A Statistical Response to Challenges in Vast Portfolio Selection

by

Danqiao Guo

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Marine Carrasco Professor, Dept. of Economics, University of Montreal
Supervisor(s):	Chengguo Weng Associate Professor
	Tony Wirjanto Professor
Internal Member:	Phelim Boyle Professor
	Mingbin Feng Assistant Professor
Internal-External Member:	Ranjini Jha Associate Professor, School of Accounting and Finance

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The thesis is written in response to emerging issues brought about by an increasing number of assets allocated in a portfolio and seeks answers to puzzling empirical findings in the portfolio management area. Over the years, researchers and practitioners working in the portfolio optimization area have been concerned with estimation errors in the first two moments of asset returns. The thesis comprises several related chapters on our statistical inquiry into this subject. Chapter 1 of the thesis contains an introduction to what will be reported in the remaining chapters.

A few well-known covariance matrix estimation methods in the literature involve adjustment of sample eigenvalues. Chapter 2 of the thesis examines the effects of sample eigenvalue adjustment on the out-of-sample performance of a portfolio constructed from the sample covariance matrix. We identify a few sample eigenvalue adjustment patterns that lead to a definite improvement in the out-of-sample portfolio Sharpe ratio when the true covariance matrix admits a high-dimensional factor model.

Chapter 3 shows that even when the covariance matrix is poorly estimated, it is still possible to obtain a robust maximum Sharpe ratio (MSR) portfolio by exploiting the uneven distribution of estimation errors across principal components. This is accomplished by approximating the vector of expected future asset returns using a few relatively accurate sample principal components. We discuss two approximation methods. The first method leads to a subtle connection to existing approaches in the literature, while the second one named the "spectral selection method" is novel and able to address main shortcomings of existing methods in the literature.

A few academic studies report an unsatisfactory performance of the optimized portfolios relative to that of the $1/N$ portfolio. Chapter 4 of the thesis reports an in-depth investigation into the reasons behind the reported superior performance of the $1/N$ portfolio. It is supported by both theoretical and empirical evidence that the success of the $1/N$ portfolio is by no means due to the failure of the portfolio optimization theory. Instead, a major reason behind the superiority of the $1/N$ portfolio is its adjacency to the mean-variance optimal portfolio.

Chapter 5 examines the performance of randomized $1/N$ stock portfolios over time. During the last four decades these portfolios outperformed the market. The construction of these portfolios implies that their constituent stocks are in general older than those in the market as a whole. We show that the differential performance can be explained by the relation between stock returns and firm age. We document a significant relation between age and returns in the US stock market. Since 1977 stock returns have been

an increasing function of age apart from the oldest ages. For this period the age effect completely dominates the size effect.

Acknowledgements

First, I would like to express my sincere gratitude to my PhD supervisors Professor Tony Wirjanto and Professor Chengguo Weng for their guidance and endless support. Research was completely new to me when I started my PhD study at the University of Waterloo. It is the training my supervisors provided to me, both consciously and unconsciously, that makes me grow into an independent researcher. I benefit a lot from their profound knowledge.

I thank my thesis committee member Professor Phelim Boyle who spent a lot of time and efforts advising my thesis research and gave me tremendous support and encouragement. I truly enjoy our stimulating discussions and exchanges of ideas over the past two years. I would like to thank the rest of my thesis committee: Professor Marine Carrasco, Professor Ranjini Jha, and Professor Ben Feng, for serving on my thesis committee and providing insightful comments on the thesis.

My sincere thanks also go to Professor Pengfei Li who taught me two core courses in statistics. I also thank Professor Ye Du and Professor Yi Li from Southwestern University of Finance and Economics for their constant support since I was an undergraduate student. I would not consider pursuing a PhD if I never met these knowledgeable and caring professors.

I would always remember the motivating discussions I had with fellow graduate students and the quality time I spent with my friends at Waterloo. They make this small town a cozy and memorable second home for me. Special thanks to Junhan and Yidan for spending a considerable amount of time with me and accompanying me through difficult times.

I am indebted to my parents. They have been very understanding and supportive on my decisions. I always feel lucky to have been blessed with them in my life. Lastly, I thank Frank for his love and support in the past five years. His comments on my research have been helpful in improving the quality of my work.

To my family

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

1.1.1 Mean-variance Optimization

Portfolio theory has remained a topic of broad and intense interest in quantitative finance since the seminal publication of Harry Markowitz ([84]). In Markowitz's mean-variance optimization framework, investment decisions are made solely based on the mean and covariance matrix of asset returns. Suppose that there are N risky assets in the market and a riskless asset with a rate of return r_f. The expected returns of risky assets form an $N \times 1$ vector $\boldsymbol{\mu}$. The $N \times N$ covariance matrix of returns on these risky assets is denoted by $\boldsymbol{\Sigma}$. According to the two fund separation theorem, a rational investor allocates her wealth between the riskless asset and the Sharpe ratio maximizing (MSR) portfolio of risky assets (also known as the tangency portfolio) levered up according to her extent of risk aversion. The Sharpe ratio of a portfolio is defined as the ratio between the expected portfolio return in excess of the riskless rate and the standard deviation of portfolio returns.

Let an $N \times 1$ vector $\mathbf{w}$ represent the weight vector of a portfolio of risky assets. All elements in $\mathbf{w}$ sum up to 1. The weight vector of the MSR portfolio $\mathbf{w}_{msr}$ can be calculated from the following optimization problem:

$$\mathbf{w}_{msr} = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T \boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}, \quad \text{s.t. } \mathbf{w}^T \mathbf{1} = 1, \tag{1.1}$$

where $\mathbf{1}$ is an $N \times 1$ vector of ones.

If all investors have a homogeneous belief on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, everyone should hold the same portfolio of risky assets. From an equilibrium perspective, the demand and supply of all securities are equated. Consequently, the portfolio that every investor holds is simply the market portfolio.

1.1.2 Estimation Errors

Debates on the practical usefulness of the mean-variance optimization are still ongoing to this date. To implement the mean-variance optimization, we need to obtain the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ which are unobservable in real applications. The common practice is to adopt a "two-step procedure" in which we first find estimators for the first two moments of asset returns and then plug these estimators into the optimization problem. From a statistical point of view, the theoretically appealing methodology is rendered less potent in practice due to the presence of estimation errors in the first step.

The vector of expected returns and the asset returns covariance matrix are treated in a static environment in Markowitz's theory. In the real world, however, it is common that an occasional exogenous event may cause a structural change in the dynamics of an asset's price, especially in its expected return. As pointed out by [86] and [62], the time-varying nature of the expected returns makes them difficult to estimate from the price history. Fortunately, a considerable progress has been made in improving the estimators for expected returns. There are a few methods shown in the literature to work better than those using the sample estimator for the expected return vector, including Bayesian approach ([63, 82]), equilibrium expected returns approach ([15]), robust portfolio approach which incorporates estimation errors into consideration ([23]), etc. The trend among professional portfolio managers is to forgo estimating expected returns from historical returns and instead resort to other resources, for example, company fundamentals, social media sentiment, analyst consensus, etc., to form a prediction, or proxy, for future asset returns. There are plenty of discussions on this topic in the empirical asset pricing literature (see [56] for a review).

Compared with the first moment, the covariance matrix is relatively more stable over time and is possible to be estimated from historical data. However the estimation errors are still a concern especially when the portfolio size is large compared with the sample size. It is a fundamental result in the random matrix theory (RMT) that under high-dimensional asymptotics and certain regularity conditions, the limiting distribution of the sample eigenvalues are more dispersed than that of the true eigenvalues ([83]). Since a calculation of the weights of a mean-variance optimal portfolio involves inverting the covariance matrix estimator, the dispersed sample eigenvalues affect the portfolio weight through their reciprocals. An undesirable consequence is that the reciprocals of the smallest sample eigenvalues are usually much larger than what they should be, leading to extreme positions in the sample-based optimal portfolio.

In any major equity market the stock universe contains hundreds or even thousands of assets but the number of observations (return vectors) for estimating the covariance matrix is usually limited. Consequently, the two-step procedure based estimator for the optimal portfolio could deviate substantially from the actual optimal one, causing an abysmal out-of-sample performance. The issue of estimation errors in a large covariance matrix estimator has been carefully discussed and tackled in the literature including [75, 76, 61, 22, 30, 39, 16, 29, 66] among others. The majority of the remedies for estimation errors in the covariance matrix estimator falls into one of the following two categories: constructing a more accurate covariance matrix estimator or imposing constraints on portfolios to avoid extreme or negative positions.

1.1.3 High-dimensional Factor Model

The high-dimensional factor model ([8, 39]) is a parsimonious yet powerful tool in characterizing cross-sectional returns of an increasing number of assets. The model assumes that cross-sectional asset returns are governed by a fixed number of common factors. The high-dimensional factor model gives rise to a spiked (population) covariance model in which only the largest few eigenvalues increase with the number of assets. The number of increasing eigenvalues is exactly the number of common factors. [98] studies the consistency of sample principal components if the population covariance model is as described above. An important finding from this study is that the distribution of estimation errors is uneven across sample principal components. If there are K common factors, only the largest K sample eigenvalues and their corresponding sample eigenvectors are consistent estimators for their population counterparts under standard high-dimensional asymptotics.

1.1.4 Portfolio Performance

A comparison between the performance of different investment strategies is frequently made in the portfolio management area. Common performance measures include the average return, standard deviation, Sharpe ratio, turnover, and certainty-equivalent return. Researchers usually pick a passive portfolio as the comparison benchmark. The most popular one is the equally weighted portfolio. The equally weighted portfolio is also known as the $1/N$ portfolio since it allocates $1/N$ of the total wealth to each of the N assets. [30] investigated 14 portfolio optimization rules and found that none is consistently better than the $1/N$ rule in terms of Sharpe ratio, certainty-equivalent return, or turnover. In recent years some investors turned to the minimum variance portfolio. The minimum variance portfolio can be viewed as a special case of the maximum Sharpe ratio portfolio when the expected return of each asset is the same. It has been documented in [62, 63, 26, 30, 61] that minimum variance equity portfolios have surprisingly high average returns and Sharpe ratios. [95] showed theoretically and empirically that minimum variance portfolios implicitly pick up risk-based asset pricing anomalies and therefore yield positive alpha.

1.2 Thesis Outline

The first two chapters of this thesis focus on mitigating the adverse effects of estimation errors in the covariance matrix estimator on out-of-sample performance of mean-variance optimized portfolios.

As has been mentioned earlier, the mainstream of the literature deals with the estimation errors in the second moment by replacing the sample covariance matrix with a better covariance estimator. Specially, a few methods that belong to this stream involve adjusting sample eigenvalues. Examples include the "shrinkage towards identity" estimator ([75]), which is a weighted average of the sample covariance matrix and a scalar multiple of the identity matrix, and the spectral cut-off method ([22]), which reconstructs the inverse covariance matrix by discarding a few principal components for regularization purposes. However, it is unclear to us a priori why these alternative covariance estimators are able to lead to an improvement in the out-of-sample performance of an optimized portfolio. In Chapter 2, we examine the effects of adjusting one or multiple sample eigenvalues on the out-of-sample Sharpe ratio of a portfolio constructed by using the two-step approach. This is accomplished by parameterizing the sample eigenvalue adjustment and evaluating the derivative of the out-of-sample Sharpe ratio with respect to the adjustment parameter. We show under high-dimensional factor model assumptions that some eigenvalue adjustment patterns could guarantee an improvement in the out-of-sample portfolio Sharpe ratio. As a result, we are able to provide an important rationale for some well-known methods that have been employed to improve the out-of-sample portfolio performance.

In Chapter 3, we study how to pick optimal portfolios by modulating the impact of estimation risk in large covariance matrices. We discover that even when the sample covariance matrix contains substantial estimation errors, it is still possible to obtain an accurate estimator for the MSR portfolio. The key to such a possibility is the geometric relationship between the expected returns vector and the linear space spanned by the sample eigenvectors. We find that if the expected returns vector is spanned by a few sample eigenvectors, the sample-based MSR portfolio is also spanned by the same set of sample eigenvectors. Due to the uneven distribution of estimation errors across different sample eigenvalues and eigenvectors, it is desirable that the portfolio estimator is spanned by a few sample eigenvectors that relatively well estimate their population counterparts. In our proposed method, the expected returns vector is approximated by a selected set of eigenvectors and the approximation is used for the construction of portfolio. As long as the approximation is close to the original vector of expected returns, we benefit from the reduced exposure to the estimation errors without much loss in the information of the expected returns. We introduce two concrete methods for approximating the expected returns, and analyze the choice of tuning parameters for both methods.

The next two chapters of the thesis provide partial explanation to the puzzling outperformance of the $1/N$ portfolio and discuss the implication of such an outperformance.

Chapter 4 focuses on the popular portfolio performance benchmark: the $1/N$ portfolio. The documented superior performance of this simple portfolio construction rule is puzzling since the $1/N$ portfolio does not involve any explicit optimization. The outperformance of the $1/N$ portfolio has been largely attributed to the inferior performance of the optimized portfolios which are contaminated by estimation errors. In our research, we find that an alternative reason behind the outperformance of the $1/N$ portfolio is that it is innately close to the true mean-variance optimal portfolio in the market. Equating the weight vector of the $1/N$ portfolio and that of the MSR portfolio gives rise to a condition about the expected returns and the covariance matrix of asset returns. This condition characterizes the market environments in which holding the $1/N$ portfolio is a favorable investment decision. Based on this condition we devise a summary market-specific measure coined the "$1/N$ favorability index" to quantify the extent to which a market favors the $1/N$ portfolio. The condition for the $1/N$ portfolio being mean-variance optimal also represents a risk-based asset pricing rule. A market-factor model predicts a positive relationship between the $1/N$ favorability index and the return on the market factor. An important implication of this finding is that the $1/N$ portfolio is more difficult to be outperformed in bullish markets. This implication can be used to explain the empirical finding in [30] that in the US equity market none of the 14 sophisticated portfolio optimization models considered in [30] is able to consistently outperform the $1/N$ portfolio.

Chapter 5 reports an interesting equity market pricing anomaly that senior stocks (in terms of their age since listing) in general have higher returns than junior stocks. We name this empirical finding the age effect. We show that over the recent 40 years the age effect dominates the size effect. The discovery of the age effect is thanks to a comparison between two types of random portfolios, a monthly rebalanced $1/N$ portfolio, whose initial set of components are randomly picked from the investment universe, and an equally weighted bootstrapped portfolio, whose components are shuffled each month and assigned equal weights. As a result, a stock that enters the $1/N$ portfolio remains in the portfolio until it is delisted and replaced by another randomly picked stock, while a stock that enters the bootstrapped portfolio very likely drops out in the next month. A simulation experiment shows that the rebalanced $1/N$ portfolio yields a higher return than the bootstrapped portfolio does. Due to the difference in the construction of the two portfolios and the steadily evolving investment universe, the stocks in the $1/N$ portfolio are in general older than those in the bootstrapped portfolio. We use a panel-data regression analysis and a comparison among factor-sorted portfolios to confirm that the difference in age distribution between these two portfolios is the reason behind the outperformance of the rebalanced $1/N$ portfolio relative to the bootstrapped portfolio.

