Eigen Portfolio Selection: A Robust Approach to Sharpe Ratio Maximization

Danqiao Guo, Phelim P. Boyle, Chengguo Weng, and Tony S. Wirjanto*

Abstract

In this paper, we study how to pick optimal portfolios by modulating the impact of estimation risk in large covariance matrices. We discover that if the expected returns vector lies in a subspace of the eigenvector space of the sample covariance matrix, the sample-based maximum Sharpe ratio portfolio also lies in the same subspace. Due to the uneven distribution of estimation errors across different sample eigenvalues and eigenvectors, it is desirable that the portfolio estimator lies in a space spanned by a few sample eigenvectors that relatively well estimate their population counterparts. Therefore, we propose approximating the expected returns vector in a lower-dimensional subspace and use the approximation for the construction of portfolio. As long as the approximation is close to the original vector, we benefit from the reduced exposure to the estimation error without much loss in the information of the expected returns. We introduce two concrete regularization methods for approximating the expected returns, and analyze the choice of tuning parameters for the methods. We conduct simulation studies and use three real-world stock returns datasets to assess the effectiveness of the two methods. Our results show that both methods mitigate the effect of the estimation error more effectively in a high-dimensional setting than a low-dimensional setting.

JEL classification: G11, G12

Keywords: Portfolio optimization, Estimation error, Approximation error, Spectral cut-off method, Spectral selection method

*We are grateful for comments from Marine Carrasco, René Garcia, Raymond Kan, and participants at FMA Annual Meeting, International Conference on Econometrics and Statistics, Actuarial Research Conference, and seminar participants at HEC Montréal, University of Montréal, and University of Toronto. However any remaining errors and omissions are ours alone. Guo: Department of Statistics and Actuarial Science, University of Waterloo, email: d22guo@uwaterloo.ca. Boyle: Lazaridis School of Business & Economics, Wilfrid Laurier University, email: pboyle@wlu.ca. Weng: Department of Statistics and Actuarial Science, University of Waterloo, email: c2weng@uwaterloo.ca. Wirjanto: Department of Statistics and Actuarial Science and the School of Accounting and Finance, University of Waterloo, email: twirjanto@uwaterloo.ca.

1 Introduction

Markowitz's mean-variance theory, despite its theoretical appeal, has not been widely used in its original form in practice. One of the main reasons (DeMiguel et al. [2009b], Kan and Zhou [2007]) is that estimation errors in expected returns and a covariance matrix have an adverse effect on a portfolio's performance. As Merton [1980] points out, estimating the expected returns from the time series of realized returns is extremely challenging, while the covariance matrix can be much more accurately estimated from historical data. This realization causes investors to dispense with the estimation of expected returns from history and instead resort to either a minimum variance (MV) portfolio or a maximum Sharpe ratio (MSR) portfolio with a better proxy for the expected returns. Over the years, hundreds of factors have been put forward to explain the cross-section of expected returns. A few examples of the firm-specific factors include idiosyncratic volatility, a collection of signals from fundamental analysis, investor sentiment, and media coverage, etc. In practice, sophisticated investors usually engage in exploiting these anomalies and building their proprietary model for the cross-sectional expected returns, while uninformed investors may adopt an equal view about all the assets and thus pursue the MV portfolio[1]. However, all investors, whether sophisticated or not, rely on the realized returns to estimate the covariance matrix. It has been well-noticed that the sample covariance matrix suffers from substantial estimation errors, especially when the number of assets is large compared with the sample size, and these errors greatly jeopardize the optimality of the resulting portfolios. Following this line of reasoning, it seems that we need to work hard on the critical issue of improving the covariance matrix estimator. However, as we proceed, it will be manifested that the problem we intend to address in this paper is not simply a covariance estimation one.

Our goal in this paper is to propose a robust approach to construct a large MSR portfolio when an investor has formed her view on the asset expected returns. Alternatively, we intend to ameliorate the portfolio optimization procedure so as to make a portfolio less contaminated

[1]If the expected return of each asset is the same, the MSR portfolio is simply the MV portfolio.

by errors in the input covariance estimator. It is important to clarify that this paper does not contribute along the direction of forming an accurate proxy for expected returns; rather, we leave the choice to the investor and then construct an MSR portfolio based on her best proxy for expected returns. Therefore, we are not concerned with the possible gap between the actual and the best proxy for expected returns. So we hereafter refer to the best proxy of expected returns simply as the expected returns for convenience purpose.

Many attempts have been made in the literature to guard portfolios against the proliferation of the estimation errors in the inputs to the portfolio selection problem. Some of these approaches are not explicitly designed to improve the MSR portfolio; instead they address a broader set of portfolio optimization problems. As we will see, they provide insightful ideas which can be applied to improve the MSR portfolios. Jagannathan and Ma [2003], for instance, point out that imposing no-short-sale constraints helps reduce portfolio risk and explain why such constraints are effective. DeMiguel et al. [2009a] provide a framework for finding portfolios that perform well out-of-sample, and this framework involves constraining the norm of portfolio weight vectors. Fan et al. [2012] introduce a gross exposure constraint which bridges the gap between the no-short-sale constrained portfolios in Jagannathan and Ma [2003] and the unconstrained problem of Markowitz [1952]. Tu and Zhou [2011] demonstrate that an optimal combination of the 1/N portfolio and a more sophisticated strategy generally outperforms the 1/N portfolio. Chen and Yuan [2016] introduce a "subspace mean-variance analysis" and show that constraining the portfolio weight vector to be within a certain linear subspace resolves the "Markowitz optimization enigma".

Another strand of research follows a "plug-in" strategy and focuses on improving the quality of the covariance matrix estimator used in the optimization. Notably, Ledoit and Wolf [2004] propose shrinking the sample covariance matrix towards a multiple of the identity matrix, so that the over-dispersed sample eigenvalues are pushed back towards their grand mean. In a related study, Ledoit and Wolf [2017] propose a more flexible covariance matrix estimator which shrinks the sample eigenvalues in a nonlinear manner. Frahm and Memmel

[2010] derive two estimators for the global minimum variance portfolio that dominate the traditional estimator with respect to the out-of-sample variance of portfolio returns. More recently, Fan et al. [2013] develop a Principal Orthogonal complement Thresholding (POET) method to deal with the estimation of a high-dimensional covariance matrix with a conditional sparse structure and fast-diverging eigenvalues. Lastly, Carrasco and Noumon [2011] investigate four regularization techniques to stabilize the inverse of the covariance matrix in a Sharpe ratio maximization problem and derive a data-driven method for selecting the tuning parameter in an optimal way.

Most existing "plug-in" methods treat the estimation of the expected returns and the covariance matrix as two separate tasks. This particular estimation strategy may explain why improving the MSR portfolio when expected returns are given has not become a concrete research topic: in this case the problem seems to simply reduce to a covariance matrix estimation one. However, it is important to notice but has been ignored by many that these two problems are not necessarily equivalent. In particular, it does not necessarily take a perfect[2] covariance matrix estimator to produce a perfect portfolio weight estimator. We use an illustrative example to further clarify this argument. For convenience, hereafter let us use the term "return" to stand for the return in excess of the riskless rate for a given asset. Suppose that we have a sample covariance matrix such that the sample eigenvector corresponding to the largest eigenvalue (hereafter referred to as the dominant eigenvector) is a perfect estimator for the population dominant eigenvector. Then, if the expected returns vector is an exact multiple of the population dominant eigenvector, we can show that the sample-based estimator for the MSR portfolio is exactly the true MSR portfolio, regardless of whether the non-dominant eigenvectors can be accurately estimated. The idea we intend to convey by this example is that not only the quality of the covariance matrix estimator but also how the expected returns vector lies in the eigenvector space matters in determining the quality of the sample-based MSR portfolio. Admittedly, the assumption here about

[2] A "perfect" estimator is one without any estimation error.

the distribution of the estimation error in the sample covariance matrix is extreme. But it has been shown by Shen et al. [2016] that estimation errors become progressively more pronounced as we move away from the dominant principal component. In particular, if the population covariance matrix admits a high-dimensional K-factor model, the largest K eigenvalues and their corresponding eigenvectors can be consistently estimated by their sample version under usual high-dimensional asymptotics.

The uneven distribution of estimation errors across principal components, together with the earlier example, suggests an "expected returns approximation" approach for improving the MSR portfolio. In this approach, we approximate the expected returns vector using a few sample eigenvectors which relatively accurately estimate their population counterparts so as to make the most of a commonly used low-quality covariance estimator. As has been mentioned earlier, the expected return vector here refers to the investor's best proxy for expected returns. This vector is known. The purpose of approximating it is not because we deem it to be inaccurate but because we intend to mitigate the estimation errors by slightly modifying it. We replace the original expected returns vector with its approximation in the "plug-in" method to obtain the portfolio weight estimator. We need to ensure that the approximation expected returns vector is close to the original vector so that little information about the expected returns will be lost. In a nutshell, the key idea behind the expected returns approximation approach is to intentionally introduce some approximation error with the goal to mitigate the estimation error. More importantly, we control the maximum amount of approximation error to be introduced and, at the same time, reduce the amount of the estimation error as much as possible.

We discuss two concrete methods that belong to the expected returns approximation approach. The first method approximates the expected returns vector using the sample eigenvectors corresponding to the largest K sample eigenvalues, where K is a parameter to be determined. These eigenvectors are hereafter referred to as the first K eigenvectors. We find an interesting equivalence, in terms of the resulting portfolios, among the first approximation

method, a "spectral cut-off method" (Carrasco et al. [2007]), and a "subspace mean-variance optimization approach" (Chen and Yuan [2016]) in the literature. Thus, we simply refer to our first method as the spectral cut-off method. The second method uses a selected set of sample eigenvectors to approximate the expected returns vector. We coin the second method as a "spectral selection method". In the spectral selection method, we solve for the approximation expected returns vector from a regression problem with varying L_1 penalty. The program is designed so that sample eigenvectors contributing more to explaining the expected returns vector as well as those corresponding to larger sample eigenvalues will enter the selected set with a higher chance. The spectral selection method generalizes the spectral cut-off method in that it is less restrictive about the approximation set. We will show that there are "blind spots" in the spectral cut-off method, in which cases we have to resort to the spectral selection method for a better portfolio performance.

It remains an important issue to determine the tuning parameters in the two spectral methods. Our intention of controlling the maximum amount of approximation error to be introduced motivates us to set an upper bound, denoted by δ, for the relative approximation error, which is defined as the normalized Euclidean norm of the difference between the original and the approximation expected returns vectors. In the spectral cut-off method, since there is a monotonic relationship between K and the implied relative approximation error, the minimal acceptable K is easily obtained once δ is set. The more sophisticated spectral selection method involves two parameters: γ specifies the severity of the aforementioned L_1 penalty, and c controls how differently we treat head and tail principal components. Moreover, γ is a function of c when δ is given. Therefore, we set an upper bound on the relative approximation error and then resort to a data-driven method to tune c, and γ is obtained consequently.

Constraining the relative approximation error is an important idea, since only when this error is well controlled can we get a portfolio closely reflecting an investor's view. This involves specifying the value of δ. Actually, even without a fine tuning procedure, we are

aware that the eligible range for δ is between 0 and 1 (since it can be viewed as a sine of a small angle) and that a reasonable value for δ should be close to 0 to ensure not too much information is discarded from the given expected returns vector. Another appealing feature of δ is that once we have decided on a value of δ, as we update the expected returns and the sample covariance matrix on portfolio rebalancing dates, the resulting K or γ automatically changes according to the spatial relationship between the updated expected returns vector and the eigenvectors of the updated covariance matrix estimator. Therefore, δ can be viewed as a dynamic selector of K or γ.

This paper makes four contributions to the literature. First, we discover the critical role of the expected returns vector in determining the possibility of improving the MSR portfolio even when the covariance matrix is poorly estimated. Second, we cast a new light on the economic interpretation of the well-known spectral cut-off method, which has been viewed as a pure stabilization technique, when it is applied to portfolio optimization problems. Third, we ameliorate the conventional procedure of tuning the parameter in the spectral cut-off method to protect an investor's view from being substantially distorted. Last but not least, we propose a novel spectral selection method for safeguarding an MSR portfolio against pervasive estimation errors in the "less informative" dimensions. In addition, we devise a tuning parameter selection procedure which bounds the approximation error and in the meanwhile learns from data.

The rest of this paper is organized as follows. Section 2 establishes a connection between an eigen portfolio and an MSR portfolio. Section 3 introduces two concrete forms of the expected returns approximation approach, discusses their respective properties, and illustrates the tuning parameter selection procedure. Section 4 provides three ways of pinning down an expected returns vector. In Section 5 we use four simulated cases to assess the effectiveness of the spectral methods. In Section 6 we use empirical returns from three major equity markets to evaluate the out-of-sample performance of different portfolios. In addition, we also justify a heuristic choice on the value for the tuning parameter δ.

Before we proceed it is useful to introduce some notations used in the rest of this paper. We denote matrices by bold capital letters, vectors by bold lower-case letters, and scalars by plain lower-case letters. Let $\mathbf{\Sigma}$ denote a $p \times p$ population covariance matrix. The equation $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ represents the eigen decomposition of $\mathbf{\Sigma}$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p\}$ is a diagonal matrix containing non-increasingly ordered eigenvalues and $\mathbf{U} = (\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(p)})$ contains the eigenvectors. Similarly, let $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{U}}^T$ denote the eigen decomposition of the sample covariance matrix $\widehat{\mathbf{\Sigma}}$, where $\widehat{\mathbf{\Lambda}} = \text{diag}\{\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_p\}$ is a diagonal matrix containing the non-increasingly ordered sample eigenvalues and $\widehat{\mathbf{U}} = (\widehat{\mathbf{u}}^{(1)}, \widehat{\mathbf{u}}^{(2)}, \ldots, \widehat{\mathbf{u}}^{(p)})$ contains the sample eigenvectors. Throughout the paper, we focus on the situations where $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$ are invertible for the sake of clarity and without loss of generality. Furthermore, we let $\|\cdot\|$ denote the spectral norm of a matrix and the L_2 norm of a vector.

2 From Eigen Portfolio to MSR Portfolio

Portfolios based on re-scaling the eigenvectors of the covariance matrix are called eigen portfolios. An appealing feature of eigen portfolios is their uncorrelatedness since the eigenvectors of the covariance matrix are mutually orthogonal. This nice property has been exploited by Steele [1995], Partovi et al. [2004], Avellaneda and Lee [2010], and Boyle [2014] among others. In this section, we set up a connection between the expected returns vector and the portfolio weight vector by viewing the eigenvectors, or equivalently the eigen portfolios, as a set of basis vectors of a vector space. The important implication conveyed in this section is that we can use a linear combination of a subset of eigenvectors to approximate the expected returns vector so as to obtain a portfolio whose weight is spanned by the same eigenvectors.

Suppose that an investor needs to make a single-period investment decision on allocating weight to p risky assets so as to maximize the end-of-period portfolio Sharpe ratio, which is an expected return to standard deviation ratio[3]. Further, the $p \times 1$ vector $\boldsymbol{\mu}$ contains the

[3]Our definition of Sharpe ratio is in accordance with the usual definition - the expected return in excess

expected returns over the investment horizon, and $\boldsymbol{\Sigma}$ is the asset returns covariance matrix. If this is the case, the investor solves for the MSR portfolio from the following problem:

$$\mathbf{w}^{msr} = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T \boldsymbol{\mu}}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}, \quad \text{s.t. } \mathbf{w}^T \mathbf{1} = 1, \tag{1}$$

where $\mathbf{1}$ is a $p \times 1$ vector of ones. It is easy to check that the solution to this problem is

$$\mathbf{w}^{msr} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}, \tag{2}$$

given that the expected return of the global minimum-variance portfolio is higher than zero[4] ($\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} > 0$). The MSR portfolio lies on the efficient frontier.

