

# Sampling error and double shrinkage estimation of minimum variance portfolios

Citation for published version (APA):

Candelon, B., Hurlin, C., & Tokpavi, S. (2011). *Sampling error and double shrinkage estimation of minimum variance portfolios*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 002 <https://doi.org/10.26481/umamet.2011002>

## Document status and date:

Published: 01/01/2011

## DOI:

[10.26481/umamet.2011002](https://doi.org/10.26481/umamet.2011002)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Bertrand Candelon, Christophe Hurlin, Sessi Tokpavi

**Sampling Error and Double Shrinkage Estimation of Minimum Variance Portfolios**

RM/11/002

**METEOR**

Maastricht University School of Business and Economics  
Maastricht Research School of Economics  
of Technology and Organization

P.O. Box 616  
NL - 6200 MD Maastricht  
The Netherlands

# Sampling Error and Double Shrinkage Estimation of Minimum Variance Portfolios

Bertrand Candelon\*, Christophe Hurlin†, Sessi Tokpavi‡

September, 2010

## Abstract

Shrinkage estimators of the covariance matrix are known to improve the stability over time of the Global Minimum Variance Portfolio (GMVP), as they are less error-prone. However, the improvement over the empirical covariance matrix is not optimal for small values of  $n$ , the estimation sample size. For typical asset allocation problems, with  $n$  small, this paper aims to introduce a new framework useful to improve the stability of the GMVP based on shrinkage estimators of the covariance matrix. First, we show analytically that the weights of any GMVP can be shrunk - within the framework of the ridge regression - towards the ones of the equally-weighted portfolio in order to reduce sampling error. Second, monte carlo simulations and empirical applications show that applying our methodology to the GMVP based on shrinkage estimators of the covariance matrix, leads to more stable portfolio weights, sharp decreases in portfolio turnovers, and often statistically lower (resp. higher) out-of-sample variances (resp. sharpe ratios). These results illustrate that double shrinkage estimation of the GMVP can be beneficial for realistic small estimation sample sizes.

*Key words* : Global Minimum Variance Portfolio; Sampling Error; Double Shrinkage; Equality Restricted Ridge Regression

---

\*b.candelon@maastrichtuniversity.nl. Maastricht University, Department of Economics. The Netherlands.

†christophe.hurlin@univ-orleans.fr, University of Orléans, Laboratoire d'Economie d'Orléans (LEO), France.

‡Corresponding author. EconomiX, University of Paris Ouest Nanterre La Défense, 200 avenue de la République, 92000 Nanterre Cedex. E-mail: sessi.tokpavi@u-paris10.fr

# 1 Introduction

Sampling error (also denoted as the estimation risk) constitutes a major drawback for optimal portfolio procedures based on the mean-variance methodology. For example, an investor implementing the well-known Markowitz procedure to determine his optimal portfolio would usually substitute for the mean and the covariance matrix of asset returns, their empirical counterparts (sample mean and covariance matrix). As these inputs are estimators it results that the estimation uncertainty usually leads to sub-optimal portfolios for which the asset weights are noisy and instable over time.<sup>1</sup>

Hence, the problem can be split into two types of uncertainty: the one resulting from the estimation of the asset returns mean and the other one stemming of the estimation of the asset returns covariance matrix. It seems that the first origin is more severe than the second one since it is more difficult to estimate means than covariance matrices (see Merton [1980], Nelson [1992], Chopra and Ziemba [1993]). Therefore, several papers propose to invest in the Global Minimum Variance Portfolio (GMVP) the left-most point on the efficient frontier. This portfolio has the appealing property to rely exclusively on estimates of the covariance matrix (see Baker and Haugen [1991], Jagannathan and Ma [2003]).

Nevertheless, the GMVP still suffers from a second source of uncertainty, i.e. the one resulting from the estimation of the covariance matrix. To even better grasp this type of sampling error, it is important to recall that the optimal allocation of the GMVP depends on the eigenvalues structure of the covariance matrix. Now it is well

---

<sup>1</sup>For a discussion on the importance of sampling error regarding the instability of mean-variance portfolios, see Jobson and Korkie [1980, 1981a], Frost and Savarino [1986, 1988], Jorion [1985, 1986], Michaud [1989], Best and Grauer [1991], Black and Litterman [1992], DeMiguel et al. [2009a].

known that for  $p/n \rightarrow c > 0$  with  $p$  the number of assets and  $n$  the sample size, the eigenvalues of the sample covariance matrix will be more dispersed than the true unobservable ones (see Marcenko and Pastur [1967]), and the eigenvectors will not be consistent (see Johnstone and Lu [2004]). Therefore, unless  $p/n \rightarrow 0$ , errors in the sample covariance matrix will affect the corresponding GMVP, leading to unstable portfolio weights. Such a drawback is quite often encountered in empirical works when the number of observations is relatively small with respect to the number of assets.

In order to tackle this problem, Ledoit and Wolf [2003, 2004a, 2004b] propose to regularize the sample covariance matrix prior to the optimization via Bayes-Stein shrinkage estimators.<sup>2</sup> The idea behind this approach consists in introducing a balance between sampling error and specification error (bias) by shrinking the sample covariance matrix towards a target characterized by a low level of estimation risk. Hence, it is possible to significantly reduce the amount of sampling error in the sample covariance matrix without adding too much specification error. However, shrinkage estimators of the GMVP are not optimal for small values of  $n$  and can under-perform in finite sample.<sup>3</sup> An illustration is given by Jagannathan and Ma [2003] who show that the GMVP computed using a randomly weighted average of the sample and the single index covariance matrices can perform as well as the shrinkage estimator of

---

<sup>2</sup>Note that other methods of reducing sampling error when computing the GMVP include the use of factor models (the single index model, the Fama and French [1993] three-factor model, etc.), the normed-constrained approach in DeMiguel et al. (2009b), or the imposition of shortsale constraints. This latter approach is widely discussed by Jagannathan and Ma [2003] who show that shrinkage of the sample covariance matrix can be achieved by imposing a shortsale constraint on the portfolio weights.

<sup>3</sup>By the term "shrinkage estimators of the GMVP", we mean the GMVP computed using Bayes-Stein shrinkage estimators of the covariance matrix.

the GMVP in Ledoit and Wolf [2003].<sup>4</sup>

Hence, for small sample sizes, this paper aims at proposing a new method to further reduce sampling error by shrinking once again traditional shrinkage estimators of the GMVP. To achieve this goal, we prove theoretically<sup>5</sup> that penalized regressions (ridge, lasso or elastic net) aiming at reducing sampling variability of parameters estimates can lower the problems induced by the estimation risk in computing shrinkage estimators of the GMVP. Therefore we introduce some new estimators of the GMVP which draw from an equality restricted ridge regression (see Grob [2003]). These estimators result from a double shrinkage operation in the sense that they shrink towards the equally-weighted portfolio traditional shrinkage estimators of the GMVP. Based on the theoretical results in Grob [2003], we give conditions under which the weights of the double shrinkage estimators of the GMVP dominate in the sense of Mean Square Error (MSE) the weights of the traditional shrinkage estimators of the GMVP.

Monte carlo simulations and empirical applications illustrate that our double shrinkage methodology leads to more stable portfolio weights, sharp decreases in portfolio turnovers, and often statistically lower (resp. higher) out-of-sample variances (resp. Sharpe ratios). The results hold for realistic small sample sizes and indicate that double shrinkage estimation of the GMVP is highly beneficial for practitioners.

The remainder of the paper is organized as follows. Section 2 introduces the

---

<sup>4</sup>This GMVP is based on a shrinkage covariance matrix obtained by combining the sample and the single index covariance matrices, using a shrinkage data-dependent weighting scheme.

<sup>5</sup>This is done via the generalization of Kempf and Memmel [2006]. We prove that the weights of any shrinkage estimator of the GMVP can be recovered through the classical least-squares regression applied to a pseudo data set.

regression-based approach for the computation of shrinkage estimators of the GMVP. In Section 3, we develop the double shrinkage methodology for the estimation of the GMVP and present the main theoretical results. Monte carlo simulations and empirical applications illustrating the relevance of our methodology are presented in section 4. The last Section concludes.