Chapter 2

Sample Eigenvalues Adjustment for Portfolio Performance Improvement under Factor Models

2.1 Introduction

Using the sample covariance matrix for portfolio optimization can be highly problematic as the number of assets becomes increasingly large relative to the sample size. Well-known remedies for this issue in the literature involve an adjustment of the sample covariance matrix, or more specifically, an adjustment of sample eigenvalues.[1] A few examples include the "shrinkage towards identity" estimator ([75]), a more general linear shrinkage estimator ([16]), a "nonlinear shrinkage" estimator ([76]), and a spectral cut-off method ([22]). All of these covariance estimators reconstructed from adjusted sample eigenvalues, when used to build an optimized portfolio, have been shown empirically to be able to boost the portfolio's out-of-sample Sharpe ratio. However, it is important to emphasize that not all of these improved estimators were designed specifically to achieve this goal. For instance, [75] and [16] minimize the expected Frobenius loss; [22] use the expected utility of a mean-variance investor as the objective function. Since improvement in these objective functions does not necessarily lead to a higher out-of-sample Sharpe ratio, it is not clear at first glance why these methods bring about an enhanced portfolio performance.

[1]There are some other existing methods for mitigating the adverse effect of estimation errors in the covariance matrix estimator, e.g., constraining portfolio weights ([61, 29, 40]), imposing factor structure ([38, 37, 39]), shrinking to a target portfolio ([17]), etc.

To address this issue, we investigate how to adjust sample eigenvalues for the purpose of improving the out-of-sample portfolio Sharpe ratio when the cross-sectional asset returns are generated from a high-dimensional factor model, which has become increasingly popular in finance and financial econometrics ([36, 8, 39]). Later, it will be seen that our findings provide a novel rationale for some of the existing methods. The discussions in this chapter can be differentiated from other work in the literature on mitigating estimation errors and improving portfolio performance in at least three important aspects.

First, while in the literature most improved covariance matrix or portfolio weight estimators optimize a particular objective function within a certain class of candidate covariance/weight estimators[2], we seek adjustment patterns which, when applied to the sample eigenvalues, could lead to a concrete improvement in terms of the out-of-sample portfolio Sharpe ratio when both the sample size and the portfolio size are sufficiently large. This pattern, if identified, will shed lights on the key factors responsible for improving a portfolio's out-of-sample performance. More importantly, each realization of the sample covariance matrix will benefit, in terms of the out-of-sample portfolio Sharpe ratio, from such an adjustment. Our study is primarily placed on the analysis of effects when there is an infinitesimal adjustment on the eigenvalues. The motivation underlying our particular focus on infinitesimal adjustments of eigenvalues is the observation that most improved estimators deviate only mildly from the sample covariance matrix. Throughout the chapter, we refer to such an effect due to an infinitesimal adjustment on eigenvalues as a marginal effect. The precise definition is given in Section 2.2.

Second, our simultaneous selection of the out-of-sample Sharpe ratio as a target and the covariance estimators reconstructed based on adjusted sample eigenvalues as candidates differentiates this chapter from its peers which adopt other objectives ([75, 22, 16]) and those which consider other candidate estimators ([75, 66, 17]). To the best of our knowledge, [76] is the only work that optimizes (the convergence limit of) the out-of-sample Sharpe ratio over the same class of estimators we consider. The key difference between our work in this chapter and theirs will be elaborated in the next paragraph.

Third, in this chapter we assume that the asset return generating process follows a high-dimensional factor model, where the returns of an increasing number of assets are governed by a fixed number of common factors. The high-dimensional factor model assumption gives rise to a spiked structure in the population covariance matrix: as the number of assets increases to infinity, the first few eigenvalues increase at the same rate

[2]Usually, the problem can be reduced to first deriving the optimal value of the decision variable and then looking for its bona fide estimator ([75]) or convergence limit ([17]). In [76], the authors search among the candidates for the one that optimizes the convergence limit of the objective function.

and the remaining ones are bounded. The spiked structure with rapidly-growing spiked eigenvalues renders many results in the random matrix theory (RMT) impotent, in that the spectral convergence results for this type of population covariance matrix have not yet been developed. This technical hurdle prevents us from deriving the optimal eigenvalues adjustment under a factor model. The few papers that borrow results from the RMT to improve portfolios ([76], [34], [17][3]) assume the population covariance matrix to have bounded eigenvalues, which contradicts conventional factor model assumptions. Thanks to the results in [98] on the consistency of principal component analysis (PCA) under a variety of assumptions on the population covariance matrix and different asymptotics, we are able to find an adjustment pattern that ensures a positive marginal effect under certain high-dimensional asymptotics.

According to the main results of this chapter, if the population covariance matrix admits a high-dimensional K-factor model, adjusting one of the first K sample eigenvalues has a diminishing effect under the high-dimensional asymptotics that the number of assets N and the sample size T both go to infinity such that T and N^{1+c} grow at the same rate for any $c > 0$. In addition, either shrinking a large but non-spiked (excluding the first K) sample eigenvalues or lifting a small one has a positive effect on the out-of-sample Sharpe ratio asymptotically. We also analyze a simultaneous adjustment of multiple eigenvalues. Let $\{\widehat{\lambda}_i\}_{i=1}^N$ denote decreasingly sorted sample eigenvalues. Then, for any value of k and any $a < 1$, simultaneously amplifying the smallest sample eigenvalues $\{\widehat{\lambda}_i : i \geq k\}$ to $\{\widehat{\lambda}_i + \lambda\widehat{\lambda}_i^a : i \geq k\}$ for some small $\lambda > 0$ leads to an improvement in the out-of-sample Sharpe ratio under the high-dimensional asymptotics. If the underlying model is a single-factor model and the factor pricing relation holds, i.e., both expected returns and covariance matrix are driven by the single-factor model (see a discussion in [82]), the aforementioned way of adjusting eigenvalues always leads to an improvement in the out-of-sample portfolio Sharpe ratio, regardless of the values of N and T.

Our results shed light on the reason for the effectiveness of the shrinkage towards identity method ($k = 1$, $a = 0$) and the spectral cut-off method ($k \geq \min\{k : \widehat{\lambda}_k < 1\}$, $a = -\infty$) adopted in real-world portfolio construction practice. A useful implication of our results is that the key to improving the out-of-sample portfolio performance is to make the eigenvalues overall less dispersed (this will be formally defined later) after the adjustment. Although [75] made a similar statement, their reasoning is based on the fact that sample eigenvalues are more dispersed compared with their population counterparts and thus should be corrected to reduce the expected Frobenius loss. By showing that

[3]Although the authors did not explicitly assume the population covariance matrix to have bounded eigenvalues, in their major technical reference ([94]), the population covariance matrix is assumed to have a bounded spectral norm.

the shrinkage towards identity type estimator has a concrete marginal effect of improving the out-of-sample Sharpe ratio under factor model assumptions, we endow the well-known method with new insights.

The remaining part of the chapter is organized as follows. Section 2.2 provides an expression for the marginal effect of sample eigenvalues adjustment on the out-of-sample Sharpe ratio and discusses what adjustment pattern would ensure a positive effect. Section 2.3 uses a simulation experiment to support the results established in Section 2.2. Section 2.4 discusses the connection of the main results in this chapter with a few existing approaches in the literature. Section 2.5 concludes the chapter.

2.2 Main Results

Throughout the chapter, we use bold capital letters to denote matrices, bold lowercase letters to denote vectors, and plain lowercase letters to denote scalars. The term "return" on a given asset denotes the asset's return in excess of the riskless rate. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$ denote an $N \times T$ matrix containing T independently and identically distributed observations on a system of $N(< T)$ asset returns with some mean vector, a positive definite covariance matrix $\mathbf{\Sigma}$, and a finite fourth moment. We denote the eigen-decomposition of the population covariance matrix by $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1 \geq \cdots \geq \lambda_N > 0\}$ and $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N)$. Let $\mathbf{S} = \frac{1}{T-1}\sum_{i=1}^{T}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$, where $\bar{\mathbf{x}} = \frac{1}{T}\sum_{i=1}^{T}\mathbf{x}_i$, denote the sample covariance matrix, whose eigen-decomposition is $\mathbf{S} = \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{U}}^T$. Similarly, we denote $\widehat{\mathbf{\Lambda}} = \text{diag}\{\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_N\}$ and $\widehat{\mathbf{U}} = (\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2, \dots, \widehat{\mathbf{u}}_N)$. Since we focus on occasions where T is greater than N, we assume that $\mathbf{S}$ is invertible throughout the chapter without loss of generality. In addition, $\|\mathbf{a}\|$ and $\|\mathbf{a}\|_\infty$ denote the Euclidean norm and the maximum norm of a vector $\mathbf{a}$ respectively. We denote $a_\tau \asymp b_\tau$ if $c_2 \leq \underline{\lim}_{\tau\to\infty}\frac{a_\tau}{b_\tau} \leq \overline{\lim}_{\tau\to\infty}\frac{a_\tau}{b_\tau} \leq c_1$ for two constants $c_1 \geq c_2 > 0$. Lastly, $\mathbf{E}_k$ denotes a conforming diagonal matrix with 1 being its kth diagonal entry and 0 elsewhere; $\mathbf{E}_{k^+}$ denotes a conforming diagonal matrix with 0 being its first $k-1$ diagonal entries and 1 being its diagonal entries at and beyond the kth.

2.2.1 Problem Setup

Suppose that an investor wishes to follow the two-step approach to construct a single-period maximum Sharpe ratio (MSR) portfolio of N risky assets. Further, suppose that the investor believes that the vector of expected returns is an exogenously given vector $\boldsymbol{\mu}$, and estimates the covariance matrix from assets' price history. We consider such a case because it has been widely noticed that estimating expected returns from price history is a major source of estimation error ([62, 61, 30]) and investors thus resort to alternative approaches to acquire the expected returns. In particular, many researchers exploit asset pricing anomalies to find better proxies for the expected returns. [56] contains a review of hundreds of cross-sectional return patterns. Given the abundance of choice for $\boldsymbol{\mu}$, we leave it to the investor to make the choice and simply take $\boldsymbol{\mu}$ as exogenously given.

When the sample covariance matrix $\mathbf{S}$ is adopted as the covariance estimator, the two-step approach based MSR portfolio can be solved from the following problem:

$$\widehat{\mathbf{w}}_{msr} := \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T \boldsymbol{\mu}}{\sqrt{\mathbf{w}^T \mathbf{S} \mathbf{w}}} \quad \text{s.t.} \quad \mathbf{w}^T \mathbf{1} = 1, \tag{2.1}$$

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

$$\widehat{\mathbf{w}}_{msr} = \frac{\mathbf{S}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T \mathbf{S}^{-1} \boldsymbol{\mu}}, \tag{2.2}$$

and the out-of-sample Sharpe ratio is

$$SR = \frac{\widehat{\mathbf{w}}_{msr}^T \boldsymbol{\mu}}{\sqrt{\widehat{\mathbf{w}}_{msr}^T \boldsymbol{\Sigma} \widehat{\mathbf{w}}_{msr}}} = \frac{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}^T \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu}}}. \tag{2.3}$$

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

In this chapter, we focus on adjusting the eigenvalues of the sample covariance matrix, which means that we consider a covariance matrix estimator that takes the form $\widetilde{\mathbf{S}} = \widehat{\mathbf{U}} \widetilde{\boldsymbol{\Lambda}} \widehat{\mathbf{U}}^T$, with $\widetilde{\boldsymbol{\Lambda}}$ being a diagonal matrix. This is the class of "rotation-equivariant" estimators introduced by [102] and considered as candidate covariance estimators by [76]. Note that $\widetilde{S}$ becomes the sample covariance matrix when $\widetilde{\Lambda}$ is taken at $\widehat{\Lambda}$. We set out our analyses by slightly pulling a number of sample eigenvalues away from their original levels and explore the marginal effect of such an adjustment on the out-of-sample Sharpe ratio. For this purpose, we use a diagonal matrix $\mathbf{V} = \operatorname{diag}\{v_1, v_2, \ldots, v_N\}$ and a scalar λ

to parameterize the sample eigenvalues adjustment:

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

The parameterization above is quite general in that when $\mathbf{V}$ takes different forms, it could reduce to a number of ways to adjust sample eigenvalues. For instance, when $\mathbf{V}$ is a scalar multiple of an identity matrix, the adjustment is similar to (but not the same as[4]) the construction of the "shrinkage towards identity" method proposed by [75]. The shrinkage towards identity estimator is usually referred to as a "linear" shrinkage in the sense that if we plot the eigenvalues after adjustment versus their original counterparts, all points lie on a straight line. This notion is in contrast to the recent development of a "nonlinear" shrinkage estimator ([76]) in which each eigenvalue is adjusted differently. Our parameterization allows for a nonlinear shrinkage since it does not restrict the v_k's to be the same. In addition, this setup also accommodates the cases in which only a subset of the sample eigenvalues are adjusted.

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

$$\widehat{\mathbf{w}}_{msr}(\mathbf{V},\lambda) = \frac{\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}. \tag{2.4}$$

It is worth pointing out that the $v_k\lambda$'s in the above expression for $\mathbf{S}_{\mathbf{V},\lambda}$ is not necessarily positive. It means a shrinkage on the kth eigenvalue for a positive $v_k\lambda$ and an amplification for a negative $v_k\lambda$. With the weight vector $\widehat{\mathbf{w}}_{msr}(\mathbf{V},\lambda)$ given in equation (2.4), the out-of-sample Sharpe ratio could also be expressed as a function of $\mathbf{V}$ and λ:

$$SR(\mathbf{V},\lambda) := \frac{\widehat{\mathbf{w}}_{msr}^T(\mathbf{V},\lambda)\boldsymbol{\mu}}{\sqrt{\widehat{\mathbf{w}}_{msr}^T(\mathbf{V},\lambda)\boldsymbol{\Sigma}\widehat{\mathbf{w}}_{msr}(\mathbf{V},\lambda)}} = \frac{\boldsymbol{\mu}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\Sigma}\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}}. \tag{2.5}$$

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

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

2.2.2 Marginal Effect of Eigenvalues Adjustment

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

Theorem 2.2.1. The marginal effect of an eigenvalues adjustment specified by a diagonal matrix $\mathbf{V}$ on the out-of-sample Sharpe ratio admits the following expression:

$$SR'_{\mathbf{V}}(0) = -\frac{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}{(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu})^{\frac{3}{2}}}\left(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu} - \boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu}\frac{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}\right). \tag{2.6}$$

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

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

2.2.3 High-Dimensional Factor Model

In this section, we base our analyses on the assumption that the asset returns are generated by a high-dimensional K-factor model, which is a well-known and extensively used model for cross-sectional asset returns ([35, 8, 38]). To be more specific, we assume that the systematic portion of the asset prices movement is driven by K common factors and the idiosyncratic returns are mutually uncorrelated[5], so that the population covariance matrix has a "low rank + diagonal" structure. However, we do not assume the factor pricing relation to hold[6], nor do we assume the factor returns and thus the asset returns to have

[5]Uncorrelatedness of residual returns is an assumption of the strict factor model. Recently, researchers have been considering a more practical approximate factor model ([37]) which allows for a sparse residual covariance matrix. But in this chapter we focus on the strict factor model for viability reasons.