Assume that we have p eigen portfolios $\mathbf{Z} = (\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(p)})$ so that their weights are multiples of the eigenvectors of the covariance matrix, i.e., $\mathbf{z}^{(i)} = \frac{\mathbf{u}^{(i)}}{\mathbf{1}^T \mathbf{u}^{(i)}}$. These portfolios are mutually orthogonal. Now consider the scenario that the expected returns vector $\boldsymbol{\mu}$ is proportional to $\mathbf{z}^{(i)}$. Since $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ have the same eigenvectors, the weight vector of this MSR portfolio is identical to $\mathbf{z}^{(i)}$. The other portfolios based on the remaining eigenvectors are orthogonal to the MSR portfolio. We can run through all the eigenvectors in the same way by selecting the expected returns vector to be proportional to each eigenvector. In each case the eigen portfolio is efficient in terms of maximizing the Sharpe ratio upon the given expected returns vector.

The next proposition provides a formal statement of the connection between the eigen portfolio and the MSR portfolio.

Proposition 2.1. If $\boldsymbol{\mu}$ is a non-zero scalar multiple of the ith eigenvector of $\boldsymbol{\Sigma}$ and sums to a positive number, i.e., $\boldsymbol{\mu} = a\mathbf{u}^{(i)}$, $a \in \{a \neq 0 : a\mathbf{1}^T \mathbf{u}^{(i)} > 0\}$, the MSR portfolio in eq. (2) is exactly the ith eigen portfolio, i.e., $\mathbf{w}^{msr} = \mathbf{z}^{(i)}$.

of the riskless rate over the standard deviation. Note that as has been mentioned earlier, in this paper we use the term "return" to denote an asset's/portfolio's return in excess of the riskless rate.

[4] If this condition is not satisfied, the MSR portfolio does not exist, and the portfolio calculated from eq. (2) corresponds to the *minimum* Sharpe ratio portfolio.

All of the proofs are given in the Appendix. We now provide some comments on this result. An interesting implication of Proposition 2.1 is that when the vector of expected returns is a scalar multiple of an eigenvector, the MSR portfolio reveals a "return preserving" property, namely, the investment in a given asset is directly proportional to its expected return. A concrete case where a Sharpe ratio maximizing investor would like to hold an eigen portfolio is when asset returns are generated from a single-factor model with a constant residual variance (MacKinlay and Pástor [2000]). If this is the case, an investor's optimal choice is to hold the dominant eigen portfolio. Proposition 2.1 can be readily extended to the scenarios where the vector of expected returns is a linear combination of a set of eigenvectors, as shown in the following proposition.

Proposition 2.2. If $\boldsymbol{\mu}$ can be expressed by eigenvectors of $\boldsymbol{\Sigma}$ as $\boldsymbol{\mu} = \sum_{i=1}^{p} a_i \mathbf{u}^{(i)}$[5] and the inequality $\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} > 0$ is satisfied, then the MSR portfolio in eq. (2) has weights given by

$$\mathbf{w}^{msr} = \sum_{i=1}^{p} \frac{\frac{a_i}{\lambda_i}}{\sum_{i=1}^{p} \frac{a_i \mathbf{1}^T \mathbf{u}^{(i)}}{\lambda_i}} \mathbf{u}^{(i)} = \sum_{i=1}^{p} \frac{\frac{a_i \mathbf{1}^T \mathbf{u}^{(i)}}{\lambda_i}}{\sum_{i=1}^{p} \frac{a_i \mathbf{1}^T \mathbf{u}^{(i)}}{\lambda_i}} \mathbf{z}^{(i)}. \tag{3}$$

Proposition 2.2 expresses the MSR portfolio as a weighted average of the eigen portfolios; in addition, it specifies how the weights are determined based on the loadings (a_i's) of the expected returns vector on the eigenvectors. An important implication of Proposition 2.2 is that the expected returns vector and the MSR portfolio lie in the same linear subspace of the eigenvector space. More specifically, if $\boldsymbol{\mu}$ is spanned by a subset of eigenvectors, the MSR portfolio will then be a weighted average of the corresponding subset of eigen portfolios. This proposition serves as a theoretical foundation and provides an intrinsic motivation for the "expected returns approximation method" which will be formally introduced in Section 3.

[5]Note that any $\boldsymbol{\mu} \in \mathbb{R}^p$ can be written as a linear combination of the eigenvectors $\{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(p)}\}$. This equation just gives a notation to the loadings of $\boldsymbol{\mu}$ on the eigenvectors.

3 Spectral Methods for Improving MSR Portfolios

In the last section, it was assumed that there was no estimation error in the inputs to the portfolio selection problem. In practice, the estimation errors in the expected returns vector and the covariance matrix are ubiquitous. There is an extensive literature on this topic, e.g., see Kolm et al. [2014] for a contemporary review on this. In this section, we explain how the eigen portfolios can be used to address the estimation risk problem in portfolio selection.

The portfolio problem we consider in this section is similar to that described in Section 2. Suppose that $\boldsymbol{\mu}$ is an investor's best proxy for the vector of expected returns of p risky assets over the investment horizon based on all information available to her; the covariance matrix is unknown and needs to be estimated from the price history. The investor wants to construct an MSR portfolio based on her best knowledge, or alternatively, a portfolio that maximizes the ex-ante Sharpe ratio. Admittedly, if the investor has a view about asset expected returns that deviates substantially from the reality, an MSR portfolio constructed based on her view will very likely yield a poor out-of-sample performance. However, it is not the goal of our paper to answer the question of how to generate an accurate expected returns proxy. Rather, we deem $\boldsymbol{\mu}$ as an exogenously given view and try to improve the MSR portfolio constructed based on this view.

The rationale for leaving the choice of the proxy for expected returns to the investor is as follows. As we have mentioned earlier, there are a variety of models to choose from for predicting returns and few investors would use the sample-based estimator as the proxy, it is thus restrictive to specify how an investor would make the decision; while investors usually rely on historical data to obtain a reasonable covariance matrix. The same argument for viewing $\boldsymbol{\mu}$ as exogenously given is used, for example, by Ledoit and Wolf [2017]. However, later we will provide a few ways to pin down a reasonable $\boldsymbol{\mu}$ to facilitate the implementation of our method.

If we adopt the sample covariance matrix as the input to the "plug-in" method, we get

the following sample-based MSR portfolio weight estimator:

$$\widehat{\mathbf{w}}^{msr} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}. \tag{4}$$

What concerns us with the above portfolio weights is that there may exist severe estimation error in $\widehat{\boldsymbol{\Sigma}}$ when p is large relative to the sample size for the estimation; further, inverting the matrix amplifies the errors, especially those in the sample eigenvectors corresponding to the smallest eigenvalues. As a consequence, the sample-based MSR portfolio could severely deviate from the true optimal portfolio. In this section, we introduce an "expected returns approximation" approach to guard the portfolio weight estimator against the estimation error in $\widehat{\boldsymbol{\Sigma}}$.

The basic idea of the expected returns approximation approach is described as follows. Following the same logic as that in Proposition 2.2, we can show that in the presence of estimation error, the expected returns vector and the weight vector of the sample-based MSR portfolio lie in the same linear subspace of the space spanned by sample eigenvectors. Therefore, if the expected returns vector can be approximated well by using a few sample eigenvectors which relatively well estimate their population counterparts, we can replace the original expected returns vector with its approximation, so that the MSR portfolio is not contaminated by the more severe estimation errors in the excluded principal components. Inevitably, we introduce an approximation error by ignoring some eigenvectors, and we discuss how to strive a good balance to achieve such a trade off.

We describe two concrete methods for approximating the vector of expected returns. The two methods are different in their choice of the approximation set, i.e., the collection of eigenvectors used to approximate the expected returns. The first method uses the first few sample eigenvectors and is shown to be equivalent to a spectral cut-off method in the literature. The second method, known as the spectral selection method, uses a selected set of sample eigenvectors as the approximation set. The selection criterion takes into consider-

ation both the contribution of a sample eigenvector to explain the expected returns and the magnitude of its corresponding eigenvalue. The two methods will be introduced in Sections 3.1 and 3.2 respectively.

3.1 Another Look at the Spectral Cut-off Method

We now describe the first approximation approach. Since we intend to approximate the vector of expected returns using a few sample eigenvectors that, compared with the remaining ones, relatively accurately estimate their population counterparts, a natural choice is to use the first few sample eigenvectors which tend to better estimate their population counterparts in terms of consistency and convergence rate (Shen et al. [2016]).

Given a feasible set for the approximation vector of expected returns, we deem an approximation to be optimal if its L_2 distance from the original vector is minimized. Suppose that we approximate $\boldsymbol{\mu}$ in the linear space spanned by the first K sample eigenvectors (the choice of K will be discussed later). We let $\sum_{i=1}^{K} a_i \widehat{\mathbf{u}}^{(i)}$ denote an approximation vector and solve for the optimal a_i's from the following minimization problem:

$$(\widehat{a}_1^{cut}, \ldots, \widehat{a}_K^{cut}) = \underset{(a_1,\ldots,a_K)}{\arg\min} \left\| \boldsymbol{\mu} - \sum_{i=1}^{K} a_i \widehat{\mathbf{u}}^{(i)} \right\|. \tag{5}$$

The following proposition presents the solution to the problem.

Proposition 3.1. The solution to the optimization problem in eq. (5) is:

$$\widehat{a}_i^{cut} = \widehat{a}_i^{ls}, \quad i = 1, 2, \ldots, K, \tag{6}$$

where $\widehat{a}_i^{ls} = \widehat{\mathbf{u}}^{(i)T} \boldsymbol{\mu}$, $i = 1, 2, \ldots, p$, is the solution to the following problem:

$$(\widehat{a}_1^{ls}, \ldots, \widehat{a}_p^{ls}) = \underset{(a_1,\ldots,a_p)}{\arg\min} \left\| \boldsymbol{\mu} - \sum_{i=1}^{p} a_i \widehat{\mathbf{u}}^{(i)} \right\|. \tag{7}$$

Note that the problem in eq. (7) differs from that in eq. (5) in their dimension. It follows

from Proposition 3.1 that the optimal approximation vector of expected returns is given by:

$$\widehat{\boldsymbol{\mu}}^{cut}(K) = \sum_{i=1}^{K} \widehat{a}_i^{cut} \widehat{\mathbf{u}}^{(i)} = \sum_{i=1}^{p} \widehat{a}_i^{ls} \mathbb{1}_{\{i \leq K\}} \widehat{\mathbf{u}}^{(i)}. \tag{8}$$

Using matrix notations, the approximation vector can be also written as $\widehat{\boldsymbol{\mu}}^{cut}(K) = \widehat{\mathbf{U}}_K \widehat{\mathbf{U}}_K^T \boldsymbol{\mu}$, where $\widehat{\mathbf{U}}_K = (\widehat{\mathbf{u}}^{(1)}, \widehat{\mathbf{u}}^{(2)}, \ldots, \widehat{\mathbf{u}}^{(K)})$. Note that $\widehat{\boldsymbol{\mu}}^{cut}(K)$ is the projection of $\boldsymbol{\mu}$ onto the linear space spanned by the first K sample eigenvectors of $\widehat{\boldsymbol{\Sigma}}$. Replacing $\boldsymbol{\mu}$ by $\widehat{\boldsymbol{\mu}}^{cut}(K)$, we obtain the following MSR portfolio:

$$\widehat{\mathbf{w}}^{cut}(K) = \frac{\widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}^{cut}(K)}{\mathbf{1}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}^{cut}(K)} = \sum_{i=1}^{p} \frac{\frac{\widehat{a}_i^{ls} \mathbb{1}_{\{i \leq K\}} \mathbf{1}^T \widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}{\sum_{i=1}^{p} \frac{\widehat{a}_i^{ls} \mathbb{1}_{\{i \leq K\}} \mathbf{1}^T \widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}} \widehat{\mathbf{z}}^{(i)}, \tag{9}$$

where $\widehat{\mathbf{z}}^{(i)} = \frac{\widehat{\mathbf{u}}^{(i)}}{\mathbf{1}^T \widehat{\mathbf{u}}^{(i)}}$ denotes the ith sample eigen portfolio. Note that the sample-based MSR portfolio is the following weighted average of sample eigen portfolios:

$$\widehat{\mathbf{w}}^{msr} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}}{\mathbf{1}^T \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}} = \sum_{i=1}^{p} \frac{\frac{\widehat{a}_i^{ls} \mathbf{1}^T \widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}{\sum_{i=1}^{p} \frac{\widehat{a}_i^{ls} \mathbf{1}^T \widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}} \widehat{\mathbf{z}}^{(i)}. \tag{10}$$

According to eqs. (9) and (10), the actual effect of approximating the expected returns using the first K sample eigenvectors on the MSR portfolio composition is to eliminate any contribution from the last $p-K$ sample eigen portfolios and reallocate the weight. Moreover, the relative weight of the first K sample eigen portfolios is not affected.

A further simplification of eq. (9) by plugging in the matrix expression for $\widehat{\boldsymbol{\mu}}^{cut}(K)$ leads to an alternative expression for the portfolio weights:

$$\widehat{\mathbf{w}}^{cut}(K) = \frac{\widehat{\boldsymbol{\Sigma}}_K^{-1} \boldsymbol{\mu}}{\mathbf{1}^T \widehat{\boldsymbol{\Sigma}}_K^{-1} \boldsymbol{\mu}}, \tag{11}$$

where $\widehat{\boldsymbol{\Sigma}}_K^{-1} = \widehat{\mathbf{U}}_K \widehat{\boldsymbol{\Lambda}}_K^{-1} \widehat{\mathbf{U}}_K^T$ and $\widehat{\boldsymbol{\Lambda}}_K = \text{diag}\{\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_K\}$. Eq. (11) conveys an interesting fact: the method of inputting the sample covariance estimator and the approxi-

mation expected returns into the "plug-in" method is equivalent to keeping the return vector unchanged and using $\widehat{\mathbf{\Sigma}}_K^{-1}$ to replace $\widehat{\mathbf{\Sigma}}^{-1}$. Note that $\widehat{\mathbf{\Sigma}}_K^{-1}$ is a modified inverse covariance matrix which discards the principal components associated with the smallest $p-K$ sample eigenvalues. This specific way of modifying an inverse covariance matrix is called the spectral cut-off method and has been discussed in Carrasco et al. [2007] and Carrasco and Noumon [2011]. The spectral cut-off method was originally introduced as a stabilizing technique to invert an ill-posed[6] sample covariance matrix.

Chen and Yuan [2016] is a related work according to which the portfolio weight vector takes the same form as that in eq. (11)[7]. The authors search for the mean-variance optimal portfolio within the linear subspace spanned by the first few eigenvectors. They show that if asset returns are generated from an approximate factor model and both expected returns and covariance matrix are estimated from data, the sample-based subspace mean-variance optimal portfolio only leads to a diminishing deterioration in Sharpe ratio compared with the actual mean-variance optimal portfolio. Actually, this nice theoretical property of the sample-based subspace optimal portfolio is guaranteed by the strong factor model assumption which implies that the expected returns vector roughly lies in the linear subspace spanned by the first few eigenvectors. When an investor's view is not in accordance with the factor model, the subspace mean-variance optimal portfolio will lead to certain information distortion for sure. This point will later be illustrated in more detail.

