T-1}\Sigma^2 + \frac{1}{T-1}tr\left(\Sigma\right)\Sigma. \tag{63}$$

$$E\left(\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\right) = \frac{(T-N-2)}{(T-N-1)(T-N-4)}\left(tr\left(\Sigma^{-1}\right)\Sigma^{-1} + (T-N-2)\Sigma^{-2}\right), \tag{64}$$

$$E\left(\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}\right) = \frac{(T-N-2)(T-2)}{(T-N-1)(T-N-4)}\Sigma^{-1}. \tag{65}$$

Proof. The proof for $E\left(\widehat{\Sigma}\widehat{\Sigma}\right)$ can be in Haff (1979), Theorem 3.1. The proof for $E\left(\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\right)$ and $E\left(\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}\right)$ are found in Haff (1979), Theorem 3.2. □

A.1 Proof of Proposition 2.1

In this part, we explain the proof Proposition 2.1. To do that, we first remind which is the objective function:

$$\min_{\alpha,\nu} \quad E\left(\left\|(1-\alpha)\widehat{\mu} + \alpha\nu\iota - \mu\right\|_2^2\right). \tag{66}$$

With some algebra, the above expression can be expressed as follows:

$$\min_{\alpha,\nu} \quad (1-\alpha)^2 E\left(\left\|\widehat{\mu} - \mu\right\|_2^2\right) + \alpha^2\left\|\nu\iota - \mu\right\|_2^2. \tag{67}$$

We observe that the optimal ν does not depend of α, and therefore we can obtain the optimal value ν_μ^* as the value that minimizes $\|\nu\iota - \mu\|_2^2$, which is $\nu_\mu^* = \overline{\mu}$. Once we have the optimal value of ν, we can obtain the optimal value of α by developing the first order conditions (FOC) of expression (67). It means that taking derivatives in (67) with respect to α and working out some elements, the optimal α will be:

$$\alpha_\mu^* = \frac{E\left(\|\widehat{\mu} - \mu\|_2^2\right)}{E\left(\|\widehat{\mu} - \mu\|_2^2\right) + \|\nu_\mu\iota - \mu\|_2^2} = \frac{E\left(\widehat{\mu}'\widehat{\mu}\right) - \mu'\mu}{E\left(\widehat{\mu}'\widehat{\mu}\right) - \mu'\mu + \|\nu_\mu\iota - \mu\|_2^2}. \tag{68}$$

Since $\widehat{\mu} \sim N\left(\mu, \frac{\Sigma}{T}\right)$ and according with Lemma A.1, the above expression is:

$$\alpha_\mu^* = \frac{(1/T)\overline{\sigma^2}}{(1/T)\overline{\sigma^2} + \overline{\mu}_2 - \overline{\mu}^2}, \tag{69}$$

and this completes the proof.

A.2 Proof of Proposition 2.2

In this part we explain the proof of Proposition 2.2. Like in the proof of Proposition 2.1, we develop the objective function in problem (9)-(10). The resulting optimization problem is the following:

$$\min_{\alpha,\nu} \quad (1-\alpha)^2 E\left(\left\|\widehat{\Sigma} - \Sigma\right\|_F^2\right) + \alpha^2 \left\|\nu I - \Sigma\right\|_F^2. \tag{70}$$

From the above expression, we observe that the optimal scale factor ν does not depend on α, thus applying the FOC's to $\|\nu I - \Sigma\|_F^2$ we obtain the optimal value of ν, which is:

$$\nu_\Sigma^* = \overline{\sigma^2}. \tag{71}$$

On the other hand, applying the FOC's in problem (70) we obtain the optimal value of α:

$$\alpha_\Sigma^* = \frac{E\left(\left\|\widehat{\Sigma} - \Sigma\right\|_F^2\right)}{E\left(\left\|\widehat{\Sigma} - \Sigma\right\|_F^2\right) + \|\nu_\Sigma^* I - \Sigma\|_F^2}. \tag{72}$$

Since $E\left(\left\|\widehat{\Sigma}-\Sigma\right\|_F^2\right) = \text{tr}\left(E\left(\widehat{\Sigma}\widehat{\Sigma}\right)\right) - \text{tr}\left(\Sigma\Sigma\right)$ and according with Lemma A.2, the optimal α takes the following form:

$$\alpha_\Sigma^* = \frac{\frac{N}{T-1}\left(\frac{\text{tr}\left(\Sigma^2\right)}{N} + N\left(\overline{\sigma^2}\right)^2\right)}{\frac{N}{T-1}\left(\frac{T}{N}\text{tr}\left(\Sigma^2\right) - (T-N-1)\left(\overline{\sigma^2}\right)^2\right)}, \tag{73}$$

and this completes the proof.

A.3 Proof of Proposition 2.3

The proof of Proposition 2.3 is exactly as the proof made for Proposition 2.2. We first develop the objective function of problem (14)-(15) for tractability. Then with we develop the FOC's to obtain the optimal values of the scale factor and the shrinkage intensity. To obtain the optimal value of the shrinkage intensity we make use of Lemma A.2 in a similar manner as it was done in the proof of Proposition 2.2.

A.4 Proof of Proposition 3.1

To make the proof of this proposition we simply develop the FOC's of the calibration functions defined by the shrinkage portfolio formed by the mean-variance and minimum-variance portfolios. In this portfolio, the scales parameter comes from the minimization of $\nu = \text{argmin}\left\{\text{E}\left\|\nu\widehat{w}_{\text{min}} - \widehat{w}_{\text{mv}}\right\|_2^2\right\}$ with respect to ν. Developing the FOC's, we obtain that the optimal scale factor is $\nu_{mv-min} = \frac{w_{min}' w_{mv}}{E\left(\widehat{w}_{min}'\widehat{w}_{min}\right)}$. First, the mean-squared loss function of the considered shrinkage portfolio is:

$$E\left(\left\|\widehat{w}_{mv-min} - w_{mv}\right\|_2^2\right) = E\left(\left\|(1-\alpha)\left(\widehat{w}_{mv} - w_{mv}\right) + \alpha\left(\nu\widehat{w}_{min} - w_{mv}\right)\right\|_2^2\right) = \tag{74}$$

$$= (1-\alpha)^2 E\left(\left\|\widehat{w}_{mv} - w_{mv}\right\|_2^2\right) + \alpha^2 E\left(\left\|\nu\widehat{w}_{min} - w_{mv}\right\|_2^2\right) + \tag{75}$$

$$+ 2(1-\alpha)\alpha E\left(\left(\widehat{w}_{mv} - w_{mv}\right)'\left(\nu\widehat{w}_{min} - w_{mv}\right)\right). \tag{76}$$

The third element in the right hand side can be simplified to:

$$E\left(\left(\widehat{w}_{mv} - w_{mv}\right)'\left(\nu\widehat{w}_{min} - w_{mv}\right)\right) = \nu\left(E\left(\widehat{w}_{mv}'\widehat{w}_{min}\right) - w_{mv}' w_{min}\right). \tag{77}$$

Therefore, developing the FOC's of $E\left(\|\widehat{w}_{mv-min}-w_{mv}\|_2^2\right)$, we obtain that the optimal α is:

$$\alpha_{msl}=\frac{E\left(\|\widehat{w}_{mv}-w_{mv}\|_2^2\right)-\tau}{E\left(\|\widehat{w}_{mv}-w_{mv}\|_2^2\right)+E\left(\|\nu\widehat{w}_{min}-w_{mv}\|_2^2\right)-2\tau}, \tag{78}$$

where $\tau=\nu\left(E\left(\widehat{w}_{min}'\widehat{w}_{mv}\right)-w_{min}'w_{mv}\right)$. In this expression, $E\left(\widehat{w}_{min}'\widehat{w}_{mv}\right)=$
$=\frac{1}{\gamma}\frac{(T-N-2)}{(T-N-1)(T-N-4)}\left(\text{tr}\left(\Sigma^{-1}\right)\iota'\Sigma^{-1}\mu+(T-N-2)\iota'\Sigma^{-2}\mu\right)$. which can be obtained from Lemma A.2.

