Journal of Machine Learning Research 25 (2024) 1-32 Submitted 11/22; Revised 6/24; Published 6/24

Two is Better Than One: Regularized Shrinkage of Large Minimum Variance Portfolios

Taras Bodnar TARAS.BODNAR@LIU.SE
Department of Management and Engineering
Linköping University
SE-581 83 Linköping Sweden

Nestor Parolya N.PAROLYA@TUDELFT.NL
Department of Applied Mathematics
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands

Erik Thorsén ERIK.THORSEN@MATH.SU.SE
Department of Mathematics
Stockholm University
Roslagsvägen 101, SE-10691 Stockholm, Sweden

Editor: Jie Peng

Abstract

In this paper, we construct a shrinkage estimator of the global minimum variance (GMV) portfolio by combining 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 for the assumption of finite moments of order $4 + \varepsilon$ for $\varepsilon > 0$. 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 maintains 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

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). 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 as very important due to the fact that the estimation error in the means is 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

$$\min_{\mathbf{w}} \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1, \tag{1}$$

and its weights are given by

$$\mathbf{w}_{GMV} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1}\mathbf{1}}. \tag{2}$$

Since the covariance matrix $\mathbf{\Sigma}$ is an unknown quantity, the GMV portfolio cannot be constructed using (2). Markowitz (1959) suggests the application of the sample estimator of $\mathbf{w}_{GMV}$ instead of (2) given by

$$\hat{\mathbf{w}}_{GMV} = \frac{\mathbf{S}_n^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{S}_n^{-1}\mathbf{1}}, \tag{3}$$

where $\mathbf{S}_n$ is the sample estimator of the covariance matrix $\mathbf{\Sigma}$ expressed as

$$\mathbf{S}_n = \frac{1}{n}\left(\mathbf{Y}_n - \bar{\mathbf{y}}_n \mathbf{1}^\top\right)\left(\mathbf{Y}_n - \bar{\mathbf{y}}_n \mathbf{1}^\top\right)^\top \quad \text{with} \quad \bar{\mathbf{y}}_n = \frac{1}{n}\mathbf{Y}_n \mathbf{1}, \tag{4}$$

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 (3) may be considered as the solution of the optimization problem (1) where the unknown covariance matrix $\mathbf{\Sigma}$ is replaced by $\mathbf{S}_n$, namely,

$$\min_{\mathbf{w}} \mathbf{w}^\top \mathbf{S}_n \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1. \tag{5}$$

There are several other estimators 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 the GMV portfolio, or they shrink the portfolio weights directly to a certain target. To the best of our knowledge, none of the existing approaches combine 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, while the new suggested 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 robust for high-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. (2023b)). This leads us to the second feature. The single source of uncertainty in the GMV portfolio is $\mathbf{S}_n$. If one can stabilize or decrease the variance of the sample covariance matrix, then the variance of the weights may decrease as well. The aim is therefore to shrink the sample covariance matrix in addition to portfolio weights. This leads to a new estimator for the weights of the GMV portfolio, derived from a double shrinkage approach. In the empirical study, we show that the new portfolio outperforms the existing approaches based on both the linear shrinkage estimator of the portfolio weights and the nonlinear shrinkage estimator of the covariance matrix. The application of the Tikhonov regularization (see, e.g., Tikhonov et al. (1995)) to the optimization problem (5) yields

$$\min_{\mathbf{w}} \mathbf{w}^\top \mathbf{S}_n \mathbf{w} + \eta \mathbf{w}^\top \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1, \tag{6}$$

where η is the regularization parameter. Similar approaches are employed in regression analysis, where the ridge regression uses Tikhonov regularization to stabilize the least squares estimator of the coefficients of the regression line (cf., Golub et al. (1999)). The solution of (6) is given by

$$\hat{\mathbf{w}}_{S;\lambda} = \frac{(\mathbf{S}_n + \eta \mathbf{I})^{-1}\mathbf{1}}{\mathbf{1}^\top (\mathbf{S}_n + \eta \mathbf{I})^{-1}\mathbf{1}}. \tag{7}$$

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 (2) results in the same solution. However, the latter presentation possesses a nice interpretation, noting that $\mathbf{S}_\lambda$ is a convex combination of the sample covariance matrix and the identity matrix. If $\lambda \to 0$ then we put all our beliefs in the identity 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), the already regularized GMV portfolio weights are shrunk towards the target portfolio $\mathbf{b}$ with $\mathbf{b}^\top \mathbf{1} = 1$ as follows

$$\hat{\mathbf{w}}_{Sh;\lambda,\psi} = \psi \hat{\mathbf{w}}_{S;\lambda} + (1 - \psi)\mathbf{b}, \tag{8}$$

where ψ is the shrinkage intensity. As such, the approach first reduces the variability in the sample covariance matrix and then further stabilizes the portfolio by shrinking the weights themselves. It also gives a way for an investor to highlight stocks they like with the target portfolio $\mathbf{b}$. In many cases, a naive portfolio $\mathbf{b} = \frac{1}{p}\mathbf{1}$ is a good choice. However, 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. (2024) 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 given by

$$L(\lambda, \psi) = \hat{\mathbf{w}}_{Sh;\lambda,\psi}^\top \mathbf{\Sigma} \hat{\mathbf{w}}_{Sh;\lambda,\psi}. \tag{9}$$

However, since $\mathbf{\Sigma}$ is not known we need to estimate the loss function.

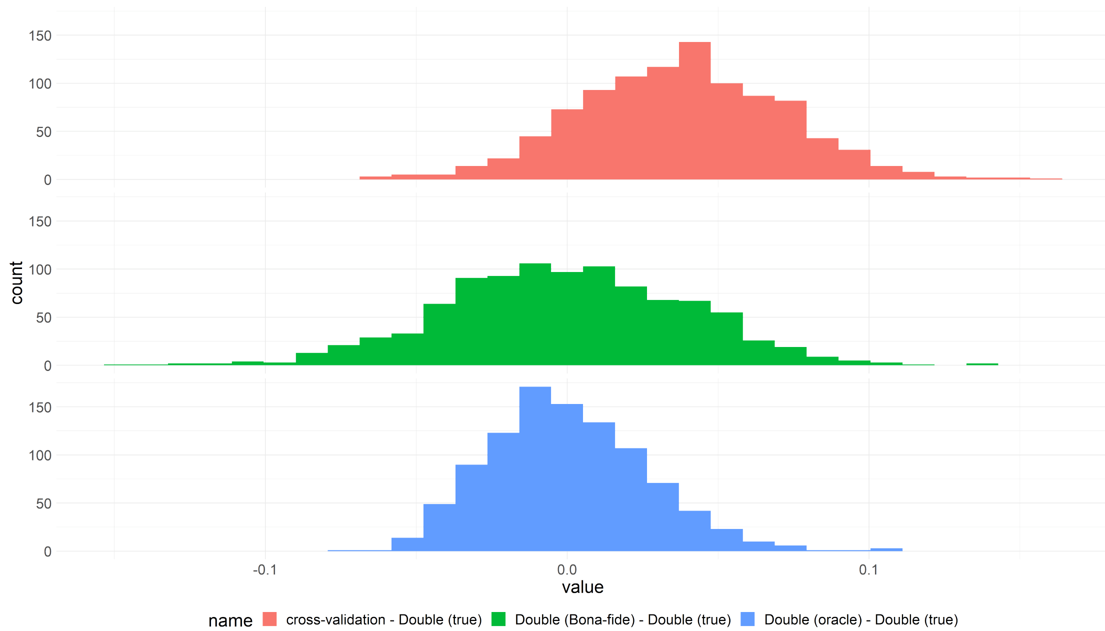


Figure 1: Estimated shrinkage intensities for λ based on the out-of-sample variance loss. The loss function (9) is determined through the cross-validation method and the two double shrinkage approaches. The bona fide loss is completely determined by data, while the oracle loss depends on the unknown quantity $\mathbf{\Sigma}$, and hence it has less variance. Each solution for λ is centered by the solution to the true loss (9).

To approximate the loss function, which depends on the unknown $\mathbf{\Sigma}$, K-fold cross-validation is usually employed with the observation data partitioned into K subsets: one of the subsets is used as the validation set while the other $K-1$ subsets constitute the training set. The data from the training set are utilized to construct a shrinkage estimator of the portfolio weights for given values of λ and ψ, while the empirical out-of-sample variance is computed by using the validation set. This procedure is then repeated K times for each possible choice of the validation set, and the resulting values of the computed loss functions are averaged. Finally, the best pair of shrinkage coefficients is determined by maximizing the average loss using a grid search. However, this approach to compute the tuning parameters introduces several obstacles. It has been established that the sample covariance matrix is a noisy estimator of the true population covariance matrix, and the validation set is usually smaller than the training set, leading to a more volatile estimator of the covariance matrix. 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)). A bona fide type loss function is constructed that consistently estimates the true loss function in the high dimensional setting. The problem reduces to a simple univariate nonlinear optimization exercise, which is easy to solve numerically.

In Figure 1, we illustrate the optimal values of $\lambda - \lambda_{true}$ where each λ is obtained by 10-fold cross-validation and the two double shrinkage approaches. The true shrinkage coefficient λ_{true} is given by the solution of the true loss function (9). In this illustration, there are fewer assets $p = 150$ than data points $n = 300$ 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 $\mathbf{\Sigma}$ does not exhibit large eigenvalues, this is a scenario where a practitioner might assume that cross-validation, in particular, should work well. However, the cross-validation appears to be considerably biased. In contrast, the optimal values of λ obtained from the derived two loss functions, denoted as double (bona fide) and double (oracle), are centered around zero. Furthermore, the values deduced from the double oracle estimator depict smaller variability as expected since the unknown covariance matrix $\mathbf{\Sigma}$ is used in the computation.

