Taking Stock of Some Recent and Notable Contribution to Research in Portfolio Analysis

Danqiao Guo, Chengguo Weng, and Tony S. Wirjanto[*]

This version: May 2021

Abstract

In this paper we provide a highly selected review and synthesis on some of the recent and notable contribution to research in portfolio analysis. A unique perspective on this development in the literature is offered in this paper by judiciously identifying a few sample eigenvalues adjustment patterns in a portfolio that leads to an improvement in the out-of-sample portfolio Sharpe ratio when the population covariance matrix admits a high-dimensional factor model. These patterns unveil a key insight into a portfolio performance improvement and shed an important light on the effectiveness of a few recently introduced "robust to estimation errors" covariance matrix estimation approaches, which were not originally designed with the goal to improve the out-of-sample portfolio performance.

Keywords: Finance, Portfolio optimization, Tail eigenvalue amplification, High-dimensional setting, Factor models.

[*]Guo: Department of Statistics and Actuarial Science, University of Waterloo, email: d22guo@uwaterloo.ca. Weng: Department of Statistics and Actuarial Science, University of Waterloo, email: c2weng@uwaterloo.ca. Wirjanto (Corresponding Author): Department of Statistics and Actuarial Science and the School of Accounting and Finance, University of Waterloo, email: twirjanto@uwaterloo.ca.

1 Introduction

Portfolio theory has remained a fascinating research topic since the publication by Markowitz (1952). However the problems of estimation error in expected returns and covariance matrix estimators have impeded an effective implementation of this theory, as has been pointed out by numerous authors, such as Merton (1980); Jagannathan and Ma (2003); Kempf and Memmel (2006); Fan et al. (2008); DeMiguel et al. (2009b) among others. Despite this difficulty, considerable progress has been made steadily in improving the quality of estimators for an expected return vector and a covariance matrix of the portfolio. Instead of simply ignoring the expected return vector and focussing on the minimum-variance portfolios[1], there are a few methods recently introduced in the literature, which seem to work better than those obtained from using a sample estimator for the expected return vector, e.g., Bayesian approach (Jorion (1986); MacKinlay and Pástor (2000)), equilibrium expected returns approach (Black and Litterman (1990)), robust portfolio approach which incorporates estimation errors (Ceria and Stubbs (2006)), etc. In addition, researchers trying to explain the cross-sectional asset returns also have uncovered a considerable amount of pricing anomalies (see Harvey et al. (2016) for a review), and such anomalies, in turn, have enabled investors to find reasonably good proxies for the expected returns on the assets in the portfolio.

Unlike expected returns covariance matrices are usually estimated from the price history, although the quality of the sample covariance matrix can be seriously compromised as the number of assets in the portfolio becomes increasingly large relative to the sample size. A few well-known approaches to address this issue have focussed on an adjustment of the sample covariance matrix, or more specifically, an adjustment of sample eigenvalues.[2] Well-known examples of this include a "shrinkage towards identity" estimator (Ledoit and Wolf (2004)), a more general linear shrinkage estimator (Bodnar et al. (2014)), a "nonlinear shrinkage" estimator (Ledoit and Wolf (2017)), and a spectral cut-off method (Carrasco and Noumon (2011)). All of these covariance estimators reconstructed from adjusted sample eigenvalues. When they are used to build an optimized port-

---

[1] The minimum-variance portfolios implicitly exploit risk-based pricing anomalies (Scherer (2010)) and have been noted for surprisingly high returns and low realized volatilities in both the US market and the global market (Jagannathan and Ma (2003); Clarke et al. (2006)).

[2] There are other methods that have been proposed to mitigate the adverse effect of estimation errors in the covariance matrix estimator, which, for instance, focus on constraining portfolio weights (Jagannathan and Ma (2003); DeMiguel et al. (2009a); Fan et al. (2012)), imposing factor structure (Fan et al. (2008, 2011, 2013)), shrinking to a target portfolio (Bodnar et al. (2018)), etc.

folio, they have been shown empirically to be able to boost the portfolio's out-of-sample Sharpe ratio. However it is important to emphasize that not all of these improved estimators were designed originally to achieve this goal. Notably, Ledoit and Wolf (2004) and Bodnar et al. (2014) minimize the expected Frobenius loss; Carrasco and Noumon (2011) use the expected utility of a mean-variance investor as the objective function. Since the improvement in these objective functions does not necessarily lead to a higher out-of-sample Sharpe ratio, it is not clear at first glance why these methods can lead to an enhanced portfolio performance.

To take stock of the relatively recent and notable contribution to research on this issue in the literature, we investigate how to judiciously adjust sample eigenvalues for the purpose of improving the out-of-sample portfolio Sharpe ratio when the underlying theoretical model for asset returns is a high-dimensional factor model. The latter has become increasingly popular in the literature (Fama and French (1993); Bai and Ng (2002); Fan et al. (2013)). Later it will be seen that how our findings can be used to provide a novel synthesis and perpsective on some of these existing methods. Specifically our focus of analysis in this paper centers on mitigating estimation errors and improving portfolio performance in at least three important aspects.

First, while in the literature most improved covariance matrix or portfolio weight estimators optimize a particular objective function within a certain class of candidate covariance/weight estimators[3], we focus our analyses in this paper on certain adjustment patterns which, when applied to the sample eigenvalues, could lead to a concrete marginal improvement[4] in terms of the out-of-sample portfolio Sharpe ratio as both the sample size and the portfolio size are sufficiently large. This pattern, if it is suitably identified, can fruitfully shed an important light on the key factors for improving a portfolio's out-of-sample performance. More importantly each realization of the sample covariance matrix benefits, in terms of the out-of-sample portfolio Sharpe ratio calculated in the two-step approach, from such an adjustment. It is important to stress that the motivation for our particular focus on the marginal effects of eigenvalues adjustment in this paper largely stems from the observation that most improved estimators typically deviate only mildly from the sample covariance matrix.

Second, our simultaneous selection of the out-of-sample Sharpe ratio as a target and the covari-

---

[3]Usually, the problem can be reduced to first deriving the optimal value of the decision variable and then looking for its bona fide estimator (Ledoit and Wolf (2004)) or convergence limit (Bodnar et al. (2018)). In Ledoit and Wolf (2017), the authors search among the candidates for the one that optimizes the convergence limit of the objective function.

[4]The marginal effect refers to the effect of an infinitesimal adjustment (according to some pattern) of the sample eigenvalues on the out-of-sample portfolio Sharpe ratio. This effect is formally quantified in Section 2.

ance estimators reconstructed based on adjusted sample eigenvalues as candidates differentiates this paper from its peers in the literature which adopt other objectives (Ledoit and Wolf (2004); Carrasco and Noumon (2011); Bodnar et al. (2014)) and those which consider other candidate estimators (Ledoit and Wolf (2004); Kan and Zhou (2007); Bodnar et al. (2018)). To the best of our knowledge Ledoit and Wolf (2017) are the only authors who optimize (convergence limit of) the out-of-sample Sharpe ratio over the same class of estimators that we consider in this paper. The main difference between our paper and theirs will be elaborated in the next paragraph.

Third, in this paper we assume in our theoretical analysis that the asset return generating process follows a high-dimensional factor model, where the returns of an increasing number of assets are governed by a fixed number of common factors. The high-dimensional factor model assumption implies a spiked structure in the population covariance matrix: as the number of assets increases to infinity, the first few eigenvalues increase at the same rate and the remaining ones are bounded. The spiked structure with rapidly-growing spiked eigenvalues renders many of the results developed within the framework of the Random Matrix Theory (RMT) practically impotent, in that the spectral convergence results for this type of population covariance matrix have not yet been fully worked out to this date. This technical hurdle prevents us from deriving the optimal eigenvalues adjustment under a factor model. The few papers that borrow results from the RMT to improve portfolios (Ledoit and Wolf (2017), Engle et al. (2017), Bodnar et al. (2018)[5]) assume the population covariance matrix to have bounded eigenvalues, which contradicts conventional factor model assumptions. Expoiting the important results obtained by Shen et al. (2016) on the consistency of Principal Component Analysis (PCA) under a variety of assumptions on the population covariance matrix and different asymptotics, we are able to find a suitable adjustment pattern that ensures a positive marginal effect under certain high-dimensional asymptotics.

According to the main theoretical results of this paper, if the population covariance matrix admits a high-dimensional K-factor model, adjusting one of the first K sample eigenvalues has a diminishing effect under the high-dimensional asymptotics that the number of assets p and the sample size n both go to infinity with $n = O(p^{1+c})$ for any $c > 0$. In addition either shrinking a large but non-spiked (excluding the first K) sample eigenvalues or lifting a small one also has a positive effect on the out-of-sample Sharpe ratio asymptotically. We also analyze a simultaneous

---

[5]It is important to point out that although the authors did not explicitly assume the population covariance matrix to have bounded eigenvalues, in their major technical reference (Rubio and Mestre (2011)), the population covariance matrix is indeed assumed to have a bounded spectral norm.

adjustment of multiple eigenvalues. Let $\{\widehat{\lambda}_i\}_{i=1}^{p}$ denote decreasingly sorted sample eigenvalues. Then, for any value of k and any $a < 1$, simultaneously amplifying the smallest sample eigenvalues $\{\widehat{\lambda}_i : i \geq k\}$ to $\{\widehat{\lambda}_i + \lambda\widehat{\lambda}_i^{a} : i \geq k\}$ for some small $\lambda > 0$ leads to an improvement in the out-of-sample Sharpe ratio under the high-dimensional asymptotics. In addition, if the underlying model is a single-factor model and the factor pricing relation is assumed to hold true, i.e., if both the expected return vector and the covariance matrix are driven by the single-factor model (see a discussion in MacKinlay and Pástor (2000)), the aforementioned way of adjusting eigenvalues also always leads to an improvement in the out-of-sample portfolio Sharpe ratio, regardless of the values of p and n.

Our results shed an important light on the reason for the effectiveness of the shrinkage-towards-identity method ($k = 1$, $a = 0$) and the spectral cut-off method ($k \geq \min\{k : \widehat{\lambda}_k < 1\}$, $a = -\infty$) in the construction of real-world portfolios in practice. A useful implication of our results is that the key to improving the out-of-sample portfolio performance is to make the eigenvalues overall less dispersed (this will be formally defined later) after the adjustment. Although Ledoit and Wolf (2004) have made a similar statement, the reasoning underlying their argument is based on the fact that sample eigenvalues are more dispersed compared with their population counterparts and, thus, should be corrected to reduce the expected Frobenius loss. By showing that the shrinkage towards identity type estimator has a concrete marginal effect of improving the out-of-sample Sharpe ratio under factor model assumptions, we provide this well-known method with a fresh new insight.

The remaining part of the paper is organized as follows. Section 2 provides an expression for the marginal effect of sample eigenvalues adjustment on the out-of-sample Sharpe ratio and discusses what type of an adjustment pattern would ensure a positive effect. Section 3 uses a simulation experiment to further illustrate the theoretical results established in Section 2. Section 4 provides synthesis on some of the recent research in portfolio analysis by highlighting an important connection of the main results in this paper with a few existing approaches recently introduced in the literature. Section 5 concludes the paper.

