Two is better than one: Regularized shrinkage of large minimum variance portfolios

Taras Bodnar[1], Nestor Parolya[2], and Erik Thorsén[1]

[1]Department of Mathematics, Stockholm University, Roslagsvägen 101, SE-10691 Stockholm, Sweden
[2]Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

August 26, 2023

Abstract

In this paper we construct a shrinkage estimator of the global minimum variance (GMV) portfolio by a combination of two techniques: Tikhonov regularization and direct shrinkage of portfolio weights. More specifically, we employ a double shrinkage approach, where the covariance matrix and portfolio weights are shrunk simultaneously. The ridge parameter controls the stability of the covariance matrix, while the portfolio shrinkage intensity shrinks the regularized portfolio weights to a predefined target. Both parameters simultaneously minimize with probability one the out-of-sample variance as the number of assets p and the sample size n tend to infinity, while their ratio p/n tends to a constant $c > 0$. This method can also be seen as the optimal combination of the well-established linear shrinkage approach of Ledoit and Wolf (2004) and the shrinkage of the portfolio weights by Bodnar et al. (2018). No specific distribution is assumed for the asset returns except of the assumption of finite $4 + \varepsilon$ moments. The performance of the double shrinkage estimator is investigated via extensive simulation and empirical studies. The suggested method significantly outperforms its predecessor (without regularization) and the nonlinear shrinkage approach in terms of the out-of-sample variance, Sharpe ratio and other empirical measures in the majority of scenarios. Moreover, it obeys the most stable portfolio weights with uniformly smallest turnover.

Keywords: Shrinkage estimator; high-dimensional covariance matrix; random matrix theory; minimum variance portfolio; parameter uncertainty; ridge regularization

JEL classification: G11, C13, C14, C55, C58, C65

1 Introduction

The global minimum variance (GMV) portfolio is the portfolio with the smallest variance among all optimal portfolios, which are the solutions to the mean-variance optimization problem suggested in the seminal paper of Harry Markowitz (see, Markowitz (1952)). This portfolio has become one of the most commonly used in the literature (see, e.g., Golosnoy and Okhrin (2007), Ledoit and Wolf (2020b) and the references therein). This is due to the fact that if all risk can be removed then a portfolio has the potential to yield an infinite Sharpe ratio. Furthermore, in practice, portfolios and their sample counterpart depend on estimates which contain estimation error. The GMV portfolio is the only optimal portfolio whose weights are solely determined by the covariance matrix of the asset returns and do not depend on the mean vector. This property has been recognized to be very important due to the fact that the estimation error in the means are several times larger than the estimation error in the variances and covariances of the asset returns (see, Merton (1980), Best and Grauer (1991), Kan et al. (2022)).

In the original optimization problem, the GMV portfolio is obtained as the solution of

$$\underset{\mathbf{w}}{\text{minimize}}\ \mathbf{w}^\top \mathbf{\Sigma}\mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1 \tag{1.1}$$

and its weights are given by

$$\mathbf{w}_{GMV} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}}. \tag{1.2}$$

Since the covariance matrix $\mathbf{\Sigma}$ is an unknown quantity, the GMV portfolio cannot be constructed by using (1.2). In Markowitz (1959) the author uses the sample estimator of $\mathbf{w}_{GMV}$ instead of (1.2). He considers

$$\hat{\mathbf{w}}_{GMV} = \frac{\mathbf{S}_n^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{S}_n^{-1}\mathbf{1}}, \tag{1.3}$$

where $\mathbf{S}_n$ is the sample estimator of the covariance matrix $\mathbf{\Sigma}$ which is given by

$$\mathbf{S}_n = \frac{1}{n}\left(\mathbf{Y}_n - \bar{\mathbf{y}}_n\mathbf{1}^\top\right)\left(\mathbf{Y}_n - \bar{\mathbf{y}}_n\mathbf{1}^\top\right)^\top \quad \text{with} \quad \bar{\mathbf{y}}_n = \frac{1}{n}\mathbf{Y}_n\mathbf{1}, \tag{1.4}$$

where $\mathbf{Y}_n = [\mathbf{y}_1, ..., \mathbf{y}_n]$ is the $p \times n$ observation matrix and $\mathbf{y}_i$, $i = 1, .., n$, is the p-dimensional vector of asset returns observed at time i. As such, the sample GMV portfolio with weights (1.3) may be considered as the solution of the optimization problem (1.1) where the unknown covariance matrix $\mathbf{\Sigma}$ is replaced by $\mathbf{S}_n$, namely,

$$\underset{\mathbf{w}}{\text{minimize}}\ \mathbf{w}^\top \mathbf{S}_n\mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1. \tag{1.5}$$

There are several other estimates of the GMV portfolio weights in the literature (see, e.g., Ledoit and Wolf (2004), Frahm and Memmel (2010), Tu and Zhou (2011), DeMiguel et al. (2013), Li et al. (2016), Ledoit and Wolf (2017), Lai et al. (2018), Bodnar et al. (2018), Lai et al. (2020) to mention a few). All of these methods either shrink the covariance matrix and use it for the estimation of GMV portfolio, or they shrink the portfolios weights directly to a certain target. To the best of our knowledge, none of the known approaches combines both procedures into one, in an applicable and theoretical framework in the high-dimensional setting.

Using $\mathbf{S}_n$ instead of $\mathbf{\Sigma}$ may produce a very noisy estimator of the portfolio weights. There are many ways to cope with this estimation uncertainty. Our approach relies on two distinct features. First, the linear shrinkage estimator from Bodnar et al. (2018) has proven to provide good results in terms of the out-of-sample variance and to be a robust for large dimensional portfolios. It does not, however, reduce the size of the positions or the variance (as measured by turnover) of the portfolio weights, (see, e.g., Bodnar et al. (2021b)). This leads us to the second feature. The single source of uncertainty in the GMV portfolio is $\mathbf{S}_n$. If we can stabilize or decrease the variance of the sample covariance matrix we may decrease the variance of the

weights. Our aim is therefore to shrink the sample covariance matrix as well. We apply the Tikhonov regularization (see, e.g., Tikhonov et al. (1995)) to the optimization problem (1.5), namely

$$\underset{\mathbf{w}}{\text{minimize}} \ \mathbf{w}^\top \mathbf{S}_n \mathbf{w} + \eta \mathbf{w}^\top \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1, \qquad (1.6)$$

where η is the regularization parameter. Similar approaches is used in the regression analysis, where the ridge regression uses the Tikhonov regularization to stabilize the least-squared estimator of the coefficients of the regression line (cf., Golub et al. (1999)). The solution of (1.6) is given by

$$\hat{\mathbf{w}}_{S;\lambda} = \frac{(\mathbf{S}_n + \eta \mathbf{I})^{-1} \mathbf{1}}{\mathbf{1}^\top (\mathbf{S}_n + \eta \mathbf{I})^{-1} \mathbf{1}}. \qquad (1.7)$$

Without loss of generality we set $\eta = \frac{1}{\lambda} - 1$ where $\lambda \in (0,1]$. Using this representation and

$$\mathbf{S}_\lambda = \lambda \mathbf{S}_n + (1-\lambda)\mathbf{I}$$

instead of $\mathbf{\Sigma}$ in (1.2) results in the same solution. However, the corner solutions are more easily understood as $\mathbf{S}_\lambda$ is a simple convex combination. If $\lambda \to 0$ then we put all our beliefs in the diagonal matrix $\mathbf{I}$, whereas if $\lambda \to 1$ then all beliefs are placed in $\mathbf{S}_n$.

Finally, combining the linear shrinkage estimator from Bodnar et al. (2018) we shrink the already regularized GMV portfolio weights as follows

$$\hat{\mathbf{w}}_{Sh;\lambda,\psi} = \psi \hat{\mathbf{w}}_{S;\lambda} + (1-\psi)\mathbf{b}, \qquad (1.8)$$

where ψ is the shrinkage intensity towards the target portfolio $\mathbf{b}$ with $\mathbf{b}^\top \mathbf{1} = 1$. This approach allows us to shrink the sample covariance matrix to decrease the variance and further decrease it by shrinking the weights themselves. It also gives a way for an investor to highlight stocks of he/she likes with the target portfolio $\mathbf{b}$. In many cases a naive portfolio $\mathbf{b} = \frac{1}{p}\mathbf{1}$ is a good choice but, in general, any deterministic target, which reflects the investment beliefs, is possible.

A common approach to determine shrinkage intensities is to use cross-validation (see, e.g., Tong et al. (2018) and Boileau et al. (2021)). That is, one aims to find the parameters λ and ψ such that some loss (or metric) $L(\lambda, \psi)$ is minimized (or maximized). There are of course different loss functions for the out-of-sample performance of optimal portfolios, see Lassance (2021) or Lassance et al. (2022) for a treatment of the out-of-sample mean-variance utility or the out-of-sample Sharpe ratio. Since our work concerns the GMV portfolio, the most natural choice of loss is the out-of-sample variance. It is given by

$$L(\lambda, \psi) = \hat{\mathbf{w}}_{Sh;\lambda,\psi}^\top \mathbf{\Sigma} \hat{\mathbf{w}}_{Sh;\lambda,\psi}. \qquad (1.9)$$

However, since $\mathbf{\Sigma}$ is not known we need to estimate it. Using K-fold cross-validation one partitions the data into train and validation sets. Using these sets one can perform a grid-search and use the validation set to estimate $\mathbf{\Sigma}$. The empirical out-of-sample variance is thereafter aggregated over fold to determine the best pair of shrinkage coefficients. However, using grid-search and cross-validation for such a problem introduces several obstacles. We have established that the sample covariance is a noisy estimate and validation sets are usually smaller than training sets. That naturally creates a more volatile estimator. Furthermore, it is not clear how big or small a grid should be. The approach we develop needs neither resampling methods nor grid search but instead relies on methods from random matrix theory (see, Bai and Silverstein (2010)). We develop a bona-fide type loss function which consistently estimates the true loss function in the high dimensional setting. The problem reduces to a simple univariate nonlinear optimization problem, which are easy to solve numerically. In Figure 1 we illustrate the optimal values of $\lambda^* - \lambda_{true}$ where each λ^* is given by the different methods that are derived in this paper and a 10-fold cross validation. The true shrinkage coefficient λ_{true} is given by the solution of the true loss function (1.9). In this illustration there are less assets $p = 150$ than data points
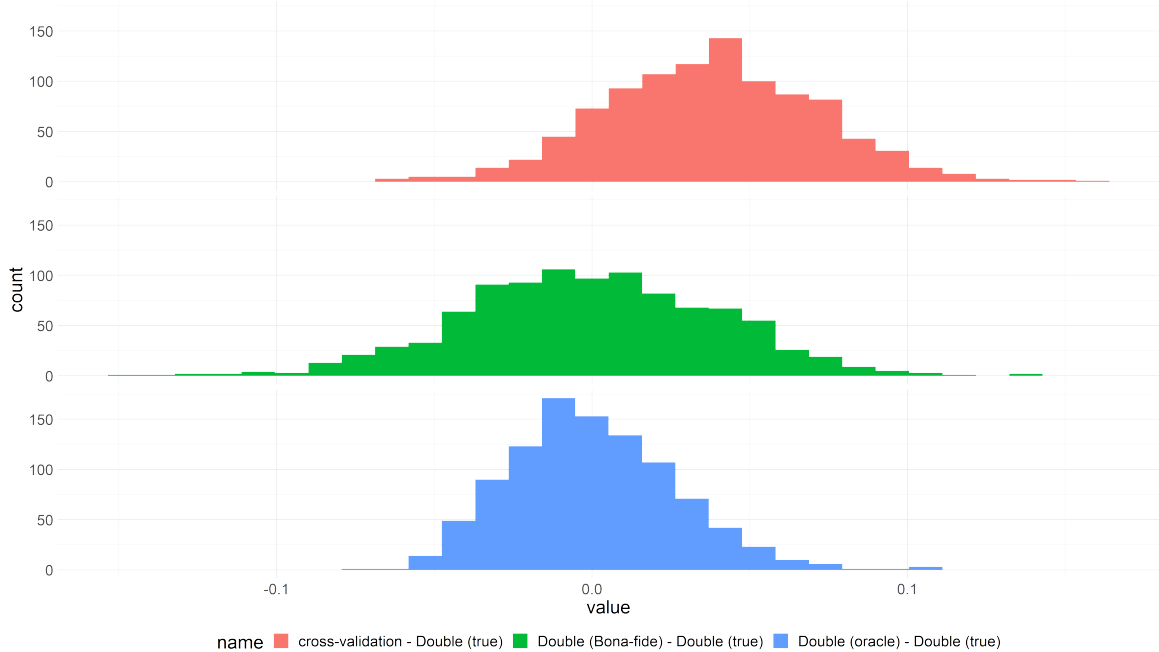
Figure 1: Estimated shrinkage intensities for λ based on the out-of-sample variance loss. The loss function (1.9) is determined through different methods. These are cross-validation and two Double shrinkage approaches. The Bona-fide loss is completely determined by data while the oracle loss depends on the unknown quantity $\mathbf{\Sigma}$, hence it has less variance. Each solution for λ is centered by the solution to the true loss (1.9).