Our work contributes to the existing literature by deriving the high-dimensional properties of the true loss function (9) as well as its bona fide counterpart. These results allow practitioners to determine the optimal value of the tuning parameter by solving numerically a simple non-linear optimization problem. Furthermore, since a bona fide estimator for the loss function is derived, one can easily extend the optimization problem to cover position restrictions, moment conditions, or potentially other restrictions on the optimization problem. The other line of our contributions leads to a new approach to constructing high-dimensional optimal portfolios that tackles the estimation uncertainty in two steps: first, by using an improved estimator of the covariance matrix, and then, by improving the resulting estimator of the portfolio weights.

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 then used to find the optimal values of the two shrinkage intensities. The results of an extensive simulation study and empirical applications are provided in Section 4. Section 5 summarizes the obtained findings. The mathematical derivations are moved to the appendix.

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 moments of order $4 + \varepsilon$ for some $\varepsilon > 0$. Assume that we observe the matrix $\mathbf{Y}_n$ according to the stochastic model

$$\mathbf{Y}_n \stackrel{d}{=} \boldsymbol{\mu}\mathbf{1}^\top + \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{X}_n \tag{10}$$

where $\mathbf{\Sigma}$ 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.

1. The obtained results can be generalized to the case with a finite number of unbounded largest eigenvalues, which would make the proofs more lengthy. The assumption of unbounded eigenvalues is needed only in the case of the centered sample covariance matrix, i.e., unknown mean vector. If $\boldsymbol{\mu}$ is known, the boundedness of eigenvalues may be ignored due to normalization presented further in (11). More details can be deduced from the proofs of the main theorems.

The aim is to estimate the shrinkage intensities λ, ψ from the following normalized optimization problem

$$\min_{\lambda,\psi} \frac{\hat{\mathbf{w}}_{Sh;\lambda,\psi}^\top \mathbf{\Sigma} \hat{\mathbf{w}}_{Sh;\lambda,\psi}}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}}. \tag{11}$$

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

$$\begin{aligned} L(\lambda,\psi) &= \left(\psi\hat{\mathbf{w}}_{S;\lambda} + (1-\psi)\mathbf{b}\right)^\top \mathbf{\Sigma} \left(\psi\hat{\mathbf{w}}_{S;\lambda} + (1-\psi)\mathbf{b}\right) \\ &= \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right) \left(\psi - \frac{\mathbf{b}^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)}{\left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)}\right)^2 \\ &\quad - \frac{\left(\mathbf{b}^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)\right)^2}{\left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)} + \mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}. \end{aligned} \tag{12}$$

For given λ, the first summand in the expression of the out-of-sample variance (12) attains its minimum value zero for[2]

$$\psi_n^*(\lambda) = \frac{\mathbf{b}^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)}{\left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)}, \tag{13}$$

while the optimal λ is found by maximizing the normalized second summand in (12), i.e.,

$$L_{n;2}(\lambda) = \frac{1}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}} \frac{\left(\mathbf{b}^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)\right)^2}{\left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)^\top \mathbf{\Sigma} \left(\mathbf{b} - \hat{\mathbf{w}}_{S;\lambda}\right)}. \tag{14}$$

Note that due to the Cauchy-Schwarz inequality the normalization $(\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b})^{-1}$ keeps $L_{n;2}$ bounded between zero and one. To determine the values of λ^* together with $\psi^*(\lambda)$ which minimize the loss function, we proceed in three steps. First, we find the deterministic equivalent to $L_{n;2}(\lambda)$ in Theorem 1 and estimate it consistently in the second step. Finally, the obtained consistent estimator is minimized in the last step.

Theorem 1 *Let $\mathbf{Y}_n$ possess the stochastic representation as in* (10). *Assume that the relative loss of the target portfolio expressed as*

$$L_{\mathbf{b}} = \frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} - \frac{1}{\mathbf{1}^\top \mathbf{\Sigma}^{-1} \mathbf{1}}}{\frac{1}{\mathbf{1}^\top \mathbf{\Sigma}^{-1} \mathbf{1}}} = \mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} \mathbf{1}^\top \mathbf{\Sigma}^{-1} \mathbf{1} - 1 \tag{15}$$

is uniformly bounded in p. Then it holds that

(i)

$$|L_{n;2}(\lambda) - L_2(\lambda)| \xrightarrow{a.s.} 0 \tag{16}$$

for $p/n \to c \in (0, \infty)$ as $n \to \infty$ with[3]

$$L_2(\lambda) = \frac{\left(1 - \frac{1}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}\right)^2}{1 - \frac{2}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{1}} + \frac{1}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}} \frac{(1 - v_2'(\eta, 0)) \mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{\Sigma} \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}{\left(\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{1}\right)^2}} \tag{17}$$

2. The authors are thankful to the Reviewer for paying attention to a typing error in the formula, which was present in the original submission.
3. This convergence is uniform if λ stays in a compact interval away from zero and one.

(ii)

$$|\psi_n^*(\lambda) - \psi^*(\lambda)| \overset{a.s.}{\to} 0 \tag{18}$$

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

$$\psi^*(\lambda) = \frac{1 - \frac{1}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}}{1 - \frac{2}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}} \frac{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{1}} + \frac{1}{\mathbf{b}^\top \mathbf{\Sigma} \mathbf{b}} \frac{(1 - v_2'(\eta, 0)) \mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{\Sigma} \mathbf{\Omega}_\lambda^{-1} \mathbf{1}}{\left(\mathbf{1}^\top \mathbf{\Omega}_\lambda^{-1} \mathbf{1}\right)^2}}, \tag{19}$$

where

$$\eta = \frac{1}{\lambda} - 1, \quad \mathbf{\Omega}_\lambda = v\,(\eta, 0)\,\lambda \mathbf{\Sigma} + (1 - \lambda)\mathbf{I}, \tag{20}$$

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

$$v(\eta, 0) = 1 - c\left(1 - \frac{\eta}{p} tr\left((v(\eta, 0)\mathbf{\Sigma} + \eta \mathbf{I})^{-1}\right)\right), \tag{21}$$

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

$$v_2'(\eta, 0) = 1 - \frac{1}{v(\eta, 0)} + \eta \frac{v_1'(\eta, 0)}{v(\eta, 0)^2}. \tag{22}$$

with

$$v_1'(\eta, 0) = v(\eta, 0) \frac{c\frac{1}{p} tr\left((v(\eta, 0)\mathbf{\Sigma} + \eta \mathbf{I})^{-1}\right) - c\eta \frac{1}{p} tr\left((v(\eta, 0)\mathbf{\Sigma} + \eta \mathbf{I})^{-2}\right)}{1 - c + 2c\eta \frac{1}{p} tr((v(\eta, 0)\mathbf{\Sigma} + \eta \mathbf{I})^{-1}) - c\eta^2 \frac{1}{p} tr((v(\eta, 0)\mathbf{\Sigma} + \eta \mathbf{I})^{-2})}. \tag{23}$$

The proof of Theorem 1 can be found in the appendix. Note that the condition (15) is fulfilled when the population covariance matrix $\mathbf{\Sigma}$ has a bounded spectrum and the elements of the target vector $\mathbf{b}$ are at most of the order $O(1/p)$. This includes many naive-like portfolios, for example, the equally weighted portfolio. It is important to observe that both the GMV and equally weighted portfolios exhibit variances of the order $O(1/p)$. Consequently, if the target portfolio $\mathbf{b}$ has a variance of a higher order, e.g., in the case of sparse portfolios, then the loss function $L_\mathbf{b}$ could potentially be unbounded. This implies that a portfolio $\mathbf{b}$ with a higher-order variance may not be reasonable as a target portfolio for minimizing the portfolio variance. Furthermore, if the elements of $\mathbf{b}$ are at most of the order $O(1/p)$, then the covariance matrix $\mathbf{\Sigma}$ can be arbitrary to some extent (sparse or dense), i.e., as long as its spectrum remains bounded. Indeed, using Ruhe's trace inequality (cf., Lemma 1 in Ruhe (1970)), we have

$$L_\mathbf{b} = \mathbf{b}^\top \mathbf{\Sigma} \mathbf{b} \cdot \mathbf{1}^\top \mathbf{\Sigma}^{-1} \mathbf{1} - 1 \le \mathbf{b}^\top \mathbf{b} \cdot \lambda_{max}(\mathbf{\Sigma}) \cdot p\lambda_{max}(\mathbf{\Sigma}^{-1}) - 1 = O\left(\frac{\lambda_{max}(\mathbf{\Sigma})}{\lambda_{min}(\mathbf{\Sigma})}\right) < \infty\,.$$

Theorem 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 (21) is a Stieltjes transform (see Rubio and Mestre (2011)) and inherits all the properties of this type of 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 $\mathbf{\Sigma}$ and $v(\eta,0)$. Fortunately, we can create consistent bona fide [4] estimators for both deterministic equivalents $L_2(\lambda)$ and $\psi^*(\lambda)$ in the 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, consistent estimators for $v(\eta,0)$, $v_1'(\eta,0)$, and $v_2'(\eta,0)$ are derived in Theorem 2, whose proof is given in the appendix.

Theorem 2 *Let $\mathbf{Y}_n$ possess the stochastic representation as in* (10)*. Then it holds that*

$$\begin{aligned} |\hat{v}(\eta,0)-v(\eta,0)| &\stackrel{a.s}{\rightarrow} 0, &(24)\\ \left|\hat{v}_1'(\eta,0)-v_1'(\eta,0)\right| &\stackrel{a.s}{\rightarrow} 0, &(25)\\ \left|\hat{v}_2'(\eta,0)-v_2'(\eta,0)\right| &\stackrel{a.s}{\rightarrow} 0, &(26)\end{aligned}$$

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

$$\begin{aligned}\hat{v}(\eta,0) &= 1-c\left(1-\eta\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)\right),\\ \hat{v}_1'(\eta,0) &= \hat{v}(\eta,0)c\left(\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)-\eta\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-2}\right)\right),\\ \hat{v}_2'(\eta,0) &= 1-\frac{1}{\hat{v}(\eta,0)}+\eta\frac{\hat{v}_1'(\eta,0)}{\hat{v}(\eta,0)^2}.\end{aligned}$$

Theorem 3 provides consistent estimators for the building blocks used in the construction of consistent estimators for $L_2(\lambda)$ and $\psi^*(\lambda)$. The proof of Theorem 3 is presented in the appendix.