2 Main Results

Throughout the paper we use bold capital letters to denote matrices, bold lowercase letters to denote vectors, and plain lowercase letters to denote scalars. The term "return" on a given asset denotes the asset's return in excess of the riskless rate. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ denote a $p \times n$ matrix containing n independently and identically distributed observations on a system of $p(< n)$ asset returns with some mean vector, a positive definite covariance matrix $\mathbf{\Sigma}$, and a finite fourth moment. We denote the eigen-decomposition of the population covariance matrix by $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1 \geq \cdots \geq \lambda_p\}$ and $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p)$. Let $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$, where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$, denote the sample covariance matrix, whose eigen-decomposition is $\mathbf{S} = \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{U}}^T$. Similarly, we denote $\widehat{\mathbf{\Lambda}} = \mathrm{diag}\{\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_p\}$ and $\widehat{\mathbf{U}} = (\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2, \ldots, \widehat{\mathbf{u}}_p)$. Since we focus on occasions where n is greater than p in the theoretical analysis, we assume that $\mathbf{S}$ is invertible throughout the paper without loss of generality. In addition, $\|\mathbf{a}\|$ and $\|\mathbf{a}\|_\infty$ denote the Euclidean norm and the maximum norm of a vector $\mathbf{a}$ respectively. Lastly, $\mathbf{E}_k$ denotes a conforming diagonal matrix with 1 being its kth diagonal entry and 0 elsewhere; $\mathbf{E}_{k+}$ denotes a conforming diagonal matrix with 0 being its first $k - 1$ diagonal entries and 1 being its diagonal entries beyond (and including) the kth entry.

2.1 Problem Setup

Suppose that an investor adopts a two-step approach to construct a single-period maximum Sharpe ratio (MSR) portfolio of p risky assets. Further suppose that the investor believes that the vector of expected returns is an exogenously given vector $\boldsymbol{\mu}$, and estimates the covariance matrix from assets' price history. We consider such a case because it has been widely noticed that estimating expected returns from price history is a major source of estimation error (Jorion (1985); Jagan-nathan and Ma (2003); DeMiguel et al. (2009b)) and investors, as a result, resort to alternative approaches to acquire the expected returns (recall the first paragraph of Section 1). In particular many researchers have exploited asset pricing anomalies to find better proxies for the expected returns. Harvey et al. (2016) contains a review of hundreds of cross-sectional return patterns. Given the abundance of choice for $\boldsymbol{\mu}$, we leave it to the investor to make her/his own individual choice and simply take $\boldsymbol{\mu}$ as exogenously given for the purpose of our analysis in this paper.

When a sample covariance matrix $\mathbf{S}$ is adopted as the covariance estimator, the two-step ap-

proach based MSR portfolio can be solved from the following optimization program:

$$\widehat{\mathbf{w}}_{MSR} := \underset{\mathbf{w}}{\arg\max} \frac{\mathbf{w}^T \boldsymbol{\mu}}{\sqrt{\mathbf{w}^T \mathbf{S} \mathbf{w}}} \quad \text{s.t.} \quad \mathbf{w}^T \mathbf{1} = 1, \tag{1}$$

where $\mathbf{1}$ is a column vector of ones. Note that in the above Sharpe ratio maximization problem, the objective function represents an ex-ante Sharpe ratio, since $\boldsymbol{\mu}$ represents the investor's subjective belief on expected asset returns over the investment horizon. The solution to the above program, if it exists, is given by

$$\widehat{\mathbf{w}}_{MSR} = \frac{\mathbf{S}^{-1} \boldsymbol{\mu}}{\mathbf{1}^T \mathbf{S}^{-1} \boldsymbol{\mu}}, \tag{2}$$

and the out-of-sample Sharpe ratio is given by

$$SR = \frac{\widehat{\mathbf{w}}_{MSR}^T \boldsymbol{\mu}}{\sqrt{\widehat{\mathbf{w}}_{MSR}^T \boldsymbol{\Sigma} \widehat{\mathbf{w}}_{MSR}}} = \frac{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu}}}. \tag{3}$$

It has become widely accepted in this literature that the out-of-sample Sharpe ratio of the two-step approach based MSR portfolio could deviate substantially from its actual maximum value, especially when p is large relative to n, due to significant estimation errors in the sample covariance matrix. Based on this reason, we exploit the possibility of improving the out-of-sample Sharpe ratio by substituting $\mathbf{S}$ with a re-constructed covariance matrix estimator $\widetilde{\mathbf{S}}$ and assign a portfolio weight as $\widetilde{\mathbf{w}}_{MSR} = \frac{\widetilde{\mathbf{S}}^{-1} \boldsymbol{\mu}}{\mathbf{1}^T \widetilde{\mathbf{S}}^{-1} \boldsymbol{\mu}}$.

In this paper we focus on adjusting the eigenvalues of the sample covariance matrix, which means that we consider a covariance matrix estimator that takes the form $\widetilde{\mathbf{S}} = \widehat{\mathbf{U}} \widetilde{\boldsymbol{\Lambda}} \widehat{\mathbf{U}}^T$, with $\widetilde{\boldsymbol{\Lambda}}$ being a diagonal matrix. This is known as a class of "rotation-equivariant" estimators which was originally introduced by Stein (1986) and which has been considered as candidate covariance estimators by Ledoit and Wolf (2017). We set out our analyses in this paper by slightly pulling a number of sample eigenvalues away from their original levels and explore the marginal effect of such an adjustment on the out-of-sample Sharpe ratio. For this purpose we use a diagonal matrix $\mathbf{V} = \text{diag}\{v_1, v_2, \ldots, v_p\}$ and a scalar λ to parameterize the sample eigenvalues adjustment

as follows:

$$\mathbf{S}_{\mathbf{V},\lambda} := \widehat{\mathbf{U}} \begin{bmatrix} \widehat{\lambda}_1 - v_1\lambda & & & & \\ & \ddots & & & \\ & & \widehat{\lambda}_k - v_k\lambda & & \\ & & & \ddots & \\ & & & & \widehat{\lambda}_p - v_p\lambda \end{bmatrix} \widehat{\mathbf{U}}^T.$$

The parameterization above is quite general in that when $\mathbf{V}$ takes different forms, it could reduce to a number of ways in which to adjust sample eigenvalues. For instance, when $\mathbf{V}$ is a scalar multiple of an identity matrix, the adjustment is similar to (but not the same as[6]) the construction of the "shrinkage towards identity" method proposed by Ledoit and Wolf (2004). The shrinkage towards identity estimator is usually referred to as a "linear" shrinkage in the sense that if we plot the eigenvalues after adjustment versus their original counterparts, all points will lie on a straight line. This notion is in contrast to the recent development of a "nonlinear" shrinkage estimator (Ledoit and Wolf (2017)) in which each eigenvalue is adjusted differently. It is important also to point out that our parameterization allows for a nonlinear shrinkage since it does not restrict the v_k's to be the same. In addition, this setup also accommodates the cases in which only a subset of the sample eigenvalues are adjusted.

The MSR portfolio constructed based on $\mathbf{S}_{\mathbf{V},\lambda}$ has a weight vector given by:

$$\widehat{\mathbf{w}}_{MSR}(\mathbf{V},\lambda) = \frac{\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}. \tag{4}$$

It is worth pointing out that the $v_k\lambda$'s in the above expression for $\mathbf{S}_{\mathbf{V},\lambda}$ is not necessarily positive. This implies a shrinkage on the kth eigenvalue for a positive $v_k\lambda$ and an amplification for a negative $v_k\lambda$. With the weight vector $\widehat{\mathbf{w}}_{MSR}(\mathbf{V},\lambda)$ given in equation (4), the out-of-sample Sharpe ratio could also be expressed as a function of $\mathbf{V}$ and λ:

$$SR(\mathbf{V},\lambda) := \frac{\widehat{\mathbf{w}}_{MSR}^T(\mathbf{V},\lambda)\boldsymbol{\mu}}{\sqrt{\widehat{\mathbf{w}}_{MSR}^T(\mathbf{V},\lambda)\boldsymbol{\Sigma}\widehat{\mathbf{w}}_{MSR}(\mathbf{V},\lambda)}} = \frac{\boldsymbol{\mu}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\Sigma}\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}}. \tag{5}$$

Therefore our goal of studying the marginal effect of adjusting sample eigenvalues on the out-of-sample Sharpe ratio can be reduced to one of studying the property of the derivative $SR'_{\mathbf{V}}(0) := \frac{\partial SR(\mathbf{V},\lambda)}{\partial \lambda}\big|_{\lambda=0}$.

---

[6]In a shrinkage towards identity estimator, the trace of the shrunk covariance matrix remains the same as the sample covariance matrix, but it is obviously not the case if $\mathbf{V}$ is a scalar multiple of an identity matrix.

2.2   Marginal Effect of Eigenvalues Adjustment

As the first step, we derive a simplified expression for $SR'_{\mathbf{V}}(0)$ as given in the following theorem.

Theorem 2.1. The marginal effect of an eigenvalues adjustment specified by a diagonal matrix $\mathbf{V}$ on the out-of-sample Sharpe ratio admits the following expression:

$$SR'_{\mathbf{V}}(0) = -\frac{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}{(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu})^{\frac{3}{2}}}\left(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu} - \boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu}\frac{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}\right). \quad (6)$$

Proof. See the Appendix.   □

Obviously $SR'_{\mathbf{V}}(0)$ is a random variable in that it depends on the sample covariance matrix. We cannot observe the realization of the random variable purely based on a sample, because the random variable also relies on the unobservable population covariance matrix $\boldsymbol{\Sigma}$. Ideally it will be intriguing to find a set of matrices, denoted by $\mathscr{M}$, such that for any $\mathbf{V} \in \mathscr{M}$, the inequality $SR'_{\mathbf{V}}(0) > 0$ holds for any value of p and n, or less preferably, holds asymptotically in certain limiting scenarios. If such a set can be identified we can improve an MSR portfolio's out-of-sample Sharpe ratio by replacing $\mathbf{S}$ with $\mathbf{S} - \lambda\mathbf{V}$ for a small positive number λ and some $\mathbf{V} \in \mathscr{M}$.

Given the complexity of the expression for $SR'_{\mathbf{V}}(0)$ it is in general quite challenging to determine its sign without imposing any explicit structural assumption on $\boldsymbol{\Sigma}$. Therefore, in the next two sections we look for the set of matrices $\mathscr{M}$ under two distinctive sets of assumptions on $\boldsymbol{\Sigma}$.

2.3   High-Dimensional Factor Model

In this section we motivate our theoretical analyses with the assumption that the asset returns are generated by a high-dimensional K-factor model, which is a well-known and extensively used model for cross-sectional financial returns (Fama and French (1992); Bai and Ng (2002); Fan et al. (2008)). To be more specific we assume that the systematic portion of the asset prices movement is driven by K common factors and the idiosyncratic returns are mutually uncorrelated[7], so that the population covariance matrix has a "low rank + diagonal" structure. However we do not assume that the factor pricing relation to hold[8], nor do we assume the factor returns and, thus, the asset

---

[7]Uncorrelatedness of residual returns is an assumption of the strict factor model. Recently researchers have been considering a more practical approximate factor model (Fan et al. (2011)) which allows for a sparse residual covariance matrix. But in this paper we focus on the strict factor model for viability reasons.