$n = 300$ and $\mathbf{\Sigma}$ has a bounded spectral norm. The returns are simulated from a t-distribution with 5 degrees of freedom. Since $p/n < 1$ and $\mathbf{\Sigma}$ is kind (does not exhibit large eigenvalues) this is a scenario where a practitioner might assume that cross-validation, in particular, should work well. However, in this simple example cross-validation is clearly biased. The optimal values from two of the loss functions we derive, denoted as Double (Bona-fide) and Double (oracle), are centered around zero. Our work contributes to the literature by deriving the high-dimensional properties of the true loss function (1.9) as well as its Bona-fide counterpart, which can be easily used by practitioners. Furthermore, since we derive a Bona-fide estimator for the loss function one can easily extend the optimization problem to cover position restrictions, moment conditions or potentially other restrictions on the optimization problem.

The rest of the paper is organized as follows. In Section 2 the asymptotic properties of the out-of-sample variance are investigated in the high dimensional setting, while Section 3 presents a bona fide estimator of the asymptotic loss, which is the n used to find the optimal values of the two shrinkage intensities. The results of en extensive simulation study and of empirical applications are provided in Section 4, while Section 5 summarizes the obtained findings. The mathematical derivations are moved to the appendix (Section 6).

2 Out-of-sample variance and shrinkage estimation

Let $\mathbf{X}_n$ be a matrix of size $p \times n$ where its elements $\{x_{ij}\}_{ij}$ are independent and identically distributed (i.i.d.) real random variables with zero mean, unit variance and finite $4 + \epsilon$ moment for some $\epsilon > 0$. Assume that we observe the matrix $\mathbf{Y}_n$ according to the stochastic model

$$\mathbf{Y}_n \overset{d}{=} \boldsymbol{\mu}\mathbf{1}^\top + \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{X}_n \tag{2.1}$$

where $\boldsymbol{\Sigma}$ is a positive definite matrix of size $p \times p$ with a bounded spectral norm (its minimum and maximum eigenvalues are uniformly bounded in p from zero and infinity, respectively)[1]. The model belongs to the location-scale family but includes many skew or bi-modal families as well. Our aim is to estimate the shrinkage intensities λ, ψ from the following normalized optimization problem

$$\min_{\lambda,\psi} \frac{\hat{\mathbf{w}}_{Sh;\lambda,\psi}^\top \boldsymbol{\Sigma} \hat{\mathbf{w}}_{Sh;\lambda,\psi}}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}}. \tag{2.2}$$

The normalization is merely a technicality. The out-of-sample variance, or the loss function $L(\lambda,\psi) = \hat{\mathbf{w}}_{Sh;\lambda,\psi}^\top \boldsymbol{\Sigma} \hat{\mathbf{w}}_{Sh;\lambda,\psi}$, can be further simplified to

$$\begin{aligned}
L(\lambda,\psi) &= (\psi\hat{\mathbf{w}}_{S;\lambda} + (1-\psi)\mathbf{b})^\top \boldsymbol{\Sigma} \left(\psi\hat{\mathbf{w}}_{S;\lambda} + (1-\psi)\mathbf{b}\right) \\
&= (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}) \left(\psi - \frac{\mathbf{b}^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})}{(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})}\right)^2 \\
&\quad - \frac{\left(\mathbf{b}^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})\right)^2}{(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})} + \mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}.
\end{aligned} \tag{2.3}$$

For given λ, the out-of-sample variance is minimized with respect to ψ at

$$\psi_n^*(\lambda) = \frac{\mathbf{b}^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})}{(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})}, \tag{2.4}$$

while the optimal λ is found by maximizing the normalized second summand in (2.3), i.e.,

$$L_{n;2}(\lambda) = \frac{1}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{\left(\mathbf{b}^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})\right)^2}{(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})^\top \boldsymbol{\Sigma} (\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda})}. \tag{2.5}$$

In order to find the value λ^* together with $\psi^*(\lambda)$, which minimize the loss function, we proceed in three steps. First, we find the deterministic equivalent to $L_{n;2}(\lambda)$, and estimate it consistently in the second step. Finally, we minimize the obtained consistent estimator in the last step.

Theorem 2.1. *Let $\mathbf{Y}_n$ possess the stochastic representation as in (2.1). Assume that the relative loss of the target portfolio expressed as*

$$L_{\mathbf{b}} = \frac{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b} - \frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}}{\frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}} = \mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b} \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} - 1 \tag{2.6}$$

is uniformly bounded in p. Then it holds that

(i)

$$|L_{n;2}(\lambda) - L_2(\lambda)| \overset{a.s.}{\to} 0 \tag{2.7}$$

for $p/n \to c \in (0,\infty)$ as $n \to \infty$ with

$$L_2(\lambda) = \frac{\left(1 - \frac{1}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \boldsymbol{\Sigma} \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}\right)^2}{1 - \frac{2}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \boldsymbol{\Sigma} \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}} + \frac{1}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{(1 - v_2'(\eta,0)) \mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}{\left(\mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}\right)^2}} \tag{2.8}$$

---

[1]In fact the obtained results can be generalized to the case with finite number of unbounded largest eigenvalues, which would make the proofs more lengthy. Moreover, one can show that this assumption is only needed in case of centered sample covariance matrix, i.e., unknown mean vector. In case $\boldsymbol{\mu}$ is known, the boundedness of eigenvalues may be ignored due to normalization presented further in (2.2). More details could be deduced from the proofs of the main theorems.

(ii)

$$|\psi_n^*(\lambda) - \psi^*(\lambda)| \overset{a.s.}{\to} 0 \tag{2.9}$$

for $p/n \to c \in (0, \infty)$ as $n \to \infty$ with

$$\psi^*(\lambda) = \frac{1 - \frac{1}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \boldsymbol{\Sigma} \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}}{1 - \frac{2}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \boldsymbol{\Sigma} \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}} + \frac{1}{\mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}} \frac{(1 - v_2'(\eta, 0)) \mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}}{\left(\mathbf{1}^\top \boldsymbol{\Omega}_\lambda^{-1} \mathbf{1}\right)^2}}, \tag{2.10}$$

where

$$\eta = \frac{1}{\lambda} - 1, \quad \boldsymbol{\Omega}_\lambda = v(\eta, 0) \lambda \boldsymbol{\Sigma} + (1 - \lambda) \mathbf{I}, \tag{2.11}$$

$v(\eta, 0)$ is the solution of the following equation

$$v(\eta, 0) = 1 - c \left(1 - \frac{\eta}{p} tr \left((v(\eta, 0) \boldsymbol{\Sigma} + \eta \mathbf{I})^{-1} \right) \right), \tag{2.12}$$

and $v_2'(\eta, 0)$ is computed by

$$v_2'(\eta, 0) = 1 - \frac{1}{v(\eta, 0)} + \eta \frac{v_1'(\eta, 0)}{v(\eta, 0)^2}. \tag{2.13}$$

with

$$v_1'(\eta, 0) = v(\eta, 0) \frac{c \frac{1}{p} tr \left((v(\eta, 0) \boldsymbol{\Sigma} + \eta \mathbf{I})^{-1} \right) - c\eta \frac{1}{p} tr \left((v(\eta, 0) \boldsymbol{\Sigma} + \eta \mathbf{I})^{-2} \right)}{1 - c + 2c\eta \frac{1}{p} tr ((v(\eta, 0) \boldsymbol{\Sigma} + \eta \mathbf{I})^{-1}) - c\eta^2 \frac{1}{p} tr ((v(\eta, 0) \boldsymbol{\Sigma} + \eta \mathbf{I})^{-2})}. \tag{2.14}$$

The proof of Theorem 2.1 can be found in the appendix. Theorem 2.1 provides the deterministic equivalents for the loss function $L_{n;2}(\lambda)$ and optimal shrinkage intensity $\psi_n^*(\lambda)$. The solution $v(\eta, 0)$ to the equation (2.12) is a Stieltjes transform (see Rubio and Mestre (2011)) and inherits all the properties of these functionals. The properties are extremely important for the consistency of the loss function. However, these deterministic equivalents are not applicable in practice since they depend on the unknown parameter $\boldsymbol{\Sigma}$ and $v(\eta, 0)$. Fortunately, we can create consistent bona-fide [2] estimators for both deterministic equivalents $L_2(\lambda)$ and $\psi^*(\lambda)$ in a high dimensional setting.

3   Bona fide estimation

In this section, we construct bona fide consistent estimators for $L_2(\lambda)$ and $\psi^*(\lambda)$ in the high-dimensional asymptotic setting. First, in Theorem 3.1 we derive the consistent estimators for $v(\eta, 0)$, $v_1'(\eta, 0)$, and $v_2'(\eta, 0)$. The proof of Theorem 3.1 is given in the appendix.

Theorem 3.1. *Let $\mathbf{Y}_n$ possess the stochastic representation as in* (2.1). *Then it holds that*

$$|\hat{v}(\eta, 0) - v(\eta, 0)| \overset{a.s}{\to} 0, \tag{3.1}$$

$$\left|\hat{v}_1'(\eta, 0) - v_1'(\eta, 0)\right| \overset{a.s}{\to} 0, \tag{3.2}$$

$$\left|\hat{v}_2'(\eta, 0) - v_2'(\eta, 0)\right| \overset{a.s}{\to} 0, \tag{3.3}$$

---

[2]With "bona-fide" we understand a concept of purely data-driven estimators, which do not depend on the unknown quantities. Thus, they are ready to be used in practice without any modifications.

for $p/n \to c \in (0, \infty)$ as $n \to \infty$ with

$$\hat{v}(\eta, 0) = 1 - c\left(1 - \eta \frac{1}{p} tr\left((\mathbf{S}_n + \eta \mathbf{I})^{-1}\right)\right),$$

$$\hat{v}'_1(\eta, 0) = \hat{v}(\eta, 0)c\left(\frac{1}{p}tr\left((\mathbf{S}_n + \eta \mathbf{I})^{-1}\right) - \eta\frac{1}{p}tr\left((\mathbf{S}_n + \eta \mathbf{I})^{-2}\right)\right),$$

$$\hat{v}'_2(\eta, 0) = 1 - \frac{1}{\hat{v}(\eta, 0)} + \eta\frac{\hat{v}'_1(\eta, 0)}{\hat{v}(\eta, 0)^2}.$$

Theorem 3.2 provides the consistent estimators for the building blocks used in the construction of the consistent estimators for $L_2(\lambda)$ and $\psi^*(\lambda)$. The proof of Theorem 3.2 is presented in the appendix.

Theorem 3.2. *Let $\mathbf{Y}_n$ possess the stochastic representation as in (2.1). Assume that the relative loss of the target portfolio given in (2.6) is uniformly bounded in p. Let $\frac{\mathbf{b}^\top \mathbf{\Sigma}^{-1}\mathbf{b}}{\mathbf{b}^\top \mathbf{\Sigma}\mathbf{b}}$ be uniformly bounded in p. Then it holds that*

$$\left|\frac{\mathbf{b}^\top \mathbf{S}\mathbf{b}}{\mathbf{b}^\top \mathbf{\Sigma}\mathbf{b}} - 1\right| \xrightarrow{a.s} 0, \tag{3.4}$$

$$\left|\frac{\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}} - \frac{\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}}\right| \xrightarrow{a.s} 0, \tag{3.5}$$

$$\left|\frac{\lambda^{-1}}{\hat{v}(\eta, 0)}\frac{1 - (1-\lambda)\mathbf{b}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}}{\sqrt{\mathbf{b}^\top \mathbf{\Sigma}\mathbf{b}\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}}} - \frac{\mathbf{b}^\top \mathbf{\Sigma}\mathbf{\Omega}_\lambda^{-1}\mathbf{1}}{\sqrt{\mathbf{b}^\top \mathbf{\Sigma}\mathbf{b}\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}}}\right| \xrightarrow{a.s} 0, \tag{3.6}$$

$$\left|\frac{1}{\lambda\hat{v}(\eta, 0)}\frac{\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}} - \frac{1-\lambda}{\lambda\hat{v}(\eta, 0)}\frac{\mathbf{1}^\top \mathbf{S}_\lambda^{-2}\mathbf{1} - \lambda^{-1}\frac{\hat{v}'_1(\eta,0)}{\hat{v}(\eta,0)}\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}\left(1 - \frac{\hat{v}'_1(\eta,0)}{\hat{v}(\eta,0)}(\frac{1}{\lambda}-1)\right)} - \frac{\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1}\mathbf{\Sigma}\mathbf{\Omega}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}}\right| \xrightarrow{a.s} 0 \tag{3.7}$$

for $p/n \to c \in (0, \infty)$ as $n \to \infty$ with $\eta = 1/\lambda - 1$.