Theorem 3 *Let $\mathbf{Y}_n$ possess the stochastic representation as in* (10)*. Assume that the relative loss of the target portfolio given in* (15) *is uniformly bounded in p. Then it holds that*

$$\left|\frac{\mathbf{b}^\top\mathbf{S}\mathbf{b}}{\mathbf{b}^\top\mathbf{\Sigma}\mathbf{b}}-1\right|\stackrel{a.s}{\rightarrow}0, \qquad (27)$$

$$\left|\frac{\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}}-\frac{\mathbf{1}^\top\mathbf{\Omega}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}}\right|\stackrel{a.s}{\rightarrow}0, \qquad (28)$$

$$\left|\frac{\lambda^{-1}}{\hat{v}(\eta,0)}\frac{1-(1-\lambda)\mathbf{b}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}{\sqrt{\mathbf{b}^\top\mathbf{\Sigma}\mathbf{b}\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}}}-\frac{\mathbf{b}^\top\mathbf{\Sigma}\mathbf{\Omega}_\lambda^{-1}\mathbf{1}}{\sqrt{\mathbf{b}^\top\mathbf{\Sigma}\mathbf{b}\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}}}\right|\stackrel{a.s}{\rightarrow}0, \qquad (29)$$

$$\left|\frac{1}{\lambda\hat{v}(\eta,0)}\frac{\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}}-\frac{1-\lambda}{\lambda\hat{v}(\eta,0)}\frac{\mathbf{1}^\top\mathbf{S}_\lambda^{-2}\mathbf{1}-\lambda^{-1}\frac{\hat{v}_1'(\eta,0)}{\hat{v}(\eta,0)}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}\left(1-\frac{\hat{v}_1'(\eta,0)}{\hat{v}(\eta,0)}(\frac{1}{\lambda}-1)\right)}-\frac{\mathbf{1}^\top\mathbf{\Omega}_\lambda^{-1}\mathbf{\Sigma}\mathbf{\Omega}_\lambda^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{\Sigma}^{-1}\mathbf{1}}\right|\stackrel{a.s}{\rightarrow}0 \qquad (30)$$

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

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

Let

$$d_1(\eta) = \frac{\lambda^{-1}}{\hat{v}(\eta,0)}\left(1-(1-\lambda)\mathbf{b}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}\right) \tag{31}$$

and

$$d_2(\eta) = \frac{1}{\lambda\hat{v}(\eta,0)}\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1} - \frac{1-\lambda}{\lambda\hat{v}(\eta,0)}\frac{\mathbf{1}^\top \mathbf{S}_\lambda^{-2}\mathbf{1} - \lambda^{-1}\frac{\hat{v}_1'(\eta,0)}{\hat{v}(\eta,0)}\mathbf{1}^\top \mathbf{S}_\lambda^{-1}\mathbf{1}}{1-\frac{\hat{v}_1'(\eta,0)}{\hat{v}(\eta,0)}(\frac{1}{\lambda}-1)}. \tag{32}$$

The application of the results derived in Theorems 2 and 3 leads to consistent bona fide estimators for $L_2(\lambda)$ and $\psi^*(\lambda)$ presented in Theorem 4.

Theorem 4 *Let $\mathbf{Y}_n$ possess the stochastic representation as in* (10)*. Assume that the relative loss of the target portfolio given in* (15) *is uniformly bounded in p. Then it holds that*

(i)

$$\left|\hat{L}_{n;2}(\lambda) - L_2(\lambda)\right| \stackrel{a.s.}{\to} 0 \tag{33}$$

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

$$\hat{L}_{n;2}(\lambda) = \frac{\left(1-\frac{1}{\mathbf{b}^\top\mathbf{S}\mathbf{b}}\frac{d_1(\eta)}{\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}\right)^2}{1-\frac{2}{\mathbf{b}^\top\mathbf{S}\mathbf{b}}\frac{d_1(\eta)}{\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}+\frac{1}{\mathbf{b}^\top\mathbf{S}\mathbf{b}}\frac{(1-\hat{v}_2'(\eta,0))d_2(\eta)}{\left(\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}\right)^2}}, \tag{34}$$

(ii)

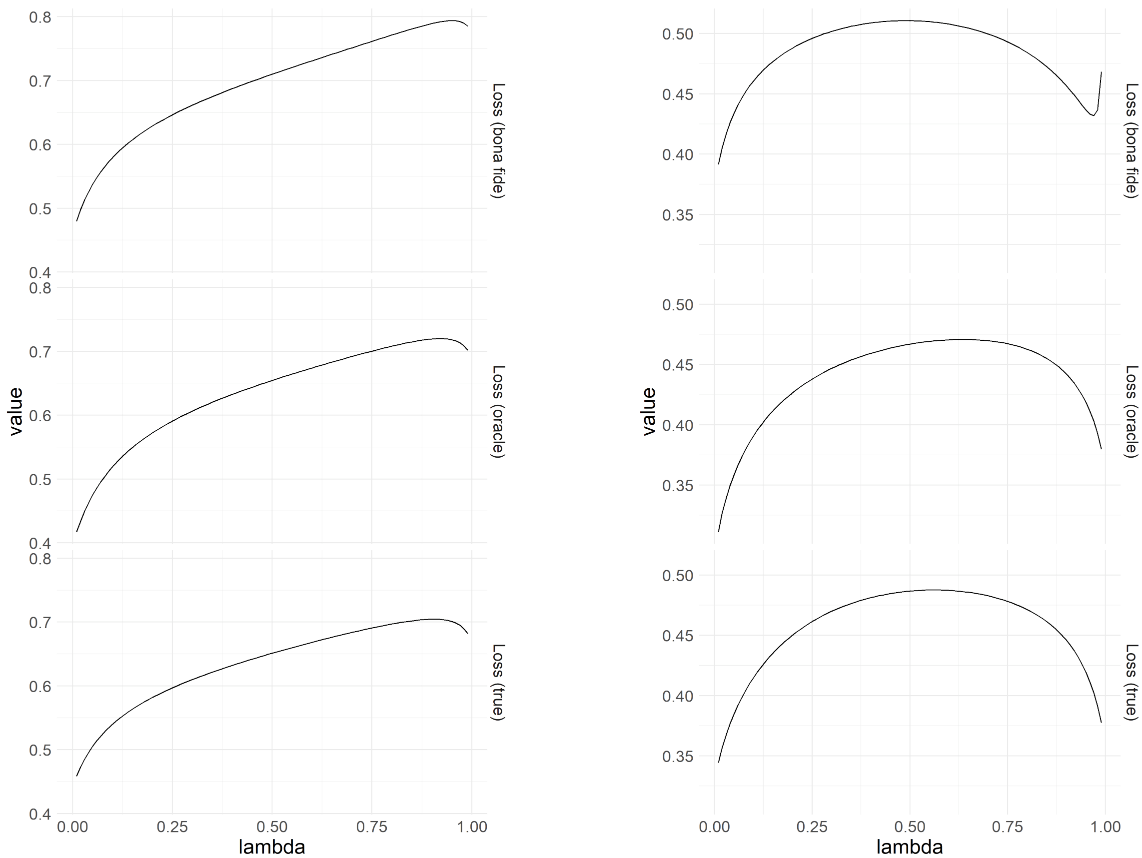
$$\left|\hat{\psi}_n^*(\lambda) - \psi^*(\lambda)\right| \stackrel{a.s.}{\to} 0 \tag{35}$$

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

$$\hat{\psi}_n^*(\lambda) = \frac{1-\frac{1}{\mathbf{b}^\top\mathbf{S}\mathbf{b}}\frac{d_1(\eta)}{\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}}{1-\frac{2}{\mathbf{b}^\top\mathbf{S}\mathbf{b}}\frac{d_1(\eta)}{\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}}+\frac{1}{\mathbf{b}^\top\mathbf{S}\mathbf{b}}\frac{(1-\hat{v}_2'(\eta,0))d_2(\eta)}{\left(\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1}\right)^2}}, \tag{36}$$

where $\eta = 1/\lambda - 1$, $\hat{v}_2'(\eta,0)$ is provided in Theorem 2, $d_1(\eta)$ and $d_2(\eta)$ are given in (31) *and* (32)*, respectively.*

The three loss functions (bona fide, oracle, and true) are illustrated in Figure 2 for two different values of p. Under the 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 is 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



(a) p=150, n=300

(b) p=450, n=300

Figure 2: The loss functions from Theorems 1 to 4 illustrated over different values of $\lambda \in (0, 1)$. The data were simulated from a t-distribution with 5 degrees of freedom and the equally weighted portfolio was used as a target.

covariance matrix explode. This can be explained by the fact that when λ is close to one, the uniform convergence of the estimated loss function to its oracle no longer holds. This issue could be repaired using a different type of ridge regularization mentioned in (Bodnar et al., 2023a, formula (2.33)), where Moore-Penrose inverse for $c > 1$ can be employed for $\lambda \to 1$. This interesting observation is left for 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 compare its behavior with existing 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 conclude this section with an empirical application of the methods on assets from the 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 (10).

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 $i = 1, ..., p$.

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

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

where the conditional covariance matrix is specified by

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

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 $\boldsymbol{\Sigma}$. The intercept $\alpha_{j,0}$ is chosen such that the unconditional covariance matrix is equal to $\boldsymbol{\Sigma}$.

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

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

for $t = 1, ..., n+m$, where $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \gamma_2, ..., \gamma_p)$ 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 $\text{vec}(\mathbb{V}ar(\mathbf{y})) = (\mathbf{I} - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})^{-1}\text{vec}(\boldsymbol{\Sigma})$ 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 **RandCovMat** from the HDShOP package (see, Bodnar et al. (2021)).