[8]If the factor pricing relation holds, $\boldsymbol{\mu}$ should be estimated according to the factor model. We refer readers to MacKinlay and Pástor (2000) for a discussion.

returns to have a mean vector independent of time. Rather we still view $\boldsymbol{\mu}$ as exogenously given. Our goal is to find a collection of $\mathbf{V}$ matrices which make $SR'_{\mathbf{V}}(0)$ asymptotically positive as both p and n go to infinity.

It is worth mentioning that we are not the first to consider manipulation of sample eigenvalues for the purpose of improving the out-of-sample Sharpe ratio under the high-dimensional asymptotics. For instance, Ledoit and Wolf (2017), after imposing a set of technical assumptions, derived a convergence limit for the out-of-sample Sharpe ratio as both the number of assets p and the sample size n go to infinity at the same rate. In the next step they look for an optimal shrinkage on each sample eigenvalue to maximize the deterministic convergence limit of the out-of-sample Sharpe ratio. However one of the technical assumptions made in Ledoit and Wolf (2017) (Assumption 2) indicates that as p increases, the eigenvalues of the population covariance matrix are contained in a compact set, or to put it simply, the largest population eigenvalue does not increase at the rate $O(p)$. This assumption indicates that the authors work under a framework seemingly at odds with the standard high-dimensional K-factor model framework. In the latter the largest K population eigenvalues increase at the rate $O(p)$. This discrepancy in assumptions, as well as the popularity of the factor model structure, makes it necessary to examine the limiting behavior of the marginal effect random variable under the high-dimensional asymptotics as well as the factor model assumptions.

Under a K-factor model the cross-sectional returns of the p assets are assumed to be driven by K common factors:

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \ldots, n, \tag{7}$$

where $\mathbf{y}_t$ is a $p \times 1$ vector of asset returns at time t, $\mathbf{B}$ is a $p \times K$ deterministic matrix of factor loadings, $\mathbf{f}_t$ is a $K \times 1$ vector of factor returns at time t, and $\boldsymbol{\varepsilon}_t$ is a $p \times 1$ noise vector independent of $\mathbf{f}_t$ with a zero mean, a covariance matrix $\boldsymbol{\Sigma}_\varepsilon$, and a finite fourth moment. Our subsequent analysis in this section is based on the following high-dimensional K-factor model assumptions.

Assumption 2.1. The autocovariance functions for both $\{\mathbf{f}_t\}_{t=1}^n$ and $\{\boldsymbol{\varepsilon}_t\}_{t=1}^n$ are independent of t.

This assumption ensures that $\boldsymbol{\Sigma} = \mathbf{B}\text{cov}(\mathbf{f}_t)\mathbf{B}^T + \boldsymbol{\Sigma}_\varepsilon$ and that $\boldsymbol{\Sigma}$ is constant across time.

Assumption 2.2. The $K \times K$ covariance matrix $\text{cov}(\mathbf{f}_t)$ is of full rank.

This assumption implies that none of the common factors can be written into a linear combination of the remaining ones. As a result the rank of the matrix $\mathbf{B}\text{cov}(\mathbf{f}_t)\mathbf{B}^T$ is K as long as the

rank of $\mathbf{B}$ is K.

Assumption 2.3. The eigenvalues of $p^{-1}\mathbf{B}^T\mathbf{B}$ are bounded away from zero for all sufficiently large p.

Fan et al. (2013) provide an insightful explanation for this assumption. The authors point out that this assumption easily holds when the factors are pervasive in the sense that a non-negligible fraction of factor loadings should be non-vanishing. This assumption, together with Assumption 2.2, ensures that the K non-zero eigenvalues of $\mathbf{B}\text{cov}(\mathbf{f}_t)\mathbf{B}^T$ diverge at rate $O(p)$.

Assumption 2.4. The residual covariance matrix $\mathbf{\Sigma}_\varepsilon$ is a constant multiple of the identity matrix, i.e. $\mathbf{\Sigma}_\varepsilon = \sigma^2\mathbf{I}$.

Admittedly this assumption serves purely as a technical purpose without which we would be unable to verify the Theorem 2.2 below. Notwithstanding its restrictiveness, this assumption captures a well-known stylized fact in finance that assets' idiosyncratic variances are of a similar scale. The same assumption is also used by MacKinlay and Pástor (2000) in a factor model.

Assumption 2.5. As $p \to \infty$, $\|\boldsymbol{\mu}\| = O(p^{1/2})$ and $\|\mathbf{U}_F^T\boldsymbol{\mu}\|_\infty = o(p^{1/2})$, where $\mathbf{U}_F = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K)$.

The first half of the assumption stipulates that a non-vanishing proportion of the assets has a non-zero expected return. This assumption is relatively straightforward and not unduly restrictive. The second half of the assumption restrains $\boldsymbol{\mu}$ from having an excessive loading on the first K eigenvectors. Typically there is no a priori reason to believe that $\boldsymbol{\mu}$ has an excessive loading on any specific eigenvector, since $\boldsymbol{\mu}$ can be obtained from multiple resources. An example of violation of this assumption is when $\boldsymbol{\mu}$ is a linear combination of the first K eigenvectors, in which case $\|\mathbf{U}_F^T\boldsymbol{\mu}\|_\infty$ has rate $O(p^{1/2})$. Guo et al. (2019) point out that in such a scenario, even directly using the sample covariance matrix still leads to a consistent portfolio weight estimator. In this paper we focus on the more common cases where manipulation of sample eigenvalues potentially leads to some improvement. So we exclude such extreme scenarios from consideration by imposing Assumption 2.5.

Theorem 2.2. Under Assumptions 2.1 - 2.5 and as p, n go to infinity with the rate of n being $n = O(p^{1+c})$ for some $c > 0$, the following results hold:

(a) For $1 \leq k \leq K$,

$$SR'_{\mathbf{E}_k}(0) \xrightarrow{a.s.} 0. \tag{8}$$

In addition there also exists an integer $K^* \in \{K+2, \ldots, p-1\}$ such that for each $k \in \{K+1, \ldots, p\}$, there exists a sequence of almost surely positive random variables denoted by $\{X_k^{(p)}\}_{p=2,3,\ldots}$, such that

$$
\begin{cases}
\frac{SR'_{\mathbf{E}_k}(0)}{X_k^{(p)}} \xrightarrow{a.s.} 1, & K+1 \leq k < K^* \\
\frac{SR'_{-\mathbf{E}_k}(0)}{X_k^{(p)}} \xrightarrow{a.s.} 1, & K^* < k \leq p
\end{cases}
. \tag{9}
$$

(b) For any $a < 1$ and any $k \in \{1, \ldots, p\}$ there exists a sequence of almost surely positive random variables denoted by $\{Y_k^{(p)}\}_{p=2,3,\ldots}$, such that

$$
\frac{SR'_{-\widehat{\mathbf{\Lambda}}^a \mathbf{E}_{k+}}(0)}{Y_k^{(p)}} \xrightarrow{a.s.} 1, \tag{10}
$$

where $\widehat{\mathbf{\Lambda}}^a$ denotes the diagonal matrix $\text{diag}\{\widehat{\lambda}_1^a, \widehat{\lambda}_2^a, \ldots, \widehat{\lambda}_p^a\}$.

Proof. See the Appendix. □

Remark 2.1. Theorem 2.2 identifies a few forms of the $\mathbf{V}$ matrix that can lead to a marginal improvement in the out-of-sample Sharpe ratio under the high-dimensional asymptotic that both p, n go to infinity with the rate of n being $n = O(p^{1+c})$ for some $c > 0$. It is important to stress that in the literature, the "high-dimensional asymptotics" term is typically used to refer to the cases when both p and n go to infinity at the same rate. The asymptotics we base our analysis on in this paper is only of a slightly lower-dimensional nature, since c can be arbitrarily small. For this reason, although the ratio p/n eventually converges to 0, we still view these asymptotics, perhaps with a slight abuse of terminology, as a high-dimensional case.

Remark 2.2. Many results in this paper hold "almost surely". Unless otherwise stated, "some random variable is positive almost surely" can be understood as that it takes the value of 0 only when $\mathbf{S} = \mathbf{\Sigma}$, an event that happens with probability 0.

Part (a) of Theorem 2.2 discusses the effect of an incremental adjustment of an individual sample eigenvalue. The result indicates that when p and n become large enough, the marginal effect of modifying one of the largest K sample eigenvalues starts to diminish (see eq. (8)). The finding of the fading marginal effect of adjusting a spiked eigenvalue is perfectly consistent with the intuition, because as Shen et al. (2016) showed correctly, under the high-dimensional K-factor model

assumptions, the largest K eigenvalues can be consistently estimated by their sample counterparts in the sense that $\frac{\widehat{\lambda}_i}{\lambda_i} \xrightarrow{a.s.} 1$, $i = 1, 2, \ldots, K$.

In addition, according to part (a) of Theorem 2.2 (see eq. (9)), both a mild shrinkage on a large but non-spiked (excluding the first K) sample eigenvalue and a mild amplification on a small one help to improve the out-of-sample Sharpe ratio. This result is in line with the intuition that the sample eigenvalues should be pushed back towards their grand mean for the purpose of improving the MSR portfolio, because sample eigenvalues are more dispersed than their population counterparts (Marčenko and Pastur (1967)). Before stating this theorem formally, we should be cautious about rushing to form such an intuition. The reason is that even though pushing sample eigenvalues back to their grand mean can help alleviating the estimation errors measured by the expected Frobenius loss $E(\|\mathbf{S} - \mathbf{\Sigma}\|_F)$ (Ledoit and Wolf (2004)), the out-of-sample Sharpe ratio will not be necessarily affected in the same direction. More discussion in this regard will be given in Section 4.

Part (b) of Theorem 2.2 focuses on a joint manipulation of multiple sample eigenvalues. The parameter a specifies the relative intensity of the adjustment on different sample eigenvalues. The result indicates that a joint amplification on a collection of tail sample eigenvalues has the marginal effect of improving the out-of-sample Sharpe ratio, as long as $a < 1$. It is worth highlighting that the eigenvalues to be adjusted must form a complete "tail" - no matter where we start the adjustment, we should adjust all of the eigenvalues beyond (smaller than) the starting point. The proposed range for a implies that the eigenvalues should become overall less dispersed after the adjustment, in the sense that the inequality $\frac{\widetilde{\lambda}_i}{\widetilde{\lambda}_j} \leq \frac{\widehat{\lambda}_i}{\widehat{\lambda}_j}$ holds for any $i < j$, where $\widetilde{\lambda}_i$ is the ith largest eigenvalue after the adjustment. When $k = 1$ and $a = 1$, the adjustment is equivalent to replacing $\mathbf{S}$ with its scalar multiple. Such an adjustment will not lead to any change in the out-of-sample Sharpe ratio. This is why a must be strictly less than 1 to ensure the positiveness of $SR'_{-\widehat{\mathbf{\Lambda}}^a \mathbf{E}_{k+}}(0)$. When $k = 1$ and $a = 0$, the adjustment is consistent with the notion of a "linear" shrinkage, since each eigenvalue is lifted by the same amount. Other values of a correspond to a "nonlinear" shrinkage. Instead of using p parameters to parameterize a nonlinear shrinkage, as in Ledoit and Wolf (2017), we use 2 by requiring that the shrinkage intensity matrix $\mathbf{V}$ is a power of the sample eigenvalue matrix times an "indicator matrix" specifying the starting point of the manipulation. This parameterization is rich in its implication despite of its parsimony.