By pointing out the equivalence between the spectral cut-off method, the subspace mean-variance analysis, and the expected returns approximation method, we provide an economic explanation for the first two: by leaving out a few tail principal components, the spectral cut-off method and the subspace mean-variance approach secretly modify an investor's view on expected returns. Thus, cutting off the last $p-K$ principal components is only desirable

[6]The sample covariance matrix can be ill-posed or even singular, especially when multicollinearity is present across investment assets or when the sample size is smaller than the number of assets.

[7]The only difference is that in Chen and Yuan [2016] $\boldsymbol{\mu}$ is replaced with its estimator, i.e., $\widehat{\mathbf{w}}^{sub}(K) = \frac{\widehat{\mathbf{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}$.

if $\boldsymbol{\mu}$ can be well approximated in the space spanned by the first K sample eigenvectors. Since the first approximation method is equivalent to the spectral cut-off method in the sense that both lead to the same portfolio, we do not coin this approximation method with a new name; instead we continue to use the term "spectral cut-off" method when referring to it.

Actually, the finding of such equivalence between a method for regularizing the covariance matrix estimator and a way of modifying the expected returns vector is not accidental. In fact, we can easily identify an "equivalent modified expected returns vector" for a few methods designed for improving the covariance matrix estimator. For example, the well-known "shrinkage towards identity" method, by making sample eigenvalues less dispersed, secretly magnifies the loading of the expected returns vector on eigenvectors corresponding to the larger eigenvalues. To illustrate this, let us suppose the shrinkage estimator $\widehat{\boldsymbol{\Sigma}}+\lambda\mathbf{I}$ is used in the "plug-in" method. Then, the resulting MSR portfolio has the following weight vector:

$$\widehat{\mathbf{w}}^{sti}(\lambda)=\frac{(\widehat{\boldsymbol{\Sigma}}+\lambda\mathbf{I})^{-1}\boldsymbol{\mu}}{\mathbf{1}^T(\widehat{\boldsymbol{\Sigma}}+\lambda\mathbf{I})^{-1}\boldsymbol{\mu}}=\sum_{i=1}^{p}\frac{\frac{\frac{\widehat{\lambda}_i}{\widehat{\lambda}_i+\lambda}\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}{\sum_{i=1}^{p}\frac{\frac{\widehat{\lambda}_i}{\widehat{\lambda}_i+\lambda}\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}\widehat{\mathbf{z}}^{(i)}. \tag{12}$$

According to the coefficient before $\widehat{\mathbf{z}}^{(i)}$, the loading of the "equivalent modified expected returns vector" on the ith sample eigenvector is $\frac{\widehat{\lambda}_i}{\widehat{\lambda}_i+\lambda}\widehat{a}_i^{ls}$. For large sample eigenvalues, this loading is hardly different from $\widehat{a}_i^{ls}$; for small sample eigenvalues however, the loading is much smaller than $\widehat{a}_i^{ls}$ in the sense of absolute value.

3.1.1 Consistency of $\widehat{\mathbf{w}}^{cut}(K)$ under a spiked covariance model

This section is devoted to showing that if the population covariance model has a spiked structure that the largest K eigenvalues increase with p while the remaining ones are bounded as p increases, then $\widehat{\mathbf{w}}^{cut}(K)$ converges almost surely to a distortion of the true optimal weight under the high-dimensional asymptotics where both p and the sample size go to infinity at the same rate.

Assumption 3.1. $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are a random sample having the distribution of

$$\mathbf{x}_i = \sum_{j=1}^{p} \lambda_j^{\frac{1}{2}} z_{i,j} \mathbf{u}^{(j)}$$

where the $z_{i,j}$'s are i.i.d. random variables with zero mean, unit variance, and finite fourth moment.

Assumption 3.1 specifies how random samples are generated from the population covariance model. Then, we calculate the sample covariance matrix from $\widehat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$, where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$.

We index all quantities, including p, by n. So, the population eigenvalues and the portfolio size will be denoted as $\lambda_j^{(n)}$ and $p(n)$ respectively throughout this section.

Assumption 3.2. As $n \to \infty$, $\lambda_1^{(n)} > \cdots > \lambda_K^{(n)} \gg \lambda_{K+1}^{(n)} \asymp \cdots \asymp \lambda_{p(n)}^{(n)} \asymp 1$.

For $i < j$, $\lambda_i^{(n)} > \lambda_j^{(n)}$ means that $\lim_{n\to\infty} \frac{\lambda_i^{(n)}}{\lambda_j^{(n)}} > 1$; $\lambda_i^{(n)} \gg \lambda_j^{(n)}$ means that $\lim_{n\to\infty} \frac{\lambda_j^{(n)}}{\lambda_i^{(n)}} = 0$; $\lambda_i^{(n)} \asymp \lambda_j^{(n)}$ means that $c_1 \leq \underline{\lim}_{n\to\infty} \frac{\lambda_i^{(n)}}{\lambda_j^{(n)}} \leq \overline{\lim}_{n\to\infty} \frac{\lambda_i^{(n)}}{\lambda_j^{(n)}} \leq c_2$ for two constants $0 < c_1 \leq c_2$. Assumption 3.2 implies that the population covariance matrix has a spiked structure: the first K eigenvalues increase as n goes to infinity, while the remaining ones are bounded. A typical asset returns model which admits such a spiked covariance structure is the high-dimensional (approximate) factor model discussed, for example, in Bai and Ng [2002] and Fan et al. [2013].[8] However, compared with the assumption made in these references, our assumption about the strength and pervasiveness of the "common factors"[9] is mild. Fan et al. [2013] assume a K-factor model in which each factor is pervasive in the sense that a non-negligible fraction of factor loadings should be non-vanishing; alternatively, the first K eigenvalues should increase at the same rate as the portfolio size. In contrast to this "strong

[8] Although the spiked covariance model we assume is implied by a factor model, we do not directly assume that asset returns are generated from a factor model, because otherwise the expected returns are also governed by the factor model and it is thus unreasonable not to use the model-based expected returns. However, our assumptions could be understood as a factor model in which only the second-order stationary of factor returns is assured.

[9] See the previous footnote.

factor" assumption, we allow the common factors to be weak in the sense that as long as the Kth eigenvalue diverges as n increases, the assumption is satisfied.

Proposition 3.2. As Assumptions 3.1 - 3.2 hold, the portfolio weight estimator $\widehat{\mathbf{w}}^{cut}(K)$ given in eq. (9) and eq. (11) converges to a distortion of the actual MSR portfolio in the sense that:

$$\frac{<\widehat{\mathbf{w}}^{cut}(K), \mathbf{w}^{cut}(K)>}{\|\widehat{\mathbf{w}}^{cut}(K)\|\|\mathbf{w}^{cut}(K)\|} \xrightarrow{a.s.} 1$$

as $n \to \infty$ and $\frac{p(n)}{n} \to c \in (0,1)$, where $\mathbf{w}^{cut}(K) = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^K}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^K}$ and $\boldsymbol{\mu}^K$ is the projection of $\boldsymbol{\mu}$ on the space span$\{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(K)}\}$.

Proposition 3.2 conveys the key idea behind the expected returns approximation method: we intentionally introduce some bias into $\boldsymbol{\mu}$ in order to ameliorate statistical properties of the portfolio estimator. The appeal of this method depends critically on how much bias we should introduce for ensuring the convergence. If $\boldsymbol{\mu}^K$ is close to $\boldsymbol{\mu}$, we do not lose much by introducing the approximation error while gain considerably from avoiding the estimation error; otherwise, the method will yield a portfolio which is far from the true optimal one. If we just choose the number of spiked eigenvalues K as the parameter in the spectral cut-off method, the quality of the resulting portfolio is no longer within our control, since we cannot ensure that $\boldsymbol{\mu}^K$ approximates $\boldsymbol{\mu}$ well. This provides a motivation to discuss how to select the tuning parameter K when using the spectral cut-off method.

We have mentioned earlier that Chen and Yuan [2016] show that when asset returns follow an approximate K factor model, the sample-based subspace mean-variance optimal portfolio $\widehat{\mathbf{w}}^{sub}(K) = \frac{\widehat{\boldsymbol{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}$, where $\widehat{\boldsymbol{\mu}}$ stands for the sample mean, only leads to a diminishing deterioration in Sharpe ratio under high-dimensional asymptotics. This is not at odds with our results, because when the approximate K factor model is assumed, the expected returns vector is roughly spanned by the first K eigenvectors, and therefore $\boldsymbol{\mu}^K$ is extremely close to $\boldsymbol{\mu}$.

3.1.2 Selection of tuning parameter

According to the discussion in the previous section, under a high-dimensional spiked covariance model, if we want a convergent portfolio weight estimator, the number of principal components we should keep is simply the number of diverging eigenvalues, or the number of factors if we take the factor model perspective (see the earlier footnote[8]).

However, even if K is known to us, using the estimator $\widehat{\mathbf{w}}^{cut}(K)$ in real-world portfolio construction tasks can create unexpected problems. This is because as we have mentioned earlier, the distortion we introduce is not within our control but implied by K. Thus the MSR portfolio estimator could converge to some highly undesired portfolio. In addition, the estimability of K relies on certain assumptions about the rate of the divergent eigenvalues. Under our mild Assumption 3.2, K is in general unidentified (Onatski [2012]). Due to the aforementioned reasons, we pursue other methods to determine the number of principal components to keep.

Our intention of reasonably well approximating the expected returns vector suggests a new way to select K. The idea is to specify the maximum relative approximation error we can tolerate and then select the minimum tolerable K. The number of principal components to keep, according to our newly proposed method, is

$$K(\delta) = \min\left\{K : \frac{\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^{cut}(K)\|}{\|\boldsymbol{\mu}\|} \leq \delta\right\}, \tag{13}$$

where δ is the aforementioned tolerance limit for the relative approximation error. Admittedly, it is possible that the resulting MSR portfolio $\widehat{\mathbf{w}}^{cut}(K(\delta))$ does not converge to any population counterpart, while it seeks a balance between a convergent but severely distorted estimator and an estimator severely contaminated by the estimation error in the sample covariance matrix. To be more specific, if we compare the portfolio composition in eq. (9) and eq. (10), it is clear that the spectral cut-off method based portfolio estimator with any $K \neq p$ shifts more weight to the convergent sample eigen portfolios.

The parameter δ is more preferable to use than K for the following reasons. By setting the value of δ, we control the amount of the approximation error to be introduced. In contrast, when the conventional cross-validation method for selecting K is used, the implied approximation error depends on the data and can be so substantial that the resulting portfolio is far from the true MSR portfolio. Another reason is that δ has both geometric and economic interpretations: it specifies the maximum sine of an angle representing the approximation error to be introduced, therefore, even without a fine tuning procedure, we are aware that an eligible range for δ is between 0 and 1, and an appropriate value for δ is close to 0 in order to avoid a substantial loss in information contained in the investor's view on expected returns. In addition, once the value of δ is determined, each time the MSR portfolio is updated according to the latest $\boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}}$, the value of K automatically changes according to the spatial relationship between the new expected returns vector and the eigenvectors of the updated covariance matrix estimator. Thus, a fixed δ is a dynamic selector of K. In contrast, if we would like to pre-determine a value of K, as $\boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}}$ are updated, we cannot control the amount of the approximation error any longer. Lastly, adopting a reasonable value of δ is an effective way to avoid the high computational cost incurred by a cross-validation method.

There is always a trade-off between the approximation error and the estimation error in choosing δ: if δ is too large, $\boldsymbol{\mu}$ will be exposed to too much distortion; if δ is too small, the resulting weight estimator will reduce to the sample covariance matrix based estimator, which is highly vulnerable to the estimation error in $\widehat{\boldsymbol{\Sigma}}$. In the empirical study section, we will discuss the use of a heuristic value for δ.

3.2 Spectral Selection Method

In this section, we discuss an alternative way of approximating the vector of expected returns. We start by presenting an illustrative example which illuminates the new approach.

Example 3.1. Suppose that the expected returns vector has such a relationship with the sample eigenvectors that $\boldsymbol{\mu} = \widehat{\mathbf{u}}^{(1)} + \widehat{\mathbf{u}}^{(p)}$ and that the spectral cut-off method is employed to

obtain a more robust MSR portfolio. We consider what will happen if only the last principal component is cut-off, i.e., when $K = p - 1$. It turns out that in this case the relative approximation error between $\boldsymbol{\mu}$ and $\widehat{\boldsymbol{\mu}}^{cut}(p-1) = \widehat{\mathbf{u}}^{(1)}$ is given by

$$\frac{\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^{cut}(p-1)\|}{\|\boldsymbol{\mu}\|} = \frac{\|\widehat{\mathbf{u}}^{(p)}\|}{\|\widehat{\mathbf{u}}^{(1)} + \widehat{\mathbf{u}}^{(p)}\|} = 0.71.$$

This relative approximation error is obviously too large to be acceptable. Therefore, even if we only cut-off a single principal component, the approximation error incurred is unacceptable, which means that under this particular specification of $\boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}}$, the spectral cut-off method fails to cut off anything.

The failure of the spectral cut-off method in Example 3.1 is attributed to the strong explanatory role of a tail eigenvector. This inspires us to use a selected set of sample eigenvectors to approximate the expected returns vector, where the selection procedure takes into account the explanatory power of each eigenvector. An ideal selection criterion should ensure that sample eigenvectors which contribute more in approximating $\boldsymbol{\mu}$ and estimate their population counterparts relatively well more likely enter the approximation set. Bearing this objective in mind, we propose the following way of approximating the expected returns vector:

$$(\widehat{a}_1^{sel}, \ldots, \widehat{a}_p^{sel}) = \underset{(a_1,\ldots,a_p)}{\arg\min} \frac{1}{2}\left\|\boldsymbol{\mu} - \sum_{i=1}^{p} a_i \widehat{\mathbf{u}}^{(i)}\right\|^2 + \gamma \sum_{i=1}^{p} \frac{|a_i|}{\widehat{\lambda}_i^c}, \tag{14}$$

where $\gamma, c > 0$ are two tuning parameters. With the L_1 penalty in eq. (14), the coefficients before those "less informative" sample eigenvectors which hardly explain the expected returns are coerced to be 0 and thus the approximation set is obtained. It is important to emphasize that unlike in many applications of the L_1 penalty where the "true model" is assumed to be sparse, our motivation for encouraging sparsity is not that we have any clue about how $\boldsymbol{\mu}$ lies in the eigenvector space; rather, we are fully aware that ignoring the penalty term would lead to a perfect fitting, but we intentionally avoid the perfect fitting for the purpose of excluding the highly erroneous sample eigenvectors from the approximation set. Since the

tail sample eigenvalues and eigenvectors are likely to be poorly estimated compared with the head ones, we penalize the a_i's differently such that the eigenvectors associated with the small sample eigenvalues less likely enter the approximation set. The value of c determines the degree of disadvantage faced by tail principal components.

Usually, an optimization problem with the L_1 penalty does not have an explicit solution and is solved through some numerical method. However, owing to the pairwise orthogonality of the sample eigenvectors, which form the "design matrix" in eq. (14), we can find an explicit solution to this optimization problem. The following proposition presents the solution.