4.2 Comparison to Benchmark Strategies

In this section, we will investigate the performance of five estimators of the GMV portfolio defined by

1. The estimator of the GMV portfolio weights from Theorem 4, abbreviated as "Double".

2. The linear shrinkage estimator of the GMV portfolio weights from Bodnar et al. (2018), denoted as "BPS".

3. The linear shrinkage estimator of the GMV portfolio weights from Frahm and Memmel (2010), abbreviated as "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 replaces $\mathbf{\Sigma}$ in the definition of the GMV portfolio weights. We will abbreviate this portfolio strategy as "LW2020".

5. The sample GMV portfolio, abbreviated as "Traditional". For $c > 1$, the Moore-Penrose inverse of $\mathbf{S}_n$ is used in the computation of the weights.

Different targets can be used in the first three estimators of the GMV portfolio weights. We consider two target portfolios in the comparison study, which are the equally weighted (ew) and the equally correlated (ec) portfolios. The first one is deterministic and does not depend on data. This is in line with the assumptions posed in Theorem 4. The second target portfolio depends on data. It assumes that all asset returns share the same correlation but have different variances. For each scenario, we will display the relative loss $V_{\mathbf{w}}/V_{GMV} - 1$, where $V_{\mathbf{w}} = \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}$ is the variance of the portfolio with weights $\mathbf{w}$. A value close to zero indicates good performance of the corresponding estimator.

Figure 3 depicts the results of the simulations under Scenario 1. Each color represents a "type" of the estimator, while the line type highlights the target portfolio used in BPS, Double, and FM. For small c, the loss is not significantly different among the considered estimators. However, as c becomes larger, the results diverge. Regardless of n, the largest loss is provided by the Traditional estimator, which is famously bad in high 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 with the equally correlated and equally weighted targets. The uncertainty from the target portfolio does not seem to impact the loss significantly. The best-performing portfolios are the Double and LW2020 approaches. Finally, the loss computed for the LW2020 estimator significantly increases around one when $n = 100$.

In Figure 4, we present the results of the simulation study conducted under Scenario 2. The ordering of the estimators remains the same as in Scenario 1. The differences are not substantial, as the inverse covariance matrix in the case of the CAPM is essentially a one-rank update of the covariance matrix considered in Scenario 1. It is important to note that the largest eigenvalue of $\mathbf{\Sigma}$ is no longer bounded in Scenario 2. There is slightly more noise, but no additional temporal or structural dependence that is unaccounted for in (10).

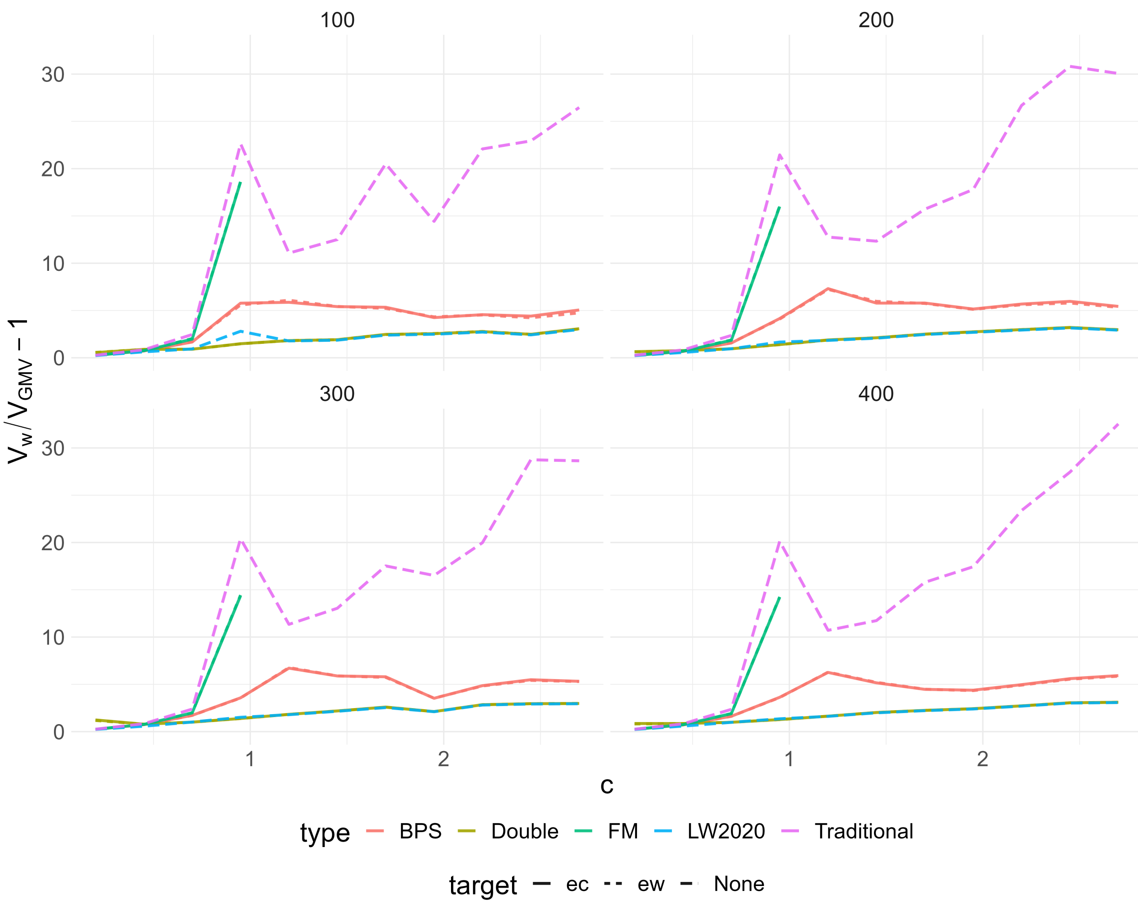


Figure 3: Relative loss $V_{\mathbf{w}}/V_{GMV} - 1$ computed for several estimators of the GMV portfolio weights under Scenario 1. Equally correlated (ec) and equally weighted (ew) targets are used in the BPS, Double, and FM estimators.

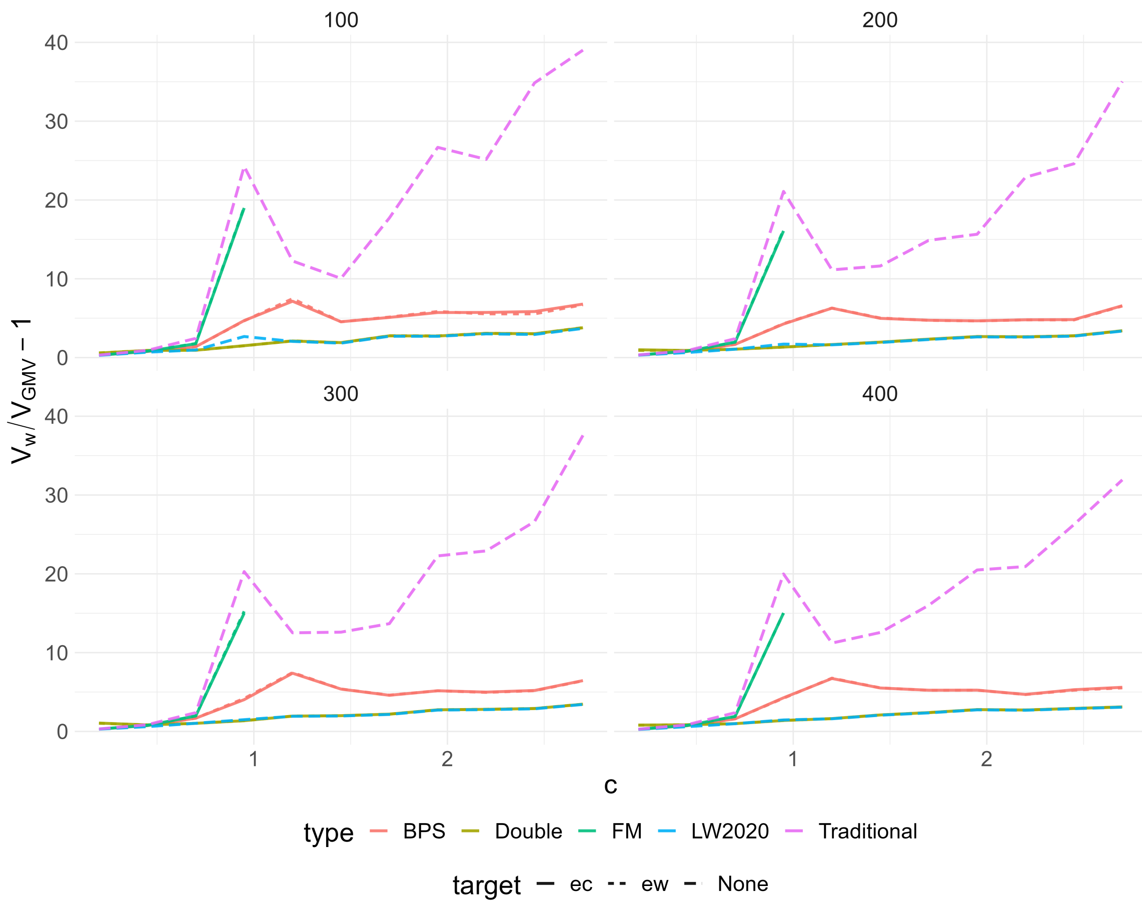


Figure 4: Relative loss $V_{\mathbf{w}}/V_{GMV} - 1$ computed for several estimators of the GMV portfolio weights under Scenario 2. Equally correlated (ec) and equally weighted (ew) targets are used in the BPS, Double, and FM estimators.