Second, the expected utility function of the shrinkage portfolio is:

$$E\left(U\left(\widehat{w}_{mv}\right)\right)=(1-\alpha)w_{mv}'\mu+\alpha\nu w_{min}'\mu- \tag{79}$$

$$-\frac{\gamma}{2}E\left((1-\alpha)^2\widehat{w}_{mv}'\Sigma\widehat{w}_{mv}+\alpha^2\nu^2\widehat{w}_{min}'\Sigma\widehat{w}_{min}+2(1-\alpha)\alpha\nu\widehat{w}_{mv}'\Sigma\widehat{w}_{min}\right). \tag{80}$$

Deriving the FOC's of the above expression, we obtain the optimal α:

$$\alpha_{ut}=\frac{E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{mv}\right)-\nu E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{min}\right)}{E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{mv}\right)+\nu^2E\left(\widehat{w}_{min}'\Sigma\widehat{w}_{min}\right)-2\nu E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{min}\right)}- \tag{81}$$

$$-\frac{w_{mv}'\mu-\nu w_{min}'\mu}{E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{mv}\right)+\nu^2E\left(\widehat{w}_{min}'\Sigma\widehat{w}_{min}\right)-2\nu E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{min}\right)}. \tag{82}$$

The optimal expression can be shorten by using the abbreviations used in Proposition 3.4:

$$\alpha_{ut}=\frac{E\left(\widehat{\sigma}_{mv}^2\right)-\nu E\left(\widehat{\rho}_{mv,min}\right)}{E\left(\widehat{\sigma}_{mv}^2\right)+\nu^2E\left(\widehat{\sigma}_{min}^2\right)-2\nu E\left(\widehat{\rho}_{mv,min}\right)}- \tag{83}$$

$$-\frac{\mu_{mv}-\nu\mu_{min}}{E\left(\widehat{\sigma}_{mv}^2\right)+\nu^2E\left(\widehat{\sigma}_{min}^2\right)-2\nu E\left(\widehat{\rho}_{mv,min}\right)}\frac{1}{\gamma}. \tag{84}$$

The proof of the variance is straightforward. From the utility framework, we assume that the investor has an infinite risk aversion level ($\gamma\approx\infty$), which means that this investor concerns only about the variance and is almost indifferent with respect to the profitability of his portfolio. In that case, the ratio $\frac{\mu_{mv}-\nu\mu_{min}}{E\left(\widehat{\sigma}_{mv}^2\right)+\nu^2E\left(\widehat{\sigma}_{min}^2\right)-2\nu E\left(\widehat{\rho}_{mv,min}\right)}\frac{1}{\gamma}$ vanishes and therefore, the optimal α is:

$$\alpha_{var}=\frac{E\left(\widehat{\sigma}_{mv}^2\right)-\nu E\left(\widehat{\rho}_{mv,min}\right)}{E\left(\widehat{\sigma}_{mv}^2\right)+\nu^2E\left(\widehat{\sigma}_{min}^2\right)-2\nu E\left(\widehat{\rho}_{mv,min}\right)}. \tag{85}$$

A.5 Proof of Proposition 3.2

We should develop the FOC's of the calibration functions defined by the shrinkage portfolio formed by the mean-variance and equally weighted portfolios. In this portfolio, the scale parameter comes from the minimization problem $\nu = \text{argmin}\left\{\mathrm{E}\left\|\nu\frac{\iota}{N} - \widehat{w}_{\mathrm{mv}}\right\|_2^2\right\}$. Developing the FOC's, we obtain that the optimal scale factor is $\nu_{mv-ew} = \frac{w_{ew}' w_{mv}}{w_{ew}' w_{ew}}$. First, the mean-squared loss function of the considered shrinkage portfolio is:

$$E\left(\|\widehat{w}_{mv-ew} - w_{mv}\|_2^2\right) = E\left(\|(1-\alpha)\left(\widehat{w}_{mv} - w_{mv}\right) + \alpha\left(\nu w_{ew} - w_{mv}\right)\|_2^2\right) = \tag{86}$$

$$= (1-\alpha)^2 E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right) + \alpha^2\left\|\nu w_{ew} - w_{mv}\right\|_2^2, \tag{87}$$

where $w_{ew} = \frac{\iota}{N}$. Now, deriving the FOC's of the above expression, we obtain the optimal α:

$$\alpha_{msl} = \frac{E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right)}{E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right) + \|\nu w_{ew} - w_{mv}\|_2^2}. \tag{88}$$

Second, the expected utility function of the shrinkage portfolio is:

$$E\left(U\left(\widehat{w}_{mv}\right)\right) = (1-\alpha)w_{mv}'\mu + \alpha\nu w_{ew}'\mu - \tag{89}$$

$$-\frac{\gamma}{2}E\left((1-\alpha)^2\widehat{w}_{mv}'\Sigma\widehat{w}_{mv} + \alpha^2\nu^2 w_{ew}'\Sigma w_{ew} + 2(1-\alpha)\alpha\nu\widehat{w}_{mv}'\Sigma w_{ew}\right). \tag{90}$$

Deriving the FOC's of the above expression, we obtain the optimal α:

$$\alpha_{ut} = \frac{E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{mv}\right) - \nu w_{mv}'\Sigma w_{ew}}{E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{mv}\right) + \nu^2 w_{ew}'\Sigma w_{ew} - 2\nu w_{mv}'\Sigma w_{ew}} - \tag{91}$$

$$-\frac{w_{mv}'\mu - \nu w_{ew}'\mu}{E\left(\widehat{w}_{mv}'\Sigma\widehat{w}_{mv}\right) + \nu^2 w_{ew}'\Sigma w_{ew} - 2\nu w_{mv}'\Sigma w_{ew}}. \tag{92}$$

The optimal expression can be shorten by using the abbreviations used in Proposition 3.4:

$$\alpha_{ut} = \frac{E\left(\widehat{\sigma}_{mv}^2\right) - \nu\rho_{mv,ew}}{E\left(\widehat{\sigma}_{mv}^2\right) + \nu^2\sigma_{ew}^2 - 2\nu\rho_{mv,ew}} - \frac{\mu_{mv} - \nu\mu_{ew}}{E\left(\widehat{\sigma}_{mv}^2\right) + \nu^2\sigma_{ew}^2 - 2\nu\rho_{mv,ew}}\frac{1}{\gamma}. \tag{93}$$

To make the proof of the variance we proceed as in the proof of Proposition 3.1. From the utility framework, we assume that the investor has an infinite risk aversion level ($\gamma \approx \infty$), which means that this investor concerns only about the variance and is almost indifferent with respect to the profitability of his portfolio. In that case, the ratio $\frac{\mu_{mv} - \nu\mu_{ew}}{E(\widehat{\sigma}_{mv}^2) + \nu^2\sigma_{ew}^2 - 2\nu\rho_{mv,ew}}\frac{1}{\gamma}$ vanishes and therefore, the optimal α is:

$$\alpha_{var} = \frac{E\left(\widehat{\sigma}_{mv}^2\right) - \nu\rho_{mv,ew}}{E\left(\widehat{\sigma}_{mv}^2\right) + \nu^2\sigma_{ew}^2 - 2\nu\rho_{mv,ew}}. \tag{94}$$