Let

$$d_1(\eta) = \frac{\lambda^{-1}}{\hat{v}(\eta, 0)}\left(1 - (1-\lambda)\mathbf{b}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}\right) \tag{3.8}$$

and

$$d_2(\eta) = \frac{1}{\lambda\hat{v}(\eta, 0)}\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1} - \frac{1-\lambda}{\lambda\hat{v}(\eta, 0)}\frac{\mathbf{1}^\top \mathbf{S}_\lambda^{-2}\mathbf{1} - \lambda^{-1}\frac{\hat{v}'_1(\eta,0)}{\hat{v}(\eta,0)}\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}}{1 - \frac{\hat{v}'_1(\eta,0)}{\hat{v}(\eta,0)}(\frac{1}{\lambda}-1)}. \tag{3.9}$$

The application of the results derived in Theorems 3.1 and 3.2 leads to a consistent bona fide estimator for $L_2(\lambda)$ and $\psi^*(\lambda)$ presented in Theorem 3.3.

Theorem 3.3. *Let $\mathbf{Y}_n$ possess the stochastic representation as in (2.1). Assume that the relative loss of the target portfolio given in (2.6) is uniformly bounded in p. Let $\frac{\mathbf{b}^\top \mathbf{\Sigma}^{-1}\mathbf{b}}{\mathbf{b}^\top \mathbf{\Sigma}\mathbf{b}}$ be uniformly bounded in p. Then it holds that*

(i)

$$\left|\hat{L}_{n;2}(\lambda) - L_2(\lambda)\right| \xrightarrow{a.s.} 0 \tag{3.10}$$

for $p/n \to c \in (0, \infty)$ as $n \to \infty$ with

$$\hat{L}_{n;2}(\lambda) = \frac{\left(1 - \frac{1}{\mathbf{b}^\top \mathbf{S}\mathbf{b}}\frac{d_1(\eta)}{\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}}\right)^2}{1 - \frac{2}{\mathbf{b}^\top \mathbf{S}\mathbf{b}}\frac{d_1(\eta)}{\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}} + \frac{1}{\mathbf{b}^\top \mathbf{S}\mathbf{b}}\frac{(1-\hat{v}'_2(\eta,0))d_2(\eta)}{(\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1})^2}}, \tag{3.11}$$

7

(ii)

$$\left|\hat{\psi}_n^*(\lambda) - \psi^*(\lambda)\right| \overset{a.s.}{\to} 0 \tag{3.12}$$

for $p/n \to c \in (0,\infty)$ as $n \to \infty$ with

$$\hat{\psi}_n^*(\lambda) = \frac{1 - \frac{1}{\mathbf{b}^\top \mathbf{Sb}} \frac{d_1(\eta)}{\mathbf{1}^\top \mathbf{S}_\lambda^{-1} \mathbf{1}}}{1 - \frac{2}{\mathbf{b}^\top \mathbf{Sb}} \frac{d_1(\eta)}{\mathbf{1}^\top \mathbf{S}_\lambda^{-1} \mathbf{1}} + \frac{1}{\mathbf{b}^\top \mathbf{Sb}} \frac{(1 - \hat{v}_2'(\eta,0))d_2(\eta)}{\left(\mathbf{1}^\top \mathbf{S}_\lambda^{-1} \mathbf{1}\right)^2}}, \tag{3.13}$$

where $\eta = 1/\lambda - 1$, $\hat{v}_2'(\eta,0)$ is provided in Theorem 3.1, $d_1(\eta)$ and $d_2(\eta)$ are given in (3.8) and (3.9), respectively.

The three loss functions (bona fide, oracle and true) are illustrated in Figure 2 for two different values of p. Under oracle loss function we understand the asymptotic equivalent of $L_{n;2}$, namely L_2. When p is equal to 150 the differences between the functions look, at least graphically, very small. The relative difference it at most 20% and the optimal λ's are extremely close to each other. We still need to face the fact that the out-of-sample variance is slightly over-estimated, even though the bona-fide estimator will be asymptotically valid. When p is equal to 450 and c is greater than one we observe a slightly different picture. For $p > n$ the bona fide loss function is not necessarily concave. This is due to the fact that as λ approaches 1, and c is greater than one, $\lambda \mathbf{S}_n + (1 - \lambda)\mathbf{I}$ becomes closer and closer to a singular matrix. Thus, the eigenvalues of the inverse of the shrunk sample covariance matrix explode. This issue could be repaired using a different type of ridge regularization mentioned in (Bodnar et al., 2022, formula (2.33)), where Moore-Penrose inverse for $c > 1$ can be employed for $\lambda \to 1$. This interesting observation is left for the future investigations.

4   Numerical study

In this section we will conduct a simulation study to assess the finite sample properties of the suggested double shrinkage estimator and to compare its behaviour with existent approaches. Due to the asymptotic nature of our procedure we first devote some attention to the finite sample properties of the suggested estimator under different data-generating processes. We end this section with an empirical application of the methods on the assets from S&P 500.

4.1   Setup of the simulation study

In the simulation study we will use the following four different stochastic models for the data-generating process:

Scenario 1: t-distribution The elements of $\mathbf{x}_t$ are drawn independently from t-distribution with 5 degrees of freedom, i.e., $x_{tj} \sim t(5)$ for $j = 1, ..., p$, while $\mathbf{y}_t$ is constructed according to (2.1).

Scenario 2: CAPM The vector of asset returns $\mathbf{y}_t$ is generated according to the CAPM (Capital Asset Pricing Model), i.e.,
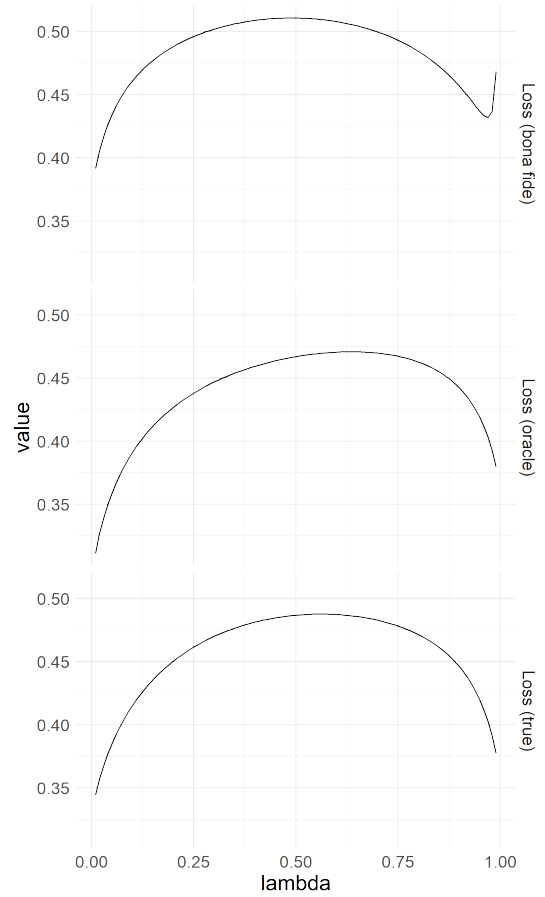
$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\beta} z_t + \boldsymbol{\Sigma}^{1/2} \mathbf{x}_t,$$

with independently distributed $z_t \sim N(0,1)$ and $\mathbf{x}_t \sim N_p(\mathbf{0}, \mathbf{I})$. The elements of vector $\boldsymbol{\beta}$ are drawn from the uniform distribution, that is $\beta_i \sim U(-1,1)$ for $i = 1, ..., p$.
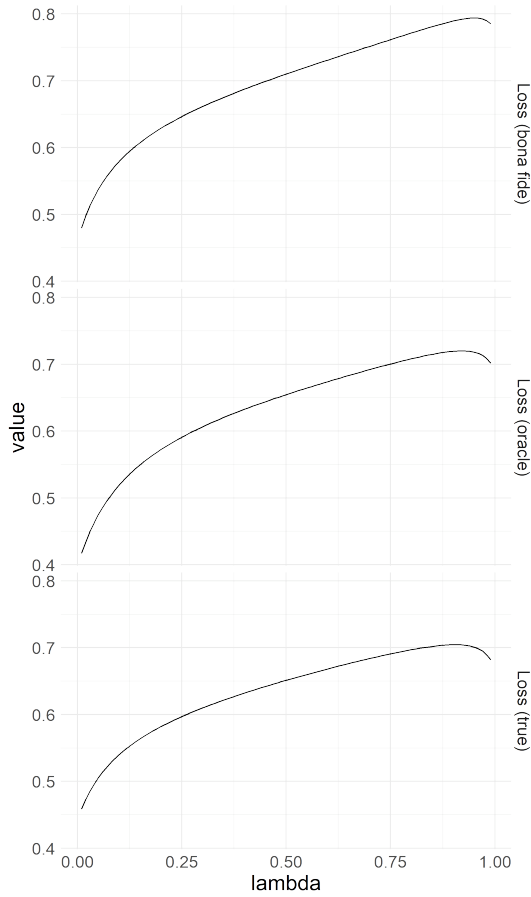
(a) p=150, n=300   (b) p=450, n=300

Figure 2: The loss functions from Theorem 2.1-3.3 illustrated over different values of $\lambda \in (0,1)$. The data were simulated from a t-distribution with 5 degrees of freedom and used the equally weighted portfolio as a target.

Scenario 3: CCC-GARCH model of Bollerslev (1990) The asset returns are simulated according to

$$\mathbf{y}_t | \mathbf{\Sigma}_t \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma}_t)$$

where the conditional covariance matrix is specified by

$$\mathbf{\Sigma}_t = \mathbf{D}_t^{1/2} \mathbf{C} \mathbf{D}_t^{1/2} \quad \text{with} \quad \mathbf{D}_t = \text{diag}(h_{1,t}, h_{2,t}, ..., h_{p,t}),$$

where

$$h_{j,t} = \alpha_{j,0} + \alpha_{j,1}(\mathbf{y}_{j,t-1} - \boldsymbol{\mu}_j)^2 + \beta_{j,1} h_{j,t-1}, \text{ for } j = 1, 2, ..., p, \text{ and } t = 1, 2, ..., n_i, \ i = 1, ..., T.$$

The coefficients of the CCC model are sampled according to $\alpha_{j,1} \sim U(0, 0.1)$ and $\beta_{j,1} \sim U(0.6, 0.7)$ which implies that the stationarity conditions, $\alpha_{j,1} + \beta_{j,1} < 1$, are always fulfilled. The constant correlation matrix $\mathbf{C}$ is induced by $\mathbf{\Sigma}$. The intercept $\alpha_{j,0}$ is chosen such that the unconditional covariance matrix is equal to $\mathbf{\Sigma}$.

Scenario 4: VARMA model The vector of asset returns $\mathbf{y}_t$ is simulated according to

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{\Gamma}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \mathbf{\Sigma}^{1/2} \mathbf{x}_t \quad \text{with} \quad \mathbf{x}_t \sim N_p(\mathbf{0}, \mathbf{I})$$

for $t = 1, ..., n + m$, where $\mathbf{\Gamma} = \text{diag}(\gamma_1, \gamma_2, ..., \gamma_p)$ with $\gamma_i \sim U(-0.9, 0.9)$ for $i = 1, ..., p$. Note that in the case of the VAR model, the covariance matrix of $\mathbf{y}_t$ is computed as $\text{vec}(\mathbb{V}ar(\mathbf{y})) = (\mathbf{I} - \mathbf{\Gamma} \otimes \mathbf{\Gamma})^{-1} \text{vec}(\mathbf{\Sigma})$ where vec denotes the vec operator. This matrix is thereafter used in the computation of the limiting objects.

We will repeat each scenario 1000 times for a number of configurations where the concentration ratio c will range from 0.25 to 2.7 and $n = 100, 200, 300, 400$. The portfolios contain at most 1080 assets which implies that we are estimating close to 600000 parameters as well as the two shrinkage coefficients. The parameters of the model are simulated in the following manner. The elements of the mean vector $\boldsymbol{\mu}$ are simulated from a uniform distribution with $\mu_i \sim U(-0.1, 0.1)$. To simulate the covariance matrix we make use of the function **RandCov-Mat** from the HDShOP package (see, Bodnar et al. (2021a)).

4.2   Comparison to benchmark strategies

In this section we will investigate the performance of five different methods. We will consider the following type of portfolios

1. The portfolio allocation problem obtained from Theorem 3.3, which we will abbreviate "Double".

2. The linear shrinkage estimator of the GMV portfolio weights from Bodnar et al. (2018), which we will abbreviate "BPS"

3. The linear shrinkage estimator of the GMV portfolio weights from Frahm and Memmel (2010), which we will abbreviate "FM". This portfolio can be constructed only for $c < 1$ following the approach suggested in Frahm and Memmel (2010).