[6]If the factor pricing relation holds, $\boldsymbol{\mu}$ should be estimated according to the factor model. We refer readers to [82] for a discussion.

a mean vector independent of time. Rather, we still view $\boldsymbol{\mu}$ as being exogenously given. Our goal is to find a collection of $\mathbf{V}$ matrices which make $SR'_{\mathbf{V}}(0)$ asymptotically positive as both N and T go to infinity.

It is worth mentioning that we are not the first to consider manipulating sample eigenvalues for the purpose of improving the out-of-sample Sharpe ratio under the high-dimensional asymptotics. [76], after imposing a set of technical assumptions, derived a convergence limit for the out-of-sample Sharpe ratio as both the number of assets N and the sample size T go to infinity at the same rate. In the next step, they look for the optimal shrinkage on each sample eigenvalue to maximize the deterministic convergence limit of the out-of-sample Sharpe ratio. Specifically, one of the technical assumptions made in [76] (Assumption 2) indicates that as N increases, the eigenvalues of the population covariance matrix are contained in a compact set, or to put it simply, the largest population eigenvalue does not increase at the rate $O(N)$. This assumption reveals that the authors work under a framework at odds with the standard high-dimensional K-factor model framework. In the latter, the largest K population eigenvalues increase at the rate $O(N)$. This discrepancy in assumptions, as well as the popularity of the factor model structure, makes it necessary to examine the limiting behavior of the marginal effect random variable under the high-dimensional asymptotics and the factor model assumptions of second-order (or weak) stationarity.

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

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \ldots, T, \tag{2.7}$$

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

Assumption 2.2.1. The autocovariance functions for both $\{\mathbf{f}_t\}_{t=1}^T$ and $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$ are independent of t.

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

Assumption 2.2.2. The $K \times K$ covariance matrix $\mathrm{cov}(\mathbf{f}_t)$ is of a full rank.

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

Assumption 2.2.3. The eigenvalues of $N^{-1}\mathbf{B}^T\mathbf{B}$ are bounded away from zero for all sufficiently large N.

[39] provide an insightful explanation for this assumption. The authors point out that this assumption easily holds when the factors are pervasive in the sense that a non-negligible fraction of factor loadings should be non-vanishing. This assumption, together with Assumption 2.2.2, ensures that the K non-zero eigenvalues of $\mathbf{B}\text{cov}(\mathbf{f}_t)\mathbf{B}^T$ diverge at a rate of $O(N)$.

Assumption 2.2.4. The residual covariance matrix $\mathbf{\Sigma}_\epsilon$ is a constant multiple of the identity matrix, i.e. $\mathbf{\Sigma}_\epsilon = \sigma^2\mathbf{I}$.

This assumption serves a technical purpose without which we would be unable to verify Theorem 2.2.2 below. Notwithstanding its restrictiveness, this assumption characterizes the stylized fact that assets' idiosyncratic variances are of a similar scale. The same assumption is also used by [82] in a factor model.

Assumption 2.2.5. As $N \to \infty$, $\|\boldsymbol{\mu}\| = O(N^{1/2})$ and $\left\|\mathbf{U}_F^T\boldsymbol{\mu}\right\|_\infty = o(N^{1/2})$, where $\mathbf{U}_F = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K)$.

The first half of the assumption stipulates that a non-vanishing proportion of the assets have a non-zero expected return. This assumption is relatively straightforward and not unduly restrictive. The second half of the assumption restrains $\boldsymbol{\mu}$ from having an excessive loading on the first K eigenvectors. Typically, there is no a priori reason to believe that $\boldsymbol{\mu}$ has an excessive loading on any specific eigenvector, since $\boldsymbol{\mu}$ can be obtained from multiple resources. A violation of this assumption occurs, for example, when $\boldsymbol{\mu}$ is a linear combination of the first K eigenvectors, in which case $\left\|\mathbf{U}_F^T\boldsymbol{\mu}\right\|_\infty$ has a rate of $O(N^{1/2})$. [51] point out that in such a scenario, even directly using the sample covariance matrix leads to a consistent portfolio weight estimator. In this chapter, we focus on the more common cases where a manipulation of sample eigenvalues potentially leads to some improvement. So we exclude such extreme scenarios by imposing Assumption 2.2.5.

Theorem 2.2.2. Under Assumptions 2.2.1 - 2.2.5 and as N, T go to infinity with $T \asymp N^{1+c}$ for some $c > 0$, the following results hold:

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

$$SR'_{\mathbf{E}_k}(0) \xrightarrow{a.s.} 0. \tag{2.8}$$

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

$$\begin{cases} \frac{SR'_{\mathbf{E}_k}(0)}{X_k^{(N)}} \xrightarrow{a.s.} 1, & K+1 \leq k < K^* \\ \frac{SR'_{-\mathbf{E}_k}(0)}{X_k^{(N)}} \xrightarrow{a.s.} 1, & K^* < k \leq N \end{cases}. \tag{2.9}$$

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

$$\frac{SR'_{-\widehat{\mathbf{\Lambda}}^a \mathbf{E}_{k+}}(0)}{Y_k^{(N)}} \xrightarrow{a.s.} 1, \tag{2.10}$$

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

Remark 2.2.1. Theorem 2.2.2 identifies a few forms of the $\mathbf{V}$ matrix that can lead to a marginal improvement in the out-of-sample Sharpe ratio under the high-dimensional asymptotic that both N and T go to infinity with $T \asymp N^{1+c}$ for some $c > 0$. In the literature, the "high-dimensional asymptotics" used to refer to the cases when N and T go to infinity at the same rate. The asymptotics we base our analysis on here is only of a slightly lower-dimensional nature, since c can be arbitrarily small. For this reason, although the ratio N/T eventually converges to 0, we still deem these asymptotics to be a high-dimensional one.

Remark 2.2.2. Many results in this chapter hold "almost surely". Unless otherwise stated, "some random variable is positive almost surely" can be understood as that it takes the value of 0 only when $\mathbf{S} = \mathbf{\Sigma}$, an event that happens with probability 0.

Part (a) of Theorem 2.2.2 discusses the effect of an incremental adjustment of an individual sample eigenvalue. The result indicates that when N and T become large enough, the marginal effect of modifying one of the largest K sample eigenvalues diminishes (see eq. (2.8)). The finding of the fading marginal effect of adjusting a spiked eigenvalue is consistent with the intuition, because as [98] showed, under the high-dimensional K-factor model assumptions, the largest K eigenvalues can be consistently estimated by their sample counterparts in the sense that $\frac{\widehat{\lambda}_i}{\lambda_i} \xrightarrow{a.s.} 1$, $i = 1, 2, \ldots, K$.

In addition, according to part (a) of Theorem 2.2.2 (see eq. (2.9)), both a mild shrinkage on a large but non-spiked (excluding the first K) sample eigenvalue and a mild amplification on a small one help to improve the out-of-sample Sharpe ratio. This result is in line with the intuition that the sample eigenvalues should be pushed back towards their grand mean for the purpose of improving the MSR portfolio, because sample eigenvalues are more dispersed than their population counterparts ([83]). Before seeing this theorem to be stated formally, readers should be cautious about adopting such an intuition. The reason is that even though pushing sample eigenvalues back to their grand mean helps reducing the estimation errors measured by the expected Frobenius loss $E(\|\mathbf{S}-\mathbf{\Sigma}\|_F)$ ([75]), there is no guarantee that the out-of-sample Sharpe ratio will be affected in the same direction. More discussion on this issue will be given in Section 2.4.

Part (b) of Theorem 2.2.2 focuses on a joint manipulation of multiple sample eigenvalues. The parameter a specifies the relative intensity of the adjustment on different sample eigenvalues. The result indicates that a joint amplification on a collection of tail sample eigenvalues has the marginal effect of improving the out-of-sample Sharpe ratio, as long as $a < 1$. It is worth highlighting that the eigenvalues to be adjusted must form a complete "tail" - no matter from where we start the adjustment, we should adjust all of the eigenvalues beyond (smaller than) the starting point. The proposed range for a implies that the eigenvalues should become overall less dispersed after the adjustment, in the sense that the inequality $\frac{\widetilde{\lambda}_i}{\widetilde{\lambda}_j} \leq \frac{\widehat{\lambda}_i}{\widehat{\lambda}_j}$ holds for any $i < j$, where $\widetilde{\lambda}_i$ is the ith largest eigenvalue after the adjustment. When $k = 1$ and $a = 1$, the adjustment is equivalent to replacing $\mathbf{S}$ with its scalar multiple. Such an adjustment will not lead to any change in the out-of-sample Sharpe ratio. This is why a must be strictly less than 1 to ensure the positiveness of $SR'_{-\widehat{\mathbf{\Lambda}}^a\mathbf{E}_{k+}}(0)$. When $k = 1$ and $a = 0$, the adjustment is consistent with the notion of a "linear" shrinkage, since each eigenvalue is lifted by the same amount. Other values of a correspond to a "nonlinear" shrinkage. Instead of using N parameters to parameterize a nonlinear shrinkage, as in [76], we use 2 by requiring that the shrinkage intensity matrix $\mathbf{V}$ is a power of the sample eigenvalue matrix times an "indicator matrix" specifying the starting point of the manipulation. This parameterization is rich in its implication despite of its parsimonity.

The out-of-sample Sharpe ratio is scale-invariant in the covariance matrix estimator. Although in a shrinkage towards identity estimator the large eigenvalues are shrunk and the small ones are lifted, the estimator has an equivalent (in the sense of leading to the same out-of-sample Sharpe ratio) rotation-equivariant counterpart each of whose eigenvalues is amplified. Therefore, our results provide a support to shrinkage estimators with trace at different levels. The connection between our results and the shrinkage estimators will be discussed in detail in Section 2.4.

2.2.4 Single-Factor Model

In this section, we assess the marginal effect of adjusting sample eigenvalue(s) under a single-factor model, which assumes that a single unobservable factor drives the price movement of all assets. [82] showed by using an "optimal orthogonal portfolio" ([81]) argument that if the exact single-factor pricing relation holds and the uncorrelated residual returns have equal variance σ^2, the true expected return (still denoted by $\boldsymbol{\mu}$) and the covariance matrix has the following relationship:

$$\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^T \frac{1}{s_h^2} + \sigma^2 \mathbf{I}, \tag{2.11}$$

where s_h denotes $\frac{\mu_h}{\sigma_h}$, the Sharpe ratio of the only factor portfolio h. A factor portfolio is a legitimate portfolio whose weight vector is proportional to the vector of factor loadings.

In the following Theorem 2.2.3, we will show that if the exact single-factor pricing relation is satisfied and if the subjective view on expected returns coincides with the true expected returns vector, we can find a set of $\mathbf{V}$ matrices that make $SR'_{\mathbf{V}}(0)$ positive even when both N and T are small. It is important to clarify that Theorem 2.2.3 could not serve a practical purpose because of the following paradox: for an arbitrary $\boldsymbol{\mu}$ supplied by some "alpha model", if the structural assumption in eq. (2.11) is not satisfied, the subsequent results in the theorem do not necessarily hold; otherwise, we would immediately obtain the population covariance matrix and the issue of estimation errors will not be an issue any longer. However, Theorem 2.2.3 has a strong theoretical implication in that at least, it identifies a set of $\mathbf{V}$ matrices that work when N and T can be any number under a reasonable economic model.

Theorem 2.2.3. If $\boldsymbol{\Sigma}$ can be expressed by the exogenously given $\boldsymbol{\mu}$ as in eq. (2.11), then the following results hold for any value of N, T, s_h, and σ^2:

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

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

Remark 2.2.3. Since the set of $\mathbf{V}$ matrices given in Theorems 2.2.2 and 2.2.3 is the same, the latter theorem may seem to be a finite sample analogous to the former, while there is a major difference between these two sets of results: the structural assumption on $\boldsymbol{\Sigma}$ in eq. (2.11) implies that $\boldsymbol{\mu}$ only has a non-zero loading on the dominant eigenvector[7] of

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

$\mathbf{\Sigma}$, but this specific scenario is excluded from the previous analyses by Assumption 2.2.5. Therefore, Theorem 2.2.3 can be viewed as a complement to Theorem 2.2.2 that impresses us with fruitful finite sample results.

Part (a) of Theorem 2.2.3 focuses on adjusting an individual sample eigenvalue. Both an incremental shrinkage on a large sample eigenvalue and an incremental amplification on a small sample eigenvalue lead to an increase in the out-of-sample Sharpe ratio. In addition, there also exists a cutting point between such large eigenvalues and small eigenvalues. This result also helps to partly justify the use of the shrinkage towards identity method. Part (b) of Theorem 2.2.3 focuses on simultaneously adjusting a few eigenvalues. The results indicate that if we apply a mild amplification on a collection of tail eigenvalues so that the eigenvalues become overall less dispersed, this will result in an improvement in the out-of-sample Sharpe ratio.

2.3 Simulation Study

In this section, we use a simulation study to substantiate the findings in Section 2.2.3. In particular, we demonstrate the property of the sample-dependent random variable $SR'_{\mathbf{V}}(0)$ for some $\mathbf{V}$ matrices we have discussed in the previous section.

2.3.1 Simulation Setup

In each experiment, we pre-specify an $N \times 1$ (subjective) expected returns vector $\boldsymbol{\mu}$ and an $N \times T$ population covariance matrix $\mathbf{\Sigma}$ and repeat the following procedure 100 times[8]:

(a) Generate $T(N) = \lfloor N^{1.5} \rfloor$ random vectors from a multivariate normal distribution with expectation $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. Calculate the sample covariance matrix $\mathbf{S}$ based on the T simulated vectors.

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

[8]Ideally, a larger number of simulated runs is needed. However, the patterns demonstrated in the quantities of interest averaged across the 100 replications are reasonably smooth and informative. Therefore we do not further increase the number of runs.

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

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

N: The value of N is either 100 or 500. We make a conjecture that when $N = 500$, the results we observe should better reflect the asymptotic results in Theorem 2.2.2. Note that we do not treat the sample size T as another parameter but simply let it be a function of N to be consistent with the asymptotics we are working with.