A.6 Proof of Proposition 3.3

We should develop the FOC's of the calibration functions defined by the shrinkage portfolio formed by the mean-variance and equally weighted portfolios. In this portfolio, the scale parameter comes from the minimization problem $\nu = \text{argmin} \left\{ \mathrm{E} \left\| \nu \frac{\iota}{N} - \widehat{w}_{\text{min}} \right\|_2^2 \right\}$. Developing the FOC's, we obtain that the optimal scale factor is $\nu_{min-ew} = \frac{w'_{ew} w_{min}}{w'_{ew} w_{ew}}$. First, the mean-squared loss function of the considered shrinkage portfolio is:

$$E\left(\|\widehat{w}_{min-ew} - w_{min}\|_2^2\right) = E\left(\|(1-\alpha)\left(\widehat{w}_{min} - w_{min}\right) + \alpha\left(\nu w_{ew} - w_{min}\right)\|_2^2\right) = \tag{95}$$

$$= (1-\alpha)^2 E\left(\|\widehat{w}_{min} - w_{min}\|_2^2\right) + \alpha^2 \|\nu w_{ew} - w_{min}\|_2^2, \tag{96}$$

where $w_{ew} = \frac{\iota}{N}$. Now, deriving the FOC's of the above expression, we obtain the optimal α:

$$\alpha_{msl} = \frac{E\left(\|\widehat{w}_{min} - w_{min}\|_2^2\right)}{E\left(\|\widehat{w}_{min} - w_{min}\|_2^2\right) + \|\nu w_{ew} - w_{min}\|_2^2} \tag{97}$$

Second, the expected utility function of the shrinkage portfolio is:

$$E\left(U\left(\widehat{w}_{min}\right)\right) = (1-\alpha) w'_{min}\mu + \alpha\nu w'_{ew}\mu - \tag{98}$$

$$- \frac{\gamma}{2} E\left((1-\alpha)^2 \widehat{w}'_{min}\Sigma\widehat{w}_{min} + \alpha^2\nu^2 w'_{ew}\Sigma w_{ew} + 2(1-\alpha)\alpha\nu\widehat{w}'_{min}\Sigma w_{ew}\right). \tag{99}$$

Deriving the FOC's of the above expression, we obtain the optimal α:

$$\alpha_{ut} = \frac{E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right) - \nu w'_{min}\Sigma w_{ew}}{E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right) + \nu^2 w'_{ew}\Sigma w_{ew} - 2\nu w'_{min}\Sigma w_{ew}} - \tag{100}$$

$$- \frac{w'_{min}\mu - \nu w'_{min}\mu}{E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right) + \nu^2 E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right) - 2\nu E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{ew}\right)}. \tag{101}$$

The optimal expression can be shorten by using the abbreviations used in Proposition 3.4:

$$\alpha_{ut} = \frac{E\left(\widehat{\sigma}^2_{min}\right) - \nu\rho_{min,ew}}{E\left(\widehat{\sigma}^2_{min}\right) + \nu^2\sigma^2_{ew} - 2\nu\rho_{min,ew}} - \frac{\mu_{min} - \nu\mu_{ew}}{E\left(\widehat{\sigma}^2_{min}\right) + \nu^2\sigma^2_{ew} - 2\nu\rho_{min,ew}} \frac{1}{\gamma} \tag{102}$$

To make the proof of the variance we proceed as in the proof of Proposition 3.2. From the utility framework, we assume that the investor has an infinite risk aversion level ($\gamma \approx \infty$), which means that this investor concerns only about the variance and is almost indifferent with respect to the profitability of his portfolio. In that case, the ratio $\frac{\mu_{min} - \nu\mu_{ew}}{E\left(\widehat{\sigma}^2_{min}\right) + \nu^2\sigma^2_{ew} - 2\nu\rho_{min,ew}} \frac{1}{\gamma}$

vanishes and therefore, the optimal α is:

$$\alpha_{var} = \frac{E\left(\widehat{\sigma}^2_{min}\right) - \nu\rho_{min,ew}}{E\left(\widehat{\sigma}^2_{min}\right) + \nu^2\sigma^2_{ew} - 2\nu\rho_{min,ew}} \tag{103}$$

A.7 Proof of proposition 3.4

Proof. Here, we illustrate how to the prove of Proposition 3.4. We develop each element mentioned in the Lemma. First, we show how to obtain $E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right)$:

$$E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right) = \frac{1}{\gamma^2}\left(E\left(\widehat{\mu}\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\widehat{\mu}\right) - \mu\Sigma^{-1}\Sigma{-1}\mu\right). \tag{104}$$

Due to the fact that returns are assumed to be independent and normally distributed, $\widehat{\mu}$ and $\widehat{\Sigma}$ are independent. Therefore, we can make use of Lemma A.1 and Lemma A.2 to compute the expected value of $E\left(\widehat{\mu}\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\widehat{\mu}\right)$. Thus, using the independence between $\widehat{\mu}$ and $\widehat{\Sigma}$, the expected value of $\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}$ given in Lemma A.2 and the expected value of quadratic forms given in Lemma A.1, we have:

$$E\left(\|\widehat{w}_{mv} - w_{mv}\|_2^2\right) = \frac{1}{\gamma^2}\left[\frac{\text{tr}\left(\Sigma^{-1}\right)(T-N-2)(T-2)}{(T-N-1)(T-N-4)T} + \right. \tag{105}$$

$$\left. + \frac{(T-N-2)}{(T-N-1)(T-N-4)}\left[\text{tr}\left(\Sigma^{-1}\right)\mu'\Sigma^{-1}\mu + (T-N-2)\mu'\Sigma^{-2}\mu\right]\right] - \tag{106}$$

$$- \frac{1}{\gamma^2}\mu'\Sigma^{-2}\mu. \tag{107}$$

The following element is $E\left(\|\nu\widehat{w}_{min} - w_{mv}\|_2^2\right)$:

$$E\left(\|\nu\widehat{w}_{min} - w_{mv}\|_2^2\right) = \nu^2 E\left(\iota'\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\iota\right) + \frac{1}{\gamma^2}\mu'\Sigma^{-2}\mu - 2\frac{\nu}{\gamma}\iota'\Sigma^{-2}\mu. \tag{108}$$

Using the value of $E\left(\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\right)$ given in Lemma A.2, we have that:

$$E\left(\|\nu\widehat{w}_{min} - w_{mv}\|_2^2\right) = \nu^2\frac{(T-N-2)}{(T-N-1)(T-N-4)}\left[\text{tr}\left(\Sigma^{-1}\right)\iota'\Sigma^{-1}\iota + \right. \tag{109}$$

$$\left. + (T-N-2)\iota'\Sigma^{-2}\iota\right] + \frac{1}{\gamma^2}\mu'\Sigma^{-2}\mu - 2\frac{\nu}{\gamma}\iota'\Sigma^{-2}\mu. \tag{110}$$

Now, we prove how to obtain the closed form expression of $E\left(\|\widehat{w}_{min} - w_{min}\|_2^2\right)$. First, we expand the expression as usual:

$$E\left(\|\widehat{w}_{min} - w_{mv}\|_2^2\right) = E\left(\iota'\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\iota\right) - \iota'\Sigma^{-2}\iota. \tag{111}$$

Again, applying the value of $E\left(\widehat{\Sigma}^{-1}\widehat{\Sigma}^{-1}\right)$ given in Lemma A.2, we obtain the following:

$$E\left(\|\widehat{w}_{min} - w_{mv}\|_2^2\right) = \frac{(T-N-2)}{(T-N-1)(T-N-4)}\left[\text{tr}\left(\Sigma^{-1}\right)\iota'\Sigma^{-1}\iota + \right. \tag{112}$$

$$\left. +(T-N-2)\iota'\Sigma^{-2}\iota\right] - \iota'\Sigma^{-2}\iota. \tag{113}$$