4. The nonlinear shrinkage estimator of the covariance matrix from Ledoit and Wolf (2020a) which is used in the GMV portfolio as a replacement for $\mathbf{\Sigma}$. We will abbreviate this portfolio strategy as "LW2020".

5. The traditional GMV portfolio which we will abbreviate "Traditional". Whenever $c > 1$ we will use the Moore-Penrose inverse of $\mathbf{S}_n$ to compute the sample weights of the GMV portfolio.
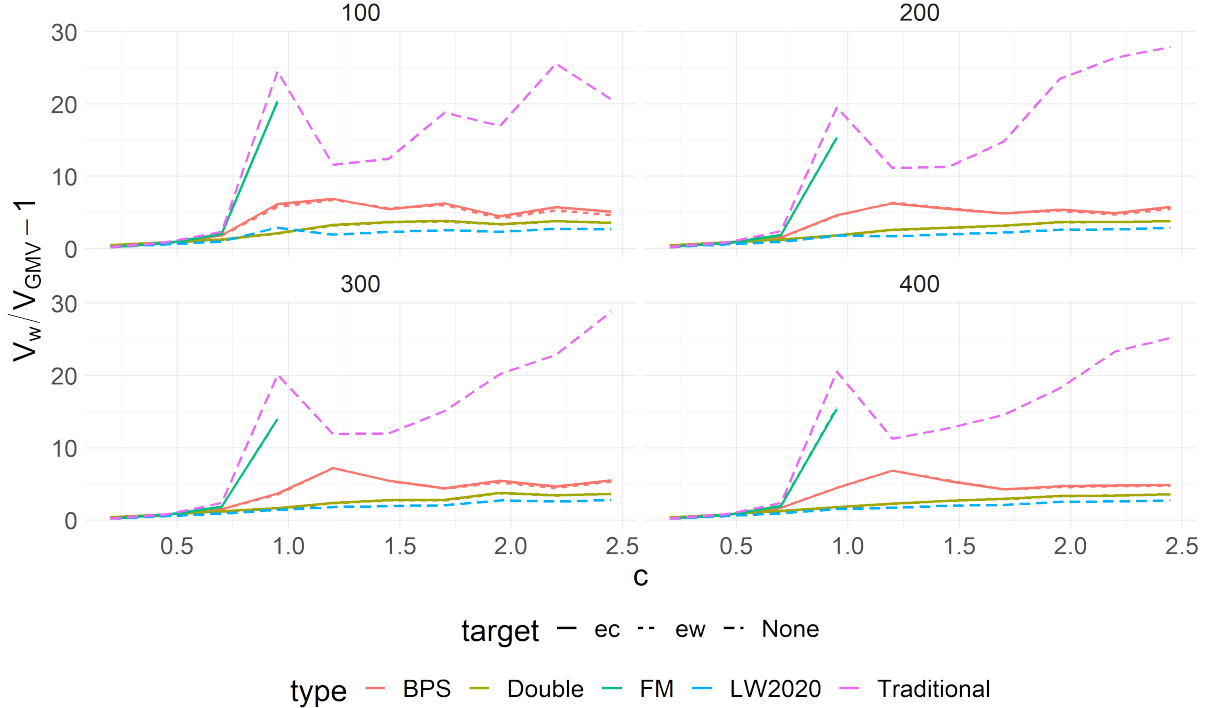
Figure 3: Relative loss $V_{\mathbf{w}}/V_{GMV} - 1$ computed for several estimator of the GMV portfolio weights under scenario 1. Notice that some methods use different targets but are of the same type.

Notice that the first three types of portfolios can take many target portfolios. We therefore include two target portfolios as well as other benchmarks in the forthcoming comparisons. These are the equally weighted portfolio and an equal correlation target. The first is deterministic and does not depend on data, which is in line with what Theorem 3.3 assumes. The second target portfolio depends on data. It assumes that the all assets share the same correlation but have different volatility. For each scenario we will display the relative loss $V_{\mathbf{w}}/V_{GMV} - 1$, where $V_{\mathbf{w}} = \mathbf{w}^{\top} \boldsymbol{\Sigma} \mathbf{w}$. A value close to zero indicates a good estimate of the in-sample loss.

In Figure 3 we can see the results of the simulations under scenario 1. Each color represent a "type" of portfolio while the linetype highlights what type of target BPS, Double and FM use. For small c the loss is not significantly different between the different types. However, as c becomes larger, the results diverge from each other. Regardless of n the largest loss is provided by Traditional portfolio. The Traditional estimator is famously bad in higher dimensions, see e.g. Bodnar et al. (2018). The FM portfolio is only defined for $c < 1$ so the loss for this method is not presented thereafter. The third best method is BPS using equal correlation and equally weighted as target portfolios. The uncertainty from the target portfolio does not seem to impact the loss a lot. It achieves the same amount of loss as the equally weighted target. The best performing portfolios are the Double and LW2020. For smaller n the difference is more pronounced, LW2020 provides the smallest loss. However, as n increases the Double portfolio with the equally weighted target provides quite similar performance. The LW2020 portfolio is 3% better (lower loss) in comparison to the Double portfolio when $n = 400$.

In Figure 4 we can see the results of the simulation study conducted under scenario 2. As with scenario 1, we can see the same type of ordering. However, in scenario 2 there is a pronounced increase for the loss of LW2020 when $n = 100$ around $c = 1$. The Double portfolios do not seem to suffer from that issue. In all other cases the loss is similar with the smallest difference between Double and LW2020 equal to 3.1%. The difference is not large since the
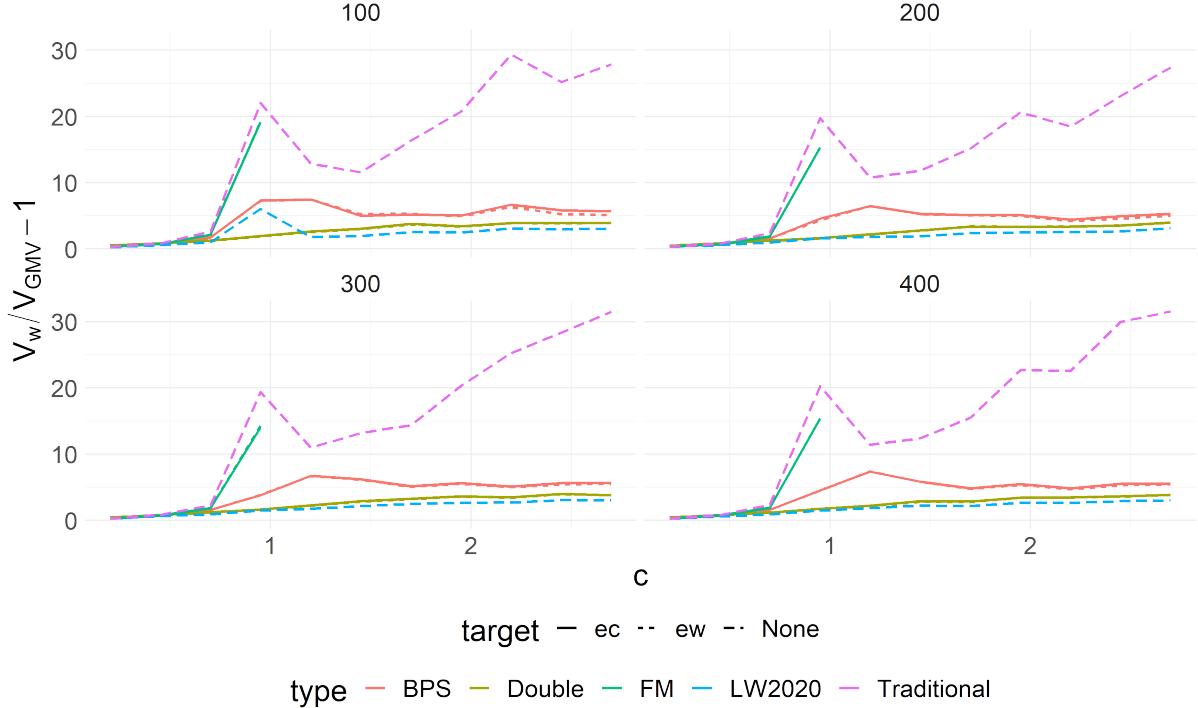
Figure 4: Relative loss $V_{\mathbf{w}}/V_{GMV} - 1$ computed for several estimator of the GMV portfolio weights under scenario 2. Notice that some methods use different targets but are of the same type.

inverse covariance matrix for CAPM is a one rank update from scenario 1. It has to be noted that in this case the largest eigenvalue of $\boldsymbol{\Sigma}$ is not bounded anymore. There is a little bit more noise, but there is no more temporal or structural dependence that we do not take care of in (2.1).

Figure 5 depicts relative loses computed for the considered estimators of the GMV portfolio under scenario 3. It displays almost exactly the same plots as shown in Figures 3 and 4. The introduction of temporal dependence is not dramatic in terms of the relative loss. We still see the same type of ordering where Double and LW2020 are almost equally good when n is sufficiently large.

In Figure 6 we can see the results obtained under scenario 4. Although the same ordering seems to hold the scale is not the same. Some methods, namely the Traditional, FM and BPS, have a much larger loss in comparison to previous scenarios. There is is a very little difference between the other methods for larger c in contrast to previous scenarios. The difference between the two best methods, the Double and LW2020 portfolios, is as small as 1% apart when $n = 400$ and c is large.

All in all, the results of this simulation experiment justify that the proposed method is at least as good as the nonlinear shrinkage technique, which is already proved to be a state-of-the-art method for the estimation of large dimensional covariance matrices. Thus, it is of high importance to test it on a real data set using some other empirical measures of performance like the out-of-sample variance, return, Sharpe ratio, turnover etc.

4.3   Empirical application

In this section we will apply the different benchmark strategies on empirical data. The data constitute of daily (log returns) for 431 assets from the S&P500 index. The out-of-sample data ranges from 2013-01-01 to 2021-11-04. The in-sample data ranges back to early 2011. We will
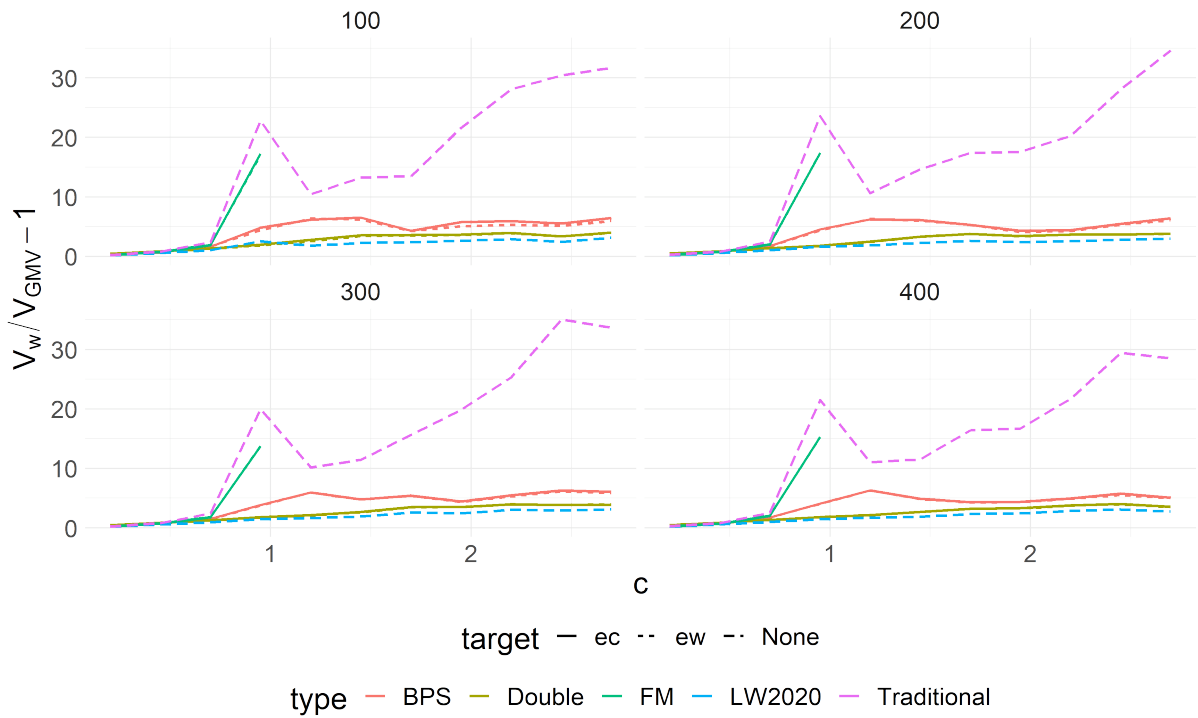
Figure 5: Relative loss $V_{\mathbf{w}}/V_{GMV} - 1$ computed for several estimator of the GMV portfolio weights under scenario 3. Notice that some methods use different targets but are of the same type.