The out-of-sample Sharpe ratio is scale-invariant in the covariance matrix estimator. Although

in a shrinkage towards identity estimator the large eigenvalues are shrunk and the small ones are lifted, the estimator has an equivalent (in the sense of leading to the same out-of-sample Sharpe ratio) rotation-equivariant correspondence each of whose eigenvalues is amplified. Therefore our results provide support to shrinkage estimators with trace at different levels. The connection between our results and the shrinkage estimators will be discussed in detail in Section 4.

2.4 Single-Factor Model

In this section we assess the marginal effect of adjusting sample eigenvalue(s) under a single-factor model, which assumes that a single unobservable factor drives the price movement of all assets. MacKinlay and Pástor (2000) showed by using an "optimal orthogonal portfolio" (MacKinlay (1995)) argument that if the exact single-factor pricing relation holds and the uncorrelated residual returns have equal variance σ^2, the true expected return (still denoted by $\boldsymbol{\mu}$) and the covariance matrix has the following relationship:

$$\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^T \frac{1}{s_h^2} + \sigma^2 \mathbf{I}, \tag{11}$$

where s_h denotes $\frac{\mu_h}{\sigma_h}$, the Sharpe ratio of the factor portfolio h.

In the following Theorem 2.3 we will show that if the exact single-factor pricing relation is satisfied and if the subjective view on expected returns coincides with the true expected returns vector, we can find a set of $\mathbf{V}$ matrices that make $SR'_{\mathbf{V}}(0)$ positive even when both p and n are small. It is important to clarify that Theorem 2.3 could not serve a practical purpose because of the following paradox: for an arbitrary $\boldsymbol{\mu}$ supplied by some "alpha model", if the structural assumption in eq. (11) is not satisfied, the subsequent results in the theorem do not necessarily hold; otherwise, we would immediately obtain the population covariance matrix and the issue of estimation errors will not be an issue any longer. However Theorem 2.3 has a strong theoretical implication in that at least, it identifies a set of $\mathbf{V}$ matrices that work when p and n can be any number under a reasonable economic model.

Theorem 2.3. If $\boldsymbol{\Sigma}$ can be expressed by the exogenously given $\boldsymbol{\mu}$ as in eq. (11), then the following results hold for any value of p, n, s_h, and σ^2:

(a) There exists an integer $K \in \{2, 3, \ldots, p-1\}$ such that with probability 1, $SR'_{\mathbf{E}_k}(0) > 0$ for all $k < K$ and $SR'_{-\mathbf{E}_k}(0) > 0$ for all $k > K$.

14

(b) For any $a < 1$ and $k \in \{1, 2 \ldots, p\}$, $SR'_{-\widehat{\boldsymbol{\Lambda}}^a \mathbf{E}_{k+}}(0) > 0$ with probability 1.

Remark 2.3. Since the set of $\mathbf{V}$ matrices given in Theorems 2.2 and 2.3 is the same, the latter theorem may seem to be a finite sample analogous to the former, while there is a major difference between these two sets of results: the structural assumption on $\boldsymbol{\Sigma}$ in eq. (11) implies that $\boldsymbol{\mu}$ only has a non-zero loading on the dominant eigenvector[9] of $\boldsymbol{\Sigma}$, but this specific scenario is excluded from the previous analyses by Assumption 2.5. Therefore Theorem 2.3 can be usefully viewed as a complement to Theorem 2.2 that impresses us with its insightful finite sample results.

Part (a) of Theorem 2.3 focuses on adjusting an individual sample eigenvalue. Both an incremental shrinkage on a large sample eigenvalue and an incremental amplification on a small sample eigenvalue lead to an increase in the out-of-sample Sharpe ratio. In addition there also exists a cutting point between such large eigenvalues and small eigenvalues. This result also helps to partly justify the use of the shrinkage towards identity method. Part (b) of Theorem 2.3 focuses on simultaneously adjusting a few eigenvalues. The results indicate that if we apply a mild amplification on a collection of tail eigenvalues, so that the eigenvalues become overall less dispersed, this will result in improvement in the out-of-sample Sharpe ratio of the protfolio.

3 Simulation Study

In this section we perform a simulation study to further illustrate the theoretical findings reported in Section 2.3. In particular we demonstrate, via simulation, the property of the sample-dependent random variable $SR'_{\mathbf{V}}(0)$ for some $\mathbf{V}$ matrices, which we have discussed in the previous section.

3.1 Simulation Setup

In each experiment we pre-specify a $p \times 1$ (subjective) expected returns vector $\boldsymbol{\mu}$ and a $p \times p$ population covariance matrix $\boldsymbol{\Sigma}$ and repeat the following procedure 100 times:

(a) Generate $n(p) = [p^{1.5}]$ random vectors from a multivariate normal distribution with an expected return vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. Calculate the sample covariance matrix $\mathbf{S}$ based on the n simulated vectors.

(b) Calculate $SR'_{\mathbf{E}_k}(0)$ and $SR'_{-\mathbf{E}_{k+}}(0)$ for $k = 1, 2, \ldots, p$ based on $\mathbf{S}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\mu}$.

---

[9]The dominant eigenvector refers to the eigenvector that corresponds to the largest eigenvalue.

Once we have collected the 100 realizations of the $2p$ marginal effect random variables, we can calculate the average value of each, denoted by $\overline{SR'_{\mathbf{E}_k}(0)}$ and $\overline{SR'_{-\mathbf{E}_{k+}}(0)}$ respectively, $k = 1, 2, \ldots, p$. Then we plot $\overline{SR'_{\mathbf{E}_k}(0)}$ as a function of k to demonstrate the marginal effect of shrinking the kth sample eigenvalue. Likewise we also plot $\overline{SR'_{-\mathbf{E}_{k+}}(0)}$ as a function of k to illustrate the marginal effect of simultaneously amplifying the smallest $p - k$ sample eigenvalues.

In this simulation study we consider 8 different combinations of p, $\boldsymbol{\Sigma}$, and $\boldsymbol{\mu}$ when specifying the true parameters. To be more specific we provide two choices for each parameter:

p: The value of p is set in this simulation exercise to either 100 or 500 in consideration of the required computational time. The idea is that when $p = 500$, the results that we obtain should better reflect the asymptotic results stated in Theorem 2.2. Note that we do not treat the sample size n as another parameter but simply let it be a function of p to be consistent with the asymptotics we are working with.

$\boldsymbol{\Sigma}$: The population covariance matrix $\boldsymbol{\Sigma}$ either strictly conforms to Assumptions 2.1 - 2.4[10] or violates Assumption 2.4 but satisfies the remaining ones. Recall that Assumption 2.4 requires all of the non-spiked population eigenvalues to be equal to each other. This assumption is indispensable for the proof of our theoretical results but is hard to be satisfied in reality. Thus we resort to the simulation study to explore to what extent the numerical results will be affected if we allow for distinct small population eigenvalues. However it is worth pointing out that either choice for $\boldsymbol{\Sigma}$ admits the K-factor model. In this study we fix the value of K to be 3.

$\boldsymbol{\mu}$: The expected return vector $\boldsymbol{\mu}$ is either an "arbitrary" one in the sense that it has a non-zero loading on each of the eigenvectors of $\boldsymbol{\Sigma}$ or a "low-rank" one which can be expressed as a linear combination of the first K eigenvectors of $\boldsymbol{\Sigma}$. To obtain an "arbitrary" $\boldsymbol{\mu}$, we simulate a $p \times 1$ vector of independent components generated from the distribution $N(0.05, 0.05^2)$. When a "low-rank" $\boldsymbol{\mu}$ is desired, we project the simulated $\boldsymbol{\mu}$ onto the subspace spanned by the first K eigenvectors of $\boldsymbol{\Sigma}$ and re-scale the projection so that it remains the same length as the "arbitrary" $\boldsymbol{\mu}$. The purpose of such a design is to highlight the finding that the marginal effect of manipulating one or more sample eigenvalues not only depends on $\boldsymbol{\Sigma}$ but also on $\boldsymbol{\mu}$.

---

[10]Note that Assumption 2.5 is not about the population covariance matrix.

Figure 1: $\overline{SR'_{\mathbf{E}_k}(0)}$ vs. k: marginal effect of shrinking the kth sample eigenvalue
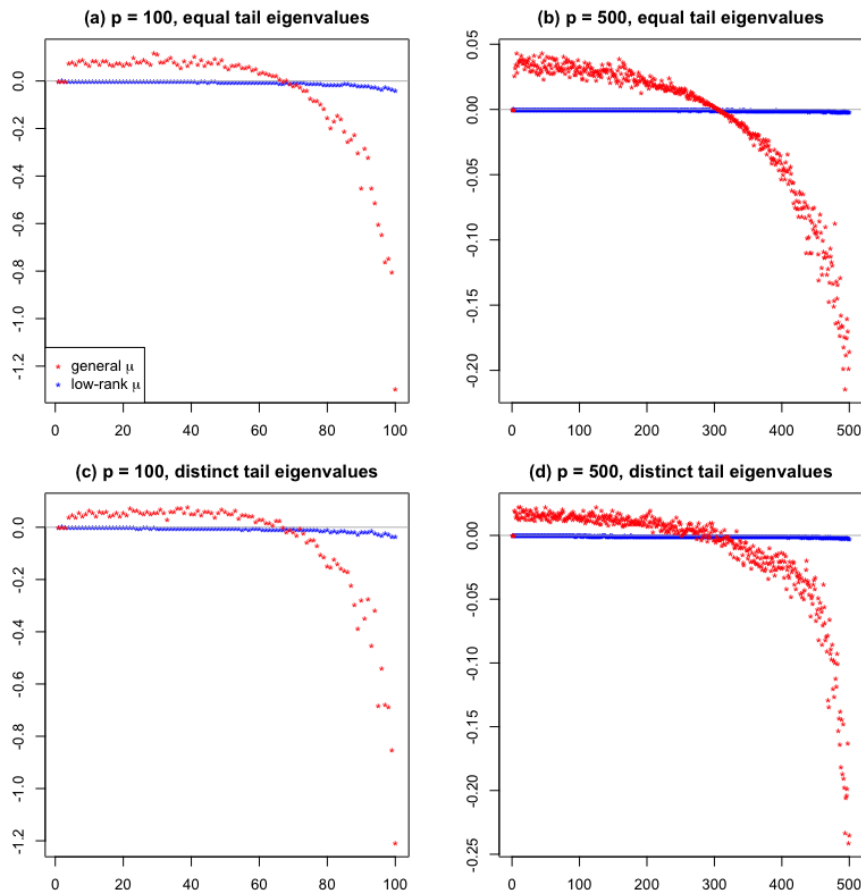
In each panel, the x-axis measures k and the y-axis measures $SR'_{\mathbf{E}_k}(0)$. In panels (a) and (c), $p = 100$ and $n = 1000$; in panels (b) and (d), $p = 500$ and $n = 11180$. In panels (a) and (b), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones equal to 0.05; in panels (c) and (d), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$. In each panel the red dots correspond to the case where $\boldsymbol{\mu}$ can have loading on all eigenvectors; the blue dots correspond to the case where $\boldsymbol{\mu}$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

3.2 Shrinkage on Individual Eigenvalue

In this section we show the simulation results on the marginal effect of shrinking a single sample eigenvalue. Specifically we illustrate how $\overline{SR'_{\mathbf{E}_k}(0)}$ varies with k in Figure 1.