The remaining three elements are very straightforward to prove. Understanding how to apply LemmaA.1 and Lemma A.2, expressions $E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{mv}\right)$, $E\left(\widehat{w}'_{min}\Sigma\widehat{w}_{min}\right)$ and $E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{min}\right)$ are simple to obtain. For instance,

$$E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{mv}\right) = \frac{1}{\gamma^2}E\left(\widehat{\mu}\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}\widehat{\mu}\right). \tag{114}$$

Since $\widehat{\mu}$ and $\widehat{\Sigma}$ are independent, using Lemma A.1 and the expression for $E\left(\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}\right)$ given in Lemma A.2, we have:

$$E\left(\widehat{w}'_{mv}\Sigma\widehat{w}_{mv}\right) = \frac{1}{\gamma^2}\left(\frac{(T-N-2)(T-2)}{(T-N-1)(T-N-4)}\left(\frac{N}{T} + \mu'\Sigma^{-1}\mu\right)\right). \tag{115}$$

The proof of the remaining two elements can be skipped having understood the steps of the previous proofs. □

B Tables

Table 1: List of Data Sets:
This table list the various data sets analyzed, the abbreviation used to identify each data set, the number of assets N contained in each data set, the time period spanned by the data set, and the source of the data. The data set of CRSP returns (SP100) is constructed in a way similar to Jagannathan and Ma (2003), with monthly rebalancing: in January of each year we randomly select 100 assets as our asset universe for the next 12 months.

[a] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[b] CRSP, The Center for Research in Security Prices

#	Data Set	Abreviation	N	Time Period	Source
1	5 Industry Portfolios representing the US stock market	5Ind	5	01/1972-06/2009	K. French[a]
2	10 Industry Portfolios representing the US stock market	10Ind	10	01/1972-06/2009	K. French
3	38 Industry Portfolios representing the U.S stock market	38IndP	38	01/1972-06/2009	K. French
4	48 Industry Portfolios representing the U.S. stock market	48Ind	48	01/1972-06/2009	K. French
5	100 Fama and French Portfolios of firms sorted by size and book to market	100FF	96	01/1972-06/2009	K. French
6	100 randomized stocks from S&P 500	SP100	100	01/1988-12/2008	CRSP[b]

Table 2: List of portfolio models:
This table lists the various portfolio strategies considered in the paper. Panel A lists the existing portfolios from the literature. Panel B list portfolios where the moments are shrunk with the methods proposed in section 2. Panel C lists the shrinkage portfolio defined in section 3. The last columns gives the abbreviation that we use to refer to the strategy.

#	Policy	Abbreviation
	Panel A: Benchmark portfolios	
1	Classical mean-variance portfolio	mv
2	Bayes-Stein mean-variance portfolio	bs
3	Kan-Zhou's (2007) three-fund portfolio	kz
4	Mixture of mean-variance and equally weighted (Tu and Zhou (2011))	tz
5	Mixture of minimum-variance and equally weighted DeMiguel et.al. (2009))	dm
6	Minimum-Variance portfolio	min
7	Minimum-variance portfolio with Ledoit and Wolf (2004) shrinkage covariance matrix	lw
8	Equally weighted portfolio	ew or 1/N
	Panel B: Portfolios estimated with new calibration procedures to shrink moments	
Shrinkage minimum-variance portfolio		
9	Finite optimal shrinkage intensity of Ledoit and Wolf's shrinkage covariance matrix	f-lw
10	Finite optimal shrinkage intensity of the inverse covariance matrix based on Ledoit and Wolf (2004) framework	i-lw
11	Finite optimal shrinkage intensity of Ledoit and Wolf's shrinkage covariance matrix considering its condition number	c-lw
Shrinkage mean-variance portfolio		
12	Mean-variance portfolio formed with the finite optimal shrinkage intensity of the vector of means	f-mv
	Panel C: Shrinkage portfolios	
13	Mixture of mean-variance and scaled minimum-variance portfolios	mv-min
14	Mixture of mean-variance and scaled equally weighted portfolios	mv-ew
15	Mixture of minimum-variance and scaled equally weighted portfolios	min-ew

Table 3: Theoretical Sharpe ratios:
This table shows an experiment where we have computed the theoretical expected values for the Sharpe Ratio of different portfolios. We consider four data sets of N=5, 10, 38, 48 and 96 assets. Moreover, we consider a window's length of T=150, which is a common estimation window's length when we deal with monthly data.

Policy	5IndP	10IndP	38IndP	48IndP	100FF
Benchmark Portfolios					
mv	0.193	0.212	0.185	0.198	0.278
min	0.237	0.263	0.240	0.210	0.242
ew	0.202	0.212	0.188	0.181	0.186
Shrinkage Portfolios MSL Minimization					
mv-min	0.237	0.266	0.253	0.247	0.288
mv-ew	0.217	0.240	0.216	0.228	0.261
min-ew	0.236	0.260	0.243	0.224	0.224
Shrinkage Portfolios Utility Maximization					
mv-min	0.236	0.266	0.252	0.247	0.288
mv-ew	0.218	0.243	0.231	0.239	0.297
min-ew	0.235	0.262	0.247	0.226	0.266
Shrinkage Portfolios Variance Minimization					
mv-min	0.225	0.266	0.240	0.210	0.242
mv-ew	0.216	0.239	0.223	0.219	0.256
min-ew	0.235	0.262	0.247	0.226	0.266
Shrinkage Portfolios Sharpe Ratio Maximization					
mv-min	0.237	0.266	0.254	0.247	0.314
mv-ew	0.218	0.243	0.233	0.244	0.320
min-ew	0.238	0.263	0.247	0.228	0.268

Table 4: Theoretical RIAL:
This table shows an experiment where we have computed the theoretical expected values for the RIAL's of different portfolios. We consider four data sets of N=5, 10, 38, 48 and 96 assets. Moreover, we consider a window's length of T=150, which is a common estimation window's length when we deal with monthly data.

Policy	5IndP	10IndP	38IndP	48IndP	100FF
Shrinkage Portfolios MSL Minimization					
mv-min	0.886	0.764	0.839	0.809	0.787
mv-ew	0.514	0.682	0.843	0.824	0.869
min-ew	0.106	0.251	0.398	0.469	0.780
Shrinkage Portfolios Utility Maximization					
mv-min	0.865	0.763	0.838	0.809	0.787
mv-ew	0.496	0.669	0.817	0.816	0.860
min-ew	0.074	0.208	0.282	0.348	0.632
Shrinkage Portfolios Variance Minimization					
mv-min	0.634	0.763	0.820	0.779	0.774
mv-ew	0.435	0.681	0.840	0.821	0.869
min-ew	0.074	0.208	0.282	0.348	0.632
Shrinkage Portfolios Sharpe Ratio Maximization					
mv-min	0.885	0.763	0.838	0.809	0.738
mv-ew	0.503	0.657	0.785	0.783	0.781
min-ew	-0.536	0.049	0.318	0.424	0.548

Table 5: Annualized Sharpe ratio with transaction costs (50 basis points) of benchmark portfolios and portfolios estimated with shrinkage moments:

This table reports the out-of-sample annualized Sharpe ratio of benchmark portfolios and portfolios calculated using shrinkage estimator calibrated with the new techniques proposed in this paper. Moreover, we adjust the Sharpe ratio with transaction costs, assuming 50 basis points.