Proposition 3.3. The solution to the optimization problem in eq. (14) is given by:

$$\widehat{a}_i^{sel} = \text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+, \quad i = 1, 2, \ldots, p, \tag{15}$$

where $\widehat{a}_i^{ls}$ is defined in eq. (7).

Therefore, the approximation expected returns vector based on the spectral selection method is:

$$\widehat{\boldsymbol{\mu}}^{sel}(\gamma, c) = \sum_{i=1}^{p} \widehat{a}_i^{sel}\widehat{\mathbf{u}}^{(i)} = \sum_{i=1}^{p} \text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+\widehat{\mathbf{u}}^{(i)}. \tag{16}$$

Proposition 3.3 explicitly presents how the spectral selection method shifts the expected returns. According to eq. (15), the spectral selection method adopts an "uneven soft thresholding" scheme to modify the loadings: to enter the approximation set, an eigenvector corresponding to a smaller eigenvalue needs to contribute more to explaining $\boldsymbol{\mu}$ to meet the higher threshold; in addition, for the eigenvectors whose contribution meets their respective thresholds, the threshold value is deducted from the original loading to form the modified loading. Therefore, compared with the spectral cut-off method (recall eq. (8)), which takes the index of an eigenvector as the single decisive factor when specifying the approximation set and keeps the loadings unchanged, the spectral selection method determines the approximation set and adjusts the loadings in a more sophisticated way. This will be discussed in more detail later.

Replacing $\boldsymbol{\mu}$ with the approximation vector $\widehat{\boldsymbol{\mu}}^{sel}(\gamma, c)$ in the "plug-in" method, we obtain the following MSR portfolio:

$$\widehat{\mathbf{w}}^{sel}(\gamma, c) = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma, c)}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma, c)} = \sum_{i=1}^{p} \frac{\frac{\text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c})_+\mathbf{1}^T\widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}{\sum_{i=1}^{p}\frac{\text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c})_+\mathbf{1}^T\widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}\widehat{\mathbf{z}}^{(i)}. \tag{17}$$

According to eq. (17), the MSR portfolio based on the spectral selection method only allocates non-zero weight to eigen portfolios which contribute enough (compared with their respective threshold) to explaining the expected returns. Further, the relative weight of two eigen portfolios that have been selected is modified. Suppose that $\widehat{\mathbf{z}}^{(i)}$ and $\widehat{\mathbf{z}}^{(j)}$ $(i \neq j)$ are two of the eigen portfolios that contribute a non-zero weight to $\widehat{\mathbf{w}}^{sel}(\gamma, c)$. Then, their relative weight in $\widehat{\mathbf{w}}^{sel}(\gamma, c)$ is:

$$\underbrace{\frac{\frac{\text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}{\frac{\text{sign}(\widehat{a}_j^{ls})(|\widehat{a}_j^{ls}|-\gamma\widehat{\lambda}_j^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(j)}}{\widehat{\lambda}_j}}}_{\text{relative weight in } \widehat{\mathbf{w}}^{sel}(\gamma,c)} = \underbrace{\frac{\frac{\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}^{(i)}}{\widehat{\lambda}_i}}{\frac{\widehat{a}_j^{ls}\mathbf{1}^T\widehat{\mathbf{u}}^{(j)}}{\widehat{\lambda}_j}}}_{\text{relative weight in } \widehat{\mathbf{w}}^{msr}} \times \underbrace{\frac{\frac{|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c}}{|\widehat{a}_i^{ls}|}}{\frac{|\widehat{a}_j^{ls}|-\gamma\widehat{\lambda}_j^{-c}}{|\widehat{a}_j^{ls}|}}}_{\text{adjustment factor}}. \tag{18}$$

Note that the first term on the RHS of eq. (18) is the relative weight of the two eigen portfolios in the sample-based MSR portfolio, and therefore the second term represents the spectral selection adjustment. It is easy to check that when $\widehat{\lambda}_i = \widehat{\lambda}_j$, the second term is greater than 1 if and only if $|\widehat{a}_i^{ls}| > |\widehat{a}_j^{ls}|$; when $|\widehat{a}_i^{ls}| = |\widehat{a}_j^{ls}|$, the second term is greater than 1 if and only if $\widehat{\lambda}_i > \widehat{\lambda}_j$. Therefore, the spectral selection method elevates weight of the head eigen portfolios[10] as well as the eigen portfolios contributing more to explaining the expected returns. These two criteria encourage a "closer to convergent" portfolio and in the meanwhile avoid the "blind spot" issue raised in Example 3.1.

[10]Head eigen portfolios refer to those corresponding to the largest eigenvalues.

3.2.1 Selection of tuning parameter

In the spectral selection method, if we fix c, then γ controls the sparsity of the solution as well as how much the approximation vector deviates from the original expected returns. When γ is zero, the resulting portfolio is just the sample-based MSR portfolio; as γ increases, the L_1 penalty encourages an increasingly sparse solution and thus an increasing approximation error. As in the spectral cut-off method, we impose an upper bound δ on the relative approximation error to be introduced. Consequently, we are able to find a critical value of γ which makes the constraint on the approximation error a binding one. Moreover, this particular γ depends on both δ and c, i.e.,

$$\gamma(\delta, c) = \max\left\{\gamma : \frac{\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^{sel}(\gamma, c)\|}{\|\boldsymbol{\mu}\|} \leq \delta\right\}. \tag{19}$$

It is notable that with the L_1 penalty in presence, the approximation vector is no longer a projection of the original vector. Thus, the relative approximation error is not the sine value of an angle anymore. Nevertheless, it is still a reasonable measure of the approximation error.

Since that γ is a function of δ and c and that δ is the tolerance limit we impose, the only remaining task is to determine c, which captures how differently we treat a head eigenvector and a tail one. We resort to the cross-validation method to determine c, because the relative magnitude of eigenvalues can be very different across asset class, study period, and portfolio size, etc. Details about the cross-validation procedure are provided in Section 6. The scheme we have proposed for tuning the parameters is an amelioration of a typical data-driven procedure: we learn how differently we are supposed to treat different eigen portfolios from the data; in the meanwhile, the approximation error to be introduced is always well controlled by δ.

Compared with the spectral cut-off approach, the spectral selection approach works in a wider range of scenarios, especially when the expected returns vector has a significant loading

on some tail sample eigenvectors. This will be illustrated by using a numerical example in Section 5.

3.2.2 Short positions and gross exposure

As we have mentioned earlier, the spectral selection method is motivated by the so-called "blind spot" scenarios where the expected return vector cannot be well approximated without some tail eigenvector and thus the spectral cut-off method fails to work. As the "blind spot" issue arises, the spectral selection method sometimes helps to suppress short positions and reduce the gross exposure, i.e., the L_1 norm, of the MSR portfolio, especially when $\boldsymbol{\mu}$ has a significant loading on both the first and a tail eigenvector. We illustrate this statement by revisiting the Example 3.1. Suppose we can at most tolerate a relative approximation error of 0.1, then, it follows that the portfolios based on the two spectral methods are

$$\widehat{\mathbf{w}}^{cut}(K(0.1)) = \widehat{\mathbf{w}}^{cut}(p) = \frac{\frac{\mathbf{1}^T\widehat{\mathbf{u}}^{(1)}}{\widehat{\lambda}_1}}{\frac{\mathbf{1}^T\widehat{\mathbf{u}}^{(1)}}{\widehat{\lambda}_1} + \frac{\mathbf{1}^T\widehat{\mathbf{u}}^{(p)}}{\widehat{\lambda}_p}}\widehat{\mathbf{z}}^{(1)} + \frac{\frac{\mathbf{1}^T\widehat{\mathbf{u}}^{(p)}}{\widehat{\lambda}_p}}{\frac{\mathbf{1}^T\widehat{\mathbf{u}}^{(1)}}{\widehat{\lambda}_1} + \frac{\mathbf{1}^T\widehat{\mathbf{u}}^{(p)}}{\widehat{\lambda}_p}}\widehat{\mathbf{z}}^{(p)}$$

and

$$\widehat{\mathbf{w}}^{sel}(\gamma(0.1,c),c) = \frac{\frac{(1-\gamma\widehat{\lambda}_1^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(1)}}{\widehat{\lambda}_1}}{\frac{(1-\gamma\widehat{\lambda}_1^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(1)}}{\widehat{\lambda}_1} + \frac{(1-\gamma\widehat{\lambda}_p^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(p)}}{\widehat{\lambda}_p}}\widehat{\mathbf{z}}^{(1)} + \frac{\frac{(1-\gamma\widehat{\lambda}_p^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(p)}}{\widehat{\lambda}_p}}{\frac{(1-\gamma\widehat{\lambda}_1^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(1)}}{\widehat{\lambda}_1} + \frac{(1-\gamma\widehat{\lambda}_p^{-c})\mathbf{1}^T\widehat{\mathbf{u}}^{(p)}}{\widehat{\lambda}_p}}\widehat{\mathbf{z}}^{(p)}$$

respectively. The spectral selection method must keep both eigenvectors because otherwise the approximation error has to exceed the threshold. Comparing the coefficients before the sample eigen portfolios, we see that the spectral selection method based portfolio assigns a higher weight to the dominant eigen portfolio. It has become consensus that the first principal component of the covariance matrix corresponds to the market factor. In addition, empirical findings have shown that when more recent stock returns data is used to estimate the covariance matrix, there is a higher chance that the dominant eigen portfolio does not have any short position (Boyle et al. [2014]). Therefore, $\widehat{\mathbf{z}}^{(1)}$ very likely has no short positions

and a gross exposure of 1; while $\widehat{\mathbf{z}}^{(p)}$ must contain short positions and thus have a greater than 1 gross exposure due to its orthogonality to $\widehat{\mathbf{z}}^{(1)}$. As a result, by shifting more weight to the dominant eigen portfolio, the spectral selection method leads to fewer short positions and a lower gross exposure. It is challenging to formally show the effectiveness of the spectral selection method in eliminating short positions and controlling the gross exposure, so we simply use the special example to illustrate this point. Empirical results in Section 6 will provide additional support for this.

4 Examples of Expected Return Proxies

In this section, we review three methods among many others for choosing a reasonably good proxy for expected returns. These three methods only require publicly available data and therefore provide investors who cannot access proprietary data and models with some options for pinning down a proxy for the expected returns. This helps facilitate the implementation of our proposed method.

4.1 Equal Expected Returns

A simple yet effective choice is to assume that each asset has the same (positive) expected returns:

$$\boldsymbol{\mu} \propto \mathbf{1}. \tag{20}$$

The minimum-variance portfolio can be viewed as an MSR portfolio in which $\mathbf{1}$ is used as the expected returns vector. The empirical finding that minimum-variance portfolios usually yield a satisfactory out-of-sample performance (Jagannathan and Ma [2003], DeMiguel et al. [2009b], Scherer [2010], Clarke et al. [2011]) implies that treating $\mathbf{1}$ as an expected returns vector proxy is not a bad choice. In the empirical studies in Section 6, we will adopt this neutral view on asset returns when constructing MSR portfolios.

4.2 Bayes-Stein Estimator

The Bayes-Stein estimator (Jorion [1986]) is a Bayesian type estimator which addresses the issue of uncertainty about parameter values. The estimator takes the form of a weighted average of the sample mean and a scalar multiple of the column vector of ones:

$$\boldsymbol{\mu} = (1 - \widehat{v})\bar{\mathbf{r}} + \widehat{v}\widehat{r}_0\mathbf{1}, \tag{21}$$

where $\bar{\mathbf{r}}$ is the sample mean vector, and $\widehat{r}_0$ and $\widehat{v}$ are estimated as follows:

$$\widehat{r}_0 = \frac{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{*-1}\bar{\mathbf{r}}}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{*-1}\mathbf{1}},$$

$$\widehat{v} = \frac{p+2}{(p+2) + (\bar{\mathbf{r}} - \widehat{r}_0\mathbf{1})^T n\widehat{\boldsymbol{\Sigma}}^{*-1}(\bar{\mathbf{r}} - \widehat{r}_0\mathbf{1})},$$

where $\widehat{\boldsymbol{\Sigma}}^* = \frac{n-1}{n-p-2}\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix. The Bayes-Stein estimator pushes the estimated returns towards their grand mean so that an investor adopting this particular $\boldsymbol{\mu}$ could avoid entering extreme positions in the assets which have gained/lost a lot in the past. Investors who follow a momentum type strategy may consider using this improved expected return proxy.

4.3 Idiosyncratic Volatility Anomaly

Researchers exploring the cross-sectional pattern of asset returns have found a number of explanatory factors. Ang et al. [2009] find that an asset's lagged idiosyncratic volatility, which is defined as the standard deviation of the residuals from the Fama-French three factor model, negatively predicts the expected return in the next period. This empirical finding suggests a modified Fama-MacBeth procedure to forecast the expected returns. In the first step, the risk factor loadings are estimated from the following regression:

$$r_i = \alpha_i + \beta_i MKT + s_i SMB + h_i HML + \epsilon_i,$$

where r_i is the daily excess return of asset i and the explanatory variables are the usual Fama-French factors. Let $\sigma_i(t-1,t)$ denote asset i's idiosyncratic volatility (standard error of ϵ_i) computed using daily data over the previous month from $t-1$ to t; let $\mathbf{b}_i(t-1,t)$ denote the vector of estimated risk factor loadings over the month $t-1$ to t. Then, the risk premium of each risk factor can be estimated from the second regression:

$$r_i(t,t+1) = c + \gamma\sigma_i(t-1,t) + \boldsymbol{\eta}_b^T\mathbf{b}_i(t-1,t) + \boldsymbol{\eta}_z^T\mathbf{z}_i(t) + u_i(t,t+1).$$

Due to the temporal difference in the explanatory variables and the response variable, we can predict future expected returns based on the regression result.

5 Simulation Study

In this section, we use a set of simulation results to assess the performance of the methods proposed in Section 3.

In the simulation study, we pre-specify the true covariance matrix $\boldsymbol{\Sigma}$ and the true expected return $\boldsymbol{\mu}$. For each set of parameters (sample size n and number of assets p), we repeat the experiment 500 times. In each replication, $2n$ random returns are independently generated from the multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We use the first n observations to train the two spectral methods and determine the MSR portfolios estimators $\widehat{\mathbf{w}}^{cut}(K(\delta))$ and $\widehat{\mathbf{w}}^{sel}(\gamma(\delta,c))$. Then we use the remaining n observations as a test set to assess these portfolios. Different portfolio methods are evaluated based on the distribution of their corresponding out-of-sample Sharpe ratios. Throughout this section, we use $\delta = 0.1$ as the maximum acceptable relative approximation error and $c = 0.5$ as the other parameter. We perform the simulation under four different (n,p) combinations to compare the performance of the spectral methods across dimensionality configurations. The $\boldsymbol{\Sigma}$ matrices specified in the simulation studies are all calibrated from daily returns of S&P 500 stocks using the well-known Fama-French three-factor model. We show how effective the spectral cut-off method

and the spectral selection method are in maximizing the portfolio Sharpe ratio under four different specifications of $\boldsymbol{\mu}$: (1) $\boldsymbol{\mu} \propto \mathbf{1}$, (2) $\boldsymbol{\mu} \propto \mathbf{u}^{(1)}$, (3) $\boldsymbol{\mu} \propto \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \cdots + \mathbf{u}^{(p)}$, and (4) $\boldsymbol{\mu}$ is randomly generated. In all the four cases, $\boldsymbol{\mu}$ is scaled so that the average annual expected return of all assets is 0.4.