## 2 Least squares regression and shrinkage estimators of the GMVP

Consider an investor facing the choice of a portfolio among the universe of  $p$  financial risky assets. At time  $t$ , denote by  $r_t = (r_{1t}, r_{2t}, \dots, r_{pt})'$  the  $(p, 1)$  vector of asset returns with covariance matrix the  $(p, p)$  matrix  $\Sigma$ . In the traditional minimum-variance optimization, the optimal portfolio weights are defined as follows

$$\varpi = \underset{x}{\operatorname{arg\,min}} \ x' \Sigma x, \quad (1)$$

under the constraint  $\vartheta x = 1$ , with  $x$  the  $(p, 1)$  vector of portfolio weights, and  $\vartheta$  a vector  $(1, p)$  whose entries are ones. The explicit solution is given by<sup>6</sup>

$$\varpi = \frac{\Sigma^{-1} \vartheta'}{\vartheta \Sigma^{-1} \vartheta'}. \quad (2)$$

For the investor, the variance  $\sigma^2$  of the GMVP equals

$$\sigma^2 = \varpi' \Sigma \varpi = \frac{1}{\vartheta \Sigma^{-1} \vartheta'}. \quad (3)$$

Theoretically, (2-3) illustrate that the weights and the variance of the GMVP depend on the covariance matrix  $\Sigma$  which is unobservable. In practice, to get a

---

<sup>6</sup>Note that equation (1) defines the solution of the GMVP which differs from the Constrained Minimum Variance Portfolio (CMVP), solution of (1) under some specified constraints. For example, prohibition of short sale or the desire of investors to fix upper bounds on portfolio weights.

feasible evaluation of these quantities, one replaces the unknown matrix  $\Sigma$  by an estimator  $\widehat{\Sigma}$ . More precisely, the two-step estimator of  $\varpi$  and  $\sigma^2$  are given by

$$\widehat{\varpi} = \frac{\widehat{\Sigma}^{-1}\vartheta'}{\vartheta\widehat{\Sigma}^{-1}\vartheta'}, \quad \widehat{\sigma}^2 = \frac{1}{\vartheta\widehat{\Sigma}^{-1}\vartheta'}. \quad (4)$$

Under the assumption that  $r_t$  is independent and identically distributed, and follows a multivariate normal distribution with mean  $\mu$  and covariance matrix  $\Sigma$ , a natural estimator (maximum likelihood) of  $\Sigma$  is the empirical or sample covariance matrix which is the most efficient estimator and defined by

$$\widehat{\Sigma} = \widehat{\Sigma}_{\text{emp}} = \frac{1}{n} \sum_{t=1}^n (r_t - \widehat{\mu})(r_t - \widehat{\mu})', \quad (5)$$

with  $n$  the available sample size and  $\widehat{\mu}$  the sample mean. However substituting the true covariance matrix by its empirical counterpart renders the GMVP highly instable as the sample covariance matrix is a very noisy estimator. Hence, most of the times in empirical applications (where the estimation sample size  $n$  is relatively low compared to the dimension of the assets  $p$ ) we observe large values for the estimated coefficients of the sample covariance matrix whereas their true values are small (see Jobson and Korkie [1980]). The Bayes-Stein shrinkage techniques constitute common solutions to reduce sampling error by combining the sample covariance matrix  $\widehat{\Sigma}_{\text{emp}}$  with a prior estimator  $\widehat{\Sigma}_{\text{prior}}$

$$\widehat{\Sigma}_{\text{bs}} = \delta\widehat{\Sigma}_{\text{prior}} + (1 - \delta)\widehat{\Sigma}_{\text{emp}}, \quad (6)$$

where  $\delta$  is a scalar shrinkage parameter lying between zero and one. The prior estimator of the covariance matrix is a structured one that is less influenced by sampling error. Ledoit and Wolf [2003, 2004a, 2004b] analyse several prior (or target)



matrices as the one derived from the single index model, the constant correlation matrix and a multiple of the identity matrix. The goal of the shrinkage technique is to reduce the mean square error of the estimator of the covariance matrix via balancing bias and variance. The resulting GMVP is obtained by replacing  $\widehat{\Sigma}_{bs}$  by  $\widehat{\Sigma}$  in (4), yielding

$$\widehat{\omega}_{bs} = \frac{\widehat{\Sigma}_{bs}^{-1} \vartheta'}{\vartheta \widehat{\Sigma}_{bs}^{-1} \vartheta'}, \quad \widehat{\sigma}_{bs}^2 = \frac{1}{\vartheta \widehat{\Sigma}_{bs}^{-1} \vartheta'}. \quad (7)$$

Although shrinkage estimators of the GMVP in (7) are less error-prone, we show in this paper that the remaining<sup>7</sup> sampling error affecting the stability of these portfolios can be further reduced via an additional shrinkage operation. The core of the double shrinkage approach relies on an alternative regression-based approach to compute  $\widehat{\omega}_{bs}$  based on a given shrinkage estimator  $\widehat{\Sigma}_{bs}$  of the covariance matrix  $\Sigma$ . To present the regression approach, let  $Z = [z_1, \dots, z_p]$  of dimension  $(p, p)$  be the square root of  $\widehat{\Sigma}_{bs}$ , that is,  $Z'Z = \widehat{\Sigma}_{bs}$ . Denote  $Z_{(i)} = [z_i - z_1, \dots, z_i - z_{i-1}, 0_p, z_i - z_{i+1}, \dots, z_i - z_p]$  the matrix of dimension  $(p, p)$  where  $0_p$  is a vector  $(p, 1)$  whose entries are zeros, with  $i \in \{1, \dots, p\}$ . Note that the  $r$ th column of  $Z_{(i)}$  is equal to the difference between  $z_i$  and the  $r$ th column of  $Z$ . Let us consider the following least squares regression

$$\widehat{\beta} = \arg \min_{\beta} u'u = \arg \min_{\beta} (Y - X\beta)'(Y - X\beta), \quad (8)$$

where  $X = \left( Z'_{(i)}, Z'_{(k)} \right)'$   $i \neq k$  is a matrix of dimension  $(2p, p)$ ,  $Y = (z'_i, z'_k)'$  a vector of dimension  $(2p, 1)$ ,  $\beta = (\beta_1, \dots, \beta_p)'$  a vector  $(p, 1)$  of parameters, and  $u$  the  $(2p, 1)$  vector of noise. The next proposition recasts the computation of the shrinkage estimators of the GMVP within the classical least squares regression problem.

---

<sup>7</sup>The remaining sampling error can be related to the lack of precision due to the scarcity of data in the estimation of the shrinkage parameter  $\delta$  in the equation (6).

**Proposition 1** *The Ordinary Least Squares (OLS) estimate of  $\beta$  in equation (8) corresponds to the weights of the shrinkage estimator of the GMVP, i.e.,*

$$\widehat{\beta} = \widehat{\omega}_{bs} \quad \text{and} \quad \sum_{s=1}^{2p} \widehat{u}_s^2 = 2\widehat{\sigma}_{bs}^2. \quad (9)$$

See Appendix A for the proof. The proposition states that an OLS regression applied to a pseudo data leads to a solution that corresponds exactly to the shrinkage estimator of the GMVP in (7). Besides, the residuals sum of squares of the regression equation (8) equals two times the variance of the shrinkage estimator of the GMVP. This result can be viewed as a generalization of proposition 2 in Kempf and Memmel [2006] who find identical results for the classical GMVP based on the empirical covariance matrix. Our result is nevertheless more general as it holds for any GMVP that relies on a given estimator of the covariance matrix which is invertible. This is in particular the case for the shrinkage estimator of the GMVP. In the next section, we rely on equation (8) to introduce our double shrinkage methodology for the estimation of the GMVP.