Policy	5IndP	10IndP	38IndP	48IndP	100FF	SP100
Panel A: Benchmark Portfolios						
Portfolios that consider the vector of means						
mv	0.593	0.599	0.004	-0.003	-0.463	-0.102
bs	0.735	0.824	0.294	0.213	-0.205	0.275
Portfolios that do not consider the vector of means						
min	0.841	0.945	0.528	0.378	-0.014	0.399
lw	0.863	0.955	0.731	0.651	1.003	0.687
Portfolios that do not make optimization						
1/N	0.761	0.780	0.695	0.688	0.712	0.328
Existing mixture of portfolios						
kz	0.714	0.807	0.284	0.208	-0.208	0.273
tz	0.673	0.704	0.301	0.360	0.083	0.195
dm	0.813	0.941	0.603	0.508	0.104	0.472
Panel B: New portfolios with shrinkage estimators						
Portfolios that consider the vector of means						
f-mv	0.809	0.880	0.402	0.269	0.017	0.257
Portfolios that do not consider the vector of means						
f-lw	0.845	0.945	0.643	0.553	0.762	0.641
i-lw	0.877	0.907	0.716	0.693	0.713	0.337
c-lw	0.852	0.947	0.805	0.766	1.152	0.700

Table 6: Annualized Sharpe ratio with transaction costs (50 basis points) of shrinkage portfolios:
This table reports the out-of-sample annualized Sharpe ratio of shrinkage portfolios. Moreover, we adjust the Sharpe ratio with transaction costs, assuming 50 basis points.

Policy	5IndP	10IndP	38IndP	48IndP	100FF	SP100
Panel A: Shrinkage portfolio with parametric calibration						
Shrinkage portfolios: MSL Minimization						
mv-min	0.702	0.812	0.313	0.274	-0.198	0.307
mv-ew	0.647	0.760	0.483	0.437	0.028	0.188
min-ew	0.844	0.954	0.658	0.587	0.435	0.509
Shrinkage portfolios: Utility Maximization						
mv-min	0.708	0.790	0.303	0.305	-0.025	0.388
mv-ew	0.691	0.738	0.386	0.436	0.205	0.234
min-ew	0.847	0.949	0.593	0.472	0.165	0.483
Shrinkage portfolios: Variance Minimization						
mv-min	0.820	0.925	0.527	0.374	-0.054	0.414
mv-ew	0.773	0.796	0.594	0.670	0.425	0.277
min-ew	0.847	0.949	0.593	0.472	0.165	0.483
Shrinkage portfolios: Sharpe Ratio Maximization						
mv-min	0.714	0.807	0.284	0.208	-0.208	0.273
mv-ew	0.685	0.715	0.249	0.254	-0.188	0.111
min-ew	0.817	0.955	0.606	0.508	0.104	0.472
Panel B: Shrinkage portfolios with bootstrap calibration						
Shrinkage portfolios: MSL Minimization						
mv-min	0.708	0.795	0.277	-0.356	-0.612	0.398
mv-ew	0.657	0.769	0.575	0.596	0.713	0.329
min-ew	0.853	0.952	0.711	0.675	0.712	0.330
Shrinkage portfolios: Utility Maximization						
mv-min	0.714	0.771	0.264	0.158	-0.471	0.399
mv-ew	0.701	0.750	0.514	0.615	0.712	0.328
min-ew	0.844	0.945	0.688	0.662	0.712	0.328
Shrinkage portfolios: Variance Minimization						
mv-min	0.821	0.931	0.528	0.374	-0.093	0.399
mv-ew	0.757	0.798	0.641	0.685	0.712	0.328
min-ew	0.855	0.948	0.684	0.660	0.712	0.328
Shrinkage portfolios: Sharpe Ratio Maximization						
mv-min	0.702	0.793	0.261	0.050	-0.471	0.153
mv-ew	0.689	0.714	0.308	0.401	0.714	0.331
min-ew	0.824	0.955	0.627	0.590	0.713	0.335

Table 7: Turnover of benchmark portfolios and portfolios estimated with shrinkage moments:
This table reports the out-of-sample Turnover of benchmark portfolios and portfolios calculated using shrinkage estimator calibrated with the new techniques proposed in this paper.

Policy	5IndP	10IndP	38IndP	48IndP	100FF	SP100
Panel A: Benchmark Portfolios						
Portfolios that consider the vector of means						
mv	0.287	0.369	1.336	2.311	8.791	2.075
bs	0.150	0.197	0.693	1.111	5.257	1.262
Portfolios that do not consider the vector of means						
min	0.081	0.124	0.395	0.567	2.775	1.156
lw	0.056	0.089	0.240	0.321	0.793	0.279
Portfolios that do not make optimization						
1/N	0.018	0.025	0.032	0.033	0.023	0.054
Existing mixture of portfolios						
kz	0.167	0.208	0.709	1.130	5.291	1.265
tz	0.175	0.242	0.717	1.009	3.626	0.708
dm	0.124	0.120	0.310	0.424	2.334	0.847
Panel B: New portfolios with shrinkage estimators						
Portfolios that consider the vector of means						
f-mv	0.096	0.149	0.491	0.807	3.686	1.271
Portfolios that do not consider the vector of means						
f-lw	0.065	0.102	0.297	0.402	1.224	0.350
i-lw	0.044	0.051	0.042	0.039	0.024	0.058
c-lw	0.051	0.080	0.186	0.237	0.453	0.216

Table 8: Turnover of shrinkage portfolios:
This table reports the out-of-sample Turnover of shrinkage portfolios.

Policy	5IndP	10IndP	38IndP	48IndP	100FF	SP100
Panel A: New shrinkage portfolios with parametric calibration						
Shrinkage portfolios: MSL Minimization						
mv-min	0.173	0.228	0.716	1.004	4.772	1.260
mv-ew	0.177	0.173	0.516	0.726	3.416	0.621
min-ew	0.074	0.103	0.279	0.350	1.389	0.580
Shrinkage portfolios: Utility Maximization						
mv-min	0.197	0.241	0.694	0.945	3.454	1.177
mv-ew	0.155	0.200	0.592	0.821	2.944	0.636
min-ew	0.072	0.112	0.340	0.474	2.165	0.822
Shrinkage portfolios: Variance Minimization						
mv-min	0.119	0.132	0.398	0.571	2.965	1.178
mv-ew	0.076	0.115	0.279	0.356	1.970	0.517
min-ew	0.072	0.112	0.340	0.474	2.165	0.822
Shrinkage portfolios: Sharpe Ratio Maximization						
mv-min	0.167	0.208	0.709	1.130	5.291	1.265
mv-ew	0.163	0.222	0.786	1.213	5.495	0.991
min-ew	0.109	0.113	0.308	0.424	2.334	0.847
Panel B: New shrinkage portfolios with bootstrap calibration						
Shrinkage portfolios: MSL Minimization						
mv-min	0.181	0.250	0.786	4.119	7.627	1.157
mv-ew	0.171	0.159	0.371	0.399	0.023	0.054
min-ew	0.071	0.093	0.192	0.181	0.023	0.054
Shrinkage portfolios: Utility Maximization						
mv-min	0.211	0.263	0.776	10.763	4.236	1.156
mv-ew	0.143	0.182	0.411	0.408	0.023	0.054
min-ew	0.093	0.097	0.228	0.222	0.023	0.054
Shrinkage portfolios: Variance Minimization						
mv-min	0.117	0.135	0.395	0.589	3.003	1.156
mv-ew	0.098	0.107	0.181	0.146	0.023	0.054
min-ew	0.090	0.097	0.230	0.221	0.023	0.054
Shrinkage portfolios: Sharpe Ratio Maximization						
mv-min	0.178	0.224	0.758	1.510	7.953	1.437
mv-ew	0.159	0.222	0.694	0.868	0.025	0.055
min-ew	0.102	0.111	0.275	0.317	0.024	0.056

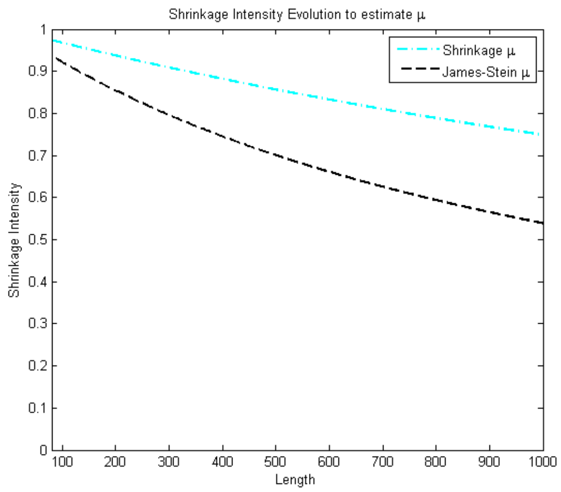
Table 9: Standard deviation of benchmark portfolios and portfolios estimated with shrinkage moments:
This table reports the out-of-sample standard deviation of benchmark portfolios and portfolios calculated using shrinkage estimator calibrated with the new techniques proposed in this paper.