Figure 6: Relative loss $V_{\mathbf{w}}/V_{GMV} - 1$ computed for several estimator of the GMV portfolio weights under scenario 4. Notice that some methods use different targets but are of the same type.
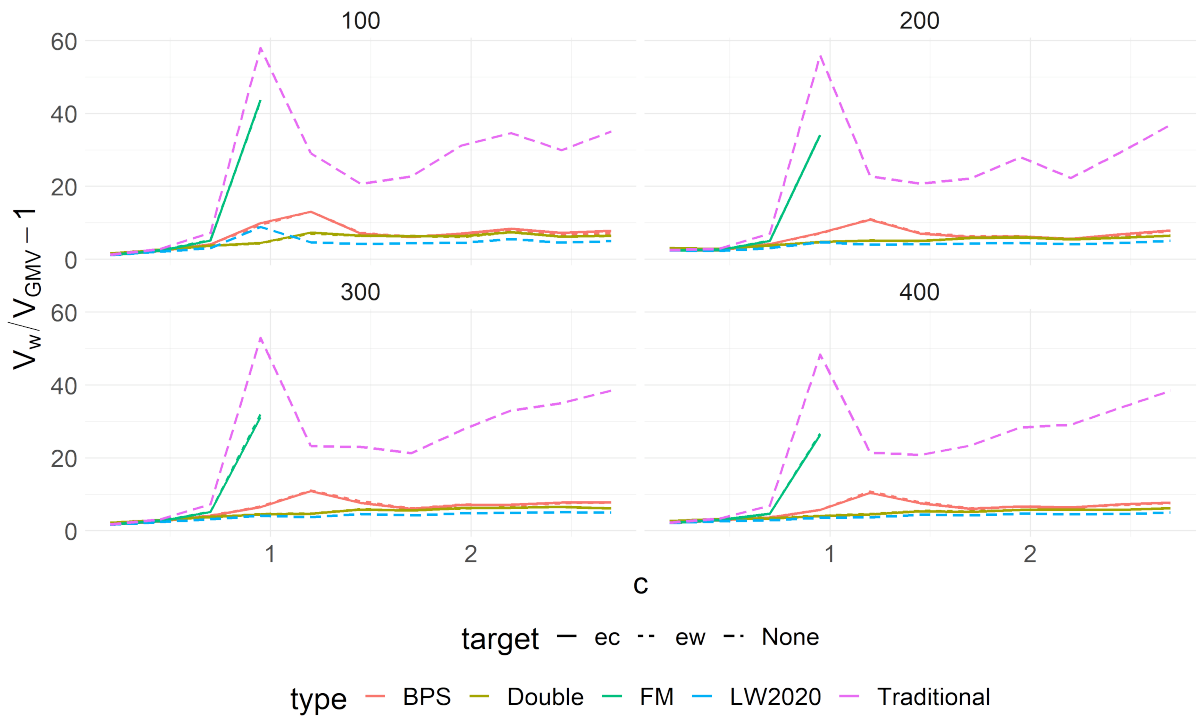
follow the previous section in that we use the equal correlation (ec) and equally weighted (ew) portfolios as targets. In this empirical application we fix the window size to $n = 250$ or $n = 500$. We thereafter change the portfolio size p. The three different portfolio sizes we consider are 260, 400 and 431. Thus, the window size $n = 250$ reflects $c \in \{1.04, 1.6, 1.724\}$ and $n = 500$ stands for $c \in \{0.52, 0.8, 0.862\}$.

All portfolios aim to minimize portfolio variance. Since the true portfolio variance is not available the most natural evaluation method should be which strategy provide the smallest out-of-sample variance, which will be denoted by $\sigma^{(k)}$. However, any portfolio is more than its volatility. A portfolio with small volatility does not necessarily provide a feasible return nor does it provide a feasible portfolio to invest in. We will therefore use the out-of-sample mean, which we denote $\bar{\mathbf{y}}_{\mathbf{w}}^{(k)}$, as well as the out-of-sample Sharpe Ratio, denoted $SR^{(k)}$, to investigate the properties of the portfolio return distribution. Moreover, the stability of the portfolio weights also reflects how risky it is. To investigate the characteristics of the portfolio weights we will also consider the following performance measures

$$|\mathbf{w}^{(k)}| = \frac{1}{Tp} \sum_{i=1}^{T} \sum_{j=1}^{p} |w_{i,j}^{(k)}|, \tag{4.1}$$

$$\max \mathbf{w}^{(k)} = \frac{1}{T} \sum_{i=1}^{T} \left(\max_{j} w_{i,j}^{(k)} \right), \tag{4.2}$$

$$\min \mathbf{w}^{(k)} = \frac{1}{T} \sum_{i=1}^{T} \left(\min_{j} w_{i,j}^{(k)} \right), \tag{4.3}$$

$$\mathbf{w}_{i}^{(k)} \mathbb{1}(\mathbf{w}_{i}^{(k)} < 0) = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{p} w_{i,j}^{(k)} \mathbb{1}(w_{i,j}^{(k)} < 0), \tag{4.4}$$

$$\mathbb{1}(\mathbf{w}_{i}^{(k)} < 0) = \frac{1}{Tp} \sum_{i=1}^{T} \sum_{j=1}^{p} \mathbb{1}(w_{i,j}^{(k)} < 0). \tag{4.5}$$

The first measure shown in equation (4.1) is equal to the average size of the portfolio positions. A large value of this measure would indicate that the portfolio takes large positions (both negative and positive). It is a common critique to ordinary mean-variance portfolios because large positions are risky themselves. The second measure, shown in equation (4.2), is equal to the average long position. It is similar to the previous measure but only considers long positions of the portfolio. As with the above, small positions are preferred to large positions. Equation (4.3) shows the average short position. With this measure we try to showcase how big short positions are. This is especially important since large short positions have potentially infinite risk. There is no limit to how much you can loose. Because of the great importance of how big short positions actually are we also include two further measures, which can be seen in (4.4) and (4.5). The former can be interpreted as the average size of the negative positions. The latter is the average proportion of short positions.

Note that for a portfolio with constant portfolio weights (4.1)-(4.5) are constant. For the equally weighted portfolio we have

$$|\mathbf{w}^{(k)}| = \max \mathbf{w}^{(k)} = \min \mathbf{w}^{(k)} = \frac{1}{p}, \; \mathbf{w}_{i}^{(k)} \mathbb{1}(\mathbf{w}_{i}^{(k)} < 0) = \mathbb{1}(\mathbf{w}_{i}^{(k)} < 0) = 0. \tag{4.6}$$

In Table 1 we display the results from the first experiment with window size equal to 250 days. Due to (4.6) we choose not to state the results for ew portfolio for (4.1)-(4.5). The FM strategy has been removed since it has a very similar performance as the Traditional strategy for $c < 1$ and it is not defined for $c > 1$. For a moderately small portfolio $p = 260$ the double shrinkage portfolio with equally correlated target provides the smallest out-of-sample variance

which is denoted boldface. The double shrinkage with equally weighted target comes in second, indicated by the *. There is a small difference between the different Double strategies using different targets, around 3 %. However, they are all of similar performance except the double shrinkage with equally weighted target. It provides more stable weights with zero short positions and smallest turnover. The portfolio using LW2020 is ranked third in terms of volatility but is dominating other strategies in terms of mean and Sharpe ratio. The portfolio weights of nonlinear shrinkage, however, are not as stable as double shrinkage with ew target. The second best Sharpe ratio is provided by the Double with ew as a target. The difference with nonlinear shrinkage is around 0.002. With the equally weighted target the Double is on par with LW2020. All others are far worse. The smallest Turnover is consistently provided by the Double shrinkage approach. The portfolio weights are very stable. Furthermore, the best performing portfolios, in terms of their characteristics, are given by the Double shrinkage approach with ec or ew as a target portfolio. It is almost always a trade between them. These two take smaller positions on average, smaller short positions and less proportion of shorted weights. The natural ordering is that Traditional is worst and BPS being second to worst in terms of the portfolio characteristics. The LW2020 method comes in second, while the Double shrinkage portfolio being the best.

When $p = 400$ the equal correlation portfolio provides the best volatility estimate and equally weighted comes in second. LW2020 is the third best when it comes to volatility but is the best in terms of return. This brings it to being the best when it comes to the Sharpe ratio. Double shrinkage with ec comes in second. In this scenario we can also see that the Traditional portfolio is the most volatile but not as much as it could be expected. This is probably the Moore-Penrose inverse, which is still working reasonably well when c is in the neighborhood of one. The Double shrinkage portfolio provides similar performance to the case $p = 260$. It is always among the best performing.

In the large dimensional case $p = 431$, the most performing portfolio is the Double with equally correlated target when it comes to volatility. The estimation uncertainty has a large effect in these dimensions. However, second to best is the nonlinear shrinkage portfolio with slightly higher Sharpe ratio as double shrinkage. In terms of the mean, however, the best is the double shrinkage portfolio with equally weighted target. The Double with ew and ec as targets are always among the best and show very stable behaviour with a tiny turnover.

Next, we perform the same experiment with $n = 500$. In Table 2 we can see the results. When $p = 260$, we are in an intermediate concentration ratio case. It is equal to $c = 260/500 = 0.52$. In this scenario the equally weighted portfolio is the best one in terms of volatility. The Double shrinkage approach with ew as a target is 5% worse. As per usual, the Traditional estimator is the worst on estimating the volatility although not far behind in this small dimensional scenario. The portfolio with best average return is given by the Double shrinkage with ec as a target. Since it is also among the best in estimating volatility it gives the largest Sharpe ratio. The Double shrinkage with ew as a target is second to best. In terms of the portfolio characteristics the Double shrinkage approach with ew as a target dominates everything. It is the best in all scenarios. Similar picture is for $p = 400$, here the best is the Double with ec as a target. The Traditional is showing surprisingly the largest return. However, it does so at a large cost (large variance and turnover).

Table 1: Characteristics of the different strategies using a moving window approach. The out-of-sample period equals to 2234 days. The window size is held fixed, equal to 250.

| name | BPS | | Double | | | | |
	ec	ew	ec	ew	ew	LW2020	Traditional
p=260							
σ^k	0.03912	0.039	**0.01104**	*0.01129**	0.01133	0.01165	0.04597
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	−0.000276	−0.00025	0.000397	0.000503	*0.000505**	**0.000542**	−0.000352
SR^k	−0.007	−0.006	0.036	*0.045**	*0.045**	**0.047**	−0.008

15

name	BPS ec	BPS ew	Double ec	Double ew	ew	LW2020	Traditional
Turnover	33267.6	33703.26	*69.74**	**3.36**		2301.52	41589.29
$|\mathbf{w}^{(k)}|$	0.0951	0.0961	*0.0066**	**0.0038**		0.0102	0.1189
$\max \mathbf{w}^{(k)}$	0.4494	0.4496	*0.0393**	**0.0046**		0.0412	0.5569
$\min \mathbf{w}^{(k)}$	-0.4209	-0.4292	*-0.0058**	**0.0028**		-0.03	-0.5327
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	-0.0929	-0.0945	*-0.0036**	**$-2e-04$**		-0.0099	-0.1171
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.491	0.488	0.385	**0**		*0.32**	0.491
p=400							
σ^k	0.01398	0.01392	**0.01106**	*0.01111**	0.01112	0.01193	0.01524
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	0.000525	0.000464	*0.00064**	0.000514	0.000509	**0.000706**	0.000458
SR^k	0.038	0.033	*0.058**	0.046	0.046	**0.059**	0.03
Turnover	3661.98	3678.92	*69.71**	**4.81**		667.62	4776.77
$|\mathbf{w}^{(k)}|$	0.0192	0.019	*0.0043**	**0.0025**		0.0107	0.0246
$\max \mathbf{w}^{(k)}$	0.0788	0.0769	*0.0256**	**0.0032**		0.0481	0.0991
$\min \mathbf{w}^{(k)}$	-0.0699	-0.07	*-0.0041**	**0.0013**		-0.0396	-0.0913
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	-0.018	-0.018	*-0.0024**	**$-6e-04$**		-0.0097	-0.0235
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.465	0.46	*0.388**	**0.004**		0.423	0.471
p=431							
σ^k	0.0121	0.01217	**0.01088**	0.01117	0.01116	*0.01105**	0.01322
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	0.000437	0.00042	0.000496	**0.000516**	0.000512	*0.000515**	0.000375
SR^k	0.036	0.034	*0.046**	*0.046**	*0.046**	**0.047**	0.028
Turnover	3165.55	3192.65	*69.99**	**5.19**		660.56	4254.45
$|\mathbf{w}^{(k)}|$	0.0164	0.0161	*0.004**	**0.0023**		0.0101	0.0214
$\max \mathbf{w}^{(k)}$	0.0663	0.0645	*0.0254**	**0.0031**		0.0459	0.0851
$\min \mathbf{w}^{(k)}$	-0.0589	-0.0591	*-0.0038**	**0.0011**		-0.0381	-0.0795
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	-0.0153	-0.0152	*-0.0022**	**$-6e-04$**		-0.0093	-0.0204
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.46	0.454	*0.39**	**0.004**		0.423	0.468

* Second to best

In the large dimensional case, where $p = 431$, we can note that Double with ec as a target is now the best in terms of volatility whereas the ew portfolio is second to best. However, the difference is small. The equally weighted is good in general, as seen in previous examples but another explanation is that the window is to large for our data-driven portfolios to cope with changes. The BPS portfolio provides the largest mean with equal correlation target. The best Sharpe ratio is provided by the LW2020 with Double just slightly behind. Traditional still provides the largest return but in terms of variance, turnover and all other measures it is the most unstable and risky portfolio.