$\boldsymbol{\Sigma}$: The population covariance matrix $\boldsymbol{\Sigma}$ either strictly conforms to Assumptions 2.2.1 - 2.2.4[9] or violates Assumption 2.2.4 but satisfies the remaining ones. Recall that Assumption 2.2.4 requires all of the non-spiked population eigenvalues to be equal to each other. This assumption is indispensable for the proof of our theoretical results but is hard to be satisfied in reality. Thus, we resort to the simulation study to explore to what extent the results will be affected if we allow for distinct small population eigenvalues. However, it is worth pointing out that either choice for $\boldsymbol{\Sigma}$ admits the K-factor model. In this study, we fix the value of K to be 3.

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

[9] Note that Assumption 2.2.5 is not about the population covariance matrix.

2.3.2 Shrinkage on Individual Eigenvalue

In this section, we show the simulation results on the marginal effect of shrinking a single sample eigenvalue. Specifically, we illustrate how $\overline{SR'_{\mathbf{E}_k}(0)}$ varies with k in Figure 2.1.

According to Figure 2.1, when Assumptions 2.2.1 - 2.2.5 are satisfied (see the red dots in panels (a) and (b)), the marginal effect of shrinking one of the first K sample eigenvalues is almost 0, shrinking a large eigenvalue beyond the Kth leads to a marginal improvement on the out-of-sample Sharpe ratio, and shrinking a small eigenvalue leads to a deterioration of it. The last observation is equivalent to the statement that amplifying a small eigenvalue has a positive marginal effect on the Sharpe ratio. Moreover, the magnitude (in the sense of an absolute value) of the marginal effect of adjusting one of the smallest eigenvalues is quite large. The population covariance matrix $\mathbf{\Sigma}$ used in panels (c) and (d) has distinct tail eigenvalues; however, we observe a very similar pattern (in the red dots) to that observed in the two panels at the top for which the tail eigenvalues are all equal to 0.05.

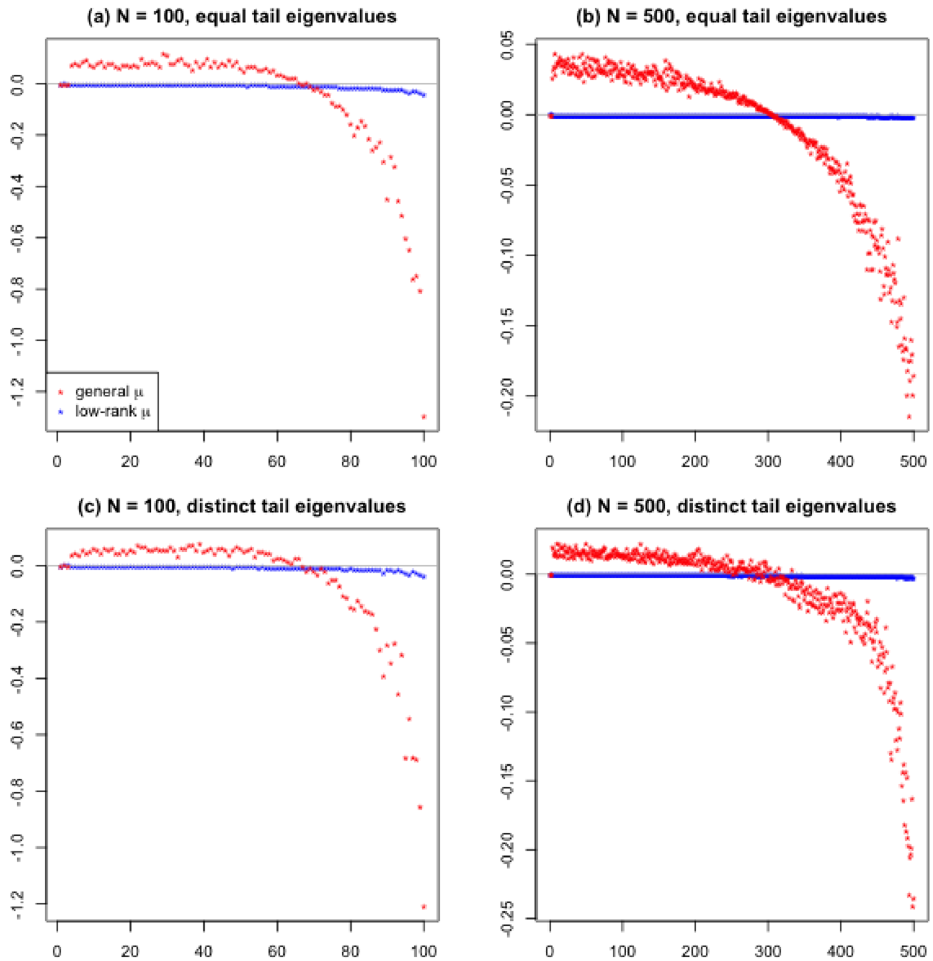
The blue dots in Figure 2.1 come from the setup where $\boldsymbol{\mu}$ is from the 3-dimensional space span$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. As mentioned in [51], in such a scenario, the sample-based MSR portfolio is a consistent estimator of the true MSR portfolio, and it thus becomes less necessary to seek an improvement via adjustment of eigenvalues. What we observe from the blue dots in Figure 2.1 supports this argument - the magnitude of the marginal effect of eigenvalues adjustment is negligible compared with the case where $\boldsymbol{\mu}$ is an "arbitrary" vector.

Table 2.1: Proportion of positive $SR'_{\mathbf{E}_k}(0)$ among the 100 replications. In this table, we only report results for the cases where $\boldsymbol{\mu}$ is an "arbitrary" vector and can have loading on all eigenvectors. In (a) and (b), the $\mathbf{\Sigma}$ has 3 spiked eigenvalues and the remaining ones equal 0.05; in (c) and (d), the $\mathbf{\Sigma}$ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$.

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

Figure 2.1 only reflects the average value of the marginal effect across the 100 replications. We are also interested in the distribution of the adjustment effect random variables, especially we are interested in what proportion among the 100 realizations of $SR'_{\mathbf{E}_k}(0)$ are positive. Table 2.1 reports such information. According to the first two rows, when the last $N - K$ population eigenvalues are equal, shrinking a large but non-spiked eigenvalue or amplifying a small one almost always leads to a marginal improvement in the out-of-

Figure 2.1: $\overline{SR'_{\mathbf{E}_k}(0)}$ vs. k: marginal effect of shrinking the kth sample eigenvalue. In each panel, the x-axis measures k and the y-axis measures $SR'_{\mathbf{E}_k}(0)$. In panels (a) and (c), $N = 100$ and $T = 1000$; in panels (b) and (d), $N = 500$ and $T = 11180$. In panels (a) and (b), the $\mathbf{\Sigma}$ has 3 spiked eigenvalues and the remaining ones equal to 0.05; in panels (c) and (d), the $\mathbf{\Sigma}$ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$. In each panel, the red dots correspond to the case where $\boldsymbol{\mu}$ can have loading on all eigenvectors; the blue dots correspond to the case where $\boldsymbol{\mu}$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.



sample Sharpe ratio. When the 'flat tail" assumption is removed, there is still a high chance (around 85% according to Table 2.1) that such a manipulation leads to an improvement.

2.3.3 Amplification on Tail Eigenvalues

In this section, we present the simulation results on the marginal effect of amplifying a collection of tail eigenvalues. In particular, we illustrate how $\overline{SR'_{-\mathbf{E}_{k^+}}(0)}$ varies with k in Figure 2.2.

Consistent with the results of Theorem 2.2.2, in panels (a) and (b) of Figure 2.2, the average marginal effect is positive for any k, regardless of the $\boldsymbol{\mu}$ used. In other words, applying a linear amplification on a few tail sample eigenvalues always has a positive marginal effect on the out-of-sample Sharpe ratio, no matter how long the tail is. Comparing the red dots with the blue ones, we conclude that when $\boldsymbol{\mu}$ lies in the subspace spanned by the first K eigenvectors, the marginal adjustment effect on the out-of-sample Sharpe ratio is minutely small in magnitude. The reason is the same as we have previously stated, i.e., this happens because the sample-based MSR portfolio is good enough to yield a Sharpe ratio close to the actual maximum one. As we move to panels (c) and (d), where the "flat tail" assumption is violated, no marked difference from the two panels at the top is observed.

In addition, even a quick glance of Figure 2.2 reminds us that there is some "optimal" k which corresponds to the strongest marginal effect. This result is expected because amplifying the largest few eigenvalues might counteract the improvement brought by amplifying the small ones; so it could be better to solely amplify the small ones. However, it is not our goal in this chapter to find the optimal k. One reason for this is the technical difficulty associated with it, and another reason is that it is not meaningful to look for the optimal k without making sure that a higher $SR'_{-\mathbf{E}_{k^+}}(0)$ leads to a higher $SR(-\mathbf{E}_{k^+}, \lambda) - SR(-\mathbf{E}_{k^+}, 0)$, where λ is a small positive number.[10] Although we do not intend to discuss how to find the optimal k, it is important to point out that the answer to this problem depends on the relationship between $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. A simple illustration on this point is that the optimal k based on the red dots and that based on the blue ones are quite different.

Figure 2.2 only shows the average value of the 100 realizations of the marginal effect random variables. As in the previous section, we are interested in what proportion of the realizations of each random variable is positive. Table 2.2 reports these proportions. According to the first two rows of Table 2.2, when all assumptions about the population

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

Figure 2.2: $\overline{SR'_{-\mathbf{E}_{k+}}(0)}$ vs. k: marginal effect of amplifying eigenvalues beyond the kth. In each panel, the x-axis measures k and the y-axis measures $SR'_{-\mathbf{E}_{k+}}(0)$. In panels (a) and (c), $N = 100$ and $T = 1000$; in panels (b) and (d), $N = 500$ and $T = 11180$. In panels (a) and (b), the $\mathbf{\Sigma}$ has 3 spiked eigenvalues and the remaining ones equal 0.05; in panels (c) and (d), the $\mathbf{\Sigma}$ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$. In each panel, the red dots correspond to the case where $\boldsymbol{\mu}$ can have loading on all eigenvectors; the blue dots correspond to the case where $\boldsymbol{\mu}$ is a linear combination of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

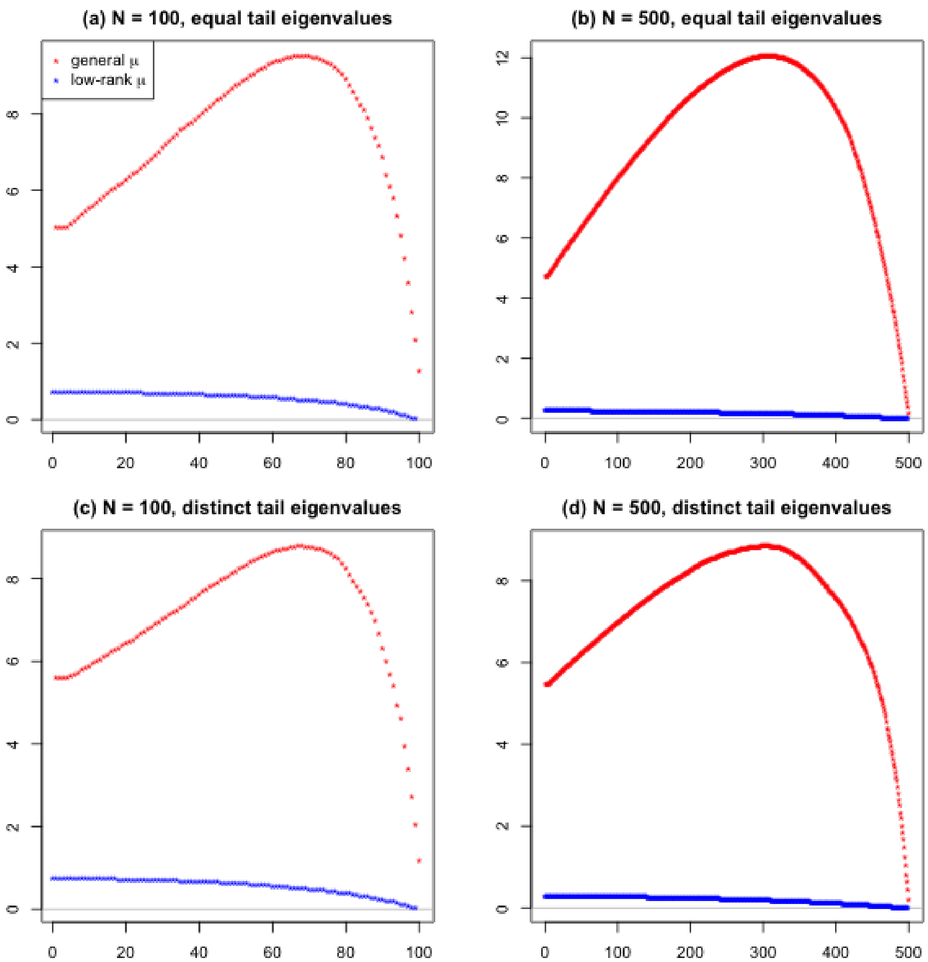


Table 2.2: Proportion of positive $SR'_{-\mathbf{E}_{k+}}(0)$ among the 100 replications. In this table, we only report results for the cases where $\boldsymbol{\mu}$ is an "arbitrary" vector and can have loading on all eigenvectors. In (a) and (b), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones equal 0.05; in (c) and (d), the $\boldsymbol{\Sigma}$ has 3 spiked eigenvalues and the remaining ones are a sorted sample generated from Uniform$(0.025, 0.075)$.

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

covariance matrix are met, the six reported marginal effect random variables are positive in all replications. When Assumption 2.2.4 is violated, as can be seen in the last two rows of Table 2.2, there are a few occasions where amplifying the last sample eigenvalue does not lead to an improvement; when it comes to amplifying the smallest two eigenvalues however, there are fewer such occasions. In all replications, amplifying the smallest six eigenvalues has a positive marginal effect on the out-of-sample Sharpe ratio.

The simulation results provide a firm support to the theoretical conclusions reached in Section 2.2. To recap, under a large-dimensional K-factor model, shrinking a large non-spiked sample eigenvalue and amplifying a small one by a small amount both lead to an improvement in the out-of-sample Sharpe ratio. In addition, a mild linear amplification on any number of tail sample eigenvalues also leads to an improvement.

2.4 Connection with Existing Methods

In this section, we comment on the connection between our theoretical results and a few existing methods in the literature for improving optimized portfolios.

2.4.1 Shrinkage towards Identity Estimator

The shrinkage towards identity estimator ([75]) is a weighted average of the sample covariance matrix and an identity matrix, i.e., $\mathbf{S}_{STI} = s_1\mathbf{S} + s_2\mathbf{I}$ for some $s_1, s_2 > 0$. The effectiveness of this estimator can be partially explained by our theoretical results. Since the out-of-sample Sharpe ratio is scale-invariant in the covariance matrix estimator, an equivalent covariance estimator is $\mathbf{S}^*_{STI} = \mathbf{S} + \frac{s_2}{s_1}\mathbf{I}$. Since both s_1 and s_2 are positive, $\mathbf{S}^*_{STI}$ conforms to our proposed way of adjusting the sample eigenvalues if we let the parameters in $\mathbf{V} = -\widehat{\boldsymbol{\Lambda}}^a\mathbf{E}_{k+}$ be $a = 0$ and $k = 1$.