Policy	5IndP	10IndP	38IndP	48IndP	100FF	SP100
Panel A: Benchmark Portfolios						
Portfolios that consider the vector of means						
mv	0.161	0.157	0.244	0.336	0.417	0.267
bs	0.142	0.133	0.159	0.201	0.276	0.176
Portfolios that do not consider the vector of means						
min	0.138	0.126	0.131	0.137	0.179	0.171
lw	0.136	0.124	0.120	0.124	0.125	0.122
Portfolios that do not make optimization						
1/N	0.154	0.148	0.166	0.165	0.174	0.169
Existing mixture of portfolios						
kz	0.144	0.135	0.161	0.203	0.277	0.176
tz	0.150	0.143	0.174	0.202	0.222	0.149
dm	0.136	0.125	0.126	0.132	0.166	0.142
Panel B: New portfolios with shrinkage estimators						
Portfolios that consider the vector of means						
f-mv	0.138	0.128	0.139	0.169	0.205	0.178
Portfolios that do not consider the vector of means						
f-lw	0.136	0.125	0.124	0.128	0.137	0.125
i-lw	0.138	0.132	0.159	0.161	0.174	0.167
c-lw	0.136	0.124	0.120	0.123	0.122	0.121

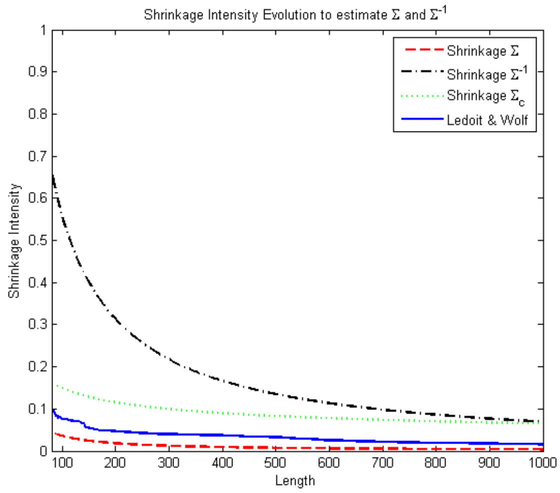
Table 10: Standard deviation of shrinkage portfolios:
This table reports the out-of-sample standard deviation of shrinkage portfolios.

Policy	5IndP	10IndP	38IndP	48IndP	100FF	SP100
Panel A: New shrinkage portfolios with parametric calibration						
Shrinkage portfolios: MSL Minimization						
mv-min	0.145	0.137	0.159	0.183	0.251	0.173
mv-ew	0.152	0.140	0.157	0.177	0.216	0.149
min-ew	0.138	0.125	0.126	0.130	0.146	0.134
Shrinkage portfolios: Utility Maximization						
mv-min	0.145	0.137	0.158	0.177	0.195	0.170
mv-ew	0.149	0.140	0.165	0.186	0.198	0.147
min-ew	0.137	0.125	0.127	0.132	0.160	0.142
Shrinkage portfolios: Variance Minimization						
mv-min	0.141	0.127	0.131	0.137	0.184	0.172
mv-ew	0.146	0.138	0.155	0.162	0.174	0.145
min-ew	0.137	0.125	0.127	0.132	0.160	0.142
Shrinkage portfolios: Sharpe Ratio Maximization						
mv-min	0.144	0.135	0.161	0.203	0.277	0.176
mv-ew	0.150	0.142	0.181	0.228	0.297	0.162
min-ew	0.136	0.124	0.126	0.132	0.166	0.142
Panel B: New shrinkage portfolios with non-parametric calibration						
Shrinkage portfolios: MSL Minimization						
mv-min	0.144	0.137	0.160	0.335	0.264	0.171
mv-ew	0.151	0.140	0.153	0.160	0.174	0.168
min-ew	0.137	0.125	0.132	0.142	0.174	0.168
Shrinkage portfolios: Utility Maximization						
mv-min	0.144	0.137	0.161	0.971	0.183	0.171
mv-ew	0.149	0.140	0.155	0.161	0.174	0.169
min-ew	0.138	0.125	0.129	0.140	0.174	0.169
Shrinkage portfolios: Variance Minimization						
mv-min	0.140	0.127	0.131	0.138	0.181	0.171
mv-ew	0.148	0.140	0.159	0.162	0.174	0.169
min-ew	0.138	0.125	0.129	0.140	0.174	0.169
Shrinkage portfolios: Sharpe Ratio Maximization						
mv-min	0.145	0.135	0.166	0.231	0.373	0.189
mv-ew	0.149	0.142	0.171	0.193	0.174	0.168
min-ew	0.137	0.124	0.127	0.135	0.174	0.167

C Figures

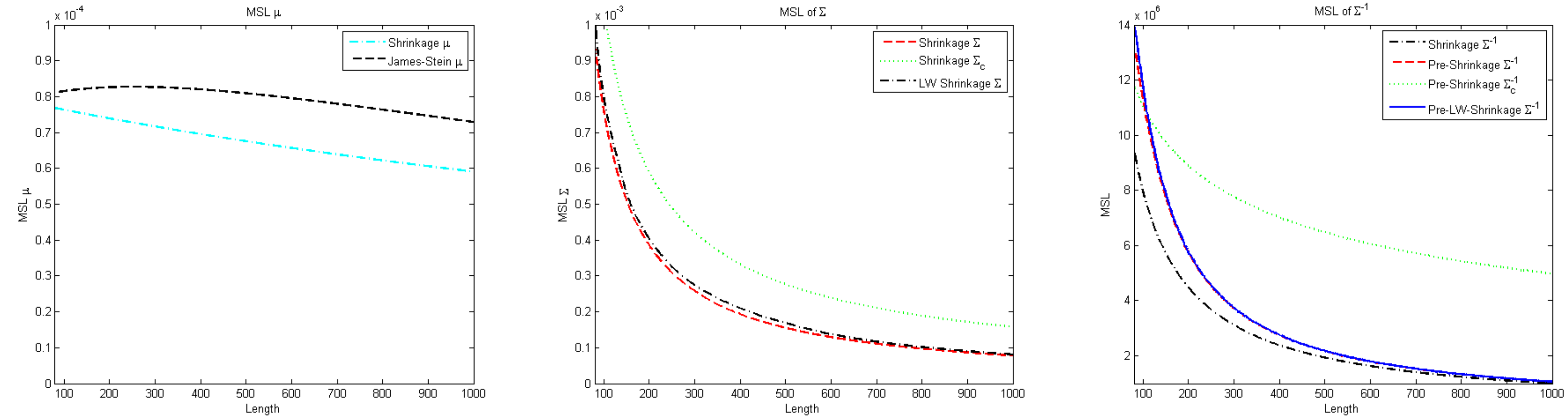


(a) Shrinkage parameter evolution for μ



(b) Shrinkage parameter evolution for Σ and Σ^{-1}

Figure 1: This plot shows the evolution of the shrinkage parameter for μ, Σ and Σ^{-1}. To calculate the shrinkage intensities we use the sample moments of a data set formed by 48 industry portfolios (48IndP). The left plot shows the evolution of the shrinkage intensities for the James-Stein shrinkage estimator (dashed line) and our proposed shrinkage estimator for the mean (dot-dashed line). The right plot shows the evolution of the shrinkage intensities for the asymptotic shrinkage estimator (solid line), the finite shrinkage estimator (dashed line), the finite shrinkage estimator for the inverse covariance matrix (dot-dashed line) and the finite shrinkage estimator for the covariance matrix accounting for the condition number (dotted line).