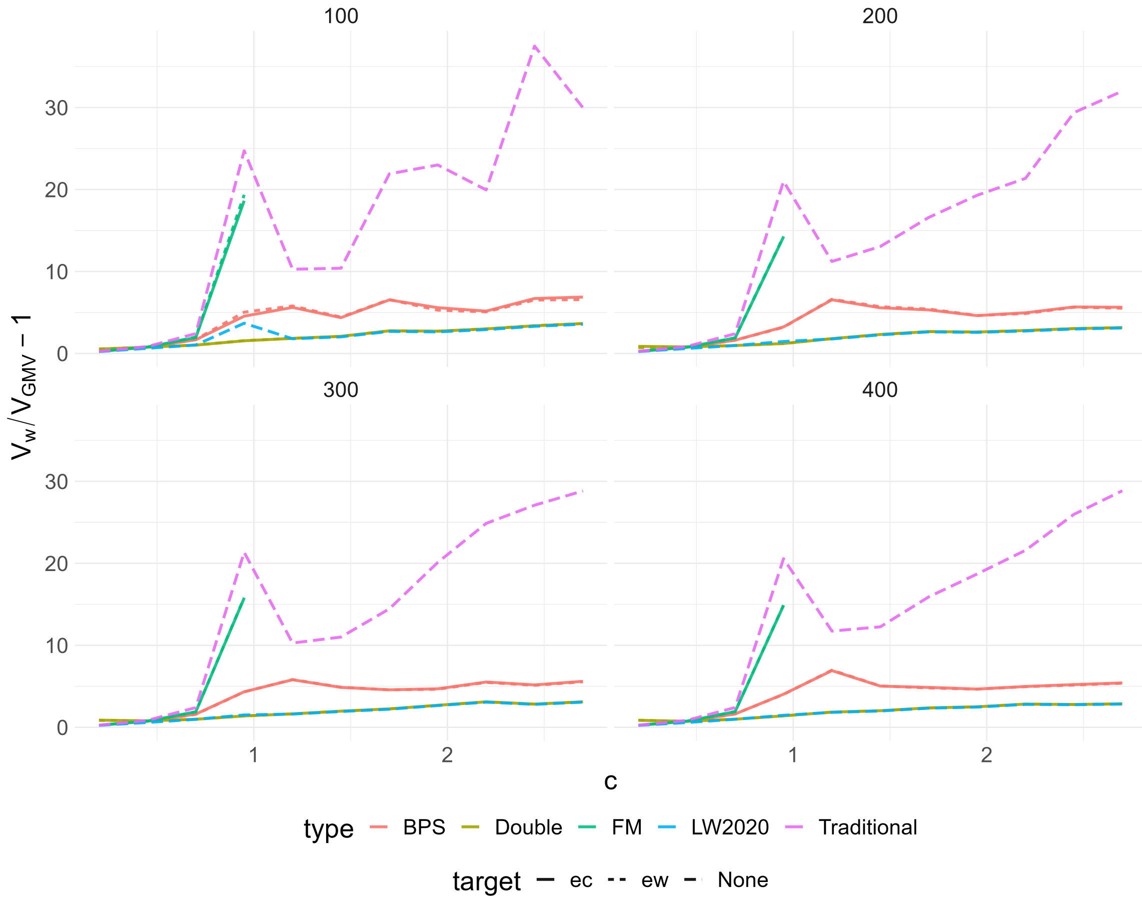


Figure 5: Relative loss $V_{\mathbf{w}}/V_{GMV}-1$ computed for several estimators of the GMV portfolio weights under Scenario 3. Equally correlated (ec) and equally weighted (ew) targets are used in the BPS, Double, and FM estimators.

Figure 5 depicts the relative losses computed for the considered estimators of the GMV portfolio under Scenario 3. It displays plots almost identical to those shown in Figures 3 and 4. The introduction of temporal dependence does not significantly impact the relative loss. We continue to observe the same type of ordering, where the Double and LW2020 approaches are almost equally effective.

In Figure 6, we present the results obtained under Scenario 4. While the same ordering of methods is maintained, the scale of results differs. Some methods, specifically the Traditional, FM, and BPS, exhibit considerably larger losses compared to previous scenarios. There are smaller differences among the other methods for large c, in contrast to earlier scenarios. The discrepancy between the two best methods, the Double and LW2020 estimators, is very slight.

Overall, the results of the simulation experiment support the conclusion that the proposed method is at least as effective as the nonlinear shrinkage technique, which is already recognized as a state-of-the-art method for estimating large dimensional covariance matri-

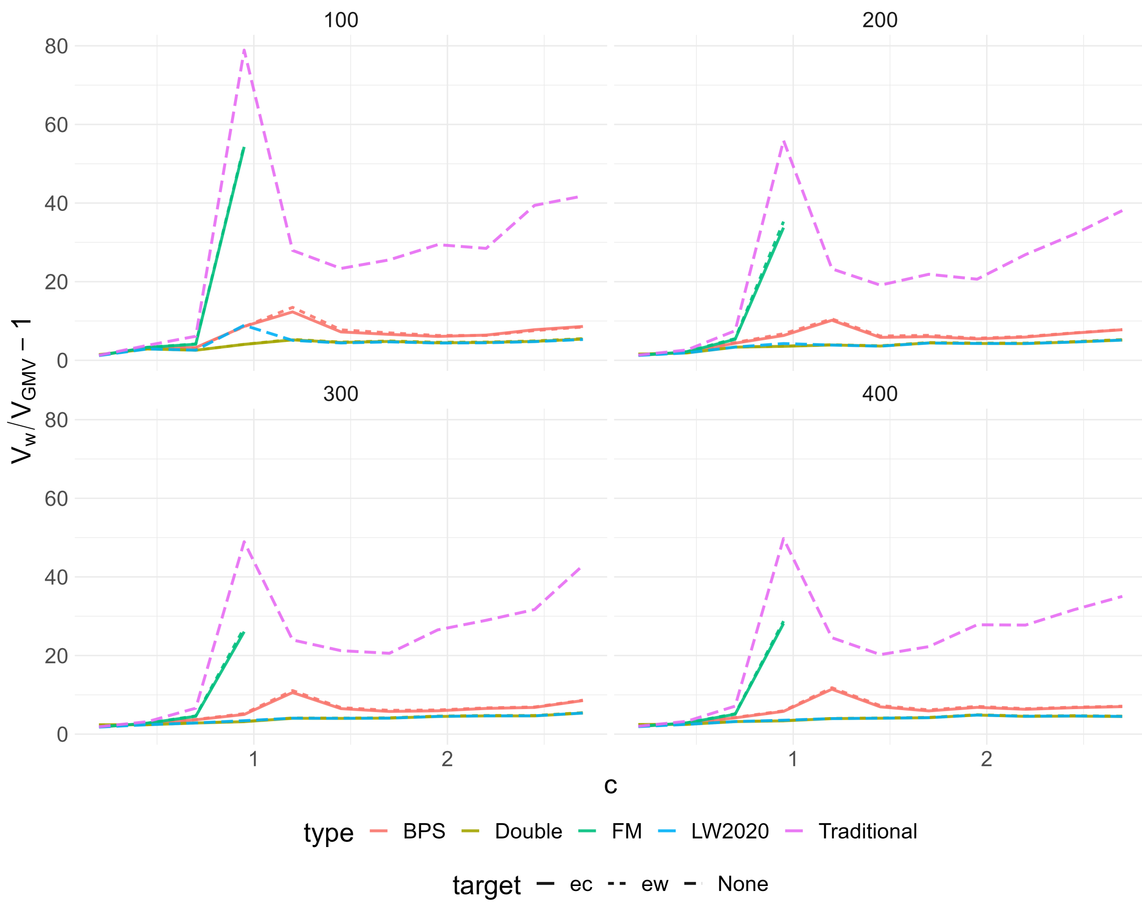


Figure 6: Relative loss $V_{\mathbf{w}}/V_{GMV} - 1$ computed for several estimators of the GMV portfolio weights under Scenario 4. Equally correlated (ec) and equally weighted (ew) targets are used in the BPS, Double, and FM estimators.

ces. Therefore, it is crucial to test these methods on a real dataset using additional empirical measures of performance, such as out-of-sample variance, return, Sharpe ratio, turnover, etc.

4.3 Empirical Application

In this section, we will apply the considered estimators of the GMV portfolio to empirical data, which constitute daily log returns on 431 assets included in the S&P 500 index. The out-of-sample data range from 2013-01-01 to 2021-11-04, while the in-sample data go back to early 2011. We will follow the approach of the previous section in that we use the equally correlated (ec) and equally weighted (ew) portfolios as targets. In this empirical application, we fix the window size to $n = 250$ or $n = 500$. The three different portfolio sizes of 260, 400, and 431 are considered. Thus, the window size $n = 250$ reflects $c \in 1.04, 1.6, 1.724$ and $n = 500$ corresponds to $c \in 0.52, 0.8, 0.862$.

All portfolios aim to minimize the portfolio variance. Since the true portfolio variance is not available, the most natural evaluation method selects the strategy that provides the smallest out-of-sample variance, denoted by $\sigma^{(k)}$. On the other hand, a portfolio is characterized not only by its volatility. A portfolio with small volatility does not necessarily provide a feasible return. Therefore, we will use the out-of-sample expected portfolio return ($\bar{\mathbf{y}}_{\mathbf{w}}^{(k)}$) and the out-of-sample Sharpe ratio ($SR^{(k)}$) to investigate the properties of the constructed estimators. Moreover, the stability of the portfolio weights also reflects the riskiness of the portfolio. To study the characteristics of the portfolio weights, we will consider the following performance measures:

$$|\mathbf{w}^{(k)}| = \frac{1}{Tp}\sum_{i=1}^{T}\sum_{j=1}^{p}|w_{i,j}^{(k)}|, \tag{37}$$

$$\max \mathbf{w}^{(k)} = \frac{1}{T}\sum_{i=1}^{T}\left(\max_j w_{i,j}^{(k)}\right), \tag{38}$$

$$\min \mathbf{w}^{(k)} = \frac{1}{T}\sum_{i=1}^{T}\left(\min_j w_{i,j}^{(k)}\right), \tag{39}$$

$$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)} < 0) = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{p} w_{i,j}^{(k)}\mathbb{1}(w_{i,j}^{(k)} < 0), \tag{40}$$

$$\mathbb{1}(\mathbf{w}_i^{(k)} < 0) = \frac{1}{Tp}\sum_{i=1}^{T}\sum_{j=1}^{p}\mathbb{1}(w_{i,j}^{(k)} < 0). \tag{41}$$

The first measure (37) represents the average size of the portfolio positions. A large value of this measure would indicate that the portfolio takes large positions, both short (negative weights) and long (positive weights). It is a common critique of classical mean-variance portfolios because large positions are risky in themselves. The second measure (38) describes the average of the largest long positions. Similar to the previous measure, it only considers long positions of the portfolio. As with the above, small positions are preferred over large ones. Equation (39) presents the average of the largest short positions.