Table 2: Characteristics of the different strategies using a moving window approach. The out-of-sample period equals to 2234 days. The window size is held fixed, equal to 500.

| | BPS | | Double | | | | |
name	ec	ew	ec	ew	ew	LW2020	Traditional
p=260							
σ^k	0.01593	0.01596	0.01131	*0.01123**	**0.01119**	0.01215	0.01821
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	0.000439	0.000372	**0.000621**	0.000556	0.000524	*0.000562**	0.00033
SR^k	0.028	0.023	**0.055**	*0.05**	0.047	0.046	0.018
Turnover	1016.3	1039.9	*224.46**	**221.3**		442.06	1385.55
$|\mathbf{w}^{(k)}|$	0.0259	0.026	*0.0117**	**0.0102**		0.0172	0.0344
$\max \mathbf{w}^{(k)}$	0.1798	0.1742	*0.0548**	**0.0429**		0.0766	0.2282
$\min \mathbf{w}^{(k)}$	-0.1213	-0.1292	*-0.0406**	**-0.0391**		-0.0679	-0.1738
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	-0.023	-0.0243	*-0.0093**	**-0.0086**		-0.0154	-0.0325
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.479	0.454	*0.419**	**0.369**		0.436	0.47
p=400							
σ^k	0.01547	0.01517	**0.01117**	0.0113	*0.01119**	0.01176	0.02124
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	*0.000779**	0.000697	0.000683	0.000627	0.000507	0.000661	**0.000879**
SR^k	0.05	0.046	**0.061**	0.055	0.045	*0.056**	0.041
Turnover	2506.41	2505.07	**283.58**	*325.38**		639.2	4605.91
$|\mathbf{w}^{(k)}|$	0.024	0.0235	**0.0082**	*0.0084**		0.0127	0.0432

$\max \mathbf{w}^{(k)}$	0.146	0.1387	*0.0409**	**0.0383**		0.0571	0.2526
$\min \mathbf{w}^{(k)}$	−0.1129	−0.118	**−0.0273**	*−0.0317**		−0.049	−0.2176
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	−0.0219	−0.0225	**−0.0069**	*−0.0073**		−0.0115	−0.0418
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.49	0.465	*0.415**	**0.403**		0.443	0.486
p=431							
σ^k	0.01649	0.01626	**0.01049**	*0.01062**	0.01116	0.01099	0.02591
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	*0.000677**	0.000627	0.000534	0.000537	0.000512	0.000573	**0.000957**
SR^k	0.041	0.039	*0.051**	*0.051**	0.046	**0.052**	0.037
Turnover	3202.26	3217.62	**292.92**	*327.96**		764.9	7195.79
$|\mathbf{w}^{(k)}|$	0.0228	0.0222	**0.0077**	*0.0078**		0.0113	0.0499
$\max \mathbf{w}^{(k)}$	0.1402	0.1306	*0.0393**	**0.036**		0.0515	0.2905
$\min \mathbf{w}^{(k)}$	−0.1076	−0.1131	**−0.0263**	*−0.0305**		−0.0443	−0.2588
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	−0.0206	−0.0213	**−0.0065**	*−0.0068**		−0.0102	−0.0483
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.498	0.466	*0.416**	**0.401**		0.439	0.492

* Second to best

4.3.1 Tracking the S&P500 index

The last setting in Table 1 and 2, where $p = 431$ includes almost all of the stocks in the S&P500 index. However, these portfolios are based on the assets in the S&P 500 index today. Since we sample different assets from the index we may have a survival bias in the experiment above. That may have a positive effect on the return. A possibly more honest method is to choose the assets that were present in the index back in 2013. However, the index evolves and includes more assets today than it did before. This puts the application in another setting entirely.

The portfolio size will need to change over time. We will therefore switch setting and track the S&P500 index, trying to target the stocks that are available in the index. We choose the assets that are part of the index at the time and based on their availability (data quality). This excludes the survivorship bias in our result. Since we target the index we choose to reallocate whenever we register a change in the market capitalization. With p changing over time, we choose two different window sizes. These are equal to $n = 240, 720$. The market capitalization together with daily log returns we have at our disposal covers 406 assets in 2013 and 447 assets in late 2021. Our aim with this experiment is to see if we can improve the index volatility while still taking reasonable positions. We will therefore exclude all portfolios that do not use a target portfolio. We will only consider the BPS and Double estimators with the index-based target, and the index itself. The FM approach is excluded since $c > 1$ for one of the scenarios.

Table 3: Out-of-sample results based on the moving window approach for the BPS and Double estimators with the index-based target, and for the index portfolio. The out-of-sample period equals to 1695 days. The portfolio size starts at 406 and is at most 447.

name	BPS	Double	index
n=240			
σ^k	**0.00834**	*0.00913**	0.01066
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	**0.000622**	*0.0006**	0.000456
SR^k	**0.075**	*0.066**	0.043
Turnover	2555.37	*185.93**	**7.04**
$|\mathbf{w}^{(k)}|$	0.0158	*0.0041**	**0.0024**
$\max \mathbf{w}^{(k)}$	0.0642	**0.0387**	*0.0506**
$\min \mathbf{w}^{(k)}$	−0.0572	*−0.0086**	**0**
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	−0.0148	*−0.0056**	**0**
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.453	*0.158**	**0**
n=720			
σ^k	**0.00839**	*0.00866**	0.01066
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	0.000231	*0.000403**	**0.000456**
SR^k	0.028	**0.047**	*0.043**
Turnover	1094.01	*143.85**	**7.04**
$|\mathbf{w}^{(k)}|$	0.0204	*0.0048**	**0.0024**
$\max \mathbf{w}^{(k)}$	0.1319	**0.039**	*0.0506**
$\min \mathbf{w}^{(k)}$	−0.106	*−0.0131**	**0**
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	−0.0193	*−0.0072**	**0**
$\mathbb{1}(\mathbf{w}_i^{(k)}<0)$	0.465	*0.165**	**0**

In Table 3 we can see the results. When $n = 240$, the portfolio will smallest volatility is BPS. It also provides the largest return. The double shrinkage approach is is the second to best while the index is the worst in both return and volatility. The same ordering holds for the Sharpe Ratio. Although the BPS provides the highest return, it does so at an extreme cost in comparison to the index and Double. If the investor is sensitive to large transitions then the Double method is a great middle-ground. It provides a very large decrease in the turnover as well as an decrease in the volatility, increase in mean and therefore an increase in the SR. The later weight characteristics displays the same behaviour as previously documented. The index has the smallest turnover. This can be explained by the fact that market cap is fairly stable and changes slowly over time. Thereafter its the Double being second to best with the exception of the largest long position. Double decreases volatility and seem to do so by taking smaller long positions, on average. It also introduces short positions, though relatively small ones. In comparison to the previous section, the Double with index as a target takes much smaller proportions of short positions.

When $n = 720$, the same ordering holds for volatility as when $n = 240$. The same does not hold for the mean. Now the index provides the best return with Double being 13% worse. However, being the second to best in both mean and volatility makes the Double have highest SR. Increasing the the window size improves the stability in the portfolio weights and therefore a decrease in the turnover.

5 Summary

In this paper we provide a novel method for investing in the GMV portfolio and a target portfolio. It uses a double shrinkage approach where the sample covariance matrix is shrunk with Thikonov regularization together with linear shrinkage of the GMV portfolio weights to a target portfolio. We construct a bona fide loss function which estimates the true loss function consistently. From that we estimate the two shrinkage coefficients given in the framework. The method is shown to be a great improvement over BPS and performs the same as LW2020 in an extensive simulation study. Furthermore, in the empirical application the method is shown to be a dominating investment strategy in majority of cases justified by different empirical performance measures. We also show that it can act as a good portfolio to track an index. In this scenario it decreases the volatility but can still provide large Sharpe ratios. Our method is opinionated. That is, it demands the investors opinion on what a target portfolio is. That in turn implies that it will work best when the target portfolio is informative, in the sense of the investors aim. However, as our investigation shows, the investor can also use non-informative target portfolios and still achieve great results.

6 Appendix

For any integer $n > 2$, we define

$$\mathbf{V}_n = \frac{1}{n}\mathbf{X}_n\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\mathbf{X}_n^\top \quad \text{and} \quad \widetilde{\mathbf{V}}_n = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^\top, \tag{6.1}$$

where $\mathbf{X}_n$ is given in (2.1). Hence,

$$\mathbf{S}_n = \boldsymbol{\Sigma}^{1/2}\mathbf{V}_n\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}^{1/2}\widetilde{\mathbf{V}}_n\boldsymbol{\Sigma}^{1/2} - \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2} \tag{6.2}$$

with $\bar{\mathbf{x}}_n = \frac{1}{n}\mathbf{X}_n\mathbf{1}_n$.

First, we present an important lemma which is a special case of Theorem 1 in Rubio and Mestre (2011). Moreover, the following result (see, e.g., Theorem 1 on page 176 in Ahlfors (1953)) will be used in a sequel together with Lemma 6.2 in the proofs of the technical lemmas.

Theorem 6.1 (Weierstrass). *Suppose that $f_n(z)$ is analytic in the region Ω_n, and that the sequence $\{f_n(z)\}$ converges to a limit function $f(z)$ in a region Ω, uniformly on every compact subset of Ω. Then $f(z)$ is analytic in Ω. Moreover, $f'(z)$ converges uniformly to $f'(z)$ on every compact subset of Ω.*

We will need two interchange the limits and derivatives many times that is why Theorem 6.1 plays a vital role here. More on the application of Weierstrass theorem can be found in the appendix of Bodnar et al. (2022).

Lemma 6.2. *Let a nonrandom $p \times p$-dimensional matrix $\boldsymbol{\Theta}_p$ possess a uniformly bounded trace norm. Then it holds that*

(i)

$$\left| tr\left(\boldsymbol{\Theta}_p \left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\boldsymbol{\Sigma}^{-1} - z\mathbf{I}_p \right)^{-1} \right) - tr\left(\boldsymbol{\Theta}_p(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1} \right) \right| \xrightarrow{a.s.} 0 \quad (6.3)$$

for $p/n \longrightarrow c \in (0, +\infty)$ as $n \to \infty$ where $v(z)$ solves the following equality

$$v(\eta, z) = \frac{1}{1 + c\frac{1}{p}tr((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1})}. \quad (6.4)$$

(ii)

$$\left| tr\left(\boldsymbol{\Theta}_p \left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\boldsymbol{\Sigma}^{-1} - z\mathbf{I}_p \right)^{-1} \boldsymbol{\Sigma}^{-1} \left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\boldsymbol{\Sigma}^{-1} - z\mathbf{I}_p \right)^{-1} \right) \right.$$
$$- \quad tr\left(\boldsymbol{\Theta}_p(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1}\boldsymbol{\Sigma}^{-1}(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1} \right)$$
$$- \quad \left. v_1'(\eta, z)tr\left(\boldsymbol{\Theta}_p(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-2} \right) \right| \xrightarrow{a.s.} 0 \quad (6.5)$$

for $p/n \longrightarrow c \in (0, +\infty)$ as $n \to \infty$ with

$$v_1'(\eta, z) = \frac{-\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1}\boldsymbol{\Sigma}^{-1}(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1} \right)}{\frac{1}{p}tr((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-2}) - c^{-1}v(\eta,z)^{-2}}. \quad (6.6)$$