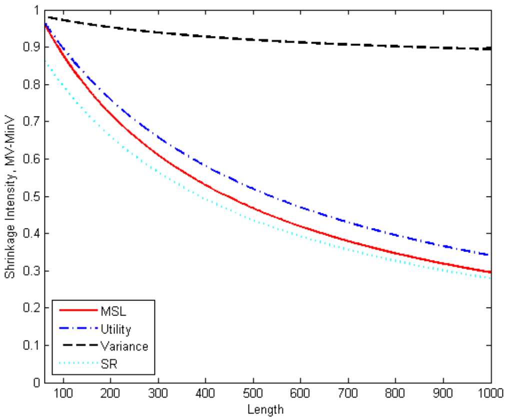


(a) Mean-Squared Loss of $\widetilde{\mu}$

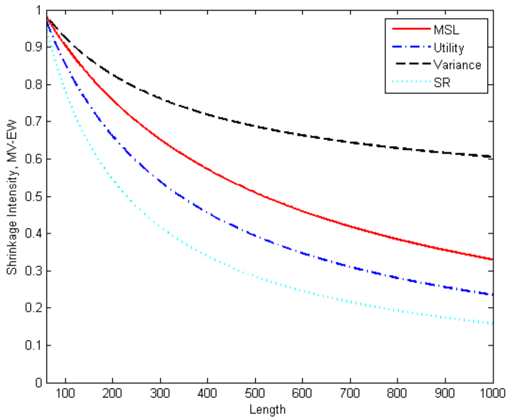
(b) Mean-Squared Loss of $\widetilde{\Sigma}$

(c) Mean-Squared Loss of $\widetilde{\Sigma}^{-1}$

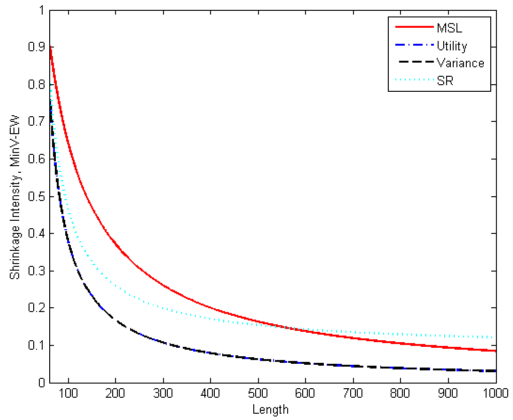
Figure 2: This plot shows the evolution of the mean-squared loss of the shrinkage estimator for μ, Σ and Σ^{-1}. To compute the mean-squared loss, we use the sample moments of a data set formed by 48 industry portfolios (48IndP). Figure 2a depicts the evolution of the mean-squared loss of the shrinkage estimators for the vector of means of the James-Stein shrinkage estimator (dashed line) and our proposed shrinkage estimator for the mean (dot-dashed line). Figure 2b depicts the evolution of the mean-squared loss of the shrinkage estimator for the covariance matrix of the asymptotic shrinkage intensity (dot-dashed line), the finite shrinkage intensity (dashed line) and the finite shrinkage intensity accounting for the condition number (dotted line). Figure 2c depicts the evolution of the mean-squared loss of estimators of the inverse covariance matrix. The dot-dashed line represents the evolution of the mean-squared loss of the shrinkage estimator for the inverse covariance matrix. The dashed line represents the evolution of the mean-squared loss for the inverse of the shrinkage estimator of the covariance matrix with finite shrinkage intensity. The solid line represents the evolution of the mean-squared loss for the inverse of the shrinkage estimator of the covariance matrix with asymptotic shrinkage intensity. The dotted line represents the evolution of the mean-squared loss for the inverse of the shrinkage estimator of the covariance matrix with finite shrinkage intensity accounting for the condition number.



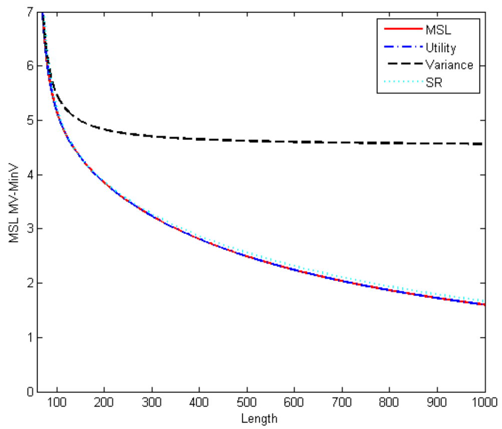
(a) Shrinkage parameter evolution for mv-min



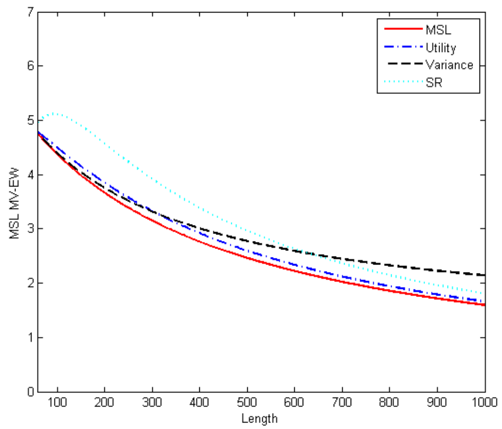
(b) Shrinkage parameter evolution for mv-ew



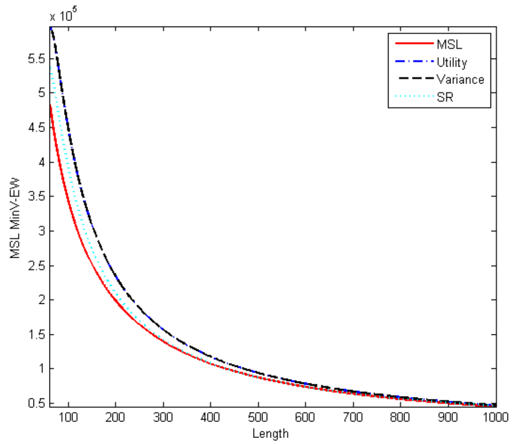
(c) Shrinkage parameter evolution for min-ew