### **3 Ridge regression and double shrinkage estimation of the GMVP**

The OLS estimator  $\widehat{\beta}$  in equation (8) is an unbiased estimator of  $\beta$  if the residual vector  $u$  satisfies the assumption  $\mathbb{E}(u) = 0$ . This estimator has also the desirable property to be of minimal variance (between the class of all unbiased estimators linear with respect to  $Y$ ) under the assumption that  $u$  is spherical (Gauss-Markov Theorem). Nevertheless, it is well known that introducing a small amount of bias in the OLS estimator can lead to a significant reduction of its variance and thus

its Mean Square Error (MSE), improving hence parameter stability and predictive performance. This is in particular the case for high dimensional problems where the size of the matrix  $X$  is large. The fall in variance is generally achieved by penalizing the least squares criterion in equation (8)

$$\bar{\beta} = \arg \min_{\beta} (Y - X\beta)'(Y - X\beta) + f_{\lambda}(\beta), \quad (10)$$

with  $f_{\lambda}(\beta)$  the penalty function. Three particular penalty functions are generally used. They correspond to the so-called ridge regression (see Hoerl and Kennard [1970]), the lasso or least absolute shrinkage and selection operator (see Tibshirani [1996]), and the elastic net (see Zou and Hastie [2005]), with the following equivalences

$$\text{Ridge: } f_{\lambda}(\beta) = \lambda \sum_{j=1}^p \beta_j^2, \quad \lambda \geq 0, \quad (11)$$

$$\text{Lasso: } f_{\lambda}(\beta) = \lambda \sum_{j=1}^p |\beta_j|, \quad \lambda \geq 0, \quad (12)$$

$$\text{Elastic net: } f_{\lambda}(\beta) = \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j|, \quad \lambda_1, \lambda_2 \geq 0, \quad (13)$$

where  $\lambda$ ,  $\lambda_1$  and  $\lambda_2$  are tuning parameters. The solution of the ridge regression has a closed form expression given by

$$\bar{\beta} = (X'X + \lambda I_p)^{-1} X'Y. \quad (14)$$

For  $\lambda = 0$  we get the OLS estimator and increasing  $\lambda$  shrinks the OLS estimator which is unbiased but noisy towards zero which is a biased but noise-free estimator. The intrinsic limit of the ridge regression is that none coefficient estimate is set exactly to zero. This characteristic is not desirable when the parsimony of the model is

required beyond the accuracy of predictions. Lasso and elastic net methods tackle this issue by setting exactly to zero some coefficients, leading thus to more parsimonious models, especially when the number of predictors is large.

Nevertheless, in this paper, we decide to focus exclusively on the ridge regularization method. Our choice is motivated by the fact that it constitutes the only method to offer an explicit solution. Besides we will show later on that it turns out to be more convenient in order to incorporate the restriction associated with optimal portfolio, i.e.  $\vartheta\bar{\beta} = 1$ , where  $\vartheta$  is the  $(1, p)$  vector whose entries are ones. As underlined by Grob [2003], when the ridge estimator is assumed to satisfy a given restriction, here  $\vartheta\bar{\beta} = 1$ , it is no longer appropriate to shrink the OLS estimator towards the zero vector (which evidently violates the restriction) but towards a vector  $\beta_0$  satisfying the restriction. This approach called the modified ridge estimation was proposed by Swindel [1976]. The estimator has the following form

$$\bar{\beta}_S = (X'X + \lambda I_p)^{-1} (X'Y + \lambda\beta_0), \quad (15)$$

where the subscript "S" refers to Swindel. In fact, this estimator corresponds to the OLS estimator for  $\lambda = 0$  and to  $\beta_0$  when  $\lambda \rightarrow \infty$ . A straightforward candidate for  $\beta_0$  is the shortest vector satisfying the restriction  $\vartheta\beta_0 = 1$ , i.e.  $\beta_0 = \vartheta'(\vartheta\vartheta')^{-1}$ . In our framework, this vector is identical to the  $(p, 1)$  vector whose entries respectively equal  $1/p$ . Therefore, the estimator  $\bar{\beta}_S$  representing the optimal portfolio weights in equation (15) shrinks the GMVP, for which the weights are given by  $\hat{\beta}$  in equation (8) towards the equally-weighted portfolio, with weights  $\beta_0$ .

One can notice that while extreme values of the ridge parameter ( $\lambda = 0$  or  $\lambda \rightarrow \infty$ )

lead to optimal portfolio weights  $\bar{\beta}_S$  that satisfy the constraint  $\vartheta\bar{\beta}_S = 1$ , there is no guarantee that it remains the case for intermediate values of  $\lambda$ . Thus, the expression of  $\bar{\beta}_S$  has to be modified in order to ensure that the constraint is met for any  $\lambda$ . To this aim, let us remark that the Swindel [1976] estimator  $\bar{\beta}_S$  is the solution of an OLS regression applied to an augmented data set. More precisely, if we denote  $X_\lambda = (X', \sqrt{\lambda}I_p)'$  and  $Y_\lambda = (Y', \beta'_0)'$ , with simple algebraic calculus, one can show that  $\bar{\beta}_S$  is the solution of the following least squares regression

$$\bar{\beta}_S = \arg \min_{\beta} u'_\lambda u_\lambda = \arg \min_{\beta} (Y_\lambda - X_\lambda \beta)' (Y_\lambda - X_\lambda \beta). \quad (16)$$

Therefore, if we denote  $\bar{\beta}_{RS}$  the corresponding restricted ridge estimator, with the subscript "RS" meaning that the Swindel estimator is restricted, it follows that

$$\bar{\beta}_{RS} = \arg \min_{\beta} (Y_\lambda - X_\lambda \beta)' (Y_\lambda - X_\lambda \beta) \quad \text{s.t. } \vartheta\beta = 1, \quad (17)$$

that is,

$$\bar{\beta}_{RS} = \bar{\beta}_S - (X'_\lambda X_\lambda)^{-1} \vartheta' \left[ \vartheta (X'_\lambda X_\lambda)^{-1} \vartheta' \right]^{-1} (\vartheta\bar{\beta}_S - 1), \quad (18)$$

applying the usual formula of the equality restricted OLS (see Draper and Smith [1981, p. 122]). It appears by replacing the augmented matrix  $X_\lambda$  by its expression that (18) reduces to

$$\bar{\beta}_{RS} = \bar{\beta}_S - S_\lambda^{-1} \vartheta' \left[ \vartheta S_\lambda^{-1} \vartheta' \right]^{-1} (\vartheta\bar{\beta}_S - 1), \quad (19)$$

with  $S_\lambda = X'X + \lambda I_p$ . The estimator  $\bar{\beta}_{RS}$  is a special case of the restricted ridge estimator in Grob [2003]. In such a case,  $\bar{\beta}_{RS}$  meets the restriction, i.e.,  $\vartheta\bar{\beta}_{RS} = 1$ ,  $\forall \lambda$ , and for  $\lambda = 0$ ,  $\bar{\beta}_{RS}$  equals  $\hat{\beta}$ , the weights of the shrinkage estimator of the GMVP in equation (9). The next proposition provides the limiting value of  $\bar{\beta}_{RS}$  for  $\lambda \rightarrow \infty$ .

**Proposition 2** *The optimal portfolio weights  $\bar{\beta}_{RS}$  in equation (19) is such that*

$$\lim_{\lambda \rightarrow \infty} \bar{\beta}_{RS} = \beta_0.$$

The proof follows from theorem 1 in Grob [2003]. The vector  $\bar{\beta}_{RS}$  is our double shrinkage estimator of the GMVP. It shrinks any shrinkage estimator of the GMVP (with weights  $\hat{\beta}$  or equivalently  $\hat{\omega}_{bs}$ ) towards the noise-free equally-weighted portfolio. Conditions under which the double shrinkage effect leads to sampling error reduction can be derived. Let us note that the MSE of  $\hat{\beta}$  equals

$$\text{MSE}(\beta, \hat{\beta}) = \text{trace} \left[ \phi^2 (X'X)^{-1} \right], \quad (20)$$

with  $\phi^2$  the variance of the noise term in equation (8). Furthermore, it can be proved (see Grob [2003]) that the MSE of the shrinkage estimator  $\bar{\beta}_{RS}$  is

$$\text{MSE}(\beta, \bar{\beta}_{RS}) = \text{trace} \left[ \phi^2 M_\lambda X'X M_\lambda + \lambda^2 M_\lambda \beta \beta' M_\lambda \right], \quad (21)$$

where  $M_\lambda = S_\lambda^{-1} - S_\lambda^{-1} \vartheta' [\vartheta S_\lambda^{-1} \vartheta']^{-1} \vartheta S_\lambda^{-1}$ . The following proposition gives conditions under which the estimator  $\bar{\beta}_{RS}$  of the optimal portfolio weights dominates in the sense of the MSE the estimator  $\hat{\beta}$ .

**Proposition 3** *For  $\bar{\beta}_{RS}$  and  $\hat{\beta}$  as defined respectively in equations (19) and (8), if  $p > 2$  and*

$$0 < \lambda \leq \frac{2\phi^2}{\beta' Q \beta},$$

*with  $Q = I_p - \vartheta' (\vartheta \vartheta')^{-1} \vartheta$ , then*

$$\text{MSE}(\beta, \bar{\beta}_{RS}) < \text{MSE}(\beta, \hat{\beta}).$$

See Appendix A for the proof. The proposition states that as soon as the number of assets is more than two and the ridge parameter  $\lambda$  lies between the interval  $]0, 2\sigma^2/\beta'Q\beta]$ , sampling error is reduced by the restricted ridge estimator  $\bar{\beta}_{RS}$ .