According to Figure 1, when Assumptions 2.1 - 2.5 are satisfied (see the red dots in panels (a) and (b)), the marginal effect of shrinking one of the first K sample eigenvalues is almost 0, shrinking a large eigenvalue beyond the Kth leads to a marginal improvement on the out-of-sample Sharpe ratio, and shrinking a small eigenvalue leads to a deterioration of it. The last observation is equivalent to the statement that amplifying a small eigenvalue has a positive marginal effect on the Sharpe ratio. Moreover the magnitude (in the sense of absolute value) of the marginal effect of adjusting one of the smallest eigenvalues is quite large. The population covariance matrix $\boldsymbol{\Sigma}$ used

in panels (c) and (d) has distinct tail eigenvalues, however, we observe a very similar pattern (in the red dots) to that observed in the two panels at the top.

The blue dots come from the setup where $\boldsymbol{\mu}$ is from the 3-dimensional space span$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. As mentioned in Guo et al. (2019), in such a scenario, the sample-based MSR portfolio is a consistent estimator of the true MSR portfolio, and it, thus, becomes less necessary to seek improvement via adjustment of eigenvalues. What we observe from the blue dots in Figure 1 supports this argument - the magnitude of the marginal effect of eigenvalues adjustment is negligible compared with the case where $\boldsymbol{\mu}$ is an "arbitrary" vector.

Table 1: Proportion of positive $SR'_{\mathbf{E}_k}(0)$ among the 100 replications

| k	K+1	K+2	K+6	p-5	p-1	p
(a) $p = 100$, equal tail	1.00	0.97	0.96	0.02	0.02	0.00
(b) $p = 500$, equal tail	0.98	0.99	1.00	0.03	0.00	0.00
(c) $p = 100$, distinct tail	0.89	0.90	0.88	0.12	0.15	0.17
(d) $p = 500$, distinct tail	0.88	0.82	0.84	0.16	0.15	0.15

In this table, we only report results for the cases where $\boldsymbol{\mu}$ is an "arbitrary" vector and can have loading on all eigenvectors. In (a) and (b), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones equal 0.05; in (c) and (d), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$.

Figure 1 only reflects the average value of the marginal effect across the 100 replications. We are also interested in the distribution of the adjustment effect random variables; especially we are interested in what proportion among the 100 realizations of $SR'_{\mathbf{E}_k}(0)$ are positive. Table 1 reports such information. According to the first two rows, when the last $p - K$ population eigenvalues are equal, shrinking a large but non-spiked eigenvalue or amplifying a small one almost always leads to a marginal improvement in the out-of-sample Sharpe ratio. When the 'flat tail' assumption is removed, there is still a high chance (around 85% according to Table 1) that such manipulation leads to further improvement.

3.3 Amplification on Tail Eigenvalues

In this section we present the simulation results on the marginal effect of amplifying a collection of tail eigenvalues. In particular we illustrate how $\overline{SR'_{-\mathbf{E}_{k+}}(0)}$ varies with k in Figure 2.

Consistent with the theoretical results of Theorem 2.2, in panels (a) and (b) of Figure 2, the average marginal effect is positive for any k, regardless of the $\boldsymbol{\mu}$ used. In other words, applying a linear amplification on a few tail sample eigenvalues always has a positive marginal effect on the out-of-sample Sharpe ratio, no matter how long the tail is. Comparing the red dots with the

Figure 2: $\overline{SR'_{-\mathbf{E}_{k+}}(0)}$ vs. k: marginal effect of amplifying eigenvalues beyond the kth
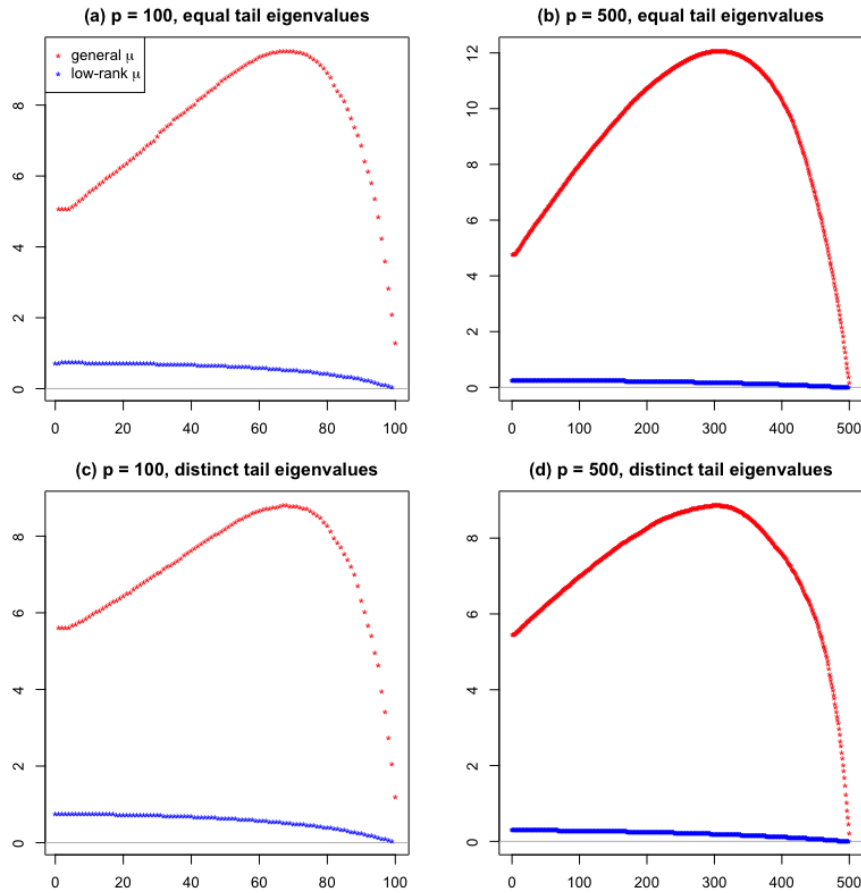
In each panel the x-axis measures k and the y-axis measures $SR'_{-\mathbf{E}_{k+}}(0)$. In panels (a) and (c) $p = 100$ and $n = 1000$; in panels (b) and (d) $p = 500$ and $n = 11180$. In panels (a) and (b) the Σ has 3 spiked eigenvalues and the remaining ones equal 0.05; in panels (c) and (d) the Σ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$. In each panel the red dots correspond to the case where $\boldsymbol{\mu}$ can have loading on all eigenvectors; the blue dots correspond to the case where $\boldsymbol{\mu}$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

blue ones we conclude that when $\boldsymbol{\mu}$ lies in the subspace spanned by the first K eigenvectors, the marginal adjustment effect on the out-of-sample Sharpe ratio is minutely small in magnitude. The reason is the same as we have previously stated, i.e., this happens because the sample-based MSR portfolio is good enough to yield a Sharpe ratio close to the actual maximum one. As we move to panels (c) and (d), where the "flat tail" assumption is violated, no marked difference from the two panels at the top is observed.

In addition, even a quick glance of Figure 2 reminds us that there is some "optimal" k which corresponds to the strongest marginal effect. This result is expected because amplifying the largest few eigenvalues might counteract the improvement brought by amplifying the small ones; so it could be better to solely amplify the small ones. However it is not our specific goal in this paper

19

to find an optimal k. One reason for this is the technical difficulty associated with it, and another reason is that it is not meaningful to look for the optimal k without making sure that a higher $SR'_{-\mathbf{E}_{k+}}(0)$ leads to a higher $SR(-\mathbf{E}_{k+}, \lambda) - SR(-\mathbf{E}_{k+}, 0)$, where λ is a small positive number.[11] Although we do not intend to discuss how to find the optimal k, it is important to point out that the answer to this problem depends on the relationship between $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. A simple illustration on this point is that the optimal k based on the red dots and that based on the blue ones are quite different.

Table 2: Proportion of positive $SR'_{-\mathbf{E}_{k+}}(0)$ among the 100 replications

k	K	K+1	K+5	p-6	p-2	p-1
(a) $p = 100$, equal tail	1.00	1.00	1.00	1.00	1.00	1.00
(b) $p = 500$, equal tail	1.00	1.00	1.00	1.00	1.00	1.00
(c) $p = 100$, distinct tail	1.00	1.00	1.00	1.00	0.94	0.83
(d) $p = 500$, distinct tail	1.00	1.00	1.00	1.00	0.96	0.85

In this table, we only report results for the cases where $\boldsymbol{\mu}$ is an "arbitrary" vector and can have loading on all eigenvectors. In (a) and (b), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones equal 0.05; in (c) and (d), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$.

Figure 2 only shows the average value of the 100 realizations of the marginal effect random variables. As in the previous section we are interested in what proportion of the realizations of each random variable is positive. Table 2 reports these proportions. According to the first two rows of Table 2, when all assumptions about the population covariance matrix are met, the six reported marginal effect random variables are positive in all replications. When Assumption 2.4 is violated, as can be seen in the last two rows of Table 2, there are a few occasions where amplifying the last sample eigenvalue does not lead to improvement; when it comes to amplifying the smallest two eigenvalues however, there are fewer such occasions. In all replications amplifying the smallest six eigenvalues has a positive marginal effect on the out-of-sample Sharpe ratio.

The simulation results provide support to the theoretical conclusions reached in Section 2. To recap under a large-dimensional K-factor model (envisaged in the theoretical analsysis), shrinking a large non-spiked sample eigenvalue and amplifying a small one by a small amount both lead to improvement in the out-of-sample Sharpe ratio. In addition a mild linear amplification on any number of tail sample eigenvalues also leads to improvement.

---

[11]Information about the higher-order derivatives is needed here.

4 Connection with Existing Methods

In this section we provide synthesis on some of the recent research in the portfolio analysis by highlighting an important connection between our theoretical results and a few recently introduced methods in the literature with the goal to improve optimized portfolios.

4.1 Shrinkage towards Identity Estimator

The shrinkage towards identity estimator (Ledoit and Wolf (2004)) is a weighted average of the sample covariance matrix and an identity matrix, i.e., $\mathbf{S}_{STI} = s_1\mathbf{S} + s_2\mathbf{I}$ for some $s_1, s_2 > 0$. The effectiveness of this estimator can be partially explained by our theoretical results. Since the out-of-sample Sharpe ratio is scale-invariant in the covariance matrix estimator, an equivalent covariance estimator is $\mathbf{S}_{STI}^* = \mathbf{S} + \frac{s_2}{s_1}\mathbf{I}$. Since both s_1 and s_2 are positive, $\mathbf{S}_{STI}^*$ conforms to our proposed way of adjusting the sample eigenvalues if we let the parameters in $\mathbf{V} = -\widehat{\mathbf{\Lambda}}^a\mathbf{E}_{k^+}$ be $a = 0$ and $k = 1$.