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

2.4.2 Nonlinear Shrinkage Estimator

The nonlinear shrinkage estimator ([76]) extends the shrinkage towards identity estimator by allowing different eigenvalues to be adjusted independently. As mentioned in Section 2.2.3, under a few assumptions about the population covariance matrix, the authors derive the convergence limit of the out-of-sample Sharpe ratio as N and T go to infinity at the same rate. Then, they search for the optimal shrinkage which maximizes the limiting out-of-sample Sharpe ratio.

Since there is no explicit formula for the optimal shrinkage, we resort to the numerical results provided in [76] to see how the sample eigenvalues are adjusted. Figure 2 in [76] shows a comparison among the eigenvalues of a nonlinear shrinkage estimator, those of a linear shrinkage estimator, and the sample eigenvalues. According to the figure, after the nonlinear adjustment, the smallest eigenvalues become overall less dispersed in the sense that $\frac{\widetilde{\lambda}_i}{\widetilde{\lambda}_j} \leq \frac{\widehat{\lambda}_i}{\widehat{\lambda}_j}$, where $\widetilde{\lambda}_i$ is the ith eigenvalue after the nonlinear shrinkage, and i, j $(i < j)$ both index some small eigenvalues. Among the large eigenvalues, there is at least one pair (i, j) with $i < j$ to satisfy $\frac{\widetilde{\lambda}_i}{\widetilde{\lambda}_j} > \frac{\widehat{\lambda}_i}{\widehat{\lambda}_j}$. As a result, there is no guarantee that after a nonlinear shrinkage, the eigenvalues become overall less dispersed. From this perspective, our results are unable to provide a clear assurance for the improvement brought by the nonlinear estimator. But readers should note that our theoretical results present only a few sufficient conditions for achieving a marginal out-of-sample Sharpe ratio improvement. So, even if the nonlinear shrinkage estimator does not make each pair of eigenvalues less dispersed, it is still possible to reach an improvement.

As has been mentioned earlier, there is some similarity between this chapter and [76], in particular in terms of the objective function used and the family of estimator considered. However, the key difference lies in the assumption about the population covariance matrix: our main results are based on a high-dimensional K-factor model under which the largest K population eigenvalues increase with N at a rate of $O(N)$; a technical assumption in [76] implies that they work under a framework where the largest population eigenvalue is bounded. Our assumption is more appropriate for financial asset returns covariance matrix, but the cost is the divergence of the out-of-sample Sharpe ratio; conversely, the assumption of bounded eigenvalues is less realistic, but ensures the convergence of the out-of-sample Sharpe ratio and thus enables the authors to find the optimal shrinkage.

2.4.3 Spectral Cut-off Method

The spectral cut-off method is a stabilization technique applied to the process of inverting an ill-posed covariance matrix in which the smallest eigenvalues are close to zero. This method reconstructs the inverse covariance estimator after discarding the eigenvectors associated with the eigenvalues that are smaller than a certain threshold. [22] propose a data-driven method for determining the threshold. A contemplation about the spectral cut-off method suggests that it can be viewed as a method which amplifies the smallest eigenvalues to infinity while keeping the large ones unchanged. From this perspective, this method falls into our framework as a polar extreme case if we let the parameters in $\mathbf{V} = -\widehat{\mathbf{\Lambda}}^a \mathbf{E}_{k+}$ be $k \geq \min\{k : \widehat{\lambda}_k < 1\}$ and $a = -\infty$. Our theoretical results in Section 2.2.2 also imply that if we simply target a marginal improvement in the out-of-sample Sharpe ratio, the number of eigenvectors to discard does not play a significant role.

2.5 Conclusion

We have discussed ways to improve the out-of-sample Sharpe ratio of an MSR portfolio constructed based on the two-step approach. To accomplish this goal, we assume a high-dimensional K-factor model and investigate how an improvement can be achieved by adjusting the sample eigenvalues according to certain patterns. Our main results show that simply adjusting one of the first K eigenvalues has a diminishing marginal effect; mildly shrinking a large but non-spiked one and amplifying a small one both lead to an improvement in the out-of-sample portfolio Sharpe ratio under the high-dimensional asymptotics. The effect of adjusting multiple eigenvalues is also studied in the chapter. Our results show that simultaneously amplifying a collection of tail eigenvalues according to a certain nonlinear pattern yields a positive effect for the out-of-sample Sharpe ratio improvement. Our theoretical results are supported by simulation studies presented in this chapter.

By identifying a few eigenvalues adjustment patterns that ensure a marginal improvement in the out-of-sample portfolio Sharpe ratio, we are able to provide an important rationale for existing methods of improving portfolio performance in the literature, such as the shrinkage towards identity method and the spectral cut-off method.

Chapter 3

Eigen Portfolio Selection: A Robust Approach to Sharpe Ratio Maximization

3.1 Introduction

As [86] 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 reasonable proxy for the expected returns. Over the years, hundreds of factors have been put forward to explain the cross-section of expected returns. However, most 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 chapter is not simply a covariance estimation one per se.

Our goal in this chapter is to propose a robust approach to construct a large MSR portfolio when an investor has formed her view on the asset expected returns. Put it differently, we intend to ameliorate the portfolio optimization procedure so as to make a portfolio less contaminated by errors in the input covariance estimator. It is important to clarify that the goal of this chapter is not to propose a good proxy for expected returns; we leave this choice to the investor and instead construct an MSR portfolio based on her choice of the most desirable proxy for expected returns. In other words, we are not concerned with the possible gap between the actual and desirable proxy for expected returns. Hereafter, we refer to the most desirable proxy of expected returns simply as the expected returns for simplicity.

Many attempts have been made in the literature to shield portfolios from 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. Nevertheless, as we will see, they provide insightful ideas which can be applied to improve the MSR portfolios. [61, 29, 40], among others, discuss the effectiveness of imposing constraints on portfolio weights in reducing estimation risk. [104] demonstrate that an optimal combination of the $1/N$ portfolio and a more sophisticated strategy generally outperforms the $1/N$ portfolio. [25] 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, [75] 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, [76] propose a more flexible covariance matrix estimator which shrinks the sample eigenvalues in a nonlinear manner. [43] 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, [39] 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, [22] 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 been fully appreciated: in this case the problem seems at first sight to simply reduce to the problem of covariance matrix estimation. However, it is important to point out that these two problems are not necessarily equivalent. In particular, it does not necessarily take a perfect[1] 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 the distribution of the estimation error in the sample covariance matrix is extreme. But it has been shown by [98] 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, points to an "expected returns approximation" approach as a fruitful way to improve 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 desired proxy for expected returns. This vector is known to us. 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 approximated 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, in this approach we are able to control the maximum amount of approximation error to be introduced and, at the same time, we are also able to reduce the amount of the estimation error as much as possible.

[1] A "perfect" estimator is defined to be one without any estimation error.

Next, 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 sample eigenvalues. These eigenvectors are hereafter referred to as "the first few eigenvectors". We find an interesting equivalence, in terms of the resulting portfolios, among the first approximation method, a "spectral cut-off method" ([22])[2], and a "subspace mean-variance optimization approach" ([25]) 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 to produce a better portfolio performance. The choice of tuning parameters for both methods are carefully discussed in this chapter.

This chapter 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 previously 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 allows us to learn from data.

The rest of this chapter is organized as follows. Section 3.2 establishes a connection between an eigen portfolio and an MSR portfolio. Section 3.3 introduces two concrete forms of the expected returns approximation approach, discusses their respective properties, and illustrates the tuning parameter selection procedure. In Section 3.4 we use four simulated cases to assess the effectiveness of the spectral methods. In Section 3.5 we use empirical

[2]The "spectral cut-off method" refers to the method of using the spectral cut-off regularized covariance matrix estimator when constructing optimized portfolios. The spectral cut-off regularization discards the principal components that are associated with eigenvalues smaller than a certain threshold.

returns from three major equity markets to evaluate the out-of-sample performance of different portfolios.

Before we proceed further with our analysis it is useful to introduce some notations used in the rest of this chapter. We denote matrices by bold capital letters, vectors by bold lower-case letters, and scalars by plain lower-case letters. Let $\mathbf{\Sigma}$ denote an $N \times N$ population covariance matrix. The equation $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ represents the eigen decomposition of $\mathbf{\Sigma}$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_N\}$ is a diagonal matrix containing non-increasingly ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N > 0$ and $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N)$ 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 $\widehat{\mathbf{\Lambda}} = \text{diag}\{\widehat{\lambda}_1, \dots, \widehat{\lambda}_N\}$ is a diagonal matrix containing the non-increasingly ordered sample eigenvalues $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_N$ and $\widehat{\mathbf{U}} = (\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2, \dots, \widehat{\mathbf{u}}_N)$ contains the sample eigenvectors. Throughout the chapter, 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.

3.2 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 [101, 89, 6, 18] 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 N 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 $N \times 1$ vector $\boldsymbol{\mu}$ contains the expected returns over the investment horizon, and $\mathbf{\Sigma}$ is the asset returns covariance matrix. If this is the case, the investor solves for the MSR portfolio from the following

[3]Our definition of Sharpe ratio is in accordance with the usual definition - the expected return in excess of the riskless rate over the standard deviation of the return. Note that as has been mentioned earlier, in this chapter we use the term "return" to denote an asset's/portfolio's return in excess of the riskless rate.

problem:

$$\mathbf{w}_{msr} = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T \boldsymbol{\mu}}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}, \quad \text{s.t. } \mathbf{w}^T \mathbf{1} = 1, \tag{3.1}$$

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

$$\mathbf{w}_{msr} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}, \tag{3.2}$$

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

Assume that we have N eigen portfolios $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$ 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 of the eigenvectors in the same way for scenarios where the expected returns vector is 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 3.2.1. 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. (3.2) is exactly the ith eigen portfolio, i.e., $\mathbf{w}_{msr} = \mathbf{z}_i$.

We now provide some comments on this result. An interesting implication of Proposition 3.2.1 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 ([82]). If this is the case, an investor's optimal choice is to hold the dominant eigen portfolio. Proposition 3.2.1 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.

[4]If this condition is not satisfied, the MSR portfolio does not exist, and the portfolio calculated from eq. (3.2) corresponds to the *minimum* Sharpe ratio portfolio.

Proposition 3.2.2. If $\boldsymbol{\mu}$ can be expressed by eigenvectors of $\boldsymbol{\Sigma}$ as $\boldsymbol{\mu} = \sum_{i=1}^{N} a_i \mathbf{u}_i$[5] and the inequality $\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} > 0$ is satisfied, then the MSR portfolio in eq. (3.2) has weights given by

$$\mathbf{w}_{msr} = \sum_{i=1}^{N} \frac{\frac{a_i}{\lambda_i}}{\sum_{i=1}^{N} \frac{a_i \mathbf{1}^T \mathbf{u}_i}{\lambda_i}} \mathbf{u}_i = \sum_{i=1}^{N} \frac{\frac{a_i \mathbf{1}^T \mathbf{u}_i}{\lambda_i}}{\sum_{i=1}^{N} \frac{a_i \mathbf{1}^T \mathbf{u}_i}{\lambda_i}} \mathbf{z}_i. \tag{3.3}$$

Proposition 3.2.2 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 3.2.2 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, then the MSR portfolio is 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 3.3 below.

3.3 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 [71] 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 a portfolio selection.

The portfolio problem we consider in this section is similar to that described in Section 3.2. Suppose that $\boldsymbol{\mu}$ is an investor's desired proxy for the vector of expected returns of N risky assets over the investment horizon based on all information available to her; but 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 our goal to answer the question of how to generate a desired expected returns proxy. Rather, we treat $\boldsymbol{\mu}$ as being exogenously given and try to improve the MSR portfolio constructed based on this view.

[5] Note that any $\boldsymbol{\mu} \in \mathbb{R}^N$ can be written as a linear combination of the eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N\}$. This equation just gives a notation to the loadings of $\boldsymbol{\mu}$ on the eigenvectors.

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 [76].

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:

$$\widehat{\mathbf{w}}_{msr} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}. \tag{3.4}$$

What concerns us with the above portfolio weights is that there may exist severe estimation error in $\widehat{\boldsymbol{\Sigma}}$ when N 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 shield the portfolio weight estimator from the proliferation of 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 3.2.2, 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 consideration 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 3.3.1 and 3.3.2 respectively.

3.3.1 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, 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 ([98]).

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:

$$(\widehat{a}_1^{cut}, \ldots, \widehat{a}_K^{cut}) = \underset{(a_1,\ldots,a_K)}{\operatorname{argmin}} \left\| \boldsymbol{\mu} - \sum_{i=1}^{K} a_i \widehat{\mathbf{u}}_i \right\|. \tag{3.5}$$

The following proposition presents the solution to the problem.

Proposition 3.3.1. The solution to the optimization problem in eq. (3.5) is:

$$\widehat{a}_i^{cut} = \widehat{a}_i^{ls}, \quad i = 1, 2, \ldots, K, \tag{3.6}$$

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

$$(\widehat{a}_1^{ls}, \ldots, \widehat{a}_N^{ls}) = \underset{(a_1,\ldots,a_N)}{\operatorname{argmin}} \left\| \boldsymbol{\mu} - \sum_{i=1}^{N} a_i \widehat{\mathbf{u}}_i \right\|. \tag{3.7}$$

Note that the problem in eq. (3.7) differs from that in eq. (3.5) in their dimension. It follows from Proposition 3.3.1 that the optimal approximation vector of expected returns is given by:

$$\widehat{\boldsymbol{\mu}}^{cut}(K) = \sum_{i=1}^{K} \widehat{a}_i^{cut} \widehat{\mathbf{u}}_i = \sum_{i=1}^{N} \widehat{a}_i^{ls} \mathbb{1}_{\{i \le K\}} \widehat{\mathbf{u}}_i. \tag{3.8}$$

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:

$$\widehat{\mathbf{w}}_{cut}(K) = \frac{\widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}^{cut}(K)}{\mathbf{1}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}^{cut}(K)} = \sum_{i=1}^{N} \frac{\frac{\widehat{a}_i^{ls} \mathbb{1}_{\{i \le K\}} \mathbf{1}^T \widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}{\sum_{i=1}^{N} \frac{\widehat{a}_i^{ls} \mathbb{1}_{\{i \le K\}} \mathbf{1}^T \widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}} \widehat{\mathbf{z}}_i, \tag{3.9}$$

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:

$$\widehat{\mathbf{w}}_{msr} = \frac{\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}} = \sum_{i=1}^{N} \frac{\frac{\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}{\sum_{i=1}^{N}\frac{\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}\widehat{\mathbf{z}}_i. \tag{3.10}$$

According to eqs. (3.9) and (3.10), 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 $N-K$ 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. (3.9) by plugging in the matrix expression for $\widehat{\boldsymbol{\mu}}^{cut}(K)$ leads to an alternative expression for the portfolio weights:

$$\widehat{\mathbf{w}}_{cut}(K) = \frac{\widehat{\mathbf{\Sigma}}_K^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}_K^{-1}\boldsymbol{\mu}}, \tag{3.11}$$

where $\widehat{\mathbf{\Sigma}}_K^{-1} = \widehat{\mathbf{U}}_K\widehat{\mathbf{\Lambda}}_K^{-1}\widehat{\mathbf{U}}_K^T$ and $\widehat{\mathbf{\Lambda}}_K = \mathrm{diag}\{\widehat{\lambda}_1,\ldots,\widehat{\lambda}_K\}$. Eq. (3.11) conveys an interesting fact: the method of inputting the sample covariance estimator and the approximation 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 $N-K$ sample eigenvalues. This specific way of modifying an inverse covariance matrix is called the spectral cut-off method and has been discussed in [21] and [22]. The spectral cut-off method was originally introduced as a stabilization technique to invert an ill-posed[6] sample covariance matrix.