The choice of the tuning parameter  $\lambda$  appears hence crucial in such a framework. This issue has been investigated within the ordinary ridge regression by several studies: the inspection of the so-called ridge trace (see Hoerl and Kennard [1970]), the examination of the Variance Inflation Factors (VIF) as a function of  $\lambda$  (see Chatterjee and Price [1977], Neter et al. [1983]), and CV or cross validation (see Allen [1971, 1974], Golub et al. [1979]). In our equality restricted ridge regression framework, an explicit solution for  $\lambda_{opt}$  the optimal value of  $\lambda$  is provided by Theorem 5 in Grob [2003]. Under the assumption that  $QX'XQ = Q$ ,  $\lambda_{opt}$  is obtained by minimizing the MSE of the restricted ridge estimator in equation (21) with respect to  $\lambda$ , yielding

$$\lambda_{opt} = \frac{(p-1)\phi^2}{\beta'Q\beta}. \quad (22)$$

An *ad hoc* estimator of  $\lambda_{opt}$  can be used even if the assumption  $QX'XQ = Q$  is not satisfied. This estimator is obtained by replacing  $\phi^2$  and  $\beta$  by suitable estimators and has the following expression

$$\hat{\lambda}_{opt} = \frac{(p-1)\hat{\phi}^2}{\hat{\beta}'Q\hat{\beta}}, \quad (23)$$

where  $\hat{\beta}$  is the weights of the classical shrinkage estimator of the GMVP, and  $\hat{\phi}^2$  the estimator of  $\phi^2$ , *i.e.*,

$$\hat{\phi}^2 = \frac{1}{p+1} \sum_{s=1}^{2p} \hat{u}_s^2. \quad (24)$$

Note that  $\hat{\lambda}_{opt}$  is an increasing function of  $\hat{\phi}^2$ . Since  $\hat{\phi}^2$  is related to  $\hat{\sigma}^2$  (see proposition 1), this means that more shrinkage is required to reduce sampling error

when the variance of the classical shrinkage estimator of the GMVP is high.<sup>8</sup> It is worth mentioning that our double shrinkage estimator  $\bar{\beta}_{RS}$  of the GMVP can be interpreted as a norm-constrained portfolio as introduced by DeMiguel et al. (2009b). The following proposition makes this statement clear.

**Proposition 4** *For a given value of  $\lambda$ , there exists a parameter  $\delta > 0$  such that the solution of the following norm-constrained optimization problem*

$$\tilde{\omega} = \underset{x}{\operatorname{argmin}} x' \widehat{\Sigma}_{bs} x, \quad (25)$$

*subject to  $(x - \beta_0)'(x - \beta_0) \leq \delta$  and  $\vartheta x = 1$ , is identical to the restricted ridge solution  $\bar{\beta}_{RS}$ , that is  $\tilde{\omega} = \bar{\beta}_{RS}$ .*

The proof is straightforward. Indeed, from proposition 1 which establishes the equality between  $\widehat{\beta}$  and  $\widehat{\omega}_{bs}$ , we can rewrite the constrained program (25) as follows

$$\tilde{\omega} = \underset{\beta}{\operatorname{argmin}} (Y - X\beta)'(Y - X\beta), \quad (26)$$

subject to  $(\beta - \beta_0)'(\beta - \beta_0) \leq \delta$  and  $\vartheta\beta = 1$ , or equivalently

$$\tilde{\omega} = \underset{\beta}{\operatorname{argmin}} (Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^p (\beta_j - \beta_{0j})^2, \quad (27)$$

subject to  $\vartheta\beta = 1$ , with  $\lambda \geq 0$  a parameter inversely related to  $\delta$ . The last program corresponds exactly to the formulation of the restricted ridge regression with solution

$\bar{\beta}_{RS}$ .

---

<sup>8</sup>As noted by Grob [2003], an iterated estimation procedure can be considered for  $\widehat{\lambda}_{opt}$ , where instead of  $\widehat{\beta}'Q\widehat{\beta}$  one uses  $\bar{\beta}'_{RS}Q\bar{\beta}_{RS}$ , and in each step  $\lambda$  is chosen as the estimate of  $\lambda$  from the previous step. However, we do not follow this strategy in the sequel, because simulations and empirical applications (not reported) show that it leads to over-shrinkage.



From proposition 4, it appears that the application of our double shrinkage methodology can be conducted by solving the program (25) using any optimization algorithm. Nevertheless, three points should be underlined to stress the relevance of the equivalent formulation via our restricted ridge regression. First, our solution  $\bar{\beta}_{RS}$  (see equation 19) has a closed-form expression and can be empirically obtained without any software optimization algorithm. Second, by recasting the solution of the program (25) within a regression framework, we address statistically the issue of bias-variance trade-off. In this line, the proposition 3 gives the necessary condition under which the double shrinkage effect leads to sampling error reduction. Third, we give an explicit solution for the choice of the tuning parameter  $\lambda$  that makes statistically the compromise between bias and variance using the MSE as criterion, whereas DeMiguel et al. [2009b] recommend the use of cross-validation (with the in-sample variance as criterion) to find the value of  $\delta$  in the program (25).

## 4 Monte carlo simulations and empirical applications

In this section, we evaluate on simulated and real data sets the relevance of our double shrinkage methodology when applied to three (3) shrinkage estimators of the GMVP. The considered GMVP are the ones that use as estimators for  $\Sigma$ , shrinkage estimators of the covariance matrix in Ledoit and Wolf [2004a, 2004b] where  $\hat{\Sigma}_{\text{prior}}$  corresponds respectively to the constant correlation matrix, the diagonal matrix and the multiple of the identity matrix.<sup>9</sup> For each of these estimators of the covariance

---

<sup>9</sup>We do not consider the shrinkage estimator in Ledoit and Wolf [2003] where  $\hat{\Sigma}_{\text{prior}}$  corresponds to the covariance matrix from the single index model. In fact, it has been shown (see Ledoit and Wolf [2004a]) that this estimator leads to similar results compared to the one with the constant correlation matrix as prior.

matrix, a "rolling-window" procedure is implemented to compare the investment strategies respectively based on the corresponding shrinkage estimator of the GMVP (with weights  $\widehat{\beta}$  or equivalently  $\widehat{\omega}_{bs}$ ) and its double shrinkage version (with weights given by  $\overline{\beta}_{RS}$ ).

#### 4.1 Results from simulated data sets

The following strategy is implemented. In a first step, we generate a data set of monthly asset returns  $\{r_{it}\}$  of dimension  $(T, p)$  where  $T$  is the number of months and  $p$  the number of assets. Following Tu and Zhou [2008], we use a real data set to calibrate the parameters. Database is extracted from Kenneth French's website and covers the 49 industry portfolios ( $p = 49$ ) from July 1963 to December 2009 (555 months). The empirical means and covariance matrix are assumed as the true parameters in the calibration, and the monthly returns data are simulated from a multivariate normal distribution with  $T = 24,000$  (2,000 years). In a second step and for each of the three shrinkage estimators of the covariance matrix, the simulated returns for the first  $n$  months (with  $n$  the length of the estimation window) are used to compute  $\widehat{\beta}$  and  $\overline{\beta}_{RS}$  which are considered as competitive allocations for the month  $n+1$ , and the two corresponding out-of-sample portfolio returns. This step is repeated by moving each time the estimation window (including the data for a new month and dropping the data for the earliest month) until the last observation. Note that at the end of the procedure and for each of the three shrinkage estimators of the covariance matrix, we have computed  $T - n$  portfolio weights  $\widehat{\beta}_t$  and  $\overline{\beta}_{RS,t}$   $t = n, \dots, T - 1$ , with corresponding out-of-sample returns  $R_t = \widehat{\beta}'_t r_{t+1}$  and  $\overline{R}_t = \overline{\beta}'_{RS,t} r_{t+1}$  where