Readers may be quick to attribute the improved out-of-sample performance of the minimum-variance portfolio to the reduced expected Frobenius loss brought by the shrinkage estimator. Actually the link between a reduced expected Frobenius loss and an increased out-of-sample Sharpe ratio is a little bit more tenuous. Although for both objective functions, $E\big(\big\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\big\|_F\big)$ and $\frac{\boldsymbol{\mu}^T\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}^T\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}}$, optimality is attained at $\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma}$, an improvement in one of them does not necessarily ameliorate the other. In this paper we adopt a more involved objective function so that the improved out-of-sample performance of the portfolio can be more clearly explained. However this achievement comes at a cost: we can only derive analytically exact results for the marginal effect of adjusting eigenvalues instead of determining the optimal amount of the adjustment. This is why it was emphasized at the beginning of this section that the effectiveness of the shrinkage towards identity method could only be *partially* explained in this paper.

4.2 Nonlinear Shrinkage Estimator

The nonlinear shrinkage estimator (Ledoit and Wolf (2017)) extends the shrinkage towards identity estimator by allowing different eigenvalues to be adjusted independently. As mentioned in Section 2.3, under a few assumptions about the population covariance matrix, the authors derive a convergence limit of the out-of-sample Sharpe ratio as p and n go to infinity at the same rate. Then

they search for the optimal shrinkage which maximizes the limiting out-of-sample Sharpe ratio.

Since there is no explicit formula for the optimal shrinkage, we resort to the numerical results provided in Ledoit and Wolf (2017) to ascertain just how the sample eigenvalues are adjusted. Figure 2 in Ledoit and Wolf (2017) shows a comparison among the eigenvalues of a nonlinear shrinkage estimator, those of a linear shrinkage estimator, and the sample eigenvalues. According to the figure, after the nonlinear adjustment, the smallest eigenvalues become overall less dispersed in the sense that $\frac{\widetilde{\lambda}_i}{\widetilde{\lambda}_j} \leq \frac{\widehat{\lambda}_i}{\widehat{\lambda}_j}$, where $\widetilde{\lambda}_i$ is the ith eigenvalue after the nonlinear shrinkage, and i, j $(i < j)$ both index some small eigenvalues. Among the large eigenvalues there is at least one i, j pair $(i < j)$ such that $\frac{\widetilde{\lambda}_i}{\widetilde{\lambda}_j} > \frac{\widehat{\lambda}_i}{\widehat{\lambda}_j}$. As a resultthere is no guarantee that after a nonlinear shrinkage, the eigenvalues become overall less dispersed. From this perspective our results are unable to provide a clear assurance for the improvement brought by the nonlinear estimator. But readers should note that our theoretical results present only a few sufficient conditions for achieving a marginal out-of-sample Sharpe ratio improvement. So, even if the nonlinear shrinkage estimator does not make each pair of eigenvalues less dispersed, it still can lead to improvement.

As has been mentioned earlier there is some similarity between this paper and Ledoit and Wolf (2017), in particular in terms of the objective function used and the family of estimator considered. However the key difference lies in the assumption about the population covariance matrix: our main theoretical results are based on a high-dimensional K-factor model under which the largest K population eigenvalues increase with p at rate $O(p)$; a technical assumption in Ledoit and Wolf (2017) implies that they work under a framework where the largest population eigenvalue is bounded. Our assumption is seemingly more appropriate for financial asset returns covariance matrix, but the cost of adopting it is the divergence of the out-of-sample Sharpe ratio; conversely, the assumption of bounded eigenvalues is less realistic, but it ensures convergence of the out-of-sample Sharpe ratio and, thus, enables the authors to find an optimal shrinkage.

4.3   Spectral Cut-off Method

The spectral cut-off method is a stabilization technique applied to the process of inverting an ill-posed covariance matrix. This method reconstructs the inverse covariance estimator after discarding the eigenvectors associated with the smallest eigenvalues. Carrasco and Noumon (2011) propose a data-driven method for determining the number of eigenvectors to discard. A contemplation about the spectral cut-off method suggests that it can be viewed as a method which

amplifies the smallest eigenvalues to infinity while keeping the large ones unchanged. From this perspective this method falls into our framework as a polar extreme case if we let the parameters in $\mathbf{V} = -\widehat{\mathbf{\Lambda}}^a \mathbf{E}_{k+}$ be $k \geq \min\{k : \widehat{\lambda}_k < 1\}$ and $a = -\infty$. Our theoretical results in Section 2.2 also imply that if we simply target a marginal improvement in the out-of-sample Sharpe ratio, the number of eigenvectors to discard does not play a significant role.

5   Conclusion

We have provided a highly selected review and synthesis on some of the recent and notable contribution to research in the portfolio analyis. A unique perspective on this development was offered in this paper with emphasis on the ways in which to improve the out-of-sample Sharpe ratio of an MSR portfolio constructed based on the two-step approach. To accomplish this goal we assume a high-dimensional *K*-factor model in our theoretical analysis and investigate how improvement can be achieved by adjusting the sample eigenvalues according to certain patterns. Our main theoretical results show that simply adjusting one of the first *K* eigenvalues has a diminishing marginal effect; mildly shrinking a large but non-spiked one and amplifying a small one both lead to an improvement in the out-of-sample portfolio Sharpe ratio under the high-dimensional asymptotics. The effect of adjusting multiple eigenvalues is also studied in the paper. Our results show that simultaneously amplifying a collection of tail eigenvalues according to certain nonlinear pattern yields a positive effect for the out-of-sample Sharpe ratio improvement. Our theoretical results are subsequently illustrated by means of a set of simulation studies performed in this paper.

Lastly, by judiciously identifying a few eigenvalues adjustment patterns in the portfolio that ensures a marginal improvement in the out-of-sample portfolio Sharpe ratio, we are able to provide a much-needed and critical synthesis on two approaches which have recently been introduced in the literature with the goal to improve the the portfolio performances. These approaches are the shrinkage towards identity method and the spectral cut-off method.

A   Proofs of Technical Results

Proof of Theorem 2.1. Let $g(\mathbf{V}, \lambda) := \frac{\boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\Sigma} \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{(\boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu})^2}$. We first calculate $g'_{\mathbf{V}}(0) \equiv \frac{\partial g(\mathbf{V},\lambda)}{\partial \lambda}\big|_{\lambda=0}$. Taking a partial derivative of $g(\mathbf{V}, \lambda)$ with respect to λ, we obtain

$$\frac{\partial g(\mathbf{V}, \lambda)}{\partial \lambda} = \left(2\boldsymbol{\Sigma} \frac{\mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}} \right)^T \frac{\partial \left(\frac{\mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}} \right)}{\partial \lambda}. \tag{12}$$

The derivative in eq. (12) can be simplified as follows:

$$\frac{\partial \left(\frac{\mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}} \right)}{\partial \lambda} = \frac{\boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu} \frac{\partial \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{\partial \lambda} - \frac{\partial \boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{\partial \lambda} \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{(\boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu})^2}. \tag{13}$$

According to the definition of $\mathbf{S}_{\mathbf{V},\lambda}$ it is easy to check that

$$\frac{\partial \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{\partial \lambda}\bigg|_{\lambda=0} = \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} \quad \text{and} \quad \frac{\partial \boldsymbol{\mu}^T \mathbf{S}_{\mathbf{V},\lambda}^{-1} \boldsymbol{\mu}}{\partial \lambda}\bigg|_{\lambda=0} = \boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu}. \tag{14}$$

Now we can readily evaluate $g'_{\mathbf{V}}(0)$ with the assistance of eq. (13) and eq. (14). After a few straightforward steps of calculation we obtain:

$$g'_{\mathbf{V}}(0) = \frac{2}{(\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu})^2} \left(\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} - \boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} \frac{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} \right). \tag{15}$$

Since $SR(\mathbf{V}, \lambda) = \frac{1}{\sqrt{g(\mathbf{V},\lambda)}}$, we get $SR'_{\mathbf{V}}(0) = -\frac{g'_{\mathbf{V}}(0)}{2[g(\mathbf{V},0)]^{\frac{3}{2}}}$. It thus follows that:

$$SR'_{\mathbf{V}}(0) = -\frac{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}}{(\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu})^{\frac{3}{2}}} \left(\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} - \boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} \frac{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} \right). \tag{16}$$

$\square$

Proof of Theorem 2.2. Before proceeding to the proof, we introduce a few necessary notations. Let $\mathbf{U}_F = (\mathbf{u}_1, \ldots, \mathbf{u}_K)$ and $\mathbf{U}_I = (\mathbf{u}_{K+1}, \ldots, \mathbf{u}_p)$ denote the eigenvectors that correspond to the factors and the idiosyncratic components respectively. Further, let $\boldsymbol{\Lambda}_F = \mathrm{diag}\{\lambda_1, \ldots, \lambda_K\}$ and $\boldsymbol{\Lambda}_I = \mathrm{diag}\{\lambda_{K+1}, \ldots, \lambda_p\}$. The sample quantities can be conformably partitioned into a factor part and an idiosyncratic part, i.e. $\widehat{\mathbf{U}}_F = (\widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_K)$, $\widehat{\mathbf{U}}_I = (\widehat{\mathbf{u}}_{K+1}, \ldots, \widehat{\mathbf{u}}_p)$, $\widehat{\boldsymbol{\Lambda}}_F = \mathrm{diag}\{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_K\}$, and $\widehat{\boldsymbol{\Lambda}}_I = \mathrm{diag}\{\widehat{\lambda}_{K+1}, \ldots, \widehat{\lambda}_p\}$. Let $\mathbf{a} = \mathbf{U}^T \boldsymbol{\mu} = (\mathbf{a}_F^T, \mathbf{a}_I^T)^T = (a_1, \ldots, a_p)^T$ denote the loading of the

24

expected returns vector on the linear space spanned by the population eigenvectors. Similarly let $\widehat{\mathbf{a}} = \widehat{\mathbf{U}}^T \boldsymbol{\mu} = (\widehat{\mathbf{a}}_F^T, \widehat{\mathbf{a}}_I^T)^T = (\widehat{a}_1, \ldots, \widehat{a}_p)^T$ denote the loading on the sample eigenvectors. Here $\mathbf{a}_F$ and $\widehat{\mathbf{a}}_F$ are $K \times 1$ vectors; $\mathbf{a}_I$ and $\widehat{\mathbf{a}}_I$ are $(p - K) \times 1$ vectors. Let $\mathbf{e}_k$ denote a conformable column vector with its kth element being 1 and everywhere else being 0. Denote $\xi_\tau = o_{a.s.}(a_\tau)$ if $\lim_{\tau \to \infty} \frac{\xi_\tau}{a_\tau} = 0$ almost surely. Denote $\xi_\tau = O_{a.s.}(a_\tau)$ if $\limsup_{\tau \to \infty} |\frac{\xi_\tau}{a_\tau}| \leq M$, where M is a positive constant.

According to Theorem 1 case (a) in Shen et al. (2016), as both p and n go to infinity with rate of n being $n = O(p^{1+c})$ for some $c > 0$, we obtain the following results:

(1) $\frac{\widehat{\lambda}_j}{\lambda_j} \xrightarrow{a.s.} 1$, for $j = 1, 2, \ldots, K$.