Case 1: $\boldsymbol{\mu} \propto \mathbf{1}$

If $\boldsymbol{\mu}$ is a (positive) scalar multiple of $\mathbf{1}$, or alternatively, each asset has the same expected rate of return, the sample-based MSR portfolio reduces to the minimum-variance portfolio, since both portfolios have a weight estimator given by $\frac{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}$. We study this case because this special $\boldsymbol{\mu}$ captures the view of an uninformed investor about $\boldsymbol{\mu}$. In addition, by setting $\boldsymbol{\mu}$ to be proportional to the vector of ones, we do not pre-assume any direct connection between the $\boldsymbol{\mu}$ and the eigenvectors of the (population) covariance matrix. We use this general case to convey an idea about how different methods perform in terms of improving out-of-sample Sharpe ratios. In the subsequent cases, we will specify a concrete relationship between $\boldsymbol{\mu}$ and the eigenvectors and study the behavior of different portfolio estimators.

Figure 1 presents the kernel density plot of the out-of-sample Sharpe ratios. Each of the four panels corresponds to a specific configuration of n and p, and each of the three colors listed in the legend represents a particular method. Note that the darker areas are caused by overlapping of the two colors. The vertical line in each panel is drawn at the true maximum Sharpe ratio. Comparing the four panels, we can make the following observations. First, as the p/n ratio becomes larger, the out-of-sample Sharpe ratios produced by the sample-based MSR portfolio move further away from the true maximum Sharpe ratio. This phenomenon reflects the vulnerability of large-scale portfolios to an estimation error. Second, in the highest-dimensional scenario, both the spectral cut-off method and the spectral selection method outperform the sample-based method, as attested by the observation on Figure 1 that the curves for both spectrum methods are located on the right to that of the sample covariance based method. This is because in a traditional big n and small p scenario, the loss incurred by introducing an approximation error is not compensated by the gain from avoiding

an estimation error, since the latter is not quite obvious. Third, the spectral selection method dominates the spectral cut-off approach under all of the dimensionality settings. The reason behind this dominance is that the spectral selection method is more capable in discarding the "useless" eigenvectors since it is less restrictive on which ones should be discarded.

Figure 1: Kernel density plot of out-of-sample Sharpe ratios for $\boldsymbol{\mu} \propto \mathbf{1}$

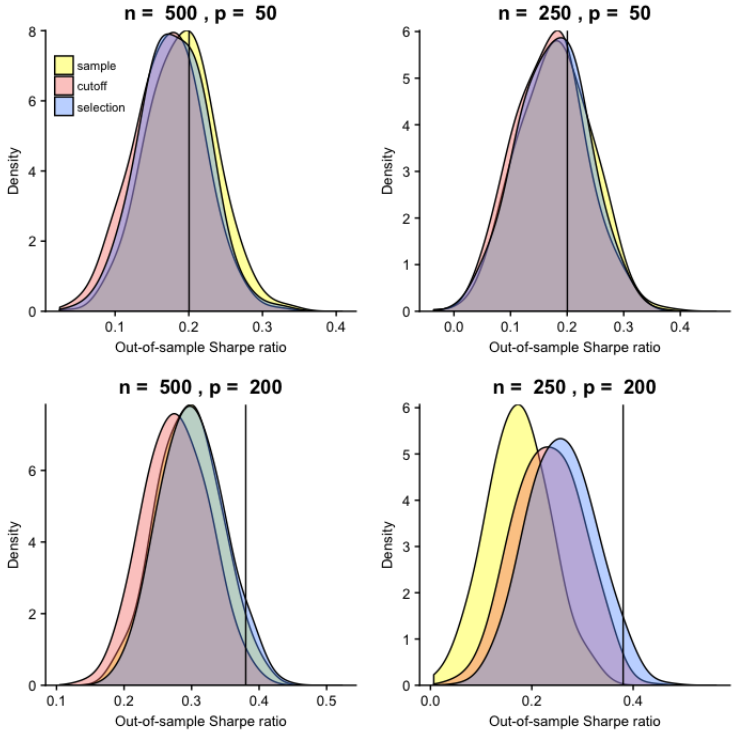
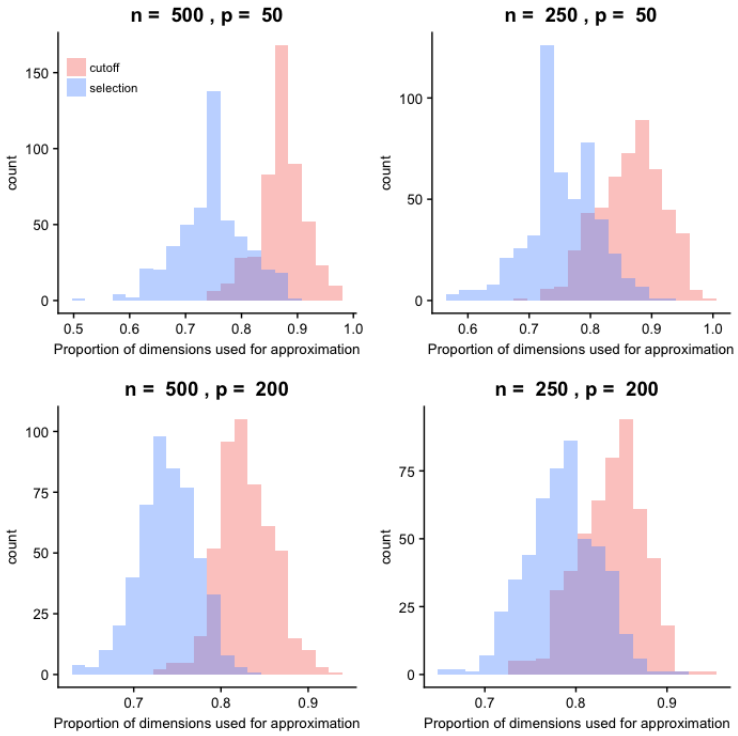


Figure 2 shows the histograms of proportion of dimensions used by the spectral methods to approximate $\boldsymbol{\mu}$. We do not include the sample-based method because it does not involve finding an approximation of $\boldsymbol{\mu}$. The darker regions are a result of overlapping of the two colors. According to the histograms, with the same tuning parameter δ, the spectral selection method clearly leads to a smaller proportion of eigenvectors used in approximating the expected returns.

Figure 2: Histogram of proportion of dimensions used to approximate $\boldsymbol{\mu}$ for $\boldsymbol{\mu} \propto \mathbf{1}$



Case 2: $\boldsymbol{\mu} \propto \mathbf{u}^{(1)}$

Starting from this case, we pre-specify a relationship between $\boldsymbol{\mu}$ and the eigenvectors of $\boldsymbol{\Sigma}$ to see whether the spectral cut-off and the spectral selection method help in improving the out-of-sample Sharpe ratios under different scenarios. In this case, we let $\boldsymbol{\mu}$ be proportional to the dominant eigenvector. This specification is consistent with a single-factor model with a constant residual variance, since under such a model, the dominant eigenvector is proportional to the factor loading, or beta, vector.

Figure 3 shows the kernel density plot of the out-of-sample Sharpe ratios. As in the previous case, as we move towards a high-dimensional setting, both spectral methods improve in terms of the out-of-sample Sharpe ratios. Another notable observation is that the histograms corresponding to the two spectral methods completely overlap. The reason is that in this case, since $\boldsymbol{\mu}$ can be perfectly explained by the dominant population eigenvector, it is highly likely that $\boldsymbol{\mu}$ can be better approximated by the dominant sample eigenvector,

given that the dominant sample eigenvector can be relatively accurately estimated.

Figure 3: Kernel density plot of out-of-sample Sharpe ratios for $\boldsymbol{\mu} \propto \mathbf{u}^{(1)}$

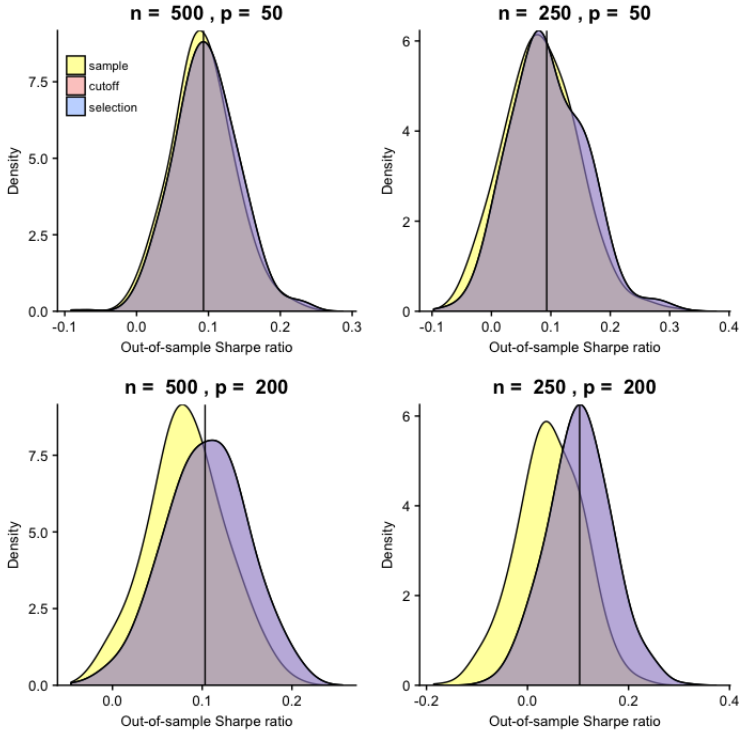
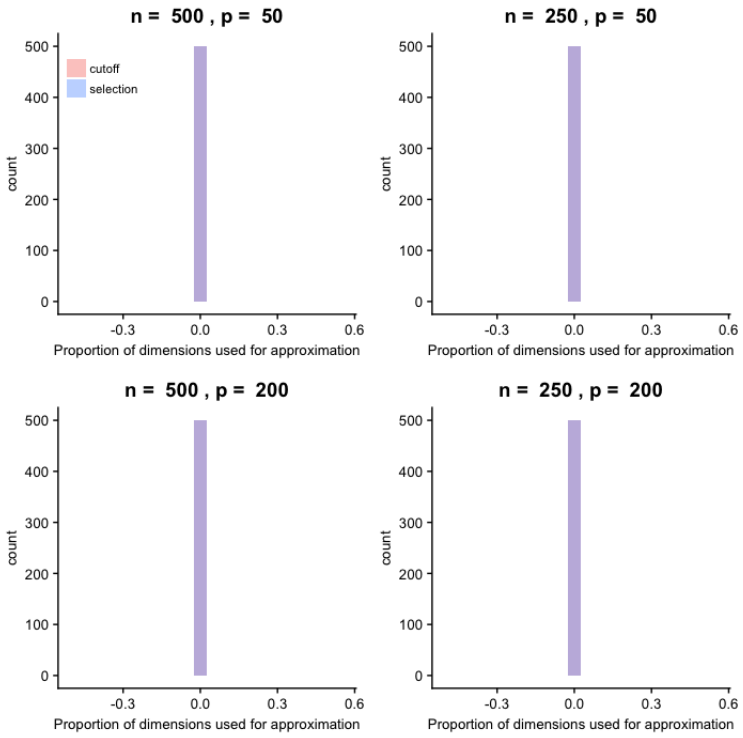


Figure 4 provides support to the above explanation for the overlapping: the histograms corresponding to both spectral methods reduce to a single bar at $\frac{100}{p}\%$, since in all of the dimensionality settings and in all of the replications, the dominant sample eigenvector approximates $\boldsymbol{\mu}$ sufficiently well.

Figure 4: Histogram of proportion of dimensions used to approximate $\boldsymbol{\mu}$ for $\boldsymbol{\mu} \propto \mathbf{u}^{(1)}$

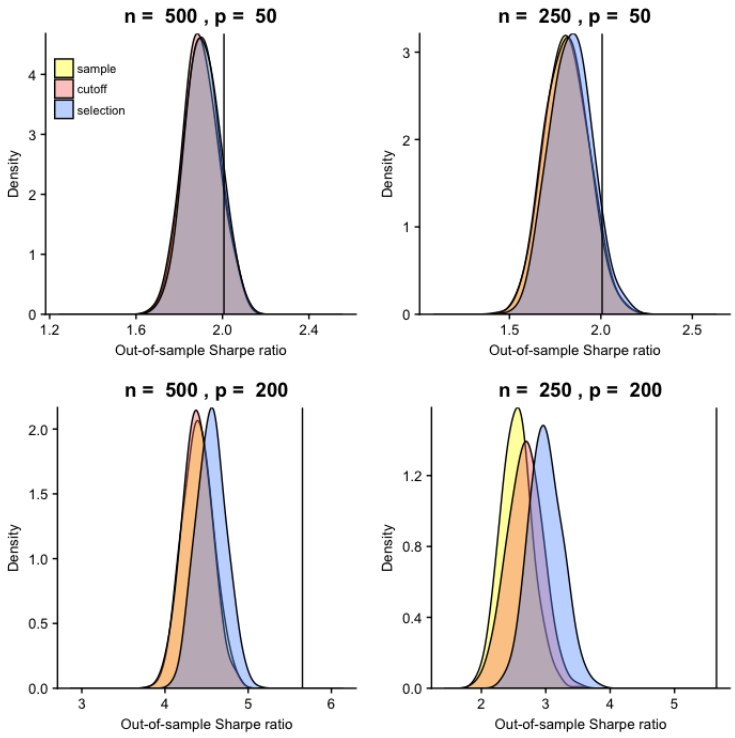


Case 3: $\boldsymbol{\mu} \propto \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \cdots + \mathbf{u}^{(p)}$

In this case, we assume that $\boldsymbol{\mu}$ has an equal loading on all of the eigenvectors of $\boldsymbol{\Sigma}$, i.e., we let $\boldsymbol{\mu} = \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \cdots + \mathbf{u}^{(p)}$. We intend to use this case to illustrate the superiority of the spectral selection method as well as to point out the scenarios where the spectral cut-off method hardly works.

Figure 5 summarizes the distribution of the out-of-sample Sharpe ratios. As in the previous two cases, in the first two "big n and small p" settings (the top panels), the difference between the spectral methods and the sample-based method is not clear cut. In the left bottom panel however, the spectral selection method demonstrates its superiority compared with the other two methods, which have a quite similar performance. The reason behind this similarity is that the equi-loading structure of $\boldsymbol{\mu}$ makes it hard for the spectral cut-off method to cut off much, since otherwise, $\boldsymbol{\mu}$ is not well approximated. As a result, the spectral cut-off method almost reduces to the sample-based method.