$r_{t+1}$  is the vector  $(p, 1)$  of asset returns at time  $t + 1$ . Four criteria are used to compare the competitive optimal portfolios: the portfolio turnovers and the boxplots of portfolio weights which provide insights on the temporal stability of each strategy, the out-of-sample variances and Sharpe ratios. The out-of-sample sharpe ratio  $\widehat{SR}_{\widehat{\beta}}$  corresponding to the estimator  $\widehat{\beta}$  of the GMVP weights is defined as

$$\widehat{SR}_{\widehat{\beta}} = \frac{\widehat{\mu}_{\widehat{\beta}}}{\widehat{\sigma}_{\widehat{\beta}}}, \quad (28)$$

where  $\widehat{\mu}_{\widehat{\beta}}$  and  $\widehat{\sigma}_{\widehat{\beta}}^2$  measure respectively the mean and the variance of  $R_t$

$$\widehat{\mu}_{\widehat{\beta}} = \frac{1}{T-n} \sum_{t=n}^{T-1} R_t, \quad (29)$$

$$\widehat{\sigma}_{\widehat{\beta}}^2 = \frac{1}{T-n-1} \sum_{t=n}^{T-1} \left( R_t - \widehat{\mu}_{\widehat{\beta}} \right)^2, \quad (30)$$

and the same statistic  $\widehat{SR}_{\overline{\beta}_{RS}}$  for the estimator  $\overline{\beta}_{RS}$  can be computed by replacing  $R_t$  by  $\overline{R}_t$  in equations (29-30). As for the portfolio turnover, it is defined as follows

$$\text{Turnover}_{\widehat{\beta}} = \frac{1}{T-n-1} \sum_{t=n}^{T-1} \sum_{j=1}^p \left( \left| \widehat{\beta}_{t+1}^{(j)} - \widehat{\beta}_{t+}^{(j)} \right| \right), \quad (31)$$

where  $\widehat{\beta}_t^{(j)}$  is the weight of the asset number  $j$  in the optimal portfolio based on  $\widehat{\beta}_t$ ,  $\widehat{\beta}_{t+}^{(j)}$  the same asset weight before rebalancing at  $t + 1$ , and  $\widehat{\beta}_{t+1}^{(j)}$  the desired asset weight at time  $t + 1$  (after rebalancing). The same expression in equation (31) is used to define the turnover of the portfolio based on the estimator  $\overline{\beta}_{RS}$ .

Panel A in Table 1 (see Appendix B) reports for several sample lengths ( $n$ ) the out-of-sample portfolio turnovers of each strategy. The first line reports the turnovers of the classical shrinkage estimator of the GMVP (with weights the vector  $\widehat{\beta}$ ) whereas the turnovers of our double shrinkage estimator of the GMVP (with weights the vector  $\overline{\beta}_{RS}$ ) are indicated below within brackets. In all cases, the turnover of the classical

shrinkage estimator of the GMVP is higher than the one obtained implementing our new estimation. Hence, it turns out that our ridge regression lowers (via the bias-variance trade-off) the time instability of the estimated portfolio weights, leading to a significant reduction of trades rebalancing across time and assets. As an illustration, for  $n = 60$ , the turnover of the classical shrinkage estimator of the GMVP that relies on the shrinkage estimator of the covariance matrix (with the constant correlation matrix as prior) equals 0.86, while it falls to 0.47 when implementing our ridge estimator. Thus, the reduction of the sampling error via the restricted ridge regression decreases the turnover by almost 50%.

This stability improvement is confirmed by the analysis of the boxplots of the estimated portfolio weights over the out-of-sample period. For example, Figure 1 in Appendix B gives the boxplots of the  $T - n$  out-of-sample optimal portfolio weights for the two competitive investment policies.<sup>10</sup> It appears that introducing a small amount of bias in the estimation process of the GMVP weights significantly increases its temporal stability. Indeed, while the weights of the classical shrinkage estimator of the GMVP range approximately from -60% to 86%, the weights of the double shrinkage estimator of the GMVP range from -30% to 60%.

It is now interesting to check how the stabilization effect affects the out-of-sample variances and sharpe ratios. To the aim, readers should consider Panels B and C in Table 1 which report these two statistics across the two competitive investment strategies. We report in parentheses the p-values of the test of equality between these

---

<sup>10</sup>To save space, we only report the boxplots for the case  $n = 60$  and for the shrinkage estimator of the covariance matrix where the prior matrix corresponds to the constant correlation matrix. All other figures are available from the authors upon request.

investment strategies. For the sharpe ratio the p-value of the difference is computed following Jobson and Korkie [1981b] and the correction proposed by Memmel [2003]. As for the variances, the comparison is achieved using the Bartlett's test for equality of variances which follows asymptotically a chi-square distribution under normality. Panel B indicates that for realistic estimation sample sizes ( $n = 60, 120$ ), the double shrinkage operation does not deteriorate the out-of-sample variances. Indeed, for these values of  $n$ , we notice a decrease in the variances, which is always significant for  $n = 60$  (5 years). Of course, as the estimation sample size  $n$  increases, the variances of the classical shrinkage estimator of the GMVP decrease, and the benefit associated with the additional shrinkage operation disappears. More precisely, even if for large values of  $n$  ( $n = 240, 720$ ) our ridge regression helps to lower the turnovers, this is at the cost of an increase in the variances of the optimal portfolios. To summarize, double shrinkage estimation of the GMVP using our ridge regression is beneficial (from the viewpoint of variances) for small sample sizes, which is the case encountered by practitioners.

Finally, Panel C of Table 1 gathers the results when considering the Sharpe ratio. It indicates that our double shrinkage approach improves the performances of the classical estimator of the GMVP. Even if it is not clear in the Table, this positive effect should decrease with the estimation sample size  $n$ , and should asymptotically disappear.

As a robustness check, we report in Table 2 the same three statistics using calibrated parameters stemming from the hundred Fama and French portfolios of firms

sorted by size and book-to-market ( $p = 100$ ) whereas the sample remains unchanged (from July 1963 to December 2009). Thus, compared to the previous simulation exercise, the asset space is larger as well as  $p/n$ . Results remain in the line of those reported in the Table 1: first, in all configurations we observe a decrease in the values of the turnovers; second, our double shrinkage methodology leads to statistically lower variances and higher sharpe ratios, except for the case  $n = 720$ , i.e. large sample size.

## 4.2 Results from empirical data sets

In the previous section, we perform a simulation exercise following DeMiguel et al. [2009a] according to whom the comparison of two competitive investment strategies with real data sets does not constitute a proof that one strategy is better than the other. Indeed, such a conclusion might be spuriously driven by the existence of small-firm effect, calendar effects, momentum, mean-reversion, fat tails, or other anomalies that have been documented in the literature. Nevertheless, practitioners always appreciate to evaluate the relevance of a new methodology with real data. Thus, we apply the "rolling-window" procedure to the earlier 49 industry portfolios, and the 100 portfolios formed on size and book-to-market.

Table 3 (see Appendix B) reports the results for the 49 industry portfolios. The presentation is similar to the ones in Tables 1 and 2 except that we report the three statistics (turnovers, variances and sharpe ratios) for the equally-weighted portfolio. There are two reasons for considering this portfolio: first, we have shown in proposition 2 that our ridge regression moves the classical shrinkage estimator of the GMVP

towards the equally-weighted portfolio. Therefore, it is interesting to compare our investment strategy to this target portfolio. Second, monte carlo simulations and empirical applications in DeMiguel et al. [2009a] provide evidence that this portfolio constitutes a relevant benchmark investment rule as many others investment strategies, even sophisticated, cannot beat it. Panel A in table 3 gives additional evidences that our double shrinkage methodology leads to large reductions in optimal portfolio turnovers. From a practical point of view, this result is useful because high turnovers lead to high transaction costs, which can detract from the overall return of a given investment policy. As expected the turnovers of the equally-weighted portfolio are very low and we benefit from this when shrinking the weights of the classical shrinkage estimator of the GMVP towards  $1/p$ .

Once again, it is important to stress that the observed drop in the optimal portfolio turnovers is not coupled with an increase (resp. reduction) in variances (resp. sharpe ratios). Indeed as reported in Panels B of Table 3, for small values of  $n$  ( $n = 60, 120$ ), the variances of the double shrinkage estimator of the GMVP fall, and in all cases except one, the differences are statistically significant.<sup>11</sup> It is a noticeable characteristic that even our ridge regression shrinks the weights of the classical shrinkage estimator of the GMVP towards  $1/p$ , the out-of-sample variance of the resulting portfolio does not lie within the variances of the previous portfolios. As an illustration, for the case  $n = 60$  and when the prior matrix is the identity matrix, the variances respectively of the classical shrinkage estimator of the GMVP and the

---

<sup>11</sup>Here, we follow DeMiguel et al. [2009b], and test the differences in variances using the (non-studentized) stationary bootstrap of Politis and Romano [1994], with  $B = 1000$  resamples and an expected block size  $b = 5$ . The p-value is computed using the methodology in Ledoit and Wolf [2008].

equally-weighted portfolio equal 0.00163 and 0.00258, while the variance of the double shrinkage estimator of the GMVP is only 0.00140. An opposite result is found for the sharpe ratios (see Panel C of Table 3) where in all configurations the sharpe ratios of the double shrinkage estimator of the GMVP is higher than the sharpe ratios of both the classical shrinkage estimator of the GMVP and the equality weighted portfolio. Note that for small samples ( $n = 60, 120$ ) the differences in sharpe ratios are always statistically significant except in one single case. These differences are tested using the studentized circular block bootstrapping methodology<sup>12</sup> in Ledoit and Wolf (2008) where the number of bootstrap replications and the size of each block are respectively set to  $B = 1000$  and  $b = 5$ .