(2) $\text{angle}(\widehat{\mathbf{u}}_j, \text{span}\{\mathbf{u}_j\}) = o_{a.s.}(1)$, for $j = 1, 2, \ldots, K$.

(3) $\text{angle}(\widehat{\mathbf{u}}_j, \text{span}\{\mathbf{u}_{K+1}, \ldots, \mathbf{u}_p\}) = o_{a.s.}\left(\frac{1}{\sqrt{p}}\right)$, for $j = K+1, K+2, \ldots, p$.

The angle between a vector and a space is defined as the angle between the vector and its projection onto the space. We first derive the convergence rate for an inner product between population eigenvectors and their sample counterparts based on the results above. According to the definition of an angle, for $j = 1, 2, \ldots, K$, we have:

$$\arccos\left(\frac{<\widehat{\mathbf{u}}_j, \mathbf{u}_j \mathbf{u}_j^T \widehat{\mathbf{u}}_j>}{\|\widehat{\mathbf{u}}_j\| \left\|\mathbf{u}_j \mathbf{u}_j^T \widehat{\mathbf{u}}_j\right\|}\right) = \arccos(\mathbf{u}_j^T \widehat{\mathbf{u}}_j) = o_{a.s.}(1).$$

Thus, by applying a Taylor expansion on both sides of the above equation, we obtain:

$$\mathbf{u}_j^T \widehat{\mathbf{u}}_j = \cos(\arccos(\mathbf{u}_j^T \widehat{\mathbf{u}}_j)) = 1 + o_{a.s.}(1), \quad j - 1, 2, \ldots, K.$$

For $j = K+1, \ldots, p$, we have:

$$\arccos\left(\frac{<\widehat{\mathbf{u}}_j, \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{u}}_j>}{\|\widehat{\mathbf{u}}_j\| \|\mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{u}}_j\|}\right) = \arccos(\sqrt{\widehat{\mathbf{u}}_j^T \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{u}}_j}) = o_{a.s.}(p^{-1/2}).$$

Applying the Taylor expansion on both sides of the above equation and taking square of the resulting expression we get:

$$\widehat{\mathbf{u}}_j^T \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{u}}_j = (1 - 0.5(o_{a.s.}(p^{-1/2}))^2)^2 = 1 + o_{a.s.}(p^{-1}), \quad j = K+1, \ldots, p. \qquad (17)$$

Since $\widehat{\mathbf{u}}_j$ is of a unit length it follows that:

$$\widehat{\mathbf{u}}_j^T \left(\sum_{i=1}^{K} \mathbf{u}_i \mathbf{u}_i^T + \mathbf{U}_I \mathbf{U}_I^T \right) \widehat{\mathbf{u}}_j = 1$$

$$\rightarrow \quad \sum_{i=1}^{K} (\mathbf{u}_i^T \widehat{\mathbf{u}}_j)^2 = o_{a.s.}(p^{-1})$$

$$\rightarrow \quad \mathbf{u}_i^T \widehat{\mathbf{u}}_j = o_{a.s.}(p^{-1/2}), \quad i = 1, \ldots, K, \, j = K+1, \ldots, p.$$

In addition we also can show that for $i, j \in \{K+1, \ldots, p\}$ and $i \neq j$,

$$\widehat{\mathbf{u}}_i^T \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{u}}_j = -\widehat{\mathbf{u}}_i^T \left(\sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^T \right) \widehat{\mathbf{u}}_j = -\sum_{k=1}^{K} \widehat{\mathbf{u}}_i^T \mathbf{u}_k \mathbf{u}_k^T \widehat{\mathbf{u}}_j = o_{a.s.}(p^{-1}). \tag{18}$$

According to eq. (17) and eq. (18), $\widehat{\mathbf{U}}_I^T \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_I = \mathbf{I}_{(p-K) \times (p-K)} + \mathbf{E}$, where $\mathbf{E}$ is a noise matrix whose elements have rate $o_{a.s.}(p^{-1})$.

Using the matrix notations introduced above the matrix $\widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}}$ can be expanded as:

$$\widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} = \begin{bmatrix} \widehat{\mathbf{U}}_F^T \\ \widehat{\mathbf{U}}_I^T \end{bmatrix} \begin{bmatrix} \mathbf{U}_F & \mathbf{U}_I \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_F & \\ & \mathbf{\Lambda}_I \end{bmatrix} \begin{bmatrix} \mathbf{U}_F^T \\ \mathbf{U}_I^T \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{U}}_F & \widehat{\mathbf{U}}_I \end{bmatrix}$$

$$= \begin{bmatrix} \widehat{\mathbf{U}}_F^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_F + \widehat{\mathbf{U}}_F^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_F & \widehat{\mathbf{U}}_F^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_I + \widehat{\mathbf{U}}_F^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_I \\ \widehat{\mathbf{U}}_I^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_F + \widehat{\mathbf{U}}_I^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_F & \widehat{\mathbf{U}}_I^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_I + \widehat{\mathbf{U}}_I^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_I \end{bmatrix}. \tag{19}$$

Let $\mathbf{b} = (\mathbf{b}_F^T, \mathbf{b}_I^T)^T$ be a $p \times 1$ random vector. According to the convergence results derived above one of the two terms in the summation in each element of eq. (19) is negligible compared with the other term. The quadratic form $\mathbf{b}^T \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} \mathbf{b}$, under the large dimensional asymptotics and Assumption 2.4, can be rewritten as:

$$\mathbf{b}^T \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} \mathbf{b} = \mathbf{b}_F^T \mathbf{\Lambda}_F \mathbf{b}_F (1 + o_{a.s.}(1)) + 2 \mathbf{b}_F^T \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_I \mathbf{b}_I (1 + o_{a.s.}(1)) + \sigma^2 \mathbf{b}_I^T \mathbf{b}_I (1 + o_{a.s.}(1)).$$

For example, when $\mathbf{b} = \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\mu} = \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{a}}$ and under the Assumption 2.5 that $\|\boldsymbol{\mu}\| = O(p^{1/2})$ and

$$\|\mathbf{a}_F\|_\infty = \left\|\mathbf{U}_F^T \boldsymbol{\mu}\right\|_\infty = o(p^{1/2}),$$

$$\mathbf{b}^T \widehat{\mathbf{U}}^T \boldsymbol{\Sigma} \widehat{\mathbf{U}} \mathbf{b}$$

$$= \boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu}$$

$$= \widehat{\mathbf{a}}_F^T \widehat{\boldsymbol{\Lambda}}_F^{-1} \boldsymbol{\Lambda}_F \widehat{\boldsymbol{\Lambda}}_F^{-1} \widehat{\mathbf{a}}_F (1 + o_{a.s.}(1)) + 2\widehat{\mathbf{a}}_F^T \widehat{\boldsymbol{\Lambda}}_F^{-1} \boldsymbol{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_I \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1))$$

$$\quad + \sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1))$$

$$= \widehat{\mathbf{a}}_F^T \widehat{\boldsymbol{\Lambda}}_F^{-1} \widehat{\mathbf{a}}_F (1 + o_{a.s.}(1)) + 2\widehat{\mathbf{a}}_F^T \mathbf{U}_F^T \widehat{\mathbf{U}}_I \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)) + \sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1))$$

$$= \sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)). \tag{20}$$

The last equality holds because the diagonal elements of $\boldsymbol{\Lambda}_F$ increase at a rate of $O(p)$ and because elements in the matrix $\mathbf{U}_F^T \widehat{\mathbf{U}}_I$ have a rate of $o_{a.s.}(p^{-1/2})$. This result will be used in the subsequent proof. Now we prove the results one-by-one.

(a) When $\mathbf{V}$ takes the form of $\mathbf{E}_k$ and as p, n go to infinity with rate of n being $n = O(p^{1+c})$ ($c > 0$), for any $k \in \{1, \ldots, p\}$, the first term inside the bracket in eq. (6) can be expanded as follows:

$$\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{E}_k \widehat{\mathbf{U}}^T \boldsymbol{\mu} = \widehat{\mathbf{a}}^T \widehat{\boldsymbol{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\Sigma} \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{E}_k \widehat{\mathbf{a}}$$

$$= \frac{\widehat{a}_k}{\widehat{\lambda}_k^2} \widehat{\mathbf{a}}^T \widehat{\boldsymbol{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\Sigma} \widehat{\mathbf{U}} \mathbf{e}_k$$

$$= \begin{cases} \dfrac{\widehat{a}_k}{\widehat{\lambda}_k^2} \widehat{\mathbf{a}}_F^T \mathbf{e}_k (1 + o_{a.s.}(1)) & k \le K \\[2mm] \dfrac{\widehat{a}_k}{\widehat{\lambda}_k^2} \sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-1} \mathbf{e}_{k-K}(1 + o_{a.s.}(1)) & k > K \end{cases}$$

$$= \begin{cases} \dfrac{\widehat{a}_k^2}{\widehat{\lambda}_k^2}(1 + o_{a.s.}(1)) & k \le K \\[2mm] \dfrac{\sigma^2 \widehat{a}_k^2}{\widehat{\lambda}_k^3}(1 + o_{a.s.}(1)) & k > K \end{cases}.$$

Since $\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu} = \widehat{\mathbf{a}}_F^T \widehat{\boldsymbol{\Lambda}}_F^{-1} \widehat{\mathbf{a}}_F + \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I = \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1))$, the marginal effect random variable is:

$$SR'_{\mathbf{E}_k}(0) = \begin{cases} -\dfrac{\widehat{a}_k^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}{\widehat{\lambda}_k^2 (\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{3}{2}}} \left(1 + o_{a.s.}(1) - \dfrac{\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}(1 + o_{a.s.}(1))\right) & k \le K \\[4mm] -\dfrac{\widehat{a}_k^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}{\widehat{\lambda}_k^2 (\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{3}{2}}} \left(\dfrac{\sigma^2}{\widehat{\lambda}_k}(1 + o_{a.s.}(1)) - \dfrac{\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}(1 + o_{a.s.}(1))\right) & k > K \end{cases}.$$

27

Next we show $SR'_{\mathbf{E}_k}(0) \xrightarrow{a.s.} 0$, for $k = 1, 2, \ldots, K$. According to the equation above, for any $k \le K$,

$$SR'_{\mathbf{E}_k}(0) = -\frac{\widehat{a}_k^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}{\widehat{\lambda}_k^2 (\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{3}{2}}}(1 + o_{a.s.}(1)) + \frac{\widehat{a}_k^2}{\widehat{\lambda}_k^2 (\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{1}{2}}}(1 + o_{a.s.}(1)) \qquad (21)$$