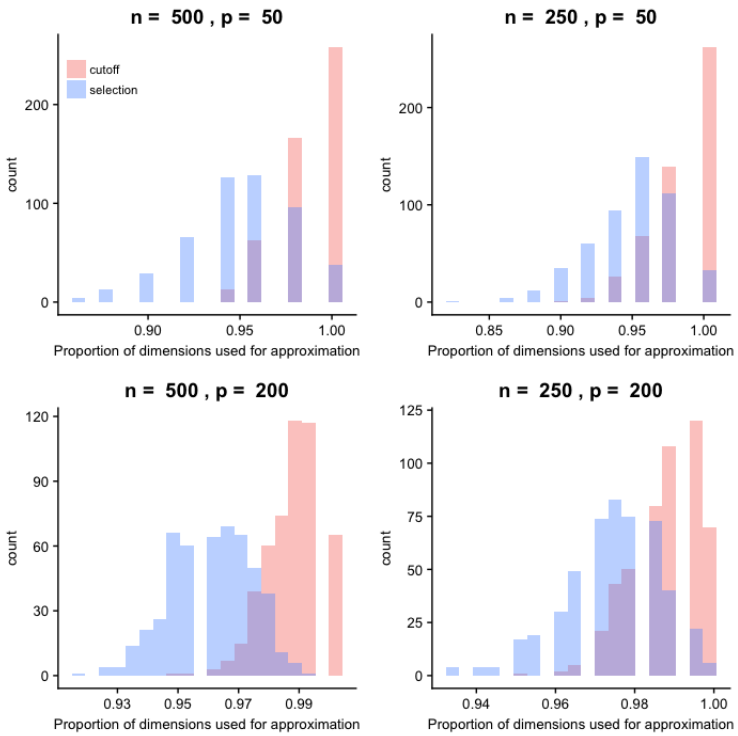
Figure 5: Kernel density plot of out-of-sample Sharpe ratios for $\boldsymbol{\mu} \propto \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \cdots + \mathbf{u}^{(p)}$



The above explanation is further supported by Figure 6. In the top two panels of Figure 6, for the spectral cut-off method, the highest spike appears at around 100%, which means that in most replications, the spectral cut-off method does not result in any dimension reduction. In the two higher-dimensional settings, the highest spike still resides very closely to 100%. While in all the four panels, the highest bar for the spectral selection method is around 97%.

Therefore, when $\boldsymbol{\mu}$ has a heavy loading on one or several of the tail eigenvectors, the spectral cut-off method fails to work and almost reduces to the sample-based method. If this is the case, we need to resort to the more generalized spectral selection method to obtain a robust MSR portfolio.

Figure 6: Histogram of proportion of dimensions used to approximate $\boldsymbol{\mu}$ for $\boldsymbol{\mu} \propto \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \cdots + \mathbf{u}^{(p)}$



Case 4: $\boldsymbol{\mu}$ is randomly generated

In the last case, we consider a $\boldsymbol{\mu}$, each of whose element is independently generated from $N(0.4, 0.4)$, so that around 16% of the assets have negative expected returns. Once $\boldsymbol{\mu}$ has been generated, we treat it as a fixed quantity representing the expected return which the investor believes in and use it across all replications.

According to Figure 7, in the two lower-dimensional settings (the top two panels), there is no significant difference among the three methods. However in the two high-dimensional settings, the spectral selection method clearly leads to the highest average out-of-sample Sharpe ratio, followed by the spectral cut-off method. In addition, it is notable that in the two panels on the right side, we observe negative out-of-sample Sharpe ratios. This happens because we sometimes enter extreme positions but turn out to make incorrect bets. Negative Sharpe ratios are highly undesirable. According to the right bottom panel, the spectral selection method results in the smallest area under the fitted density curve in the

negative half of the x-axis. Therefore, we conclude that the the spectral selection method is the most effective method in producing robust MSR portfolios.

Figure 7: Kernel density plot of out-of-sample Sharpe ratios for randomly generated $\boldsymbol{\mu}$

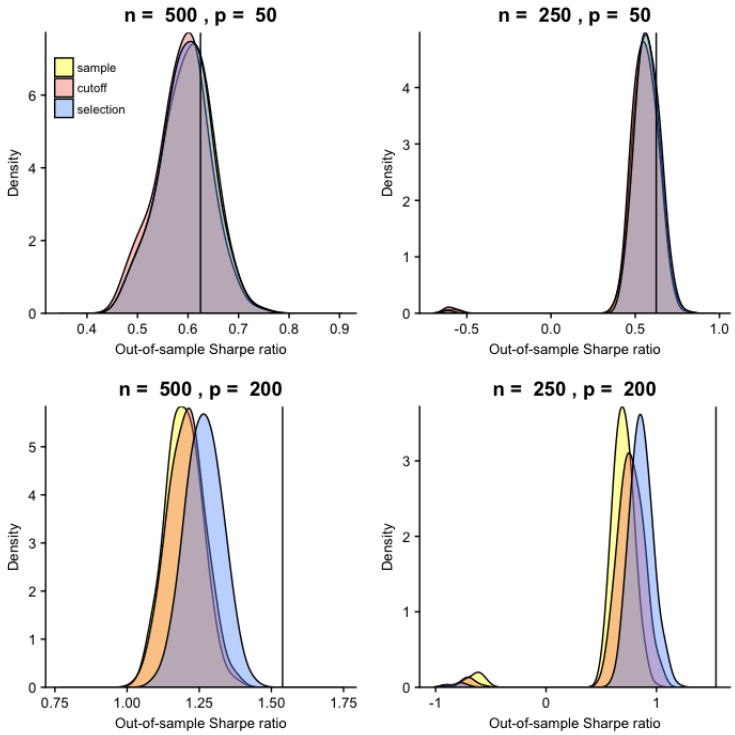
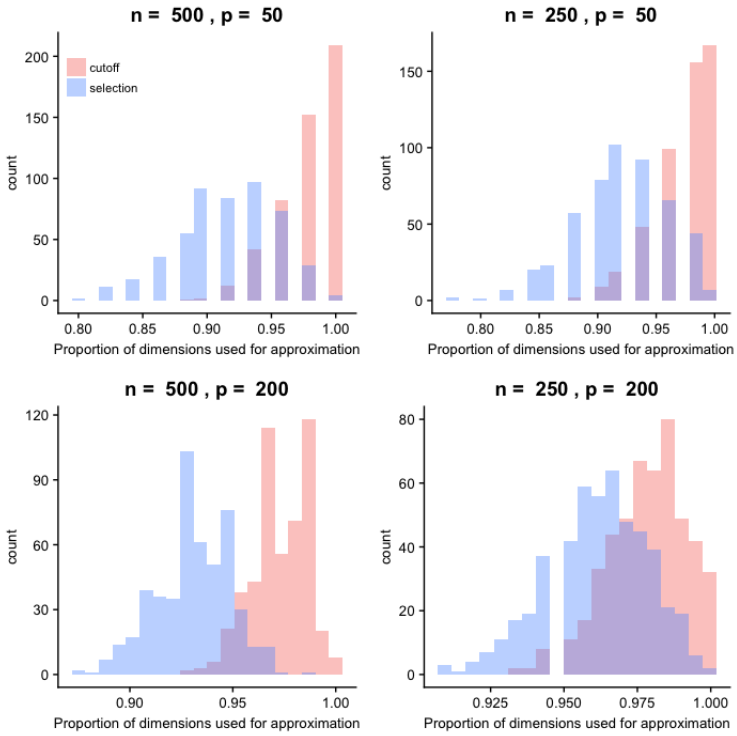


Figure 8 looks quite similar to the histogram in the previous case. The spectral selection method almost always leads to some dimension reduction.

Figure 8: Histogram of proportion of dimensions used to approximate $\boldsymbol{\mu}$ for randomly generated $\boldsymbol{\mu}$



According to the four cases we discuss in this simulation study, by approximating $\boldsymbol{\mu}$ by a number of selected eigen portfolios and allowing the relative approximation error to be less than 10%, the spectral selection method does not lead to a substantial deterioration in out-of-sample Sharpe ratios in the "big n and small p" settings and in addition significantly improve the average out-of-sample Sharpe ratio in a high-dimensional setting. The spectral selection method also effectively reduces the occurrence of negative Sharpe ratios.

6 Empirical Analysis

6.1 Out-of-sample Performance of Portfolios

In this section, we use real world stock returns data from different markets around the world to assess the effectiveness of the spectral cut-off and spectral selection methods in improving portfolio Sharpe ratios.

6.1.1 Data and procedure

We have assumed $\boldsymbol{\mu}$ to be an investor's best proxy for expected returns over the investment horizon and have presented a few potential choices of $\boldsymbol{\mu}$, but in the empirical study, we have to pin down a single desirable expected returns proxy before proceeding to the portfolio construction process. Only when the proxy is close to the unobservable true vector of expected returns, the comparison of different portfolios in terms of their out-of-sample Sharpe ratio would reasonably reflect the effectiveness of each method. The lasting success of the 1/N portfolio, especially in terms of returns (DeMiguel et al. [2009b]), leads us to use $\boldsymbol{\mu} \propto \mathbf{1}$ as the best proxy for expected returns, given that we do not possess any additional information to predict expected returns.

For each market, we select a representative stock index, for instance, S&P 500 index for the US market, and use its constituent stocks to construct portfolios. The same procedure is repeated each year from the starting year of the dataset till $t = 2011$. We use adjusted returns data from the first trading day of year t to the last trading day of year $t + 4$ (n is around 1260) to estimate a stock returns covariance matrix. The stocks that enter the portfolio are those that (1) belong to the index on the last trading day of year $t + 4$ and (2) have at least five years' complete price history. Then on the first trading day in January of year $t + 5$ we build an MSR portfolio based on the estimated covariance matrix and our best proxy for expected returns over the holding period. We hold this portfolio until the last trading day of December of year $t + 5$, at which time we liquidate the portfolio and start the process all over again. We use daily returns of the S&P 500 component stocks from January 1984 to December 2016 to back-test the performance of different portfolio methods in the US market. As a result, each portfolio has a 28-year holding period from January 1989 to December 2016. The other two datasets we use are daily returns of the S&P United Kingdom index constituents and those of the Japanese Nikkei 225 index constituents, both running from January 2001 to December 2016.

6.1.2 Comparison methods

We compare in total eight portfolio methods in the empirical study. Besides the spectral cut-off and spectral selection methods, the comparison methods include the well-known equally weighted portfolio (also known as the $1/N$ portfolio), the sample-based MSR portfolio, the POET-based (Fan et al. [2013]) MSR portfolio[11], the shrinkage towards identity method based MSR portfolio (Ledoit and Wolf [2004]), as well as the no-short-selling MSR portfolio (Jagannathan and Ma [2003]), whose outperformance among optimized portfolios[12] has been documented in DeMiguel et al. [2009b]. All of the MSR portfolios are constructed based on the same given $\boldsymbol{\mu}$.

For the spectral cut-off method, we consider two ways of selecting the parameter δ. One is to determine the optimal δ at each portfolio rebalancing date via a cross-validation procedure which is described as follows.

- Partition the available returns data into a training set and a cross validation set. The cross validation set contains the most recent 20% of the data. Calculate the returns sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$ using the training data.

- For each value of δ of interest, construct an MSR portfolio using the spectral cut-off method with parameter δ, i.e., $\widehat{\mathbf{w}}^{cut}(K(\delta)) = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{cut}(K(\delta))}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{cut}(K(\delta))}$, and compute $SR(\delta)$, which is the Sharpe ratio of the portfolio on the cross validation set. The optimal δ is

$$\delta_{CV} = \underset{\delta}{\operatorname{argmax}}\, SR(\delta).$$

- As the portfolio is frequently rebalanced we record a sequence of δ_{CV}. Let $\bar{\delta}_{CV}$ stand for the average of the sequence.

[11] When using the POET method to estimate the covariance matrix, we use Bai and Ng's method to determine the number of common factors and use the recommended parameter $C = 0.5$ to estimate the residual covariance matrix using the soft thresholding method.

[12] We refer to the portfolios that involve an explicit optimization procedure as "optimized portfolios". Examples include the minimum-variance portfolio, the MSR portfolio, etc. In contrast, portfolios such as the equally weighted portfolio or the value weighted portfolio are not optimized portfolios.

This method of selecting δ is equivalent to the conventional method of using a data-driven method to select K each time. As discussed before, adopting the cross-validation approach could cause the approximation error to be out of control. The second way is to directly set $\delta = 0.15$. This approach ensures that the approximation error never exceeds our tolerance limit.

For the spectral selection method, we only consider a single choice of δ: we simply use $\delta = 0.15$. The reason is that in this method we have an additional tuning parameter c to determine via the following cross-validation procedure:

- Partition the available returns data into a training set and a cross validation set. The cross validation set contains the most recent 20% of the data. Calculate the returns sample covariance matrix $\widehat{\mathbf{\Sigma}}$ using the training data.

- For each value of c of interest, construct an MSR portfolio using the spectral selection method with parameter c and $\delta = 0.15$, i.e., $\widehat{\mathbf{w}}^{sel}(\gamma(\delta, c), c) = \frac{\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma(\delta,c),c)}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma(\delta,c),c)}$, and compute $SR(c)$, which is the Sharpe ratio of the portfolio on the cross validation set. The optimal c is

$$c_{CV} = \underset{c}{\operatorname{argmax}}\, SR(c).$$

6.1.3 Performance measures and results

The following holding period performance measures are recorded for each portfolio: annualized standard deviation, annualized return, Sharpe ratio, average percentage of short position, average turnover, average gross exposure (Fan et al. [2012]), as well as the percentage of dimensions used to approximate $\boldsymbol{\mu}$ (for the two spectral methods). All the reported performance measures are adjusted for transaction costs. We set the proportional transaction costs equal to 50 basis points per transaction as assumed in Balduzzi and Lynch [1999] and in DeMiguel et al. [2009b]. If we denote by s the proportional transaction cost, then the

evolution of wealth for a portfolio strategy k is

$$W_{k,t+1} = W_{k,t}(1 + R_{k,t+1})\left(1 - s\sum_{j=1}^{N}|\hat{w}_{k,j,t+1} - \hat{w}_{k,j,t}|\right),$$

where $R_{k,t+1}$ is the portfolio return under strategy k during the period from t to $t+1$, and $\hat{w}_{k,j,t}$ is the weight of the jth asset at time t according to strategy k. The transaction cost adjusted Sharpe ratio is the performance measure we highlight in this paper.

Tables 1 - 3 summarize the holding period performance measures of all the portfolios. According to these tables, in all the three markets, the spectral selection method with $\delta = 0.15$ leads to a holding period Sharpe ratio higher than that of the equally weighted portfolio, the sample-based MSR portfolio, the POET-based MSR portfolio, the shrinkage method based MSR portfolio, the no-short-selling MSR portfolio, as well as the spectral cut-off method with $\delta = 0.15$. In addition, in two out of the three markets, the spectral selection method with $\delta = 0.15$ results in a higher Sharpe ratio compared with the conventional spectral cut-off method ($\delta = \delta_{CV}$), which has an unstable performance in different markets. In a nutshell, the spectral selection method with $\delta = 0.15$ yields the most satisfying performance and is robust to the choice of country, study period, and more importantly, the dimensionality of the training sample.

Recall that on each portfolio balancing date, we always use the past five years' data to estimate the covariance matrix and construct the portfolios, but the portfolio size varies across markets since different indices have different numbers of component stocks: the S&P 500 index has 500; the S&P United Kingdom index has around 90; the Nikkei 225 index has 225. Thus, according to the relative magnitude of the portfolio size and the sample size of the training data, the optimized portfolios from the UK market are faced with a less severe estimation error problem. This is evidenced by the high Sharpe ratio yielded by the sample-based MSR portfolio in the UK market (see Table 2). However, even in this case, introducing a 0.15 relative approximation error by neither spectral method causes an

obvious deterioration (compared with the sample-based MSR portfolio) in the holding period Sharpe ratio. Therefore, $\delta = 0.15$ is a reasonable upper bound to impose on the relative approximation error.