Finally, Table 4 reports the same statistics as in table 3 but for our second data set, that is, the 100 portfolios formed on size and book-to-market. The results are identical : our new investment strategy helps to reduce the turnovers, while improving the out-of-sample variances and sharpe ratios for small samples.

## 5 Conclusion

This paper contributes to the literature of sampling error reduction in the context of assets allocation. Sampling error is generally known to offset the benefit of diversification due to its impact on the instability of the optimal portfolio weights. In the minimum variance allocation framework, Bayes-Stein shrinkage estimators of the covariance matrix are known to improve the stability of the Global Minimum Variance

---

<sup>12</sup>Beyond the Jobson and Korkie [1981b] test used above (see section on monte carlo simulations) to test for the equality of sharpe ratio, this test procedure is known to be robust to departures from normality and i.i.d. data. We thank Mickael Wolf for making available the matlab codes (for the implementation of this test) at <http://www.iew.uzh.ch/institute/people/wolf/publication>.



Portfolio (GMVP). Yet the improvement is not optimal over small samples. By recasting the computation of the traditional shrinkage estimators of the GMVP within the linear regression framework, we prove that an additional shrinkage operation via a restricted ridge regression reduces the variability of these portfolios. More precisely, we show that the weights of the traditional shrinkage estimators of the GMVP can be recovered through an ordinary least squares regression applied to a pseudo data set. Thus stabilization of the weights is obtained by minimizing the residual sum of squares of this regression subject to a bound on the L2-norm of the coefficients.

Monte carlo simulations and empirical applications confirm the advantages of the double shrinkage estimation of the GMVP. Indeed, our results suggest that shrinking once again the traditional shrinkage estimators of the GMVP helps to stabilize the optimal portfolio weights, while improving for small estimation sample sizes, the out-of-sample variances and sharpe ratios. Therefore, it strikes us that double shrinkage estimation of the GMVP weights constitutes an appealing method for investors in their quest for optimal portfolio diversification.

## Appendix A: Proofs of propositions

**Proof of Proposition 1.** First, note that for a given Bayes-Stein shrinkage covariance matrix  $\widehat{\Sigma}_{bs}$ , the corresponding GMVP is defined by

$$\begin{aligned}\widehat{\omega}_{bs} &= \arg \min_x x' \widehat{\Sigma}_{bs} x \quad \text{sc} \quad \vartheta x = 1 \\ &= \arg \min_x x' Z' Z x \quad \text{sc} \quad \vartheta x = 1.\end{aligned}\tag{32}$$

Second,  $\widehat{\beta}$  is the solution of the program given by the equation (8)

$$\widehat{\beta} = \arg \min_{\beta} \sum_{s=1}^{2p} u_s^2 = \arg \min_{\beta} \sum_{s=1}^{2p} (Y_s - X_s \beta)^2.\tag{33}$$

Since  $X_s = (Z'_{(i)s}, Z'_{(k)s})'$  and  $Y_s = (z'_{is}, z'_{ks})'$ , we have

$$\sum_{s=1}^{2p} u_s^2 = \sum_{s=1}^p (z_{is} - Z_{(i)s} \beta)^2 + \sum_{s=1}^p (z_{ks} - Z_{(k)s} \beta)^2.\tag{34}$$

By definition,  $Z_{(i)s}$  and  $Z_{(k)s}$  are respectively equal to

$$Z_{(i)s} = [z_{is} - z_{1s}, \dots, z_{is} - z_{i-1s}, 0, z_{is} - z_{i+1s}, \dots, z_{is} - z_{ps}],\tag{35}$$

$$Z_{(k)s} = [z_{ks} - z_{1s}, \dots, z_{ks} - z_{k-1s}, 0, z_{ks} - z_{k+1s}, \dots, z_{ks} - z_{ps}].\tag{36}$$

Replacing these expressions in (34) and after simple calculus, one gets

$$\sum_{s=1}^{2p} u_s^2 = \sum_{s=1}^p \left( \left( 1 - \sum_{\substack{j=1 \\ j \neq i}}^p \beta_j \right) z_{is} + \sum_{\substack{j=1 \\ j \neq i}}^p \beta_j z_{js}^2 \right) + \sum_{s=1}^p \left( \left( 1 - \sum_{\substack{j=1 \\ j \neq k}}^p \beta_j \right) z_{ks} + \sum_{\substack{j=1 \\ j \neq k}}^p \beta_j z_{js}^2 \right).\tag{37}$$

Inserting the equality  $\vartheta \beta = 1$  in the above equation leads to

$$\begin{aligned}\sum_{s=1}^{2p} u_s^2 &= \sum_{s=1}^p (\beta_1 z_{1s} + \dots + \beta_p z_{ps})^2 + \sum_{s=1}^p (\beta_1 z_{1s} + \dots + \beta_p z_{ps})^2 \\ &= \beta' Z' Z \beta + \beta' Z' Z \beta \\ &= 2\beta' Z' Z \beta.\end{aligned}\tag{38}$$

It follows that  $\widehat{\beta}$  is the solution of the minimisation program

$$\widehat{\beta} = \arg \min_{\beta} \beta' Z' Z \beta, \quad (39)$$

which is equivalent to the program (32) and  $\widehat{\beta} = \widehat{\omega}_{bs}$ . Moreover, from equation (38)

we have

$$\begin{aligned} \sum_{s=1}^{2p} \widehat{u}_s^2 &= 2\widehat{\beta}' Z' Z \widehat{\beta} \\ &= 2\widehat{\omega}'_{bs} \widehat{\Sigma} \widehat{\omega}_{bs} \\ &= 2\widehat{\sigma}_{bs}^2. \end{aligned} \quad (40)$$

■

**Proof of Proposition 3.** From proposition 1, we know that the estimator  $\widehat{\beta}$  in equation (8) is equal to the weights of the shrinkage estimator of the GMVP in (7), *i.e.*,  $\widehat{\beta} = \widehat{\omega}_{bs}$ . Thus,  $\widehat{\beta} = (X'X)^{-1} X'Y$  satisfies the restriction  $\vartheta\widehat{\beta} = 1$ . It follows that  $\widehat{\beta} = \widehat{\beta}_R$ , with  $\widehat{\beta}_R$  the corresponding restricted OLS estimator

$$\widehat{\beta}_R = \widehat{\beta} - (X'X)^{-1} \vartheta' \left[ \vartheta (X'X)^{-1} \vartheta' \right]^{-1} (\vartheta\widehat{\beta} - 1). \quad (41)$$

Then, comparing  $\text{MSE}(\beta, \widehat{\beta})$  and  $\text{MSE}(\beta, \overline{\beta}_{RS})$  amounts to compare  $\text{MSE}(\beta, \widehat{\beta}_R)$  and  $\text{MSE}(\beta, \overline{\beta}_{RS})$ . The rest of the proof follows from theorem 4 in Grob (2003). ■

## Appendix B: Tables and figures

Figure 1: Boxplots of GMVP weights based on the shrinkage (towards the constant correlation matrix) estimator of the covariance matrix