[25] is a related work according to which the portfolio weight vector takes the same form as that in eq. (3.11)[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

[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 [25] $\boldsymbol{\mu}$ is replaced with its estimator, i.e., $\widehat{\mathbf{w}}_{sub}(K) = \frac{\widehat{\mathbf{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}$.

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. 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 subtly modify an investor's view on expected returns. Thus, cutting off the last $N-K$ principal components is only desirable 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, tacitly magnifies the loading of the expected returns vector on eigenvectors corresponding to the larger eigenvalues. To illustrate this, let us suppose that 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:

$$\widehat{\mathbf{w}}_{sti}(\lambda) = \frac{(\widehat{\boldsymbol{\Sigma}}+\lambda\mathbf{I})^{-1}\boldsymbol{\mu}}{\mathbf{1}^T(\widehat{\boldsymbol{\Sigma}}+\lambda\mathbf{I})^{-1}\boldsymbol{\mu}} = \sum_{i=1}^{N} \frac{\frac{\frac{\widehat{\lambda}_i}{\widehat{\lambda}_i+\lambda}\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}{\sum_{i=1}^{N}\frac{\frac{\widehat{\lambda}_i}{\widehat{\lambda}_i+\lambda}\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}\widehat{\mathbf{z}}_i. \tag{3.12}$$

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 an absolute value.

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 N while the remaining ones are bounded as N 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 N and the sample size go to infinity at the same rate.

Assumption 3.3.1. $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ is a random sample with the distribution of

$$\mathbf{x}_i = \sum_{j=1}^{N} \lambda_j^{\frac{1}{2}} z_{i,j} \mathbf{u}_j$$

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

Assumption 3.3.1 specifies how random samples are generated from the population covariance model. Then, we calculate the sample covariance matrix from $\widehat{\mathbf{\Sigma}} = \frac{1}{T}\mathbf{X}\mathbf{X}^T$, where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$.

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

Assumption 3.3.2. As $T \to \infty$, $\lambda_1^{(T)} > \cdots > \lambda_K^{(T)} \gg \lambda_{K+1}^{(T)} \asymp \cdots \asymp \lambda_{N(T)}^{(T)} \asymp 1$.

For $i < j$, $\lambda_i^{(T)} > \lambda_j^{(T)}$ means that $\lim_{T\to\infty} \frac{\lambda_i^{(T)}}{\lambda_j^{(T)}} > 1$; $\lambda_i^{(T)} \gg \lambda_j^{(T)}$ means that $\lim_{T\to\infty} \frac{\lambda_j^{(T)}}{\lambda_i^{(T)}} = 0$; $\lambda_i^{(T)} \asymp \lambda_j^{(T)}$ means that $c_1 \le \underline{\lim}_{T\to\infty} \frac{\lambda_i^{(T)}}{\lambda_j^{(T)}} \le \overline{\lim}_{T\to\infty} \frac{\lambda_i^{(T)}}{\lambda_j^{(T)}} \le c_2$ for two constants $0 < c_1 \le c_2$. Assumption 3.3.2 implies that the population covariance matrix has a spiked structure: the first K eigenvalues increase as T 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 [8] and [39].[8] However, compared with the assumption made in these references, our assumption about the strength and pervasiveness of the "common factors"[9] is mild. [39] 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 factor" assumption,

[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.

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

Proposition 3.3.2. As Assumptions 3.3.1 - 3.3.2 hold, the portfolio weight estimator $\widehat{\mathbf{w}}_{cut}(K)$ given in eq. (3.9) and eq. (3.11) 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 $T \to \infty$ and $\frac{N(T)}{T} \to c \in (0, 1)$, where $\mathbf{w}_{cut}(K) = \frac{\mathbf{\Sigma}^{-1}\boldsymbol{\mu}^K}{\mathbf{1}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}^K}$ 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 3.3.2 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 critically depends 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 [25] show that when asset returns follow an approximate K factor model, the sample-based subspace mean-variance optimal portfolio $\widehat{\mathbf{w}}_{sub}(K) = \frac{\widehat{\mathbf{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}_K^{-1}\widehat{\boldsymbol{\mu}}}$, 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 $\boldsymbol{\mu}$.

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 undesirable portfolio. In addition, the estimability of K relies on certain assumptions about the rate of the divergent eigenvalues. Under our mild Assumption 3.3.2, K is in general unidentified ([88]). 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

$$K(\delta) = \min \left\{ K : \frac{\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^{cut}(K)\|}{\|\boldsymbol{\mu}\|} \leq \delta \right\}, \tag{3.13}$$

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. (3.9) and eq. (3.10), it is clear that the spectral cut-off method based portfolio estimator with any $K \neq N$ 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.

3.3.2 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 3.3.1. Suppose that the expected returns vector has such a relationship with the sample eigenvectors that $\boldsymbol{\mu} = \widehat{\mathbf{u}}_1 + \widehat{\mathbf{u}}_N$ 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

$$\frac{\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^{cut}(p-1)\|}{\|\boldsymbol{\mu}\|} = \frac{\|\widehat{\mathbf{u}}_N\|}{\|\widehat{\mathbf{u}}_1 + \widehat{\mathbf{u}}_N\|} = 0.71.$$

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 3.3.1 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:

$$(\widehat{a}_1^{sel}, \ldots, \widehat{a}_N^{sel}) = \underset{(a_1,\ldots,a_N)}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{\mu} - \sum_{i=1}^{N} a_i \widehat{\mathbf{u}}_i\right\|^2 + \gamma \sum_{i=1}^{N} \frac{|a_i|}{\widehat{\lambda}_i^c}, \tag{3.14}$$

where $\gamma, c > 0$ are two tuning parameters. With the L_1 penalty in eq. (3.14), 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. (3.14), we can find an explicit solution to this optimization problem. The following proposition presents the solution.

Proposition 3.3.3. The solution to the optimization problem in eq. (3.14) is given by:

$$\widehat{a}_i^{sel} = \text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+, \quad i = 1, 2, \ldots, N, \tag{3.15}$$

where $\widehat{a}_i^{ls}$ is defined in eq. (3.7).

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

$$\widehat{\boldsymbol{\mu}}^{sel}(\gamma, c) = \sum_{i=1}^{N} \widehat{a}_i^{sel}\widehat{\mathbf{u}}_i = \sum_{i=1}^{N} \text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}| - \gamma\widehat{\lambda}_i^{-c})_+\widehat{\mathbf{u}}_i. \tag{3.16}$$

Proposition 3.3.3 explicitly presents how the spectral selection method shifts the expected returns. According to eq. (3.15), 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. (3.8)), 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:

$$\widehat{\mathbf{w}}_{sel}(\gamma, c) = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma, c)}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma, c)} = \sum_{i=1}^{N} \frac{\frac{\text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c})_+\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}{\sum_{i=1}^{N}\frac{\text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c})_+\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}\widehat{\mathbf{z}}_i. \tag{3.17}$$

According to eq. (3.17), 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:

$$\underbrace{\frac{\frac{\text{sign}(\widehat{a}_i^{ls})(|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c})\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}{\frac{\text{sign}(\widehat{a}_j^{ls})(|\widehat{a}_j^{ls}|-\gamma\widehat{\lambda}_j^{-c})\mathbf{1}^T\widehat{\mathbf{u}}_j}{\widehat{\lambda}_j}}}_{\text{relative weight in } \widehat{\mathbf{w}}_{sel}(\gamma,c)} = \underbrace{\frac{\frac{\widehat{a}_i^{ls}\mathbf{1}^T\widehat{\mathbf{u}}_i}{\widehat{\lambda}_i}}{\frac{\widehat{a}_j^{ls}\mathbf{1}^T\widehat{\mathbf{u}}_j}{\widehat{\lambda}_j}}}_{\text{relative weight in } \widehat{\mathbf{w}}_{msr}} \times \underbrace{\frac{\frac{|\widehat{a}_i^{ls}|-\gamma\widehat{\lambda}_i^{-c}}{|\widehat{a}_i^{ls}|}}{\frac{|\widehat{a}_j^{ls}|-\gamma\widehat{\lambda}_j^{-c}}{|\widehat{a}_j^{ls}|}}}_{\text{adjustment factor}}. \tag{3.18}$$

Note that the first term on the RHS of eq. (3.18) 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 the 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 3.3.1.

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

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

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.,

$$\gamma(\delta, c) = \max\left\{\gamma : \frac{\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^{sel}(\gamma, c)\|}{\|\boldsymbol{\mu}\|} \leq \delta\right\}. \tag{3.19}$$

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 on the approximation error, 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 3.5. 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 mean time, 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 3.4.

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 a head and a tail eigenvector. We illustrate this statement by revisiting the Example 3.3.1. Suppose that 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

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

and

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

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 ([19]). Therefore, $\widehat{\mathbf{z}}_1$ very likely has no short positions and a gross exposure of 1; while $\widehat{\mathbf{z}}_N$ 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 specific example to illustrate this point. Empirical results in Section 3.5 will provide further evidence in support for this.

3.4 Simulation Study

In this section, we use a set of simulation results to assess the performance of the methods proposed in Section 3.3.

In the simulation study, we pre-specify the true covariance matrix $\mathbf{\Sigma}$ and the true expected return $\boldsymbol{\mu}$. For each set of parameters (sample size T and number of assets N), we repeat the experiment 500 times. In each replication, $2T$ random returns are independently generated from the multivariate normal distribution $N(\boldsymbol{\mu}, \mathbf{\Sigma})$. We use the first T 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 T 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 combinations of (T, N) to compare the performance of the spectral methods across dimensionality configurations. The $\mathbf{\Sigma}$ 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) $\boldsymbol{\mu} \propto \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_N$, and (4) $\boldsymbol{\mu}$ is randomly generated. In all of 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 $\frac{\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^T\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}}$. 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 3.1 presents the kernel density plot of the out-of-sample Sharpe ratios. Each of the four panels corresponds to a specific configuration of T and N, 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 N/T 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 extent of 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 3.1 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 T and small N 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 outperforms 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 3.1: Kernel density plot of out-of-sample Sharpe ratios for $\boldsymbol{\mu} \propto \mathbf{1}$

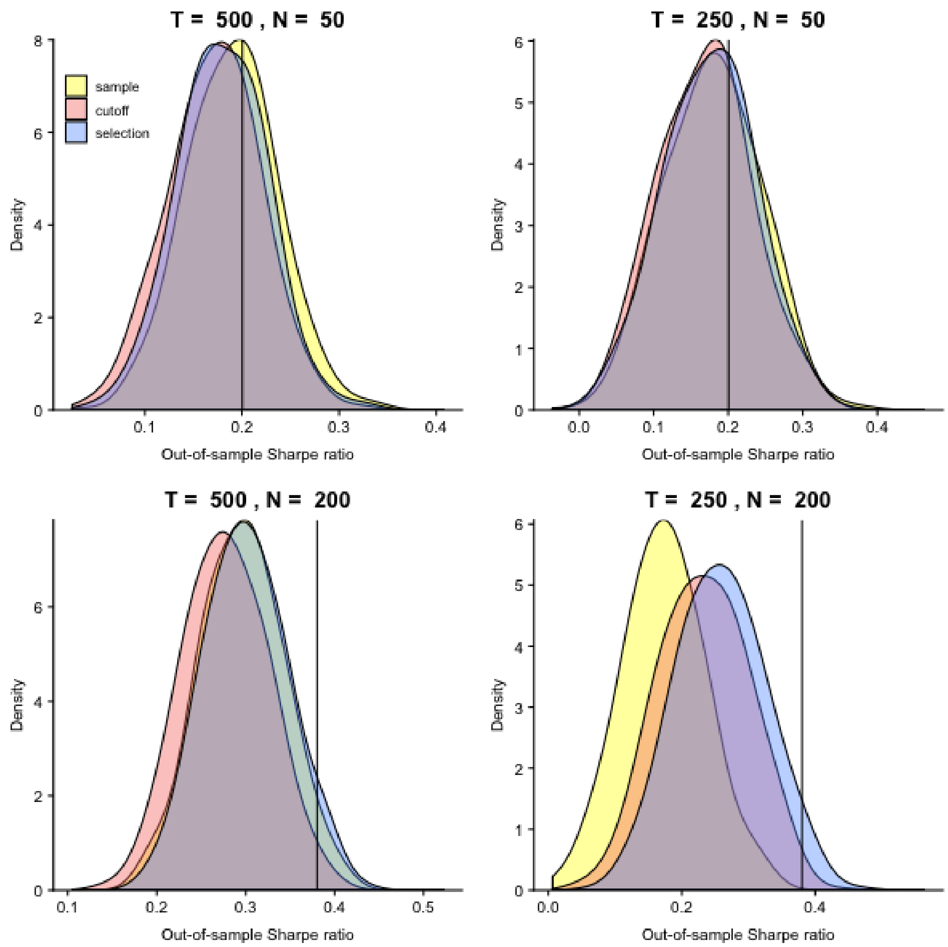
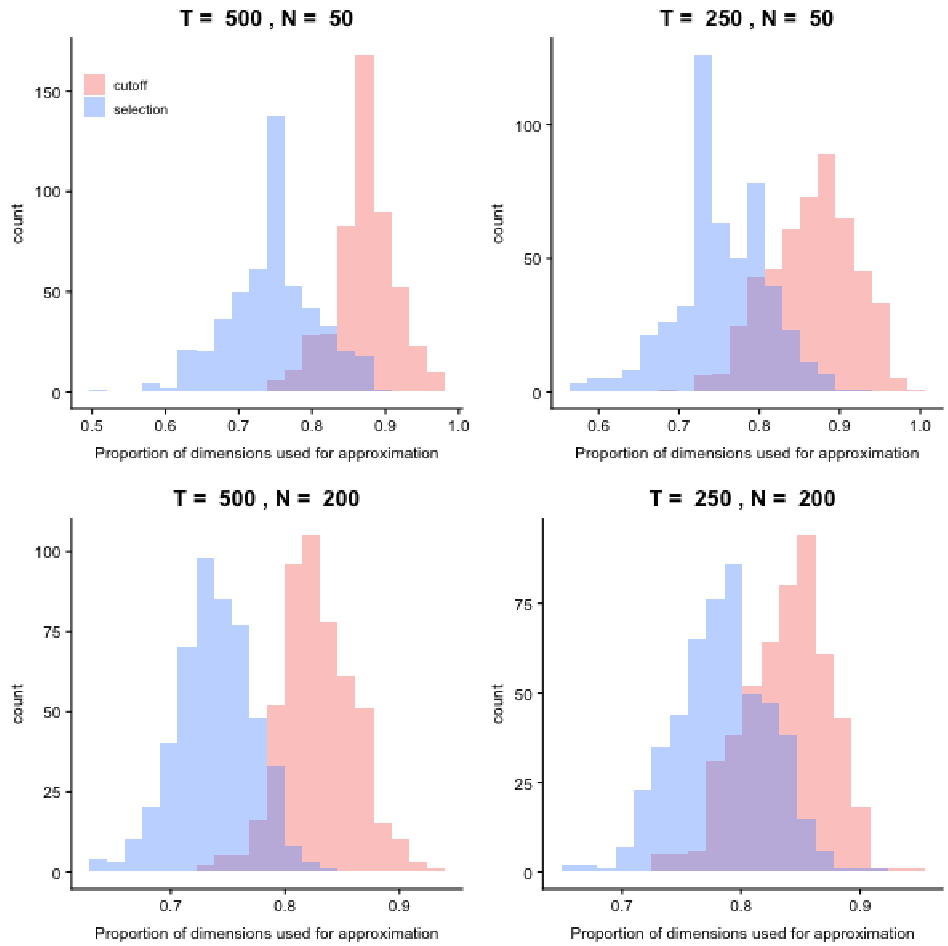