This measure is particularly important since large short positions have potentially infinite risk, i.e., there is no limit to how much the investor can lose. Owing to the significant importance of the size of short positions, we also include two additional measures (40) and (41). The former can be interpreted as the average size of the negative positions, while the latter represents the average proportion of short positions.

	BPS		Double			
name	BPS_{ec}	BPS_{ew}	Double_{ec}	Double_{ew}	LW2020	Traditional
p=260						
σ^k	0.0345	0.03535	**0.0116**	_0.01165_*	0.01225	0.04103
$\bar{\mathbf{y}}^k_{\mathbf{w}}$	0.000362	0.000307	_0.000566_*	**0.000581**	0.00056	0.000263
SR^k	0.01	0.009	_0.049_*	**0.05**	0.046	0.006
$\lvert\mathbf{w}^{(k)}\rvert$	0.0905	0.0938	**0.0061**	_0.0062_*	0.01	0.114
$\max \mathbf{w}^{(k)}$	0.4249	0.4368	0.0381	**0.0177**	_0.0377_*	0.5309
$\min \mathbf{w}^{(k)}$	−0.3981	−0.417	**−0.0052**	_−0.0181_*	−0.0292	−0.5083
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	−0.0883	−0.0923	**−0.0029**	_−0.0049_*	−0.0101	−0.1123
250 $\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	0.491	0.487	0.382	**0.246**	_0.303_*	0.491
Turnover	39754.41	41248.72	**88.18**	_100.27_*	2824.26	50073
p=400						
σ^k	0.01438	0.01448	_0.01189_*	**0.01186**	0.01261	0.0157
$\bar{\mathbf{y}}^k_{\mathbf{w}}$	0.000634	**0.000649**	0.000513	0.000536	0.000534	_0.000646_*
SR^k	_0.044_*	**0.045**	0.043	**0.045**	0.042	0.041
$\lvert\mathbf{w}^{(k)}\rvert$	0.0182	0.0187	**0.004**	**0.004**	_0.0105_*	0.0239
$\max \mathbf{w}^{(k)}$	0.0746	0.0748	_0.0296_*	**0.0121**	0.047	0.095
$\min \mathbf{w}^{(k)}$	−0.0663	−0.0692	**−0.0031**	_−0.0133_*	−0.0398	−0.0892
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	−0.017	−0.0177	**−0.0019**	_−0.0031_*	−0.0095	−0.0228
$\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	0.462	0.457	_0.397_*	**0.241**	0.419	0.468
Turnover	4257.59	4473.48	**88.34**	_97.33_*	793.77	5719.24
p=431						
σ^k	0.01471	0.01484	**0.0115**	_0.01186_*	0.01286	0.01596
$\bar{\mathbf{y}}^k_{\mathbf{w}}$	0.000476	0.000472	_0.000514_*	**0.000517**	0.000427	0.000444
SR^k	0.032	0.032	**0.045**	_0.044_*	0.033	0.028
$\lvert\mathbf{w}^{(k)}\rvert$	0.0158	0.0162	**0.0037**	**0.0037**	_0.0098_*	0.0209
$\max \mathbf{w}^{(k)}$	0.0652	0.0654	_0.0276_*	**0.0111**	0.0446	0.0841
$\min \mathbf{w}^{(k)}$	−0.0569	−0.0597	**−0.0029**	_−0.0121_*	−0.0373	−0.078
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	−0.0147	−0.0153	**−0.0018**	_−0.0029_*	−0.0089	−0.02
$\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	0.459	0.455	_0.399_*	**0.24**	0.421	0.466
Turnover	3651.43	3854.04	**87.92**	_96.66_*	788.84	4998.56

* Second to best

Table 1: Performance of the considered estimators based on the moving-window approach. The out-of-sample period is 2684 days with window size equal to 250.

Table 1 displays the results from the first experiment with a window size equal to 250 days. The best values in each row are denoted in boldface, while the second-best values are highlighted with the symbol *. The FM strategy is not considered in the study since it demonstrates very similar performance to the Traditional strategy for $c < 1$ and is not defined for $c > 1$.

For a moderately small portfolio size of $p = 260$, the double shrinkage estimator with the equally correlated target yields the smallest out-of-sample variance[5], while the double shrinkage estimator with the equally weighted target ranks second. The same ordering between the two double strategies is observed in terms of turnover, while the opposite situation occurs when comparing based on the out-of-sample expected return and the out-of-sample Sharpe ratio. In general, there is a small difference between the Double strategies with the considered two targets, and they both are ranked in the first two places. The portfolio based on the LW2020 approach comes third in terms of the out-of-sample variance; however, it considerably outperforms the other remaining strategies. The portfolio weights of the nonlinear shrinkage are not as stable as the weights of the two double shrinkage approaches. In particular, the turnover of the LW2020 strategy is approximately 30 times larger than the turnover of the two double shrinkage methods. Finally, the BPS strategy with both targets is ranked in the fourth and fifth places, followed by the Traditional approach.

	BPS		Double			
name	BPS_{ec}	BPS_{ew}	Double_{ec}	Double_{ew}	LW2020	Traditional
p=260						
σ^k	0.01843	0.01831	_0.01154_*	**0.0115**	0.0146	0.02051
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	_0.000603_*	0.000558	0.000595	0.000562	**0.000604**	0.00058
SR^k	0.033	0.03	**0.052**	_0.049_*	0.041	0.028
$\|\mathbf{w}^{(k)}\|$	0.0248	0.0256	_0.0063_*	**0.0061**	0.017	0.0328
$\max \mathbf{w}^{(k)}$	0.1535	0.1532	_0.0393_*	**0.0179**	0.0769	0.1927
$\min \mathbf{w}^{(k)}$	−0.1021	−0.1123	**−0.0054**	_−0.0172_*	−0.0635	−0.1462
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	−0.022	−0.0238	**−0.0031**	_−0.0047_*	−0.015	−0.0308
$\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	0.478	0.457	_0.4_*	**0.243**	0.437	0.471
Turnover	1103.98	1166.8	**56.42**	_60.66_	487.87	1503.39
p=400						
σ^k	0.01645	0.01621	_0.01088_*	**0.01087**	0.0121	0.02233
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	0.000706	_0.000728_*	0.000562	0.000546	0.000468	**0.000819**
SR^k	0.043	0.045	_0.052_*	**0.05**	0.039	0.037
$\|\mathbf{w}^{(k)}\|$	0.0229	0.0234	**0.004**	**0.004**	_0.0124_*	0.0421
$\max \mathbf{w}^{(k)}$	0.1451	0.142	0.0277	**0.0117**	0.0583	_0.2516_*
$\min \mathbf{w}^{(k)}$	−0.1105	−0.1206	**−0.0033**	_−0.0128_*	−0.0508	−0.2207
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	−0.0209	−0.0225	**−0.0019**	_−0.0031_*	−0.0113	−0.0407
$\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	0.489	0.465	_0.392_*	**0.244**	0.438	0.485
Turnover	2683.1	2824.09	**53.65**	_61.04_*	706.3	5027.35
p=431						
σ^k	0.01677	0.01707	**0.0108**	_0.01121_*	0.01221	0.02613
$\bar{\mathbf{y}}_{\mathbf{w}}^k$	0.000447	0.000434	**0.000578**	_0.00056_*	0.000543	0.000311
SR^k	0.027	0.025	**0.054**	_0.05_*	0.044	0.012
$\|\mathbf{w}^{(k)}\|$	0.0222	0.0231	**0.0037**	_0.0038_*	0.0112	0.0487
$\max \mathbf{w}^{(k)}$	0.1392	0.1359	_0.0258_*	**0.0112**	0.0521	0.2853
$\min \mathbf{w}^{(k)}$	−0.1065	−0.1192	**−0.0031**	_−0.012_*	−0.0449	−0.2563
$\mathbf{w}_i^{(k)}\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	−0.0202	−0.0223	**−0.0018**	_−0.0029_*	−0.0101	−0.0475
$\mathbb{1}(\mathbf{w}_i^{(k)} < 0)$	0.493	0.465	_0.395_*	**0.247**	0.438	0.489

5. The results regarding the variance are highly significant, which is justified (all p-values are virtually equal to zero) by the HAC test provided by Ledoit and Wolf (2011).

Turnover	3374.12	3630.07	**53.71**	*61.65**	847.26	7585.94

* Second to best

Table 2: Performance of the considered estimators based on the moving-window approach. The out-of-sample period is 2434 days with window size equal to 500.

The same ranking between the strategies is also observed for $p = 400$ and $p = 431$. Both double shrinkage approaches are the best two strategies in terms of almost all the considered performance measures. The only exception is in the case of $p = 400$, where the BPS approach with the equally weighted target is the best in terms of the out-of-sample expected return and the out-of-sample Sharpe ratio. Similar to the case of $p = 260$, the LW2020 strategy comes third, followed by the BPS approach. Finally, the Traditional GMV portfolio performs considerably worse than its competitors. It is noted that the turnover of the LW2020 method is significantly reduced when $p = 400$ and $p = 431$, but it is still approximately nine times larger than the turnover of the double shrinkage approaches.