(iii)

$$\left| tr\left(\boldsymbol{\Theta}_p \left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\boldsymbol{\Sigma}^{-1} - z\mathbf{I}_p \right)^{-2} \right) \right.$$
$$- \quad \left. (1 - v_2'(\eta, z))tr\left(\boldsymbol{\Theta}_p(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-2} \right) \right| \xrightarrow{a.s.} 0 \quad (6.7)$$

for $p/n \longrightarrow c \in (0, +\infty)$ as $n \to \infty$ with

$$v_2'(\eta, z) = \frac{\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-2} \right)}{\frac{1}{p}tr((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-2}) - c^{-1}v(\eta,z)^{-2}}. \quad (6.8)$$

Proof of Lemma 6.2: (i) The application of Theorem 1 in Rubio and Mestre (2011) leads to (6.3) where $v(\eta, z)$ is a unique solution in $\mathbb{C}^+$ of the following equation

$$\frac{1}{v(\eta, z)} - 1 = \frac{c}{p}\mathrm{tr}\left((\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-1}\right). \tag{6.9}$$

(ii) For the second result of the lemma we get that

$$\mathrm{tr}\left(\mathbf{\Theta}_p\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-1}\mathbf{\Sigma}^{-1}\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-1}\right)$$
$$= -\frac{\partial}{\partial\eta}\mathrm{tr}\left(\mathbf{\Theta}_p\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-1}\right),$$

which almost surely converges to

$$-\frac{\partial}{\partial\eta}\mathrm{tr}\left(\mathbf{\Theta}_p(\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-1}\right)$$
$$= \mathrm{tr}\left(\mathbf{\Theta}_p(\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-1}(\mathbf{\Sigma}^{-1} + v_1'(\eta, z)\mathbf{I})(\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-1}\right)$$

following Theorem 6.1. The first-order partial derivative $v_1'(\eta, z)$ is obtained from (6.9) as

$$-\frac{v_1'(\eta, z)}{v(\eta, z)^2} = -\frac{c}{p}\mathrm{tr}\left((\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-1}\mathbf{\Sigma}^{-1}(\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-1}\right)$$
$$- v_1'(\eta, z)\frac{c}{p}\mathrm{tr}\left((\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-2}\right),$$

from which (6.6) is deduced.

(iii) For the third assertion of the lemma we note that

$$\mathrm{tr}\left(\mathbf{\Theta}_p\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-2}\right) = \frac{\partial}{\partial z}\mathrm{tr}\left(\mathbf{\Theta}_p\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-1}\right),$$

which almost surely tends to

$$\frac{\partial}{\partial z}\mathrm{tr}\left(\mathbf{\Theta}_p(\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-1}\right) = (1 - v_2'(\eta, z))\mathrm{tr}\left(\mathbf{\Theta}_p(\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-2}\right)$$

following Theorem 6.1. Moreover, $v_2'(\eta, z)$ is computed from (6.9) and it is obtained from the following equation

$$-\frac{v_2'(\eta, z)}{v(\eta, z)^2} = (1 - v_2'(\eta, z))\frac{c}{p}\mathrm{tr}\left((\eta\mathbf{\Sigma}^{-1} + (v(\eta, z) - z)\mathbf{I})^{-2}\right).$$

This completes the proof of the lemma.

$\square$

Lemma 6.3. *Let $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ be universal nonrandom vectors with bounded Euclidean norms. Then it holds that*

$$\left|\boldsymbol{\xi}'\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1}\right)^{-1}\boldsymbol{\theta} - \boldsymbol{\xi}'(\eta\mathbf{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{6.10}$$

$$\left|\boldsymbol{\xi}'\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1}\right)^{-1}\mathbf{\Sigma}^{-1}\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1}\right)^{-1}\boldsymbol{\theta}\right. \tag{6.11}$$

$$\left.-\boldsymbol{\xi}'(\eta\mathbf{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-1}\mathbf{\Sigma}^{-1}(\eta\mathbf{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-1}\boldsymbol{\theta} - v_1'(\eta, 0)\boldsymbol{\xi}'(\eta\mathbf{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-2}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0$$

$$\left|\boldsymbol{\xi}'\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1}\right)^{-2}\boldsymbol{\theta} - (1 - v_2'(\eta, 0))\boldsymbol{\xi}'(\eta\mathbf{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-2}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0 \tag{6.12}$$

for $p/n \longrightarrow c \in (0, \infty)$ as $n \to \infty$ where $v(\eta, 0)$ is the solution of

$$v(\eta, 0) = 1 - c\left(1 - \frac{\eta}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right)\right), \tag{6.13}$$

and $v_1'(\eta, 0)$ and $v_2'(\eta, 0)$ are computed by

$$v_1'(\eta, 0) = v(\eta, 0)\frac{c\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right) - c\eta\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-2}\right)}{1 - c + 2c\eta\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right) - c\eta^2\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-2}\right)} \tag{6.14}$$

and

$$v_2'(\eta, 0) = 1 - \frac{1}{v(\eta, 0)} + \eta\frac{v_1'(\eta, 0)}{v(\eta, 0)^2}. \tag{6.15}$$

Proof of Lemma 6.3: Since the trace norm of $\boldsymbol{\theta\xi}'$ is uniformly bounded, i.e.,

$$||\boldsymbol{\theta\xi}'||_{tr} \leq \sqrt{\boldsymbol{\theta}'\boldsymbol{\theta}}\sqrt{\boldsymbol{\xi}'\boldsymbol{\xi}} < \infty,$$

the application of Lemma 6.2 leads to (6.10), (6.11), and (6.12) where $v(\eta, 0)$ satisfies the following equality

$$\frac{1}{v(\eta, 0)} - 1 = \frac{c}{p}tr\left(\left(\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I}\right)^{-1}\right) = \frac{c}{v(\eta, 0)}\left(1 - \frac{\eta}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right)\right),$$

which results in (6.13).

The application of (6.6) leads to

$$
\begin{aligned}
v_1'(\eta, 0) &= \frac{-\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-1}\boldsymbol{\Sigma}^{-1}(\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-1}\right)}{\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-2}\right) - c^{-1}v(\eta, 0)^{-2}} \\
&= v(\eta, 0)\frac{c\frac{1}{p}tr\left(v(\eta, 0)(\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-1}\boldsymbol{\Sigma}^{-1}(\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-1}\right)}{1 - c\frac{1}{p}tr\left(v(\eta, 0)^2(\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-2}\right)} \\
&= v(\eta, 0)\frac{c\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right) - c\eta\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-2}\right)}{1 - c + 2c\eta\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right) - c\eta^2\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-2}\right)}.
\end{aligned}
$$

Finally, using (6.8), we get

$$
\begin{aligned}
v_2'(\eta, 0) &= \frac{\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-2}\right)}{\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-2}\right) - c^{-1}v(\eta, 0)^{-2}} \\
&= 1 - \frac{1}{1 - c\frac{1}{p}tr\left(v(\eta, 0)^2(\eta\boldsymbol{\Sigma}^{-1} + v(\eta, 0)\mathbf{I})^{-2}\right)} \\
&= 1 - \frac{1}{1 - c + 2c\eta\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right) - c\eta^2\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-2}\right)} \\
&= 1 - \frac{1}{v(\eta, 0) + \eta\left(c\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-1}\right) - c\eta\frac{1}{p}tr\left((v(\eta, 0)\boldsymbol{\Sigma} + \eta\mathbf{I})^{-2}\right)\right)} \\
&= 1 - \frac{1}{v(\eta, 0) + \eta\frac{v(\eta, 0)v_1'(\eta, 0)}{v(\eta, 0) - \eta v_1'(\eta, 0)}} = 1 - \frac{1}{v(\eta, 0)} + \eta\frac{v_1'(\eta, 0)}{v(\eta, 0)^2}.
\end{aligned}
$$

$\square$

Lemma 6.4. *Let $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ be universal nonrandom vectors such that $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\xi}$ have bounded Euclidean norms. Then it holds that*

$$\left|\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-1}\boldsymbol{\theta} - \boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{6.16}$$

$$\left|\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-2}\boldsymbol{\theta} - \boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-2}\boldsymbol{\theta} - v_1'(\eta,0)\boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{6.17}$$

$$\left|\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-1}\boldsymbol{\Sigma}\mathbf{S}_\lambda^{-1}\boldsymbol{\theta} - (1 - v_2'(\eta,0))\boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0 \tag{6.18}$$

for $p/n \longrightarrow c \in (0,\infty)$ as $n \to \infty$ with $\eta = 1/\lambda - 1$,

$$\boldsymbol{\Omega}_\lambda = v(\eta,0)\lambda\boldsymbol{\Sigma} + (1-\lambda)\mathbf{I},$$

and $v(\eta,0)$, $v_1'(\eta,0)$ and $v_2'(\eta,0)$ given in Lemma 6.3.

Proof of Lemma 6.4: Let $\widetilde{\mathbf{S}}_n = \boldsymbol{\Sigma}^{1/2}\widetilde{\mathbf{V}}_n\boldsymbol{\Sigma}^{1/2}$. Using (6.1) and (6.2) and the formula for the 1-rank update of inverse matrix (see, e.g., Horn and Johnsohn (1985)), we get

$$\lambda\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-1}\boldsymbol{\theta} = \boldsymbol{\xi}^\top \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I} - \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\right)^{-1}\boldsymbol{\theta}$$

$$= \boldsymbol{\xi}^\top\boldsymbol{\Sigma}^{-1/2}\left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1\right)\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$$

$$+ \frac{\boldsymbol{\xi}^\top\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\theta}}{1 - \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n},$$

where

$$\left|\boldsymbol{\xi}^\top\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\right| \xrightarrow{a.s.} 0 \tag{6.19}$$

for $\lambda \in (0,1]$ by Pan (2014, p. 673). Furthermore, the quantity

$$\frac{1}{1 - \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n} \tag{6.20}$$

is bounded following Pan (2014, Eq. (2.28)). Hence, the application of Lemma 6.3 leads to the first statement of Lemma 6.4.

We compute

$$\lambda^2\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-2}\boldsymbol{\theta} = \boldsymbol{\xi}^\top\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I} - \bar{\boldsymbol{\Sigma}}^{1/2}\mathbf{x}_n\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\right)^{-2}\boldsymbol{\theta}$$

$$= \boldsymbol{\xi}^\top\boldsymbol{\Sigma}^{-1/2}\left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1\right)\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1}\left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1\right)\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$$

$$+ \frac{\boldsymbol{\xi}^\top\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-2}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\theta}}{1 - \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n}$$

$$+ \frac{\boldsymbol{\xi}^\top\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-2}\boldsymbol{\theta}}{1 - \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n}$$

$$+ \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-2}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n$$

$$\times \frac{\boldsymbol{\xi}^\top\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\theta}}{\left(1 - \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\right)^2},$$

where

$$\bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-2} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n$$

$$\leq \left(\frac{1}{\lambda} - 1 \right)^{-1} \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-1} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n < \infty$$

and

$$\boldsymbol{\xi}^\top \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-2} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n$$

$$\leq \sqrt{ \boldsymbol{\xi}^\top \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-2} \boldsymbol{\xi} } \sqrt{ \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-2} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n } < \infty$$

For the third statement of the lemma we consider

$$\lambda^2 \boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-1} \mathbf{\Sigma} \mathbf{S}_\lambda^{-1} \boldsymbol{\theta} = \boldsymbol{\xi}^\top \mathbf{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} - \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^\top \right)^{-2} \mathbf{\Sigma}^{-1/2} \boldsymbol{\theta}$$

$$= \boldsymbol{\xi}^\top \mathbf{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \mathbf{\Sigma}^{-1/2} \boldsymbol{\theta}$$

$$+ \frac{ \boldsymbol{\xi}^\top \mathbf{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-1} \mathbf{\Sigma}^{-1/2} \boldsymbol{\theta} }{ 1 - \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-1} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n }$$

$$+ \frac{ \boldsymbol{\xi}^\top \mathbf{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \mathbf{\Sigma}^{-1/2} \boldsymbol{\theta} }{ 1 - \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-1} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n }$$

$$+ \bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \bar{\mathbf{x}}_n$$

$$\times \frac{ \boldsymbol{\xi}^\top \mathbf{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-1} \bar{\mathbf{x}}_n \bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-1} \mathbf{\Sigma}^{-1/2} \boldsymbol{\theta} }{ \left(1 - \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-1} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n \right)^2 }.$$

Next, we prove that $\bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \bar{\mathbf{x}}_n$ is bounded for $p/n \longrightarrow c \in (0, \infty)$ as $n \to \infty$.

$$\bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \bar{\mathbf{x}}_n$$

$$= \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\tilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-1} \mathbf{\Sigma} \left(\tilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-1} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n$$

$$\leq \lambda_{\max}(\mathbf{\Sigma}) \cdot \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\tilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-2} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n$$

$$\leq \lambda_{\max}(\mathbf{\Sigma}) \left(\frac{1}{\lambda} - 1 \right)^{-1} \bar{\mathbf{x}}_n^\top \mathbf{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{I} \right)^{-1} \mathbf{\Sigma}^{1/2} \bar{\mathbf{x}}_n < \infty \qquad (6.21)$$