Figure 3.2 shows the histograms of proportion of dimensions used by the spectral methods to approximate $\boldsymbol{\mu}$ from these 500 simulations. We do not include the sample-based method because it does not involve finding an approximation of $\boldsymbol{\mu}$. The darker regions are a result of overlapping of 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 3.2: 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 to improve 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 3.3 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 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 well approximated by the dominant sample eigenvector, given that the dominant sample eigenvector can be more accurately estimated.

Figure 3.4 reinforces the above explanation for the overlapping: the histograms corresponding to both spectral methods reduce to a single bar at $\frac{100}{N}\%$, since in all of the dimensionality settings and in all of the replications, the dominant sample eigenvector approximates $\boldsymbol{\mu}$ sufficiently well.

Case 3: $\boldsymbol{\mu} \propto \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_N$

In this case, we assume that $\boldsymbol{\mu}$ has an equal loading on all of the eigenvectors of $\boldsymbol{\Sigma}$, i.e., we let $\boldsymbol{\mu} = \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_N$. 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 3.5 summarizes the distribution of the out-of-sample Sharpe ratios. As in the previous two cases, in the first two "big T and small N" 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 to 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 in this case.

The above explanation is further supported by Figure 3.6. In the top two panels of Figure 3.6, 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

Figure 3.3: Kernel density plot of out-of-sample Sharpe ratios for $\boldsymbol{\mu} \propto \mathbf{u}_1$

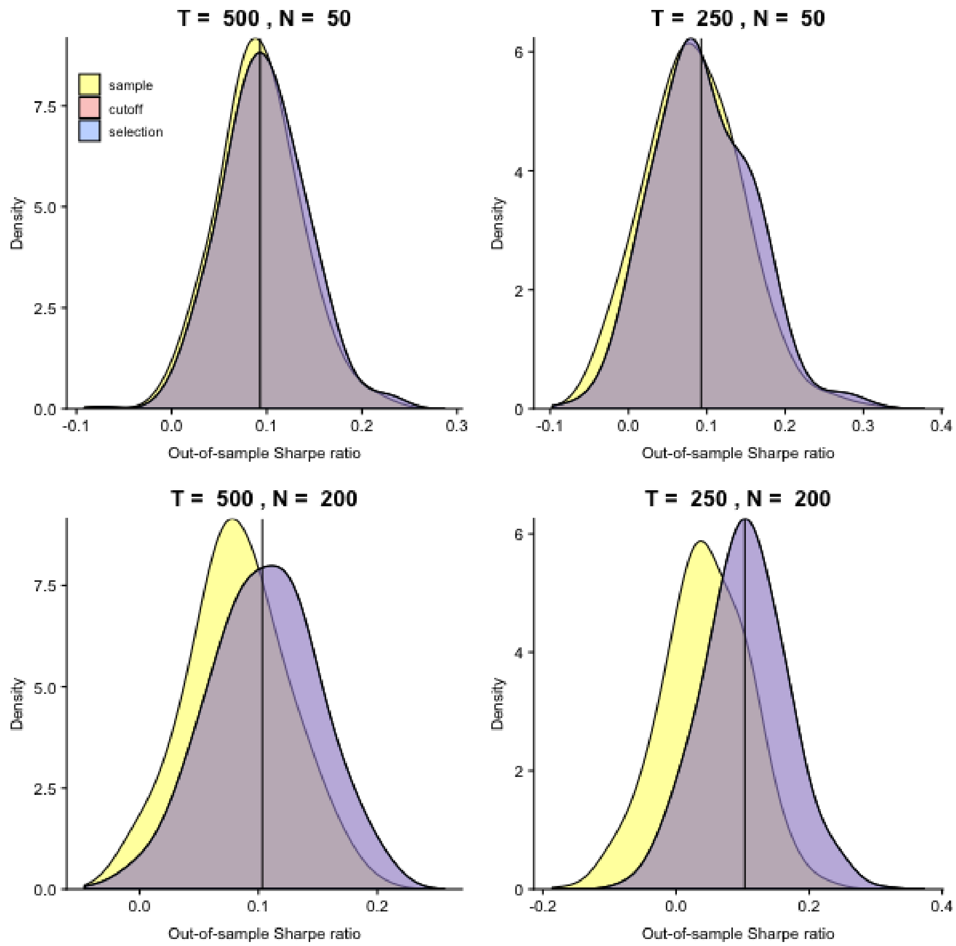


Figure 3.4: Histogram of proportion of dimensions used to approximate $\boldsymbol{\mu}$ for $\boldsymbol{\mu} \propto \mathbf{u}_1$

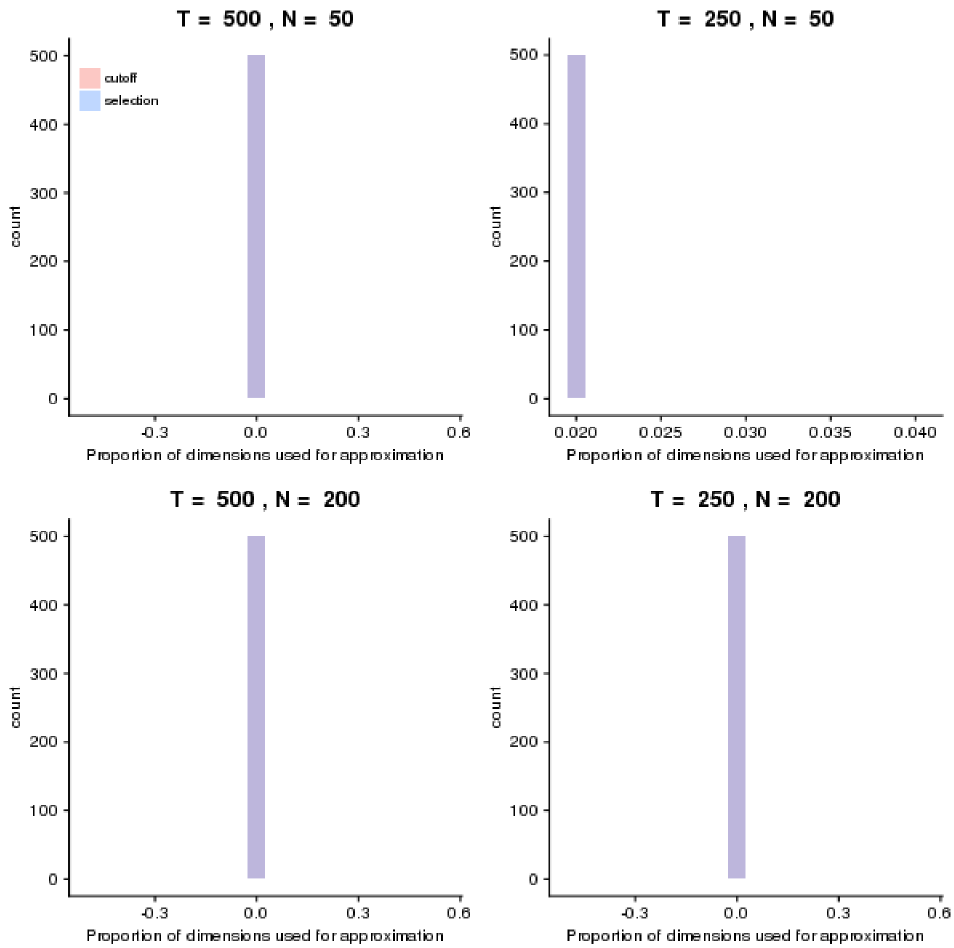
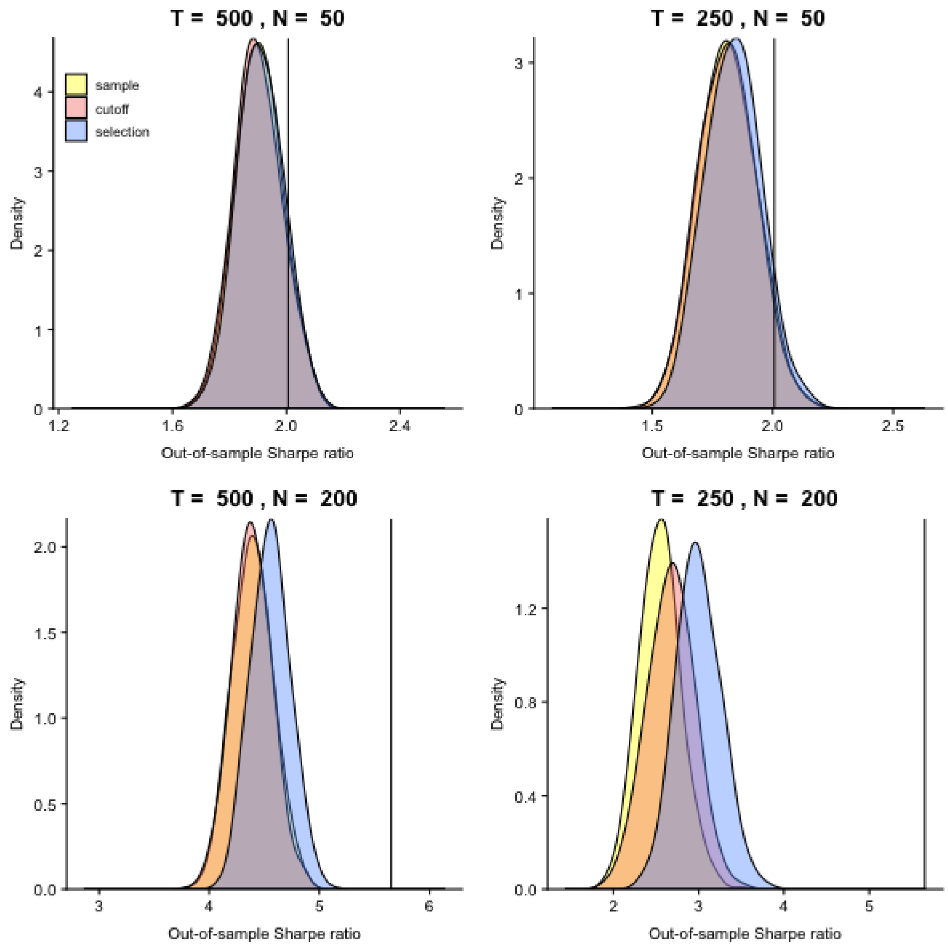


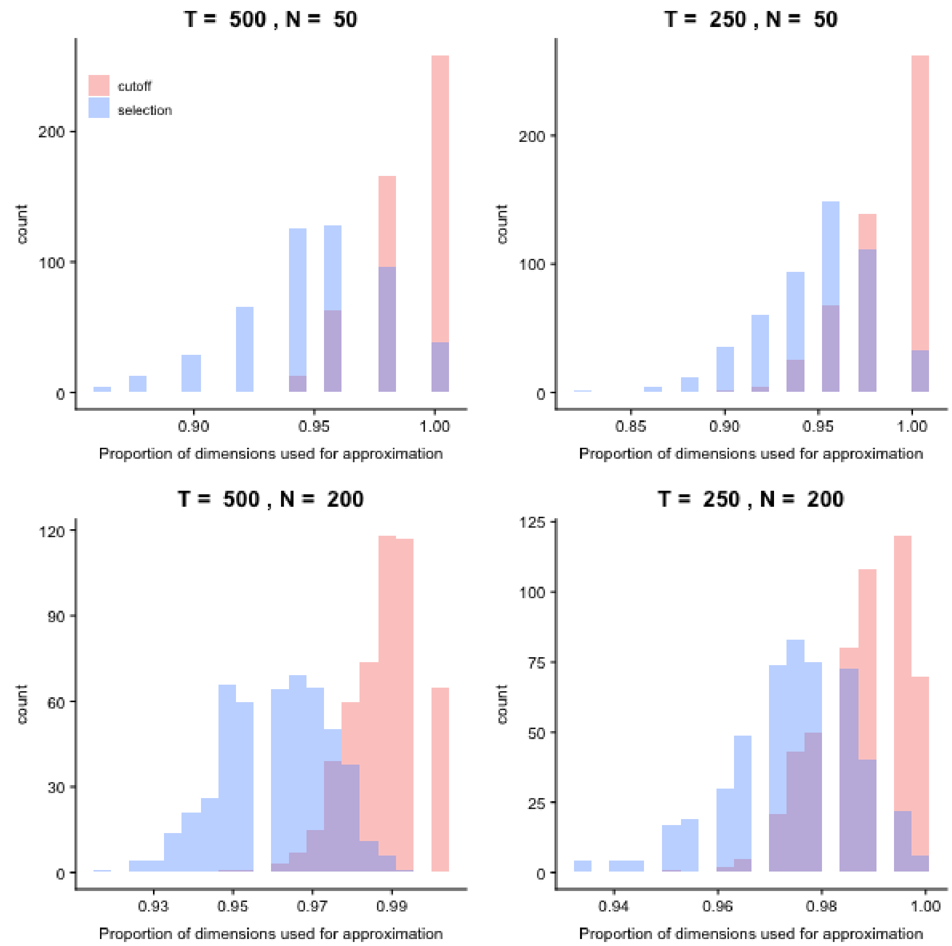
Figure 3.5: Kernel density plot of out-of-sample Sharpe ratios for $\boldsymbol{\mu} \propto \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_N$



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 general spectral selection method to obtain a robust MSR portfolio.