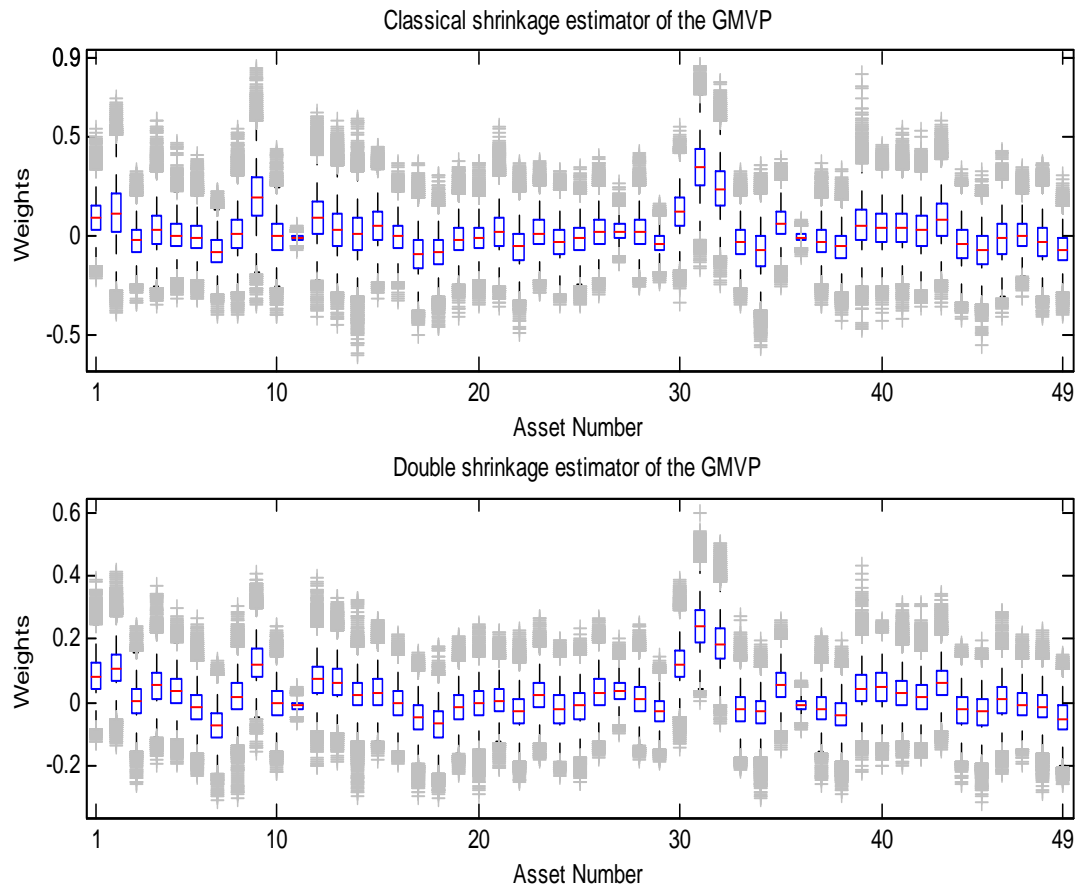


Table 1: Results with simulated data (calibration with the 49-Industry portfolios data)

Prior matrix	$n = 60$	$n = 120$	$n = 180$	$n = 240$	$n = 720$
Panel A: Portfolios turnovers					
Const. Corr.	0.8648 [0.4718]	0.5144 [0.2483]	0.3729 [0.1737]	0.3031 [0.1426]	0.1840 [0.0902]
Diagonal	0.5778 [0.3163]	0.4068 [0.2008]	0.3194 [0.1516]	0.2723 [0.1300]	0.1778 [0.0876]
Identity	0.6204 [0.3250]	0.4229 [0.2026]	0.3274 [0.1520]	0.2775 [0.1303]	0.1787 [0.0876]
Panel B : Portfolios variances					
Const. Corr.	0.00165 [0.00141] (0.00)	0.00132 [0.00119] (0.00)	0.00118 [0.00114] (0.00)	0.00110 [0.00111] (0.64)	0.00097 [0.00107] (0.00)
Diagonal	0.00139 [0.00135] (0.01)	0.00122 [0.00120] (0.31)	0.00113 [0.00116] (0.10)	0.00108 [0.00113] (0.00)	0.00096 [0.00108] (0.00)
Identity	0.00148 [0.00140] (0.00)	0.00124 [0.00122] (0.08)	0.00115 [0.00117] (0.17)	0.00108 [0.00113] (0.00)	0.00096 [0.00108] (0.00)
Panel C : Portfolios sharpe ratios					
Const. Corr.	0.2382 [0.2760] (0.09)	0.2721 [0.3058] (0.07)	0.2755 [0.3113] (0.00)	0.2839 [0.3154] (0.00)	0.2926 [0.3091] (0.00)
Diagonal	0.2349 [0.2560] (0.00)	0.2705 [0.2904] (0.00)	0.2747 [0.3004] (0.00)	0.2826 [0.3062] (0.00)	0.2917 [0.3058] (0.00)
Identity	0.2351 [0.2542] (0.02)	0.2712 [0.2899] (0.00)	0.2756 [0.3002] (0.00)	0.2845 [0.3067] (0.00)	0.2923 [0.3058] (0.00)

Notes: For different values of  $n$  the estimation sample length, this table reports the monthly turnovers, variances and sharpe ratios for the GMVP that rely on the shrinkage estimators of the covariance matrix in Ledoit and Wolf [2004a, 2004b]. For each of these estimators identified by the corresponding prior matrix, we first report the statistics for the traditional shrinkage estimator of the GMVP, followed in bracket by the same statistic for the double shrinkage estimator of the GMVP. Concerning the variances and the sharpe ratios, the third reported value (in parentheses) is the p-value of the difference between the two reported values.

Table 2: Results with simulated data (calibration with the 100 portfolios sorted on size and book-to-market data)

Prior matrix	$n = 60$	$n = 120$	$n = 180$	$n = 240$	$n = 720$
Panel A: Portfolios turnovers					
Const. Corr.	2.1721 [1.5547]	2.0497 [1.0864]	1.4204 [0.6747]	1.0581 [0.4828]	0.4832 [0.2166]
Diagonal	1.3661 [0.8587]	1.1972 [0.6224]	0.9762 [0.4656]	0.8168 [0.3744]	0.4536 [0.2032]
Identity	1.6179 [1.0746]	1.4804 [0.7818]	1.1560 [0.5534]	0.9307 [0.4267]	0.4696 [0.2102]
Panel B : Portfolios variances					
Const. Corr.	0.00188 [0.00161] (0.00)	0.00192 [0.00140] (0.00)	0.00158 [0.00121] (0.00)	0.00138 [0.00113] (0.00)	0.00100 [0.00102] (0.09)
Diagonal	0.00154 [0.00141] (0.00)	0.00146 [0.00125] (0.00)	0.00134 [0.00117] (0.00)	0.00125 [0.00113] (0.00)	0.00098 [0.00103] (0.00)
Identity	0.00164 [0.00146] (0.00)	0.00160 [0.00129] (0.00)	0.00143 [0.00118] (0.00)	0.00131 [0.00113] (0.00)	0.00099 [0.00102] (0.00)
Panel C : Portfolios sharpe ratios					
Const. Corr.	0.3080 [0.3286] (0.00)	0.2996 [0.3393] (0.00)	0.3344 [0.3671] (0.00)	0.3644 [0.3905] (0.00)	0.4125 [0.3886] (0.00)
Diagonal	0.3056 [0.3181] (0.00)	0.3232 [0.3370] (0.00)	0.3489 [0.3583] (0.00)	0.3762 [0.3793] (0.00)	0.4122 [0.3823] (0.00)
Identity	0.3082 [0.3208] (0.00)	0.3150 [0.3387] (0.00)	0.3420 [0.3615] (0.00)	0.3696 [0.3841] (0.00)	0.4117 [0.3850] (0.00)

Notes: For different values of  $n$  the estimation sample length, this table reports the monthly turnovers, variances and sharpe ratios for the GMVP that rely on the shrinkage estimators of the covariance matrix in Ledoit and Wolf [2004a, 2004b]. For each of these estimators identified by the corresponding prior matrix, we first report the statistics for the traditional shrinkage estimator of the GMVP, followed in bracket by the same statistic for the double shrinkage estimator of the GMVP. Concerning the variances and the sharpe ratios, the third reported value (in parentheses) is the p-value of the difference between the two reported values.

Table 3: Empirical Results with the 49 Industry portfolios data

Prior matrix	$n = 60$	$n = 120$	$n = 180$
Panel A: Portfolios turnovers			
Const. Corr.	0.5412 [0.3171]	0.3873 [0.1950]	0.2999 [0.1425]
Diagonal	0.5871 [0.3371]	0.4191 [0.2049]	0.3242 [0.1535]
Identity	0.8696 [0.4876]	0.5561 [0.2553]	0.3950 [0.1790]
Eq. Weighted	0.0333	0.0327	0.0328
Panel B : Portfolios variances			
Const. Corr.	0.00135 [0.00133] (0.70)	0.00147 [0.00135] (0.04)	0.00123 [0.00119] (0.67)
Diagonal	0.00133 [0.00124] (0.03)	0.00146 [0.00135] (0.01)	0.00122 [0.00119] (0.67)
Identity	0.00163 [0.00140] (0.00)	0.00160 [0.00138] (0.00)	0.00129 [0.00120] (0.13)
Eq. Weighted	0.00258	0.00253	0.00236
Panel C : Portfolios sharpe ratios			
Const. Corr.	0.2072 [0.2250] (0.33)	0.2013 [0.2450] (0.02)	0.2774 [0.3051] (0.26)
Diagonal	0.2100 [0.2436] (0.02)	0.2070 [0.2489] (0.01)	0.2746 [0.3075] (0.14)
Identity	0.1692 [0.2077] (0.00)	0.2000 [0.2441] (0.00)	0.2634 [0.3038] (0.08)
Eq. Weighted	0.1921	0.2126	0.2268

Notes: For different values of  $n$  the estimation sample length, this table reports the monthly turnovers, variances and sharpe ratios for the GMVP that rely on the shrinkage estimators of the covariance matrix in Ledoit and Wolf [2004a, 2004b]. For each of these estimators identified by the corresponding prior matrix, we first report the statistics for the traditional shrinkage estimator of the GMVP, followed in bracket by the same statistic for the double shrinkage estimator of the GMVP. Concerning the variances and the sharpe ratios, the third reported value (in parentheses) is the p-value of the difference between the two reported values. Eq. Weighted refers to the benchmark equality weighted portfolio.