Using (6.21) we get that

$$\left| \boldsymbol{\xi}^\top \mathbf{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \bar{\mathbf{x}}_n \right|$$

$$\leq \sqrt{ \boldsymbol{\xi}^\top \mathbf{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \mathbf{\Sigma}^{-1/2} \boldsymbol{\xi} } \sqrt{ \bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1 \right) \mathbf{\Sigma}^{-1} \right)^{-2} \bar{\mathbf{x}}_n } < \infty .$$

Hence, the application of (6.19), (6.20), and Lemma 6.3 completes the proof of the lemma. $\square$

Proof of Theorem 2.1: Let $V_{GMV} = 1/(\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1})$. The application of the results of Lemma 6.4 with $\mathbf{\xi} = \mathbf{\Sigma}\mathbf{b}/\sqrt{\mathbf{b}^\top\mathbf{\Sigma}\mathbf{b}}$ and $\mathbf{\theta} = \mathbf{1}/\sqrt{\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}}$ leads to

$$\left| (L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1} - (L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\mathbf{\Sigma}\mathbf{\Omega}_\lambda^{-1}\mathbf{1} \right| \xrightarrow{a.s.} 0, \tag{6.22}$$

$$\left| V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1} - \lambda^{-1}V_{GMV}\mathbf{1}^\top\mathbf{\Omega}_\lambda^{-1}\mathbf{1} \right| \xrightarrow{a.s.} 0, \tag{6.23}$$

$$\left| V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1} - V_{GMV}(1 - v_2'(\eta,0))\mathbf{1}'\mathbf{\Omega}_\lambda^{-1}\mathbf{\Sigma}\mathbf{\Omega}_\lambda^{-1}\mathbf{1} \right| \xrightarrow{a.s.} 0 \tag{6.24}$$

Using (6.22)-(6.24) and the equality

$$L_{n;2}(\lambda) = \frac{\left(1 - \frac{1}{\sqrt{L_\mathbf{b}+1}}\frac{(L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1}}{V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}\right)^2}{1 - \frac{2}{\sqrt{L_\mathbf{b}+1}}\frac{(L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1}}{V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}} + \frac{1}{L_\mathbf{b}+1}\frac{V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1}}{\left(V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}\right)^2}}$$

we get the statement of part (i) of the theorem, while the application of the equality

$$\psi_n^*(\lambda) = \frac{1 - \frac{1}{\sqrt{L_\mathbf{b}+1}}\frac{(L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1}}{V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}}{1 - \frac{2}{\sqrt{L_\mathbf{b}+1}}\frac{(L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1}}{V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}} + \frac{1}{L_\mathbf{b}+1}\frac{V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1}}{\left(V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}\right)^2}}$$

yields the second statement of the theorem. □

Lemma 6.5. *Let* $\frac{1}{p}\mathbf{\Sigma}^{-1}$ *possess a bounded trace norm. Then it holds that*

$$\left| \frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right) - \frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right) \right| \xrightarrow{a.s} 0 \tag{6.25}$$

$$\left| \frac{\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-2}\right)\left(1 - c + 2c\eta\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)\right) - c\left[\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)\right]^2}{1 - c + c\eta^2\frac{1}{p}tr((\mathbf{S}_n+\eta\mathbf{I})^{-2})} \right.$$

$$\left. - \frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right) \right| \xrightarrow{a.s} 0 \tag{6.26}$$

for $p/n \to c \in (0,\infty)$ *as* $n \to \infty$.

Proof of Lemma 6.5: From part (i) of Lemma 6.2 with $\mathbf{\Theta}_p = \frac{1}{p}\mathbf{\Sigma}^{-1}$ and the proof of Lemma 6.4 we obtain that $\frac{1}{p}\text{tr}((\mathbf{S}_n+\eta\mathbf{I})^{-1})$ is consistent for $\frac{1}{p}\text{tr}\left((\eta\mathbf{I}+v(\eta,0)\mathbf{\Sigma})^{-1}\right)$ in the high-dimensional setting.

Furthermore, applying part (ii) of Lemma 6.2 with $\mathbf{\Theta}_p = \frac{1}{p}\mathbf{\Sigma}^{-1}$ and following the proof of Lemma 6.4 we get that

$$\frac{1}{p}\text{tr}\left((\mathbf{S}_n+\eta\mathbf{I})^{-2}\right) \xrightarrow{a.s.} \frac{1}{p}\text{tr}\left(\mathbf{\Sigma}^{-1}\left(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I}\right)^{-1}\mathbf{\Sigma}^{-1}\left(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I}\right)^{-1}\right)$$

$$+ v_1'(\eta,0)\frac{1}{p}\text{tr}\left(\mathbf{\Sigma}^{-1}\left(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I}\right)^{-2}\right)$$

$$= \frac{1}{p}\text{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right) + v_1'(\eta,0)\frac{1}{p}\text{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\mathbf{\Sigma}(v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right)$$

$$= \frac{v_1'(\eta,0)}{v(\eta,0)}\frac{1}{p}\text{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right) + \left(1 - \frac{v_1'(\eta,0)}{v(\eta,0)}\eta\right)\frac{1}{p}\text{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right), \tag{6.27}$$

24

where the application of (6.14) leads to

$$\frac{v_1'(\eta,0)}{v(\eta,0)} = \frac{c\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right) - c\eta\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)}{1 - c + 2c\eta\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right) - c\eta^2\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)}.$$

Thus, $\frac{1}{p}\mathrm{tr}\left((\mathbf{S}_n+\eta\mathbf{I})^{-2}\right)$ converges almost surely to

$$\frac{c\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right) - c\eta\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)}{1 - c + 2c\eta\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right) - c\eta^2\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)}$$
$$\times\quad \frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right)$$
$$+\quad \frac{1 - c + c\eta\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right)}{1 - c + 2c\eta\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right) - c\eta^2\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)}$$
$$\times\quad \frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)$$
$$=\quad \frac{c\left[\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right)\right]^2 + (1-c)\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)}{1 - c + 2c\eta\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-1}\right) - c\eta^2\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\boldsymbol{\Sigma}+\eta\mathbf{I})^{-2}\right)},$$

which together with (6.25) leads to the second statement of the lemma. $\square$

Proof of Theorem 3.1: The result (3.1) is a direct consequence of (6.13) in Lemma 6.3 and (6.25) in Lemma 6.5.

Let $t_1 = \frac{1}{p}\mathrm{tr}\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)$ and $t_2 = \frac{1}{p}\mathrm{tr}\left((\mathbf{S}_n+\eta\mathbf{I})^{-2}\right)$. Then, the application of (6.14) in Lemma 6.3 and the results of Lemma 6.5 leads to a consistent estimator of $v_1'(\eta,0)$ expressed as

$$\begin{aligned}
\hat{v}_1'(\eta,0) &= \hat{v}(\eta,0)\frac{ct_1 - c\eta\frac{t_2(1-c+2c\eta t_1)-ct_1^2}{1-c+c\eta^2 t_2}}{1 - c + 2c\eta t_1 - c\eta^2\frac{t_2(1-c+2c\eta t_1)-ct_1^2}{1-c+c\eta^2 t_2}}\\
&= \hat{v}(\eta,0)\frac{(1-c)c(t_1-\eta t_2) + c^2\eta t_1(t_1-\eta t_2)}{(1-c)^2 + 2(1-c)c\eta t_1 + c^2\eta^2 t_1^2} = \hat{v}(\eta,0)c\frac{(1-c+c\eta t_1)(t_1-\eta t_2)}{(1-c+c\eta t_1)^2}\\
&= \hat{v}(\eta,0)c\frac{t_1-\eta t_2}{(1-c+c\eta t_1)} = \hat{v}(\eta,0)c\frac{t_1-\eta t_2}{\hat{v}(\eta,0)}.
\end{aligned}$$

Finally, the result (3.3) follows from (3.1) and (3.2) together with (6.15) in Lemma 6.3. $\square$

Proof of Theorem 3.2: The result (3.4) is a special case of Theorem 3.2 in Bodnar et al. (2014). Equation (3.5) follows from Lemma 6.4 with $\boldsymbol{\xi} = \boldsymbol{\theta} = 1/\sqrt{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$. For the derivation of (3.6) we note that

$$\mathbf{b}^\top\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1} = \frac{1}{\lambda v(\eta,0)}\left(1 - (1-\lambda)\mathbf{b}^\top\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1}\right) \tag{6.28}$$

and apply Lemma 6.4 with $\boldsymbol{\xi} = \mathbf{b}/\sqrt{\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}}$ and $\boldsymbol{\theta} = 1/\sqrt{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$.

Finally, (3.7) is obtained by noting that

$$\mathbf{1}^\top\boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1} = \frac{1}{\lambda v(\eta,0)}\mathbf{1}^\top\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1} - \frac{1-\lambda}{\lambda v(\eta,0)}\mathbf{1}^\top\boldsymbol{\Omega}_\lambda^{-2}\mathbf{1},$$

where (3.5) is used for the first summand and (6.17) of Lemma 6.4 with $\boldsymbol{\xi} = \boldsymbol{\theta} = 1/\sqrt{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$ for the second one. $\square$

References

Ahlfors, L. V. (1953). *Complex analysis: An introduction to the theory of analytic functions of one complex variable.* International series in pure and applied mathematics. McGraw-Hill, New York.

Bai, Z. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*, volume 20. Springer.

Best, M. J. and Grauer, R. R. (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(2):315–342.

Bodnar, T., Dmytriv, S., Okhrin, Y., Otryakhin, D., and Parolya, N. (2021a). *HDShOP: high-dimensional shrinkage optimal portfolios.* R package version 0.1.2.

Bodnar, T., Gupta, A. K., and Parolya, N. (2014). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *Journal of Multivariate Analysis*, 132:215–228.

Bodnar, T., Okhrin, Y., and Parolya, N. (2022). Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, pages 1–17, to appear.

Bodnar, T., Parolya, N., and Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research*, 266(1):371–390.

Bodnar, T., Parolya, N., and Thorsen, E. (2021b). Dynamic shrinkage estimation of the high-dimensional minimum-variance portfolio. *arXiv preprint arXiv:2106.02131*.

Boileau, P., Hejazi, N. S., Collica, B., van der Laan, M. J., and Dudoit, S. (2021). 'cvcovest': Cross-validated covariance matrix estimator selection and evaluation in 'R'. *Journal of Open Source Software*, 6(63):3273.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72(3):498–505.

DeMiguel, V., Martin-Utrera, A., and Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034.

Frahm, G. and Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, 159:289–302.

Golosnoy, V. and Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *The European Journal of Finance*, 13(5):441–458.

Golub, G. H., Hansen, P. C., and O'Leary, D. P. (1999). Tikhonov regularization and total least squares. *SIAM journal on matrix analysis and applications*, 21(1):185–194.

Horn, R. A. and Johnsohn, C. R. (1985). *Matrix Analysis.* Cambridge University Press, Cambridge.

Kan, R., Wang, X., and Zhou, G. (2022). Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science*, 68(3):2047–2068.

Lai, Z.-R., Tan, L., Wu, X., and Fang, L. (2020). Loss control with rank-one covariance estimate for short-term portfolio optimization. *Journal of Machine Learning Research*, 21(97):1–37.

Lai, Z.-R., Yang, P.-Y., Fang, L., and Wu, X. (2018). Short-term sparse portfolio optimization based on alternating direction method of multipliers. *Journal of Machine Learning Research*, 19(63):1–28.

Lassance, N. (2021). Maximizing the out-of-sample sharpe ratio. *Available at SSRN 3959708*.

Lassance, N., Martín-Utrera, A., and Simaan, M. (2022). The risk of out-of-sample portfolio performance.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2017). Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.

Ledoit, O. and Wolf, M. (2020a). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48(5):3043–3065.

Ledoit, O. and Wolf, M. (2020b). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics*, 20(1):187–218.

Li, B., Sahoo, D., and Hoi, S. C. (2016). Olps: A toolbox for on-line portfolio selection. *Journal of Machine Learning Research*, 17(35):1–5.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1).

Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. John Wiley.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Pan, G. (2014). Comparison between two types of large sample covariance matrices. *Annales de l'IHP Probabilités et statistiques*, 50(2):655–677.

Rubio, F. and Mestre, X. (2011). Spectral convergence for a general class of random matrices. *Statistics & Probability Letters*, 81(5):592–602.

Tikhonov, A. N., Goncharsky, A., Stepanov, V., and Yagola, A. G. (1995). *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media.

Tong, J., Hu, R., Xi, J., Xiao, Z., Guo, Q., and Yu, Y. (2018). Linear shrinkage estimation of covariance matrices using low-complexity cross-validation. *Signal Processing*, 148:223–233.

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