Next, we perform the same experiment with $n = 500$, with the results presented in Table 2. Again, both double shrinkage approaches considerably outperform the other competitors in terms of almost all the considered performance measures. In particular, these two strategies possess the smallest values of the out-of-sample variance, which is statistically justified by the HAC test of Ledoit and Wolf (2011). When $p = 260$ and $p = 400$, the double shrinkage estimator with equally weighted target is the best one, while the double approach is superior when $p = 431$. The LW2020 strategy ranks third, while the BPS method with both targets is in fourth and fifth places. Finally, the Traditional estimator performs very poorly, especially when $p = 400$ and 431. Interestingly, the application of the double shrinkage strategies leads to very stable portfolio weights. The turnovers of these two strategies are considerably smaller than those of the other competitors. Even though the increase in window size leads to considerably smaller turnovers for all the considered methods, they are still significantly larger than those obtained by the double shrinkage approach, regardless of the used target portfolios.

5. Summary

In this paper, we introduce a novel method for investing in the GMV portfolio and a target portfolio. This method utilizes a double shrinkage approach, where the sample covariance matrix is shrunk with Tikhonov regularization, coupled with linear shrinkage of the GMV portfolio weights towards a target portfolio. We construct a bona fide loss function that consistently estimates the true loss function. From this, we derive the two shrinkage coefficients within the framework. The method is demonstrated to be a significant improvement over both the nonlinear shrinkage approach and the linear shrinkage approach applied directly to the portfolio weights. Furthermore, the method is shown to be a dominant investment strategy in the majority of cases, as justified by various performance measures in the empirical application. Finally, the proposed approach is opinionated, requiring the investor's view on what constitutes a target portfolio. This implies that it works best when the target portfolio is informative, aligning with the investor's objectives. However, the investor can also employ non-informative target portfolios and still achieve excellent results, as documented in the empirical illustration.

Acknowledgments

The authors would like to thank Professor Jie Peng and the two anonymous Reviewers for their constructive comments that improved the quality of this paper. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) partially funded by the Swedish Research Council through grant agreement no. 2022-06725

Appendix A.

For any integer $n > 2$, we define

$$\mathbf{V}_n = \frac{1}{n}\mathbf{X}_n\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\mathbf{X}_n^\top \quad \text{and} \quad \widetilde{\mathbf{V}}_n = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^\top, \tag{42}$$

where $\mathbf{X}_n$ is given in (10). Hence,

$$\mathbf{S}_n = \mathbf{\Sigma}^{1/2}\mathbf{V}_n\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}^{1/2}\widetilde{\mathbf{V}}_n\mathbf{\Sigma}^{1/2} - \mathbf{\Sigma}^{1/2}\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^\top\mathbf{\Sigma}^{1/2} \tag{43}$$

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

First, we present the Weierstrass theorem (see, e.g., Theorem 1 on page 176 in Ahlfors (1953)) and an important lemma which is a special case of Theorem 1 in Rubio and Mestre (2011). These results will be used in the proofs throughout the appendix.

Theorem 5 (Weierstrass theorem) *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 5 plays a vital role here. More on the application of Weierstrass theorem can be found in the appendix of Bodnar et al. (2023a).

Lemma 6 *Let a nonrandom $p \times p$-dimensional matrix $\mathbf{\Theta}_p$ possess a uniformly bounded trace norm. Then it holds that*

(i)

$$\left| tr\left(\mathbf{\Theta}_p\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\mathbf{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-1}\right) - tr\left(\mathbf{\Theta}_p(\eta\mathbf{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1}\right)\right| \xrightarrow{a.s.} 0 \tag{44}$$

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

$$v(\eta, z) = \frac{1}{1 + c\frac{1}{p}tr\left((\eta\mathbf{\Sigma}^{-1} + (v(\eta,z) - z)\mathbf{I})^{-1}\right)}. \tag{45}$$

(ii)

$$\Bigg| tr\left(\boldsymbol{\Theta}_p\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\boldsymbol{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-1}\boldsymbol{\Sigma}^{-1}\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\boldsymbol{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-1}\right)$$
$$- \quad tr\left(\boldsymbol{\Theta}_p(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-1}\boldsymbol{\Sigma}^{-1}(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-1}\right)$$
$$- \quad v_1'(\eta,z)\,tr\left(\boldsymbol{\Theta}_p(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-2}\right)\Bigg| \xrightarrow{a.s.} 0 \tag{46}$$

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

$$v_1'(\eta,z) = \frac{-\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-1}\boldsymbol{\Sigma}^{-1}(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-1}\right)}{\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-2}\right) - c^{-1}v(\eta,z)^{-2}}. \tag{47}$$

(iii)

$$\Bigg| tr\left(\boldsymbol{\Theta}_p\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n' + \eta\boldsymbol{\Sigma}^{-1} - z\mathbf{I}_p\right)^{-2}\right)$$
$$- \quad (1 - v_2'(\eta,z))\,tr\left(\boldsymbol{\Theta}_p(\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-2}\right)\Bigg| \xrightarrow{a.s.} 0 \tag{48}$$

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

$$v_2'(\eta,z) = \frac{\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-2}\right)}{\frac{1}{p}tr\left((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-2}\right) - c^{-1}v(\eta,z)^{-2}}. \tag{49}$$

Proof of Lemma 6.

(i) The application of Theorem 1 in Rubio and Mestre (2011) leads to (44) where $v(\eta,z)$ is a unique solution in $\mathbb{C}^+$ of the following equation

$$\frac{1}{v(\eta,z)} - 1 = \frac{c}{p}\mathrm{tr}\left((\eta\boldsymbol{\Sigma}^{-1} + (v(\eta,z)-z)\mathbf{I})^{-1}\right). \tag{50}$$

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

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

which almost surely converges to

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

following Theorem 5. The first-order partial derivative $v_1'(\eta,z)$ is obtained from (50) as

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

from which (47) 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 5. Moreover, $v_2'(\eta,z)$ is computed from (50) 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.

■

Lemma 7 *Let $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ be universal nonrandom vectors with bounded Euclidean norms. Then it holds that*

$$\left|\boldsymbol{\xi}'\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n'+\eta\mathbf{\Sigma}^{-1}\right)^{-1}\boldsymbol{\theta}-\boldsymbol{\xi}'(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I})^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{51}$$

$$\left|\boldsymbol{\xi}'\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n'+\eta\mathbf{\Sigma}^{-1}\right)^{-1}\mathbf{\Sigma}^{-1}\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n'+\eta\mathbf{\Sigma}^{-1}\right)^{-1}\boldsymbol{\theta}\right. \tag{52}$$

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

$$\left|\boldsymbol{\xi}'\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n'+\eta\mathbf{\Sigma}^{-1}\right)^{-2}\boldsymbol{\theta}-(1-v_2'(\eta,0))\boldsymbol{\xi}'(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I})^{-2}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0 \tag{53}$$

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

$$v(\eta,0) = 1-c\left(1-\frac{\eta}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right)\right), \tag{54}$$

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

$$v_1'(\eta,0)=v(\eta,0)\frac{c\frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right)-c\eta\frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right)}{1-c+2c\eta\frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right)-c\eta^2\frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right)} \tag{55}$$

and

$$v_2'(\eta,0)=1-\frac{1}{v(\eta,0)}+\eta\frac{v_1'(\eta,0)}{v(\eta,0)^2}. \tag{56}$$

Proof of Lemma 7. Since the trace norm of $\boldsymbol{\theta}\boldsymbol{\xi}'$ is uniformly bounded, i.e.,

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

the application of Lemma 6 leads to (51), (52), and (53) where $v(\eta,0)$ satisfies the following equality

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

which results in (54).

The application of (47) leads to

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

Finally, using (49), we get

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

■

Lemma 8 *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*

$$\left|\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-1}\boldsymbol{\theta} - \boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{57}$$

$$\left|\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-2}\boldsymbol{\theta} - \boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-2}\boldsymbol{\theta} - v_1'(\eta,0)\boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0, \tag{58}$$

$$\left|\boldsymbol{\xi}^\top \mathbf{S}_\lambda^{-1}\boldsymbol{\Sigma}\mathbf{S}_\lambda^{-1}\boldsymbol{\theta} - (1 - v_2'(\eta,0))\boldsymbol{\xi}^\top \boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\theta}\right| \xrightarrow{a.s.} 0 \tag{59}$$

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

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

and $v(\eta,0)$, $v_1'(\eta,0)$ and $v_2'(\eta,0)$ given in Lemma 7.

Proof of Lemma 8. Let $\widetilde{\mathbf{S}}_n = \boldsymbol{\Sigma}^{1/2}\widetilde{\mathbf{V}}_n\boldsymbol{\Sigma}^{1/2}$. Using (42) and (43) and the formula for the 1-rank update of inverse matrix (see, e.g., Horn and Johnsohn (1985)), we get

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

where

$$\left|\boldsymbol{\xi}^\top\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda}-1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n\right| \xrightarrow{a.s.} 0 \tag{60}$$

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

$$\frac{1}{1 - \bar{\mathbf{x}}_n^\top\boldsymbol{\Sigma}^{1/2}\left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda}-1\right)\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n} \tag{61}$$

is bounded following Pan (2014, Eq. (2.28)). Hence, the application of Lemma 7 leads to the first statement of Lemma 8.