Since $\frac{\widehat{\mathbf{a}}_I^F \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_I^F \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I} \in [\widehat{\lambda}_p, \widehat{\lambda}_{K+1}]$ and $\frac{1}{(\widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{1}{2}}} \le \frac{\widehat{\lambda}_{K+1}}{(\widehat{\mathbf{a}}_I^T \widehat{\mathbf{a}}_I)^{\frac{1}{2}}}$, the absolute value of the first term (without the $(1 + o_{a.s.}(1))$ part) on the RHS of eq. (21) satisfies

$$\frac{\widehat{a}_k^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}{\widehat{\lambda}_k^2 (\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{3}{2}}} \le \frac{\widehat{a}_k^2 \widehat{\lambda}_{K+1}^2}{\sigma^3 \widehat{\lambda}_k^2 (\widehat{\mathbf{a}}_I^T \widehat{\mathbf{a}}_I)^{\frac{1}{2}}} \le \frac{\widehat{a}_k^2}{\sigma^3 (\widehat{\mathbf{a}}_I^T \widehat{\mathbf{a}}_I)^{\frac{1}{2}}} = o_{a.s.}(1).$$

Similarly the absolute value of the second term satisfies

$$\frac{\widehat{a}_k^2}{\widehat{\lambda}_k^2 (\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{1}{2}}} \le \frac{\widehat{a}_k^2 \widehat{\lambda}_{K+1}}{\sigma \widehat{\lambda}_k^2 (\widehat{\mathbf{a}}_I^T \widehat{\mathbf{a}}_I)^{\frac{1}{2}}} \le \frac{\widehat{a}_k^2}{\sigma \widehat{\lambda}_k (\widehat{\mathbf{a}}_I^T \widehat{\mathbf{a}}_I)^{\frac{1}{2}}} = o_{a.s.}(1).$$

It thus follows that $SR'_{\mathbf{E}_k}(0) \xrightarrow{a.s.} 0$ for $k = 1, 2, \ldots, K$. For $k = K+1, K+2, \ldots, p$, define a random variable

$$X_k^{(p)*} = -\frac{\widehat{a}_k^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}{\widehat{\lambda}_k^2 (\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I)^{\frac{3}{2}}}\left(\frac{\sigma^2}{\widehat{\lambda}_k} - \frac{\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I}\right).$$

Since $\frac{\widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I} \in \left(\frac{1}{\widehat{\lambda}_{K+1}}, \frac{1}{\widehat{\lambda}_p}\right)$ with probability 1, there exists a $K^* \in \{K+2, \ldots, p-1\}$, such that with probability 1, $X_k^{(p)*} > 0$ for any $K+1 \le k < K^*$ and $X_k^{(p)*} < 0$ for any $K^* < k \le p$. We complete the proof by letting $X_k^{(p)} = \text{sign}(X_k^{(p)*})X_k^{(p)*}$.

(b) When $\mathbf{V}$ takes the form of $\mathbf{E}_{k+}$ and as p, n go to infinity with rate of n being $n = O(p^{1+c})$ $(c > 0)$, for any $k \in \{1, 2, \ldots, p\}$, the first term inside the bracket in eq. (6) can be expanded as follows:

$$-\boldsymbol{\mu}^T \mathbf{S}^{-1} \mathbf{\Sigma} \widehat{\mathbf{U}} \widehat{\mathbf{\Lambda}}^{-2} \widehat{\mathbf{\Lambda}}^a \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu} = -\widehat{\mathbf{a}}^T \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} \widehat{\mathbf{\Lambda}}^{-2} \widehat{\mathbf{\Lambda}}^a \mathbf{E}_{k+} \widehat{\mathbf{a}}$$

$$= \begin{cases} -\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-3+a} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)) & k \le K \\ -\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-3+a} \mathbf{E}_{(k-K)^+} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)) & k > K \end{cases}$$

$$= -\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-3+a} \mathbf{E}_{\max\{k-K,1\}^+} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)).$$

The second term inside the bracket in eq. ([6](6)) can be written as:

$$\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \widehat{\boldsymbol{\Lambda}}^a \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu} \frac{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} = \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2+a} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I \frac{\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I} (1 + o_{a.s.}(1)).$$

Let

$$Y_k^{(p)} = \frac{\sigma^2 \boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu} \widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2+a} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I}{(\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu})^{\frac{3}{2}}} \left(\frac{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-3+a} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2+a} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I} - \frac{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I} \right).$$

Then, under the high-dimensional asymptotics, $\frac{SR'_{-\widehat{\boldsymbol{\Lambda}}^a \mathbf{E}_{k+}}(0)}{Y_k^{(p)}} \xrightarrow{a.s.} 1$. It is easy to obtain:

$$\frac{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_T \widehat{\boldsymbol{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I} \leq \frac{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_T \widehat{\boldsymbol{\Lambda}}_I^{-1} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I}, \quad k = 1, 2, \ldots, p.$$

When $a < 1$, with probability 1, we have:

$$\frac{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-3+a} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_T \widehat{\boldsymbol{\Lambda}}_I^{-2+a} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I} > \frac{\widehat{\mathbf{a}}_I^T \widehat{\boldsymbol{\Lambda}}_I^{-2} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_T \widehat{\boldsymbol{\Lambda}}_I^{-1} \mathbf{E}_{\max\{k-K,1\}+} \widehat{\mathbf{a}}_I}, \quad k = 1, 2, \ldots, p.$$

Thus $Y_k^{(p)} > 0$ with probability 1 for any p and $k \in \{1, 2, \ldots, p\}$.

$\square$

Proof of Theorem [2.3](2.3). With the additional assumption that $\boldsymbol{\Sigma} = \boldsymbol{\mu} \boldsymbol{\mu}^T \frac{1}{s_h^2} + \sigma^2 \mathbf{I}$, the terms inside the bracket in eq. ([6](6)) become:

$$\Delta \equiv \boldsymbol{\mu}^T \mathbf{S}^{-1} (\boldsymbol{\mu} \boldsymbol{\mu}^T \frac{1}{s_h^2} + \sigma^2 \mathbf{I}) \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} - \boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} \frac{\boldsymbol{\mu}^T \mathbf{S}^{-1} (\boldsymbol{\mu} \boldsymbol{\mu}^T \frac{1}{s_h^2} + \sigma^2 \mathbf{I}) \mathbf{S}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}}$$

$$= \sigma^2 \left(\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-3} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} - \boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{V} \widehat{\mathbf{U}}^T \boldsymbol{\mu} \frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} \right).$$

Next we prove the two conclusions one-by-one.

(a) We can find the lower bound of $\frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}}$ through the following inequality:

$$\frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} = \frac{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\mu}} \geq \frac{\widehat{\lambda}_1^{-1} \boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\mu}} = \frac{1}{\widehat{\lambda}_1}.$$

The equality is attained only when $\widehat{\lambda}_1 = \cdots = \widehat{\lambda}_p$ or when $\widehat{\mathbf{U}}^T \boldsymbol{\mu} = \mathbf{0}$. Thus, $\frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} > \frac{1}{\widehat{\lambda}_1}$ with probability 1. Similarly we can show that $\frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} < \frac{1}{\widehat{\lambda}_p}$ with probability 1. When $\mathbf{V}$ takes the form of $\mathbf{E}_k$ for some $k \in \{1, \ldots, p\}$, the expression for Δ can be further simplified:

$$\Delta = \frac{\sigma^2 (\widehat{\mathbf{u}}_k^T \boldsymbol{\mu})^2}{\widehat{\lambda}_k^2} \left(\frac{1}{\widehat{\lambda}_k} - \frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} \right).$$

Thus the expression for $SR'_{\mathbf{E}_k}(0)$ becomes:

$$SR'_{\mathbf{E}_k}(0) = - \frac{\sigma^2 (\widehat{\mathbf{u}}_k^T \boldsymbol{\mu})^2 \boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}}{\widehat{\lambda}_k^2 (\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu})^{\frac{3}{2}}} \left(\frac{1}{\widehat{\lambda}_k} - \frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} \right).$$

Since the sequence $\{1/\widehat{\lambda}_k\}_{k=1,\ldots,p}$ increases in k and $\frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} \in \left(\frac{1}{\widehat{\lambda}_1}, \frac{1}{\widehat{\lambda}_p} \right)$ with probability 1, there exists a $K \in \{2, 3, \ldots, p-1\}$, such that for all $k < K$, $SR'_{\mathbf{E}_k}(0) > 0$, and for all $k > K$, $SR'_{\mathbf{E}_k}(0) < 0$, or equivalently, $SR'_{-\mathbf{E}_k}(0) > 0$, with probability 1.

(b) When $\mathbf{V}$ takes the form of $-\widehat{\boldsymbol{\Lambda}}^a \mathbf{E}_{k+}$ the marginal effect random variable becomes:

$$SR'_{-\widehat{\boldsymbol{\Lambda}}^a \mathbf{E}_{k+}}(0) = \frac{\sigma^2 \boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2+a} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{(\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu})^{\frac{3}{2}}} \left(\frac{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-3+a} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2+a} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}} - \frac{\boldsymbol{\mu}^T \mathbf{S}^{-2} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}} \right).$$

$$(22)$$

The terms inside the bracket in eq. (22), denoted as Δ, can be further simplified as:

$$\begin{aligned}
\Delta &= \frac{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-3+a} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2+a} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}} - \frac{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\mu}} \\
&\geq \frac{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-3+a} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2+a} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}} - \frac{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-2} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}{\boldsymbol{\mu}^T \widehat{\mathbf{U}} \widehat{\boldsymbol{\Lambda}}^{-1} \mathbf{E}_{k+} \widehat{\mathbf{U}}^T \boldsymbol{\mu}}.
\end{aligned}$$

Δ is positive with probability 1 only when $|3 - a - 1| > |2 - a - 2|$, which is equivalent to $a < 1$. Therefore, $SR'_{-\widehat{\boldsymbol{\Lambda}}^a \mathbf{E}_{k+}}(0) > 0$ with probability 1, for any $a < 1$ and any $k \in \{1, 2, \ldots, p\}$.

$\square$

References

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Black, F. and Litterman, R. (1990). Asset allocation: combining investor views with market equilibrium. Technical report, Discussion paper, Goldman, Sachs & Co.

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., 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.

Carrasco, M. and Noumon, N. (2011). Optimal portfolio selection using regularization. Technical report, Citeseer.

Ceria, S. and Stubbs, R. A. (2006). Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, 7(2):109–127.

Clarke, R. G., De Silva, H., and Thorley, S. (2006). Minimum-variance portfolios in the us equity market. *The Journal of Portfolio Management*, 33(1):10–24.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.

Engle, R. F., Ledoit, O., and Wolf, M. (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, pages 1–13.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, 39(6):3320.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.

Guo, D., Boyle, P., Weng, C., and Wirjanto, T. (2019). Eigen portfolio selection: A robust approach to sharpe ratio maximization. https://ssrn.com/abstract=3070416.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, pages 259–278.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Kempf, A. and Memmel, C. (2006). Estimating the global minimum variance portfolio. *Schmalenbach Business Review*, 58(4):332–348.

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.

MacKinlay, A. C. (1995). Multifactor models do not explain deviations from the capm. *Journal of Financial Economics*, 38(1):3–28.

MacKinlay, C. A. and Pástor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of financial studies*, 13(4):883–916.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

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

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

Scherer, B. (2010). A new look at minimum variance investing. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1681306.

Shen, D., Shen, H., and Marron, J. (2016). A general framework for consistency of principal component analysis. *The Journal of Machine Learning Research*, 17(1):5218–5251.

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Soviet Mathematics*, 34(1):1373–1403.