The average turnover is another important performance measure and is closely associated with another two performance measures: the percentage of short positions and the gross exposure. Unsurprisingly, among all of the methods, the equally weighted portfolio always has the lowest average turnover, followed by the no-short-selling MSR portfolio, since these two portfolios have positive weights and do not undergo a massive adjustment on portfolio rebalancing dates. The spectral cut-off method with $\delta = 0.15$ leads to the third lowest turnover, followed by the spectral selection method with $\delta = 0.15$. This is because the cross-validation procedure for determining the tuning parameter c in the spectral selection method brings in fluctuation in the parameter. But the turnover is not drastically increased thanks to the upper bound on the approximation error; otherwise the approximation vector of expected returns could vary substantially from time to time, leading to an intolerable turnover. The POET method and the shrinkage to identity method help eliminate extreme positions resulted from the sample-based method by regularizing the sample covariance matrix and thus pull down the value of the turnover a bit. The spectral cut-off method with $\delta = \delta_{CV}$ results in a high average turnover, comparable to or even higher than that of the sample-based MSR portfolio. This high turnover arises because the cross-validation procedure performed on each portfolio rebalancing date brings in extra instability to the parameter.

So far we can see that the spectral selection method with $\delta = 0.15$ is the most preferable one among all of the portfolio methods. But before formally recommending the spectral methods with this specific parameter, in the following section, we use more datasets of different dimensionalities to assess whether this parameter works well in other scenarios.

Table 1: Holding period (Jan 1989 - Dec 2016) performance of different portfolios of S&P 500 index component stocks

Method	std dev	return	Sharpe	% short position	turnover	gross exp	% dim used
equally weighted	17.62%	8.44%	0.48	0.00%	0.14	1.00	—
sample-based	12.67%	1.81%	0.14	47.48%	2.84	6.80	—
spectral cut-off ($\delta = \delta_{CV}$)	13.18%	3.37%	0.26	36.74%	2.47	3.95	45.21%
spectral cut-off ($\delta = 0.15$)	13.33%	6.20%	0.46	30.11%	0.61	1.95	4.45%
spectral selection ($\delta = 0.15$)	12.70%	6.33%	0.50	29.59%	0.70	1.94	7.67%
POET	11.48%	3.51%	0.31	45.58%	1.27	4.02	—
shrink to identity	12.22%	2.23%	0.18	46.85%	2.53	6.29	—
sample-based no short	18.97%	8.44%	0.44	0.02%	0.14	1.00	—

Table 2: Holding period (Jan 2006 - Dec 2016) performance of different portfolios of S&P United Kingdom index component stocks

Method	std dev	return	Sharpe	% short position	turnover	gross exp	% dim used
equally weighted	20.14%	4.80%	0.24	0.00%	0.12	1.00	—
sample-based	13.53%	3.76%	0.28	40.29%	0.71	2.34	—
spectral cut-off ($\delta = \delta_{CV}$)	14.99%	4.95%	0.33	31.93%	1.14	2.02	47.50%
spectral cut-off ($\delta = 0.15$)	16.06%	4.16%	0.26	28.39%	0.47	1.72	10.72%
spectral selection ($\delta = 0.15$)	14.83%	4.37%	0.29	27.05%	0.47	1.59	13.23%
POET	16.49%	2.11%	0.13	41.26%	0.63	2.25	—
shrink to identity	13.50%	3.82%	0.28	39.83%	0.69	2.31	—
sample-based no short	21.88%	4.33%	0.20	0.00%	0.12	1.00	—

Table 3: Holding period (Jan 2006 - Dec 2016) performance of different portfolios of Nikkei 225 index component stocks

Method	std dev	return	Sharpe	% short position	turnover	gross exp	% dim used
equally weighted	25.20%	2.26%	0.09	0.00%	0.11	1.00	—
sample-based	18.37%	0.78%	0.04	45.77%	1.80	5.18	—
spectral cut-off ($\delta = \delta_{CV}$)	19.40%	0.52%	0.03	38.68%	2.10	3.07	39.72%
spectral cut-off ($\delta = 0.15$)	21.65%	4.31%	0.20	28.95%	0.37	1.68	1.85%
spectral selection ($\delta = 0.15$)	21.20%	4.55%	0.21	29.86%	0.44	1.71	4.33%
POET	18.37%	2.03%	0.11	44.81%	1.05	3.77	—
shrink to identity	18.26%	0.92%	0.05	45.72%	1.73	5.05	—
sample-based no short	26.92%	1.61%	0.06	0.00%	0.11	1.00	—

6.2 A Rule of Thumb for Selecting δ

Since the cross validation method for selecting δ does not necessarily outperform the less complicated method of directly adopting a reasonable value of δ, as can be seen from Tables 1 - 3, we recommend a value, $\delta = 0.15$, as a rule of thumb. The rest of this section is devoted to using real-world datasets of different sizes to show that this δ works well in a wide range of dimensionality configurations, as well as in different markets.

Tables 4, 6, and 8 summarize the portfolios' average holding period performance measures when different number of stocks enter the portfolios and the training sets contain different number of days, with each table corresponding to a market. The portfolio construction procedure is the same as that described in Section 6.1, except that instead of always using

the past five years' daily returns data to construct portfolios, we use the past $n/252$ years' data. In addition, within each of Tables 4, 6, and 8, we make the holding period under each of the four dimensionality configurations the same for a comparison purpose. Since there are multiple ways to choose p stocks from an index's component stocks, for each pair of (p, n), we repeat the random draw 20 times, perform the portfolio construction procedure each time when p stocks are drawn, and record the twenty-time average holding period annualized standard deviation, return, Sharpe ratio, and average turnover when each portfolio method is used. As in the previous section, all performance measures are reported after adjusted for transaction costs. In addition, for each (p, n) combination, we also record the average $\bar{\delta}_{CV}$ of the 20 replications.

According to the results in Tables 4, 6, and 8, the outperformance of the spectral selection method with $\delta = 0.15$ is quite consistent across different dimensionality configurations, almost always ranking the first or the second among all of the portfolio methods. Even in the few cases where the spectral selection method with $\delta = 0.15$ is not the top performer, it does not cause obvious deterioration in the Sharpe ratio. In addition, the spectral selection method with $\delta = 0.15$ almost always leads to a low turnover, only higher than the two all-positive portfolios. Moreover, the superiority of the spectral methods in high-dimensional settings is manifested in the right bottom block of each of Tables 4, 6, and 8. A closer scrutiny of the portfolios' performance in high-dimensional settings warns us against using the sample-based method, since it may lead to a low portfolio return. A possible reason for the low return is that inverting the large sample covariance matrix brings in extreme long and short positions, and incorrect bets on these positions cause severe loss.

Tables 5, 7, and 9 summarize the average $\bar{\delta}_{CV}$ obtained from the cross-validation procedure for determining δ in the spectral cut-off method with $\delta = \delta_{CV}$ under different dimensionality configurations, each table corresponding to a market. The numbers fluctuate between 0.08 and 0.13, which provides an additional support for using $\delta = 0.15$ as a rule of thumb. In addition, these empirical results demonstrate that the amount of approximation

error that should be introduced is not very sensitive to the relative magnitude of n and p, therefore we simply recommend a parameter value independent of n and p. It should be noticed that even if for a (p, n) pair, the average $\bar{\delta}_{CV}$ is exactly 0.15, the spectral cut-off method with $\delta = \delta_{CV}$ and that with $\delta = 0.15$ are fundamentally different, because in the former method the δ_{CV} obtained at each portfolio rebalancing date is different, and 0.15 is just the mean of the sequence of δ_{CV}.

Table 4: Average holding period (Jan 1995 - Dec 2016) Sharpe ratio of different portfolios when p stocks are used to construct portfolios. All stocks are S&P 500 constituents.

	$p = 50$, $n = 1260$				$p = 50$, $n = 504$			
Method	std	return	Sharpe	turnover	std	return	Sharpe	turnover
equally weighted	18.10%	8.46%	0.47	0.14	18.12%	8.62%	0.48	0.14
sample-based	13.83%	6.00%	0.43	0.43	14.02%	5.45%	0.39	0.77
spectral cut-off ($\delta = \delta_{CV}$)	15.13%	6.46%	0.43	0.70	14.81%	6.05%	0.41	0.85
spectral cut-off ($\delta = 0.15$)	15.64%	7.51%	0.48	0.40	15.07%	6.89%	0.46	0.57
spectral selection ($\delta = 0.15$)	15.20%	7.90%	0.52	0.35	14.54%	7.19%	0.49	0.47
POET	14.38%	5.64%	0.39	0.40	14.25%	5.12%	0.36	0.61
shrink to identity	13.78%	6.09%	0.44	0.42	13.85%	5.65%	0.41	0.71
sample-based no short	19.96%	8.34%	0.42	0.14	20.27%	8.52%	0.42	0.16
	$p = 100$, $n = 1260$				$p = 100$, $n = 504$			
Method	std	return	Sharpe	turnover	std	return	Sharpe	turnover
equally weighted	18.08%	8.30%	0.46	0.14	17.84%	8.40%	0.47	0.14
sample-based	13.03%	5.11%	0.39	0.69	13.33%	4.28%	0.32	1.35
spectral cut-off ($\delta = \delta_{CV}$)	14.77%	5.61%	0.38	0.98	14.17%	5.47%	0.39	1.26
spectral cut-off ($\delta = 0.15$)	14.87%	7.65%	0.51	0.46	13.85%	6.45%	0.47	0.71
spectral selection ($\delta = 0.15$)	14.56%	7.68%	0.53	0.42	13.42%	6.67%	0.50	0.61
POET	13.54%	5.46%	0.40	0.54	12.90%	4.16%	0.32	0.82
shrink to identity	12.97%	5.25%	0.41	0.67	12.99%	4.56%	0.35	1.20
sample-based no short	19.70%	8.35%	0.42	0.14	19.84%	8.10%	0.41	0.16

Table 5: Average $\bar{\delta}_{CV}$ under different dimensionality configurations. All stocks are S&P 500 index constituents.

(p, n)	δ_{CV}
$(50, 1260)$	0.12
$(50, 504)$	0.12
$(100, 1260)$	0.12
$(100, 504)$	0.11

Table 6: Average holding period (Jan 2006 - Dec 2016) Sharpe ratio of different portfolios when p stocks are used to construct portfolios. All stocks are S&P United Kingdom index constituents.

	$p = 40$, $n = 1260$				$p = 40$, $n = 252$			
Method	std	return	Sharpe	turnover	std	return	Sharpe	turnover
equally weighted	20.78%	4.74%	0.23	0.12	20.25%	4.66%	0.23	0.12
sample-based	15.23%	4.24%	0.28	0.42	15.58%	2.99%	0.19	1.14
spectral cut-off ($\delta = \delta_{CV}$)	16.60%	4.07%	0.25	0.68	16.08%	3.61%	0.22	0.98
spectral cut-off ($\delta = 0.15$)	17.22%	5.42%	0.32	0.43	16.32%	3.18%	0.20	0.92
spectral selection ($\delta = 0.15$)	16.69%	5.35%	0.32	0.34	15.02%	3.53%	0.23	0.70
POET	16.52%	3.32%	0.20	0.45	16.14%	3.71%	0.23	0.80
shrink to identity	15.20%	4.26%	0.28	0.41	15.27%	3.17%	0.21	1.04
sample-based no short	22.57%	4.26%	0.19	0.13	22.39%	4.67%	0.21	0.16
	$p = 80$, $n = 1260$				$p = 80$, $n = 252$			
Method	std	return	Sharpe	turnover	std	return	Sharpe	turnover
equally weighted	20.15%	4.71%	0.23	0.12	20.09%	4.78%	0.24	0.12
sample-based	13.62%	3.75%	0.28	0.65	14.92%	1.95%	0.13	1.93
spectral cut-off ($\delta = \delta_{CV}$)	14.42%	3.84%	0.27	0.89	15.10%	3.34%	0.22	1.47
spectral cut-off ($\delta = 0.15$)	15.86%	4.51%	0.28	0.45	15.05%	2.74%	0.18	1.20
spectral selection ($\delta = 0.15$)	15.03%	4.74%	0.32	0.42	14.01%	2.58%	0.18	1.12
POET	16.26%	2.06%	0.13	0.60	14.97%	3.66%	0.24	0.98
shrink to identity	13.59%	3.80%	0.28	0.64	14.25%	2.54%	0.18	1.64
sample-based no short	21.86%	4.30%	0.20	0.12	22.10%	4.71%	0.21	0.16

Table 7: Average $\bar{\delta}_{CV}$ under different dimensionality configurations. All stocks are S&P United Kingdom index constituents.

(p, n)	δ_{CV}
$(40, 1260)$	0.12
$(40, 252)$	0.12
$(80, 1260)$	0.12
$(80, 252)$	0.13

Table 8: Average holding period (Jan 2006 - Dec 2016) Sharpe ratio of different portfolios when p stocks are used to construct portfolios. All stocks are Japanese Nikkei 225 index constituents.

	$p = 50$, $n = 1260$				$p = 50$, $n = 252$			
Method	std	return	Sharpe	turnover	std	return	Sharpe	turnover
equally weighted	25.57%	2.22%	0.09	0.11	25.41%	1.98%	0.08	0.11
sample-based	19.90%	2.94%	0.15	0.48	20.92%	2.71%	0.13	1.75
spectral cut-off ($\delta = \delta_{CV}$)	21.01%	2.81%	0.13	0.83	21.29%	2.64%	0.12	1.43
spectral cut-off ($\delta = 0.15$)	22.88%	3.74%	0.16	0.33	21.97%	2.21%	0.10	0.75
spectral selection ($\delta = 0.15$)	22.62%	3.74%	0.17	0.27	21.74%	2.57%	0.12	0.59
POET	20.47%	2.82%	0.14	0.49	20.20%	3.43%	0.17	1.17
shrink to identity	19.88%	2.99%	0.15	0.47	20.48%	2.99%	0.15	1.54
sample-based no short	27.20%	1.57%	0.06	0.11	27.34%	1.41%	0.05	0.14
	$p = 100$, $n = 1260$				$p = 100$, $n = 252$			
Method	std	return	Sharpe	turnover	std	return	Sharpe	turnover
equally weighted	25.36%	2.40%	0.09	0.11	25.39%	1.92%	0.08	0.11
sample-based	18.77%	2.05%	0.11	0.85	21.85%	1.30%	0.06	3.31
spectral cut-off ($\delta = \delta_{CV}$)	20.00%	2.72%	0.14	1.16	21.64%	0.85%	0.04	2.33
spectral cut-off ($\delta = 0.15$)	22.27%	3.62%	0.16	0.35	20.91%	2.40%	0.11	0.88
spectral selection ($\delta = 0.15$)	21.85%	3.90%	0.18	0.34	20.63%	2.87%	0.14	0.78
POET	19.41%	1.86%	0.10	0.68	18.90%	2.76%	0.15	1.54
shrink to identity	18.73%	2.14%	0.11	0.82	20.55%	2.05%	0.10	2.70
sample-based no short	27.09%	1.75%	0.06	0.12	27.33%	1.36%	0.05	0.14

Table 9: Average $\bar{\delta}_{CV}$ under different dimensionality configurations. All stocks are Japanese Nikkei 225 index constituents.

(p, n)	$\bar{\delta}_{CV}$
$(40, 1260)$	0.08
$(40, 252)$	0.09
$(80, 1260)$	0.08
$(80, 252)$	0.09

Up to this point, we have shown the suitability of $\delta = 0.15$ as the parameter value in the spectral cut-off and spectral selection methods under different dimensionality configurations and in different markets. However, this δ is only recommended when $\boldsymbol{\mu}$ is set to $\boldsymbol{\mu} \propto \mathbf{1}$, alternatively, when we do not possess any additional information to forecast $\boldsymbol{\mu}$. If an investor has a view on asset returns other than $\boldsymbol{\mu} \propto \mathbf{1}$, a good choice is to obtain an average value of cross-validated parameter, $\bar{\delta}_{CV}$, from the historical data and then use this $\bar{\delta}_{CV}$ as the fixed upper bound for the relative approximation error when constructing portfolios. When the sample size is sufficiently large compared with the number of assets, we recommend a smaller δ as well because in such a scenario the estimation error does not pose a severe problem.