We compute

$$
\begin{aligned}
\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},
\end{aligned}
$$

where

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

and

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

For the third statement of the lemma we consider

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

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

$$\begin{aligned}
&\bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1\right)\boldsymbol{\Sigma}^{-1}\right)^{-2} \bar{\mathbf{x}}_n \\
&= \bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}^{1/2} \left(\tilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1} \boldsymbol{\Sigma} \left(\tilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1} \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n \\
&\le \lambda_{\max}(\boldsymbol{\Sigma}) \cdot \bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}^{1/2} \left(\tilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-2} \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n \\
&\le \lambda_{\max}(\boldsymbol{\Sigma}) \left(\frac{1}{\lambda} - 1\right)^{-1} \bar{\mathbf{x}}_n^\top \boldsymbol{\Sigma}^{1/2} \left(\widetilde{\mathbf{S}}_n + \left(\frac{1}{\lambda} - 1\right)\mathbf{I}\right)^{-1} \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{x}}_n < \infty
\end{aligned} \tag{62}$$

Using (62) we get that

$$\begin{aligned}
&\left|\boldsymbol{\xi}^\top \boldsymbol{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1\right)\boldsymbol{\Sigma}^{-1}\right)^{-2} \bar{\mathbf{x}}_n\right| \\
\le\ &\sqrt{\boldsymbol{\xi}^\top \boldsymbol{\Sigma}^{-1/2} \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1\right)\boldsymbol{\Sigma}^{-1}\right)^{-2} \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\xi}} \sqrt{\bar{\mathbf{x}}_n^\top \left(\widetilde{\mathbf{V}}_n + \left(\frac{1}{\lambda} - 1\right)\boldsymbol{\Sigma}^{-1}\right)^{-2} \bar{\mathbf{x}}_n} < \infty\,.
\end{aligned}$$

Hence, the application of (60), (61), and Lemma 7 completes the proof of the lemma. ■

Proof of Theorem 1. Let $V_{GMV} = 1/(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1})$. The application of the results of Lemma 8 with $\boldsymbol{\xi} = \boldsymbol{\Sigma}\mathbf{b}/\sqrt{\mathbf{b}^\top\boldsymbol{\Sigma}\mathbf{b}}$ and $\boldsymbol{\theta} = \mathbf{1}/\sqrt{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$ leads to

$$\left|(L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\boldsymbol{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1} - (L_\mathbf{b}+1)^{-1/2}\mathbf{b}^\top\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1}\right| \xrightarrow{a.s.} 0, \tag{63}$$

$$\left|V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\mathbf{1} - \lambda^{-1}V_{GMV}\mathbf{1}^\top\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1}\right| \xrightarrow{a.s.} 0, \tag{64}$$

$$\left|V_{GMV}\mathbf{1}^\top\mathbf{S}_\lambda^{-1}\boldsymbol{\Sigma}\mathbf{S}_\lambda^{-1}\mathbf{1} - V_{GMV}(1 - v_2'(\eta, 0))\mathbf{1}'\boldsymbol{\Omega}_\lambda^{-1}\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1}\right| \xrightarrow{a.s.} 0 \tag{65}$$

Using (63)-(65) and the equality

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

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

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

yields the second statement of the theorem. ■

Lemma 9 *Let $\frac{1}{p}\mathbf{\Sigma}^{-1}$ possess a bounded trace norm. Then it holds that*

$$\left|\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)-\frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right)\right|\stackrel{a.s}{\rightarrow}0 \tag{66}$$

$$\left|\frac{\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-2}\right)\left(1-c+2c\eta\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)\right)-c\left[\frac{1}{p}tr\left((\mathbf{S}_n+\eta\mathbf{I})^{-1}\right)\right]^2}{1-c+c\eta^2\frac{1}{p}tr((\mathbf{S}_n+\eta\mathbf{I})^{-2})}\right.$$
$$\left.-\frac{1}{p}tr\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right)\right|\stackrel{a.s}{\rightarrow}0 \tag{67}$$

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

Proof of Lemma 9. From part (i) of Lemma 6 with $\mathbf{\Theta}_p=\frac{1}{p}\mathbf{\Sigma}^{-1}$ and the proof of Lemma 8 we obtain that $\frac{1}{p}\mathrm{tr}((\mathbf{S}_n+\eta\mathbf{I})^{-1})$ is consistent for $\frac{1}{p}\mathrm{tr}\left((\eta\mathbf{I}+v(\eta,0)\mathbf{\Sigma})^{-1}\right)$ in the high dimensional setting.

Furthermore, applying part (ii) of Lemma 6 with $\mathbf{\Theta}_p=\frac{1}{p}\mathbf{\Sigma}^{-1}$ and following the proof of Lemma 8 we get that

$$\begin{aligned}
&\frac{1}{p}\mathrm{tr}\left((\mathbf{S}_n+\eta\mathbf{I})^{-2}\right)\stackrel{a.s.}{\rightarrow}\frac{1}{p}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\left(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I}\right)^{-1}\mathbf{\Sigma}^{-1}\left(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I}\right)^{-1}\right)\\
&+v_1'(\eta,0)\frac{1}{p}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\left(\eta\mathbf{\Sigma}^{-1}+v(\eta,0)\mathbf{I}\right)^{-2}\right)\\
&=\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right)+v_1'(\eta,0)\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\mathbf{\Sigma}\left(v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I}\right)^{-1}\right)\\
=&\frac{v_1'(\eta,0)}{v(\eta,0)}\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-1}\right)+\left(1-\frac{v_1'(\eta,0)}{v(\eta,0)}\eta\right)\frac{1}{p}\mathrm{tr}\left((v(\eta,0)\mathbf{\Sigma}+\eta\mathbf{I})^{-2}\right),
\end{aligned} \tag{68}$$

where the application of (55) leads to

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

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

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

which together with (66) leads to the second statement of the lemma. ■

Proof of Theorem 2. The result (24) is a direct consequence of (54) in Lemma 7 and (66) in Lemma 9.

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 (55) in Lemma 7 and the results of Lemma 9 leads to a consistent estimator of $v_1'(\eta, 0)$ expressed as

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

Finally, the result (26) follows from (24) and (25) together with (56) in Lemma 7. ■

Proof of Theorem 3. The result (27) is a special case of Theorem 3.2 in Bodnar et al. (2014). Equation (28) follows from Lemma 8 with $\boldsymbol{\xi} = \boldsymbol{\theta} = \mathbf{1}/\sqrt{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$. For the derivation of (29) we note that

$$\mathbf{b}^\top\boldsymbol{\Sigma}\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1} = \frac{1}{\lambda v(\eta,0)}\left(1-(1-\lambda)\mathbf{b}^\top\boldsymbol{\Omega}_\lambda^{-1}\mathbf{1}\right) \tag{69}$$

and apply Lemma 8 with $\boldsymbol{\xi} = \mathbf{b}/\sqrt{\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}}$ and $\boldsymbol{\theta} = \mathbf{1}/\sqrt{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$.

Finally, (30) 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 (28) is used for the first summand and (58) of Lemma 8 with $\boldsymbol{\xi} = \boldsymbol{\theta} = \mathbf{1}/\sqrt{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$ for the second one. ■

References

Lars V. Ahlfors. *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, 1953.

Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

Michael J Best and Robert R Grauer. 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, 1991.

Taras Bodnar, Arjun K Gupta, and Nestor Parolya. On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *Journal of Multivariate Analysis*, 132:215–228, 2014.

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

Taras Bodnar, Solomiia Dmytriv, Yarema Okhrin, Dmitry Otryakhin, and Nestor Parolya. *HDShOP: high-dimensional shrinkage optimal portfolios*, 2021. URL `https://CRAN.R-project.org/package=HDShOP`. R package version 0.1.2.

Taras Bodnar, Yarema Okhrin, and Nestor Parolya. Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, 41(1):140–156, 2023a.

Taras Bodnar, Nestor Parolya, and Erik Thorsén. Dynamic shrinkage estimation of the high-dimensional minimum-variance portfolio. *IEEE Transactions on Signal Processing*, 71:1334–1349, 2023b.

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

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

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

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

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

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

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

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

Zhao-Rong Lai, Pei-Yi Yang, Liangda Fang, and Xiaotian Wu. Short-term sparse portfolio optimization based on alternating direction method of multipliers. *Journal of Machine Learning Research*, 19(63):1–28, 2018. URL `http://jmlr.org/papers/v19/17-558.html`.

Zhao-Rong Lai, Liming Tan, Xiaotian Wu, and Liangda Fang. Loss control with rank-one covariance estimate for short-term portfolio optimization. *Journal of Machine Learning Research*, 21(97):1–37, 2020. URL `http://jmlr.org/papers/v21/19-959.html`.

Nathan Lassance. Maximizing the out-of-sample Sharpe ratio. *Available at SSRN 3959708*, 2021.

Nathan Lassance, Alberto Martin-Utrera, and Majeed Simaan. The risk of expected utility under parameter uncertainty. *Management Science, forthcoming*, 2024.

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

Olivier Ledoit and Michael Wolf. Robust performances hypothesis testing with the variance. *Wilmott*, 2011(55):86–89, 2011. doi: https://doi.org/10.1002/wilm.10036. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/wilm.10036`.

Olivier Ledoit and Michael Wolf. Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, 06 2017.

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

Olivier Ledoit and Michael Wolf. The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics*, 20(1):187–218, 06 2020b.

Bin Li, Doyen Sahoo, and Steven C.H. Hoi. Olps: A toolbox for on-line portfolio selection. *Journal of Machine Learning Research*, 17(35):1–5, 2016. URL `http://jmlr.org/papers/v17/15-317.html`.

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

Harry M Markowitz. *Portfolio selection: Efficient diversification of investments.* John Wiley, 1959.

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

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

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

Axel Ruhe. Perturbation bounds for means of eigenvalues and invariant subspaces. *BIT Numerical Mathematics*, 10:343–354, 1970.

Andrei Nikolaevich Tikhonov, AV Goncharsky, VV Stepanov, and Anatoly G Yagola. *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media, 1995.

Jun Tong, Rui Hu, Jiangtao Xi, Zhitao Xiao, Qinghua Guo, and Yanguang Yu. Linear shrinkage estimation of covariance matrices using low-complexity cross-validation. *Signal Processing*, 148:223–233, 2018.

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