(d) Mean-Squared Loss of mv-min

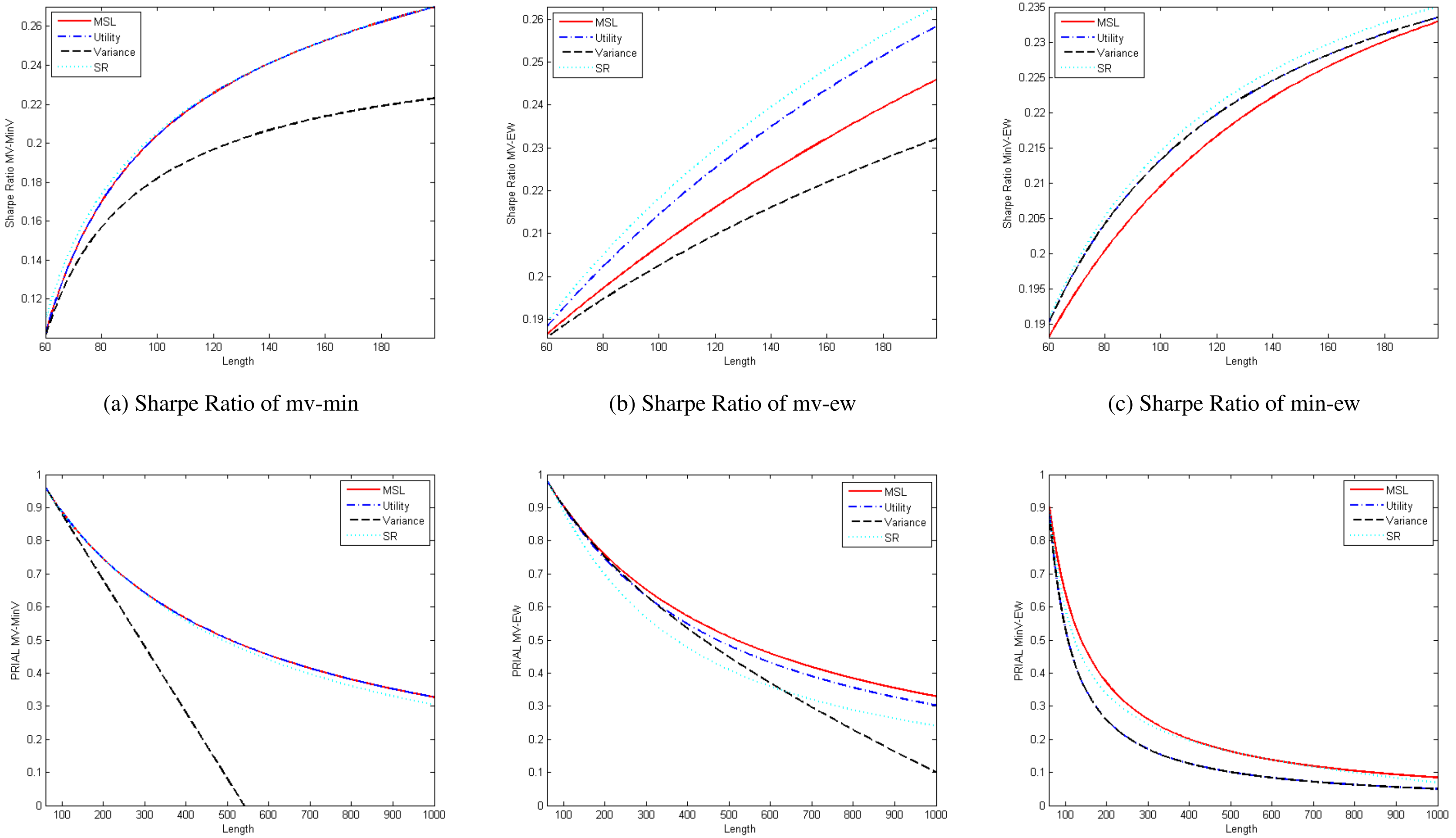


(e) Mean-Squared Loss of mv-ew



(f) Mean-Squared Loss of min-ew

Figure 3: These plots show the evolution of the optimal shrinkage parameters and the Mean-Squared Loss (MSL) against sample's length. To calculate the shrinkage intensities and MSL's, we use the sample moments of a data set formed by 48 industry portfolios (48IndP). In the first row, we have the plots of the shrinkage intensities for the shrinkage portfolios considered in the study. In the second row, we have the plots of the MSL's for the shrinkage portfolios considered in the study. In every figure, the solid line depicts the value of interest when the portfolio is calibrated via mean-squared loss minimization, the dot-dashed line depicts the value of interest when the portfolio model is calibrated via expected utility maximization, the dashed line depicts the value of interest when the portfolio model is calibrated via portfolio variance minimization and the dotted line depicts the value of interest when the portfolio model is calibrated via Sharpe ratio maximization.



(a) Sharpe Ratio of mv-min

(b) Sharpe Ratio of mv-ew

(c) Sharpe Ratio of min-ew

(d) RIAL of mv-min

(e) RIAL of mv-ew

(f) RIAL of min-ew

Figure 4: These plots show the evolution of the expected Sharpe Ratios (SR) and the *relative improvement in average loss* (RIAL) against sample's length. To calculate the SR's and RIAL's we use the sample moments of a data set formed by 48 industry portfolios (48IndP). In the first row, we have the plots of the SR's for the shrinkage portfolios considered in the study. In the second row, we have the plots of the RIAL's for the shrinkage portfolios considered in the study. In every figure, the solid line depicts the value of interest when the portfolio is calibrated via mean-squared loss minimization, the dot-dashed line depicts the value of interest when the portfolio model is calibrated via expected utility maximization, the dashed line depicts the value of interest when the portfolio model is calibrated via portfolio variance minimization and the dotted line depicts the value of interest when the portfolio model is calibrated via Sharpe ratio maximization.

References

[1] Best, M. J. and R. R. Grauer (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *The Review of Financial Studies‘ 4*, 315–342.

[2] Brandt, M. W. (2004). *Portfolio Choice Problems*. Handbook of Financial Econometrics.

[3] Britten-Jones, M. (1999). The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *Journal of Finance 54*(2), 655–671.

[4] DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A Generalized Approach to Portfolio Optimization: Improving Performance By Constraining Portfolio Norms. *Management Science 55*, 798–812.

[5] DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies 22*(5), 1915–1953.

[6] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics 7*(1), pp. 1–26.

[7] Fama, E. and F. French (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance 47*, 427–465.

[8] Frost, P. A. and J. E. Savarino (1986). An empirical bayes approach to efficient portfolio selection. *The Journal of Financial and Quantitative Analysis 21*(3), pp. 293–305.

[9] Haff, L. R. (1979). An Identity for the Wishart Distribution with Applications. *Journal of Multivariate Analysis 9*, 531–544.

[10] Jagannathan, R. and T. Ma (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance 58*, 1651–1684.

[11] James, W. and J. Stein (1961). Estimation with Quadratic Loss. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pp. 361–379. University of California Press.

[12] Jobson, J., B. Korkie, and V. Ratti (1979). Improved estimation for Markowitz portfolios using James-Stein type estimators. *Proceedings of the American Statistical Association*, 279–284.

[13] Jobson, J. D. and B. Korkie (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association 75*, 544–554.

[14] Jobson, J. D. and B. Korkie (1981). Putting Markowitz theory to work. *Journal of Portfolio Management 7*, 70–74.

[15] Jorion, P. (1985). International Portfolio Diversification with Estimation Risk. *Journal of Business 58*, 259–278.

[16] Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis 21*(3), 279–292.

[17] Kan, R. and D. R. Smith (2008). The Distribution of the Sample Minimum-Variance Frontier. *Management Science 54*, 1364–1380.

[18] Kan, R. and G. Zhou (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis 42*, 621–656.

[19] Ledoit, O. and M. Wolf (2003). Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance 10*, 603–621.

[20] Ledoit, O. and M. Wolf (2004a). A well-conditioned estimaor for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 365–411.

[21] Ledoit, O. and M. Wolf (2004b). Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management 30*, 110–119.

[22] MacKinlay, A. C. and L. Pastor (2000). Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Review of Financial studies 13*, 883–916.

[23] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance 7*(1), 77–91.

[24] Merton, R. C. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics 8*, 323–361.

[25] Okhrin, Y. and W. Schmid (2006). Distributional properties of portfolio weights. *Journal of Econometrics*, 235–256.

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

[27] Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications in Statistics*. Wiley.

[28] Siegel, A. F. and A. Woodgate (2007). Performance of Portfolios Optimized with Estimation Error. *Management Science 53*, 1005–1015.

[29] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium*.

[30] Tu, J. and G. Zhou (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics 99*(1), 204 – 215.