7 Conclusion

In this paper, we discover that if the expected returns vector lies in a subspace of the eigenvector space of the sample covariance matrix, the sample-based MSR portfolio also lies in the same subspace. Due to the uneven distribution of estimation errors across different sample eigenvalues and eigenvectors, it is desirable that the portfolio estimator lies in a space spanned by a few sample eigenvectors that relatively well estimate their population counterparts. Therefore, we propose the idea of approximating the expected returns vector in a lower-dimensional subspace. We then use this approximation to replace the original expected returns vector. As long as the approximation is close to the original vector, we benefit from the reduced exposure to the estimation error without much loss, although in the process our vector of expected returns is slighted distorted.

We introduce two concrete methods for approximating the expected returns vector. The first one, which has been shown to be equivalent to the spectral cut-off method in the literature, uses the first K sample eigenvectors to approximate the expected returns. This particular choice of the approximation set is due to the fact that the leading eigenvalues and their corresponding eigenvectors are relatively precisely estimated. The second one, namely the spectral selection method, uses a selected set of sample eigenvectors to approximate the expected returns. The selected sample eigenvectors tend to be more useful in explaining the expected returns and correspond to a larger sample eigenvalue.

In both spectral methods, we specify an upper bound δ for the approximation error to be introduced. There are a few advantages of treating δ as the parameter. The most important one is that by constraining the approximation error, our view on the expected returns vector is not severely distorted. Such a turning parameter δ offers a convenient scheme for striving the trade-off between the approximation error and the estimation error, as it enables us to set a limit for the former and make our best effort to reduce the latter.

Extensive simulation studies are conducted to demonstrate the superiority of the spectral methods. Both spectral methods mitigate the effect of the estimation error more effectively in a high-dimensional setting. The reason behind this is that when the sample size is sufficiently large compared with the number of assets, the estimation error is not so ubiquitous. In this scenario, the introduction of an approximation error is not warranted.

We use three real-world stock returns datasets to assess the effectiveness of the two spectral methods. These datasets are of different dimensionality configurations and from different markets around the world. Since the $1/N$ portfolio usually yields impressive returns, we deduce that the weight vector of the equally weighted portfolio and the expected returns vector usually form an acute angle. Therefore, we set the vector of expected returns to be $\mathbf{1}$. It turns out that the spectral selection method with $\delta = 0.15$ yields better transaction costs adjusted holding period Sharpe ratios even compared with the renowned $1/N$ portfolio, with only few exceptions. The suitability of this δ under different dimensionality settings and in

different markets is evidenced by the numerical results presented in the paper.

References

Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, 91(1):1–23, 2009.

Marco Avellaneda and Jeong-Hyun Lee. Statistical arbitrage in the us equities market. *Quantitative Finance*, 10(7):761–782, 2010.

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

Pierluigi Balduzzi and Anthony W Lynch. Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, 52(1):47–78, 1999.

Phelim Boyle. Positive weights on the efficient frontier. *North American Actuarial Journal*, 18(4):462–477, 2014.

Phelim P Boyle, Shui Feng, David Melkuev, and Johnew Zhang. Correlation matrices with the perron-frobenius property. 2014.

Marine Carrasco and Nérée Noumon. Optimal portfolio selection using regularization. Technical report, Discussion paper, 2011.

Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics*, 6:5633–5751, 2007.

Jiaqin Chen and Ming Yuan. Efficient portfolio selection in a large market. *Journal of Financial Econometrics*, 14(3):496–524, 2016.

Roger Clarke, Harindra De Silva, and Steven Thorley. Minimum-variance portfolio composition. *The Journal of Portfolio Management*, 37(2):31–45, 2011.

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

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5), 2009b.

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

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

Gabriel Frahm and Christoph Memmel. Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, 159(2):289–302, 2010.

Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683, 2003.

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

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

Petter N Kolm, Reha Tütüncü, and Frank J Fabozzi. 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, 234 (2):356–371, 2014.

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

Olivier Ledoit and Michael Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, 2017.

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

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

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

Alexei Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2012.

M Hossein Partovi, Michael Caputo, et al. Principal portfolios: Recasting the efficient frontier. *Economics Bulletin*, 7(3):1–10, 2004.

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

Dan Shen, Haipeng Shen, and JS Marron. A general framework for consistency of principal component analysis. *Journal of Machine Learning Research*, 17(150):1–34, 2016.

A Steele. On the eigen structure of the mean variance efficient set. *Journal of Business Finance & Accounting*, 22(2):245–255, 1995.

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

A Proofs of the technical results in Section 2

Proof of Proposition 2.1. To prove this proposition, we need to first verify that the global minimum-variance portfolio has a positive expected return and then plug in the vector of expected returns into eq. (2) and see whether the resulting MSR portfolio matches the corresponding eigen portfolio.

Let $\boldsymbol{\mu} = a\mathbf{u}^{(i)}$, $a \in \{a \neq 0 : a\mathbf{1}^T\mathbf{u}^{(i)} > 0\}$, then the expected return of the global minimum-variance portfolio $\mathbf{w}^{mv}$ is

$$\begin{aligned}\mu_{mv} = \mathbf{w}^{mvT}\boldsymbol{\mu} = \frac{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} = \frac{a\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}^{(i)}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} = \frac{a\mathbf{1}^T\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{u}^{(i)}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} \\ = \frac{a\mathbf{1}^T\mathbf{U}\boldsymbol{\Lambda}^{-1}(0,\ldots,0,\overset{ith}{1},0,\ldots,0)^T}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} = \frac{\frac{1}{\lambda_i}a\mathbf{1}^T\mathbf{u}^{(i)}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} > 0.\end{aligned}$$

Since it is verified that the global minimum-variance portfolio has a positive expected return, it follows that the MSR portfolio is determined by eq. (2):

$$\begin{aligned}\mathbf{w}^{msr} &= \frac{a\boldsymbol{\Sigma}^{-1}\mathbf{u}^{(i)}}{a\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{u}^{(i)}} = \frac{\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{u}^{(i)}}{\mathbf{1}^T\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{u}^{(i)}} \\ &= \frac{\mathbf{U}\Lambda^{-1}(0,\ldots,0,\overset{ith}{1},0,\ldots,0)^T}{\mathbf{1}^T\mathbf{U}\boldsymbol{\Lambda}^{-1}(0,\ldots,0,\underset{ith}{1},0,\ldots,0)^T} = \frac{\frac{1}{\lambda_i}\mathbf{u}^{(i)}}{\frac{1}{\lambda_i}\mathbf{1}^T\mathbf{u}^{(i)}} = \mathbf{z}^{(i)}.\end{aligned}$$

Thus we have shown that the MSR portfolio is exactly the ith eigen portfolio. □

Proof of Proposition 2.2. We first verify the existence of the MSR portfolio. The expected return of the minimum-variance portfolio is:

$$\mu_{mv} = \mathbf{w}^{mvT}\boldsymbol{\mu} = \frac{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}} > 0.$$

Therefore, the weight of the MSR portfolio is:

$$\mathbf{w}^{msr} = \frac{\boldsymbol{\Sigma}^{-1}\sum_{i=1}^p a_i\mathbf{u}^{(i)}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\sum_{i=1}^p a_i\mathbf{u}^{(i)}} = \frac{\sum_{i=1}^p \frac{a_i}{\lambda_i}\mathbf{u}^{(i)}}{\sum_{i=1}^p \frac{a_i}{\lambda_i}\mathbf{1}^T\mathbf{u}^{(i)}} = \sum_{i=1}^p \frac{\frac{a_i\mathbf{1}^T\mathbf{u}^{(i)}}{\lambda_i}}{\sum_{i=1}^p \frac{a_i\mathbf{1}^T\mathbf{u}^{(i)}}{\lambda_i}}\mathbf{z}^{(i)}.$$

□

B Proofs of the technical results in Section 3

Proof of Proposition 3.2. Before proving the convergence, we introduce a few new notations. We let $\mathbf{U}_K = (\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(K)})$ denote the matrix of the first K population eigenvectors and let $\mathbf{\Lambda}_K = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_K\}$ denote the diagonal matrix of the first K population eigenvalues. With the new notations, $\boldsymbol{\mu}^K$ can be written as $\boldsymbol{\mu}^K = \mathbf{U}_K \mathbf{U}_K^T \boldsymbol{\mu}$. In the argument of equivalence, we show that the spectral cut-off method with parameter K leads to a weight vector which is expressed as

$$\widehat{\mathbf{w}}^{cut}(K) = \frac{\widehat{\mathbf{\Sigma}}_K^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}_K^{-1}\boldsymbol{\mu}}.$$

Similarly, we can show that the distorted MSR portfolio can be written as

$$\mathbf{w}^{cut}(K) = \frac{\mathbf{\Sigma}_K^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\mathbf{\Sigma}_K^{-1}\boldsymbol{\mu}}.$$

Next, we prove the convergence of $\widehat{\mathbf{w}}^{cut}(K)$ by appealing to some results on the limiting behavior of sample eigenvalues and eigenvectors. According to Theorem 1 in Shen et al. [2016], under Assumptions 3.1 and 3.2, the first K sample eigenvalues and eigenvectors have such a limiting behavior that $\frac{\widehat{\lambda}_j}{\lambda_j^{(n)}} \xrightarrow{a.s.} 1$ and $\langle \widehat{\mathbf{u}}^{(j)}, \mathbf{u}^{(j)} \rangle \xrightarrow{a.s.} 1$, $j = 1, 2, \ldots, K$, as $n \to \infty$ and $p(n)/n \to c \in (0, 1)$. Therefore, we obtain:

$$\frac{< \widehat{\mathbf{w}}^{cut}(K), \mathbf{w}^{cut}(K) >}{\|\widehat{\mathbf{w}}^{cut}(K)\|\|\mathbf{w}^{cut}(K)\|} = \frac{\boldsymbol{\mu}^T\widehat{\mathbf{\Sigma}}_K^{-1}\mathbf{\Sigma}_K^{-1}\boldsymbol{\mu}}{\|\widehat{\mathbf{\Sigma}}_K^{-1}\boldsymbol{\mu}\|\|\mathbf{\Sigma}_K^{-1}\boldsymbol{\mu}\|} \xrightarrow{a.s.} 1$$

as $n \to \infty$ and $p(n)/n \to c \in (0, 1)$. □

Proof of Proposition 3.3. It is straightforward to check that the solution to the least square problem in eq. (7) is $\widehat{\mathbf{a}}^{ls} = (\widehat{a}_1^{ls}, \ldots, \widehat{a}_p^{ls})^T = (\widehat{\mathbf{U}}^T\widehat{\mathbf{U}})^{-1}\widehat{\mathbf{U}}^T\boldsymbol{\mu} = \widehat{\mathbf{U}}^T\boldsymbol{\mu}$. Next, we solve the optimization problem in eq. (14). Let $\mathbf{a} = (a_1, \ldots, a_p)^T$ denote the decision vector and

define another vector $\mathbf{b} = (b_1, \ldots, b_p)^T \coloneqq \widehat{\mathbf{\Lambda}}^{-c}\mathbf{a}$. Then, the optimization problem can be rewritten in terms of $\mathbf{b}$:

$$\min_{\mathbf{b}} \frac{1}{2}(\boldsymbol{\mu} - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}^c\mathbf{b})^T(\boldsymbol{\mu} - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}^c\mathbf{b}) + \gamma\|\mathbf{b}\|_1.$$

Here, $\|\mathbf{b}\|_1$ is the L_1 norm of $\mathbf{b}$. Expanding out the first term we get $\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\mu} - \boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}^c\mathbf{b} + \frac{1}{2}\mathbf{b}^T\widehat{\mathbf{\Lambda}}^{2c}\mathbf{b}$ and since $\boldsymbol{\mu}^T\boldsymbol{\mu}$ does not contain the variable of interest, we can discard it and consider an equivalent problem:

$$\min_{\mathbf{b}} -\boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}^c\mathbf{b} + \frac{1}{2}\mathbf{b}^T\widehat{\mathbf{\Lambda}}^{2c}\mathbf{b} + \gamma\|\mathbf{b}\|_1.$$

Since $\widehat{\mathbf{a}}^{ls} = \widehat{\mathbf{U}}^T\boldsymbol{\mu}$, the following problem can be written as:

$$\min_{\mathbf{b}} \sum_{i=1}^{p} -\widehat{a}_i^{ls}\widehat{\lambda}_i^c b_i + \frac{1}{2}\widehat{\lambda}_i^{2c}b_i^2 + \gamma|b_i|.$$

Our objective function is now a sum of objectives, each corresponding to a separate variable b_i, so they may each be solved individually. Fix a certain i. Then, we minimize

$$\mathcal{L}_i = -\widehat{a}_i^{ls}\widehat{\lambda}_i^c b_i + \frac{1}{2}\widehat{\lambda}_i^{2c}b_i^2 + \gamma|b_i|.$$

If $\widehat{a}_i^{ls} > 0$, then we must have $b_i \geq 0$ since otherwise we could flip its sign and get a lower value for the objective function. Likewise if $\widehat{a}_i^{ls} < 0$, then we must choose $b_i < 0$.

Case 1: $\widehat{a}_i^{ls} > 0$. Since $b_i \geq 0$, $\mathcal{L}_i = -\widehat{a}_i^{ls}\widehat{\lambda}_i^c b_i + \frac{1}{2}\widehat{\lambda}_i^{2c}b_i^2 + \gamma b_i$. Differentiating with respect to b_i and setting equal to zero, $b_i = \widehat{\lambda}_i^{-c}(\widehat{a}_i^{ls} - \gamma\widehat{\lambda}_i^{-c})$ and this is only feasible if the right-hand side is non-negative, so the actual solution is $b_i = \widehat{\lambda}_i^{-c}(\widehat{a}_i^{ls} - \gamma\widehat{\lambda}_i^{-c})_+ = \text{sign}(\widehat{a}_i^{ls})\widehat{\lambda}_i^{-c}(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+$. Thus, in this case we obtain:

$$\widehat{a}_i^{sel} = \widehat{\lambda}_i^c b_i = \text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+.$$

Case 2: $\widehat{a}_i^{ls} \leq 0$. This implies we must have $b_i \leq 0$ and so $\mathcal{L}_i = -\widehat{a}_i^{ls}\widehat{\lambda}_i^c b_i + \frac{1}{2}\widehat{\lambda}_i^{2c} b_i^2 - \gamma b_i$. Differentiating with respect to b_i and setting equal to zero, $b_i = \widehat{\lambda}_i^{-c}(\widehat{a}_i^{ls} + \gamma\widehat{\lambda}_i^{-c}) = \text{sign}(\widehat{a}_i^{ls})\widehat{\lambda}_i^{-c}(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})$. Again, the solution is only feasible if the right-hand side is non-positive, so the actual solution is $b_i = \text{sign}(\widehat{a}_i^{ls})\widehat{\lambda}_i^{-c}(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+$. Thus, in this case we also obtain:

$$\widehat{a}_i^{sel} = \widehat{\lambda}_i^c b_i = \text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+.$$

Obtaining the same solution in both cases completes the proof. □