Table 4: Empirical Results with the 100 portfolios (sorted on size and book-to-market) data

Prior matrix	$n = 60$	$n = 120$	$n = 180$
Panel A: Portfolios turnovers			
Const. Corr.	1.1930 [0.8511]	1.4473 [0.7794]	1.1174 [0.5510]
Diagonal	1.1759 [0.7707]	0.9099 [0.4539]	0.7207 [0.3295]
Identity	1.5870 [1.1335]	1.3837 [0.6933]	0.9126 [0.4107]
Eq. Weighted	0.0257	0.0258	0.0247
Panel B : Portfolios variances			
Const. Corr.	0.00200 [0.00179] (0.00)	0.00244 [0.00186] (0.00)	0.00168 [0.00137] (0.00)
Diagonal	0.00169 [0.00154] (0.00)	0.00184 [0.00156] (0.00)	0.00137 [0.00127] (0.16)
Identity	0.00186 [0.00169] (0.00)	0.00222 [0.00176] (0.00)	0.00155 [0.00136] (0.06)
Eq. Weighted	0.00281	0.00277	0.00258
Panel C : Portfolios sharpe ratios			
Const. Corr.	0.2742 [0.2915] (0.04)	0.2455 [0.2958] (0.00)	0.3336 [0.3583] (0.34)
Diagonal	0.2358 [0.2574] (0.03)	0.2496 [0.2816] (0.06)	0.3383 [0.3423] (0.85)
Identity	0.2382 [0.2632] (0.00)	0.2375 [0.2803] (0.01)	0.3280 [0.3403] (0.63)
Eq. Weighted	0.1659	0.1944	0.2148

Notes: For different values of  $n$  the estimation sample length, this table reports the monthly turnovers, variances and sharpe ratios for the GMVP that rely on the shrinkage estimators of the covariance matrix in Ledoit and Wolf [2004a, 2004b]. For each of these estimators identified by the corresponding prior matrix, we first report the statistics for the traditional shrinkage estimator of the GMVP, followed in bracket by the same statistic for the double shrinkage estimator of the GMVP. Concerning the variances and the sharpe ratios, the third reported value (in parentheses) is the p-value of the difference between the two reported values. Eq. Weighted refers to the benchmark equality weighted portfolio.



## References

- ALLEN, D. M. (1971), "Mean Square Error of Prediction as a Criterion for Selecting Variables", *Technometrics*, 13, 469-475.
- ALLEN, D. M. (1974), "The Relationship Between Variable Selection, Data Augmentation and a Method of Prediction", *Technometrics*, 16, 125-127.
- BAKER, N. AND R. HAUGEN (1991), "The Efficient Market Inefficiency of Capitalization-Weighted Stock Portfolios", *Journal of Portfolio Management*, 17(1), 35-40.
- 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", *Review of Financial Studies*, 4, 315-342.
- BLACK, F. AND R. LITTERMAN (1992), "Global Portfolio Optimization", *Financial Analysts Journal*, 48, 28-43.
- CHATTERJEE, S. AND B. PRICE (1977), *Regression Analysis by Example*, New York: John Wiley.
- CHOPRA, V. K. AND W. T. ZIEMBA (1993), "The Effects of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice", *Journal of Portfolio Management*, 4, 6-11.
- DEMIGUEL, V., GARLAPPI, L. AND R. UPPAL (2009a), "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?", *The Review of Financial Studies*, 22, 1915-1953.
- DEMIGUEL, V., GARLAPPI, L., NOGALES, F. J. AND R. UPPAL (2009b), "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms", *Management Science*, 55, 5, 798-812.
- DRAPER, N. R. AND H. SMITH (1981), *Applied Regression Analysis*, New York: John Wiley.
- FAMA, E. F. AND K. R. FRENCH (1992), "The cross-section of expected stock returns", *Journal of Finance*, 47, 427-465.
- FROST, P. A. AND J. E. SAVARINO (1986), "An Empirical Bayes Approach to Efficient Portfolio Selection", *Journal of Financial and Quantitative Analysis*, 21, 293-305.
- FROST, P. A. AND J. E. SAVARINO (1988), "For Better Performance Constrain Portfolio Weights", *Journal of Portfolio Management*, 15, 29-34.
- GOLUB, G. H., HEATH, M. AND G. WAHBA (1979), "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter", *Technometrics*, 21, 215-223.
- GROB, J. (2003), "Restricted ridge estimation", *Statistics & Probability Letters*, 65, 57-64.
- HOERL, A. E. AND R. W. KENNARD (1970), "Ridge Regression: Biased Estimation of Non-Orthogonal Problems", *Technometrics*, 12, 55-67.

- JAGANNATHAN, R. AND T. MA (2003), "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps", *Journal of Finance*, 58, 1651-1684.
- JOBSON, J. D. AND B. KORKIE (1980), "Estimation for Markowitz Efficient Portfolios", *Journal of the American Statistical Association*, 75, 544-554.
- JOBSON, J. D. AND B. KORKIE (1981a), "Putting Markowitz Theory at work", *Journal of Portfolio Management*, 7(4), 70-74.
- JOBSON, J. D. AND B. KORKIE (1981b), "Performance Hypothesis Testing with the Sharpe and Treynor Measures", *Journal of Finance*, 36, 889-908.
- JOHNSTONE, I. M. AND A. Y. LU (2004), "Sparse principal components analysis", *J. Amer. Statist. Assoc.* To appear.
- JORION, P. (1985), "International Portfolio Diversification with Estimation Risk", *Journal of Business*, 58, 259-278.
- JORION, P. (1986), "Bayes-Stein Estimation for Portfolio Analysis", *Journal of Financial and Quantitative Analysis*, 21, 279-292.
- KEMPF, A. AND C. MEMMEL (2006), "Estimating the global minimum variance portfolio", *Schmalenbach Business Review*, 58, 332-348.
- 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.
- LEDOIT, O. AND M. WOLF (2004a), "Honey, I Shrunk the Sample Covariance Matrix", *Journal of Portfolio Management*, 31, 110-119.
- LEDOIT, O. AND M. WOLF (2004b), "A well-conditioned estimator for large-dimensional covariance matrices", *Journal of Multivariate Analysis*, 88, 365-411.
- MARCENKO, V. A. AND L. A. PASTUR (1967), "Distributions of eigenvalues of some sets of random matrices", *Math. USSR-Sb*, 1:507-536.
- MEMMEL, C. (2003), "Performance Hypothesis Testing with the Sharpe Ratio", *Finance Letters*, 1, 21-23.
- MERTON, R. C. (1980), "On Estimating the Expected Return on the Market: An Exploratory Investigation", *Journal of Financial Economics*, 8, 323-361.
- MICHAUD, R. O. (1989), "The Markowitz optimization enigma: Is optimized optimal?", *Financial Analysts Journal*, 45, 31-42.
- NELSON, D. (1992), "Filtering and forecasting with misspecified ARCH models: Getting the right variance with the wrong model", *Journal of Econometrics*, 52, 61-90.
- NETER, J., WASSERMAN, W. AND M. A. KUTNER (1983), *Applied Linear Regression Models*, Homewood, IL: Richard D. Irwin.
- SWINDEL, B. F. (1976), "Good estimators based on prior information", *Comm. Statist. Theory Methods*, 5, 1065-1075.

TIBSHIRANI, R. (1996), "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society*, B, 58, 267–288.

ZOU, H. AND T. HASTIE (2005), "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society*, Series B, 67, 301-320.

ZHOU, G. AND J. TU (2010), "Being Naive about Naive Diversification: Can Investment Theory be Consistently Useful ", *Journal of Financial Economics* (forthcoming).