Figure 3.6: Histogram of proportion of dimensions used to approximate $\boldsymbol{\mu}$ for $\boldsymbol{\mu} \propto \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_N$



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

In the last case, we consider a $\boldsymbol{\mu}$, for which each 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 3.7, 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 observe that the the spectral selection method is the most effective method in producing robust MSR portfolios.

Figure 3.8 looks quite similar to the histogram in the previous case. The spectral selection method almost always leads to some dimension reduction.

In summary, 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 T and small N" settings and in addition significantly improves 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.

3.5 Empirical Analysis

3.5.1 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 Sharpe ratios of portfolio.

Figure 3.7: Kernel density plot of out-of-sample Sharpe ratios for randomly generated $\boldsymbol{\mu}$

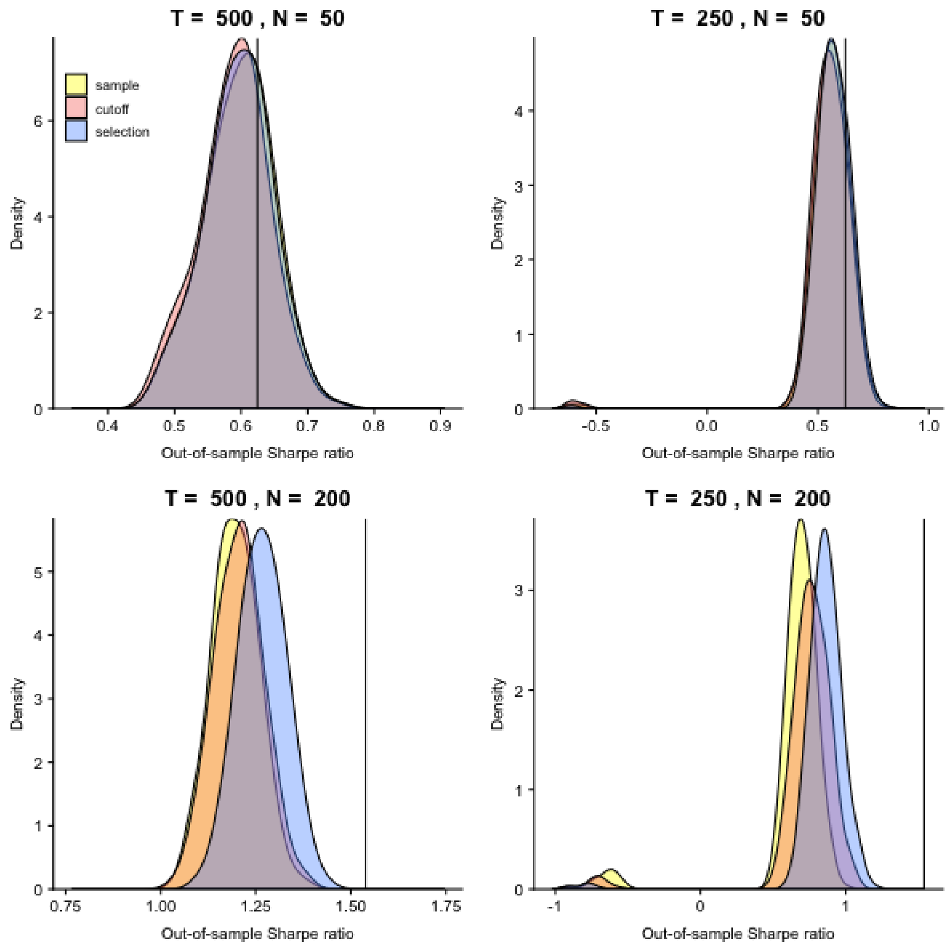
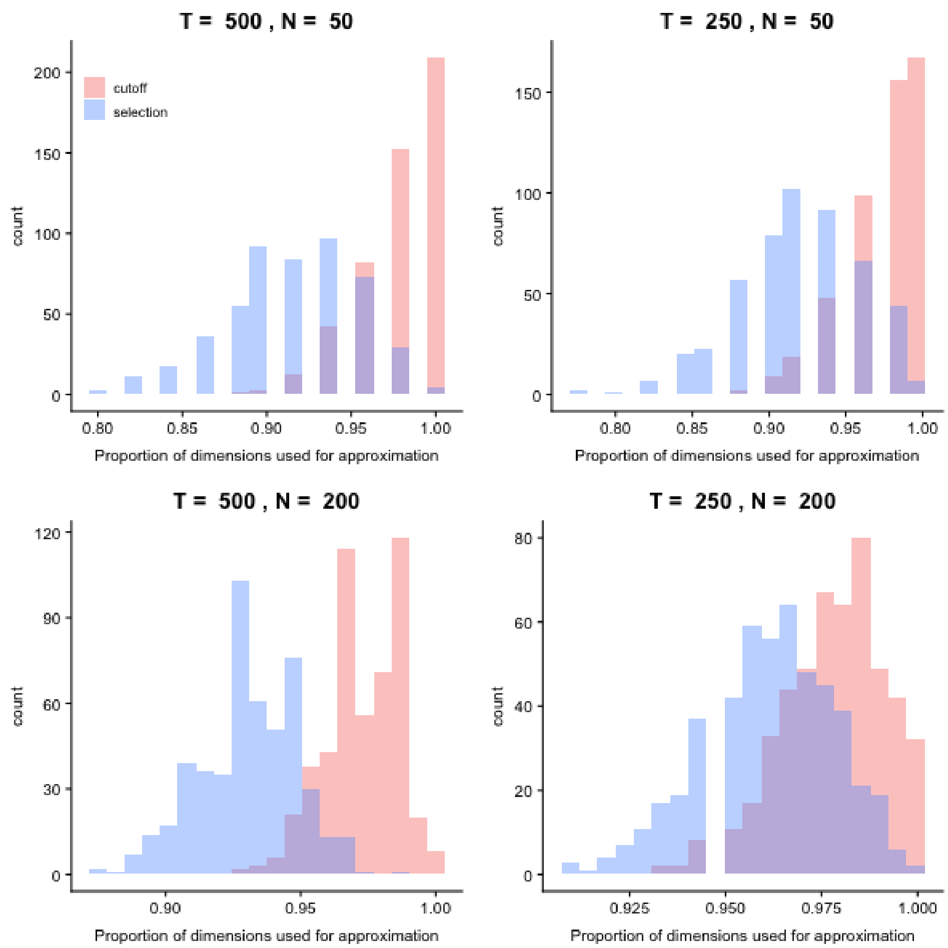


Figure 3.8: Histogram of proportion of dimensions used to approximate $\boldsymbol{\mu}$ for randomly generated $\boldsymbol{\mu}$



Data and procedure

We have taken $\boldsymbol{\mu}$ as an investor's desired 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 ([30]), leads us to use $\boldsymbol{\mu} \propto \mathbf{1}$ as the most desirable 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 $t + 4$ (T 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 desired 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.

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 $1/N$ portfolio, the sample-based MSR portfolio, the POET-based ([39]) MSR portfolio[11],

[11] When using the POET method to estimate the covariance matrix, we use the method in [8] to determine the number of common factors and use the recommended parameter $C = 0.5$ to estimate the residual covariance matrix by the soft thresholding method.

the shrinkage towards identity method based MSR portfolio ([75]), as well as the no-short-selling MSR portfolio ([61]), for which the outperformance among optimized portfolios[12] has been documented in [30]. 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 calculated as
$$\delta_{CV} = \operatorname*{argmax}_{\delta} SR(\delta).$$
- As the portfolio is frequently rebalanced we record a sequence of δ_{CV}. Let $\bar{\delta}_{CV}$ stand for the average of the sequence.

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 $\widehat{\boldsymbol{\Sigma}}$ using the training data.

[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.

- 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., $\widehat{\mathbf{w}}_{sel}(\gamma(\delta, c), c) = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma(\delta,c),c)}{\mathbf{1}^T\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{sel}(\gamma(\delta,c),c)}$, and compute $SR(c)$, which is the Sharpe ratio of the portfolio on the cross-validation set. The optimal c is chosen as
$$c_{CV} = \underset{c}{\operatorname{argmax}}\, SR(c).$$

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 ([40]), as well as the percentage of dimensions used to approximate $\boldsymbol{\mu}$ (for the two spectral methods). All of 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 [10] and in [30]. 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 chapter.

Tables 3.1 - 3.3 summarize the holding period performance measures of all of the portfolios. According to these tables, in all of 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 3.2). 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 among all of the portfolio methods. 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 also works well in other scenarios.

3.5.2 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

Table 3.1: 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	0.00%	0.14	1.00	—

Table 3.2: 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 3.3: 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	—

3.1 - 3.3, 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 also works well in a wide range of dimensionality configurations, as well as in different markets.

Tables 3.4, 3.6, and 3.8 summarize the portfolios' average holding period performance measures when different numbers of stocks enter the portfolios and the training sets contain different numbers of days, with each table corresponding to a market. The portfolio construction procedure is the same as that described in Section 3.5.1, except that instead of always using the past five years' daily returns data to construct portfolios, we use the past $T/252$ years' data for a couple of different T. In addition, within each of Tables 3.4, 3.6, and 3.8, 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 N stocks from an index's component stocks, for each pair of (N, T), we repeat the random draw 20 times, perform the portfolio construction procedure each time when N 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 (N, T) combination, we also record the average $\bar{\delta}_{CV}$ of the 20 replications.

According to the results in Tables 3.4, 3.6, and 3.8, the outperformance of the spectral selection method with $\delta = 0.15$ is quite consistent across different dimensionality configurations, almost always being ranked 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 any 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 3.4, 3.6, and 3.8. 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 3.5, 3.7, and 3.9 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 T and N, therefore we simply recommend a parameter value independent of T and N. It should be noticed that even if for a (N, T) 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 3.4: Average holding period (Jan 1995 - Dec 2016) performance of different portfolios when N stocks are used to construct portfolios and T daily returns are used to train portfolios. All stocks are S&P 500 constituents.

	$N = 50$, $T = 1260$				$N = 50$, $T = 504$			
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
	$N = 100$, $T = 1260$				$N = 100$, $T = 504$			
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 3.5: Average $\bar{\delta}_{CV}$ under different dimensionality configurations. All stocks are S&P 500 index constituents.

(N, T)	$\bar{\delta}_{CV}$
$(50, 1260)$	0.12
$(50, 504)$	0.12
$(100, 1260)$	0.12
$(100, 504)$	0.11

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

Table 3.6: Average holding period (Jan 2006 - Dec 2016) performance of different portfolios when N stocks are used to construct portfolios and T daily returns are used to train portfolios. All stocks are S&P United Kingdom index constituents.

	$N = 40$, $T = 1260$				$N = 40$, $T = 252$			
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
	$N = 80$, $T = 1260$				$N = 80$, $T = 252$			
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 3.7: Average $\bar{\delta}_{CV}$ under different dimensionality configurations. All stocks are S&P United Kingdom index constituents.

(N, T)	$\bar{\delta}_{CV}$
$(40, 1260)$	0.12
$(40, 252)$	0.12
$(80, 1260)$	0.12
$(80, 252)$	0.13

Table 3.8: Average holding period (Jan 2006 - Dec 2016) performance of different portfolios when N stocks are used to construct portfolios and T daily returns are used to train portfolios. All stocks are Japanese Nikkei 225 index constituents.

	$N = 50$, $T = 1260$				$N = 50$, $T = 252$			
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
	$N = 100$, $T = 1260$				$N = 100$, $T = 252$			
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 3.9: Average $\bar{\delta}_{CV}$ under different dimensionality configurations. All stocks are Japanese Nikkei 225 index constituents.

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

$\boldsymbol{\mu} \propto \mathbf{1}$, alternatively, when we do not possess any additional information to forecast $\boldsymbol{\mu}$. If an investor has a certain 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 because in such a scenario the estimation error does not pose a severe problem.

3.6 Conclusion

In this chapter, we show 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 show that it is possible to benefit from the reduced exposure to the estimation error without incurring much loss, although in the process our vector of expected returns will be slightly 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 can be consistently estimated under reasonable assumptions. 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 on the expected return vector. 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 will not be severely distorted. Such a turning parameter δ offers a convenient scheme for striving a sensible 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 than in a low-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 severe. In this scenario, the benefit of introducing 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 chapter.

Chapter 4

When Does the $1/N$ Rule Work?

4.1 Introduction

The $1/N$ rule is a portfolio allocation scheme which allocates an equal share of wealth to each of N available assets on a portfolio rebalancing date. This $1/N$ rule is seen to be at odds with the more explicit portfolio optimization theory of [84], since this rule neither requires any estimation of the input nor involves any explicit optimization process and is therefore deemed as a passive investment strategy. However, there is evidence (see [13]) that many market participants use this simple heuristic when allocating their wealth across different asset classes. Further, [30] published a seminal study on the $1/N$ rule. After conducting an extensive evaluation of the performance of various portfolio methods, including the $1/N$ rule and a few other methods designed to mitigate the effect of estimation errors, the authors reported that "of the 14 portfolio methods evaluated across seven empirical datasets from the US market, none is consistently better than the $1/N$ portfolio in terms of the Sharpe ratio, certainty-equivalent return, or turnover". To complement the influential study of [30], [60] focused on global diversification in the stock market from the perspective of a Eurozone investor and found that none of the optimized portfolios is able to consistently outperform the $1/N$ rule out-of-sample. However, when [42] replicated the experiment of [30] using data from the UK market, he found that a number of optimal portfolio strategies significantly outperform the $1/N$ portfolio, even after an adjustment for a trading cost. Thus, [42] provides a support in defence of the optimized portfolios.

Previous studies reach disparate conclusions about the performance of the $1/N$ portfolio in different countries. [50] also noticed that the benchmark portfolio[1] is not equally efficient in different countries. Therefore, it is intriguing to scrutinize the reason behind such discrepancy. We hypothesize that, beyond luck, there must be some underlying features that make it profitable to follow the $1/N$ rule in a market. A simple thought in support of our conjecture is that, during a crisis period when the price for most stocks falls, there is a faint chance for the $1/N$ rule to yield a positive return, but it is still possible for a well-managed long-short portfolio to earn a profit. This chapter attempts to identify market-specific characteristics which make a market favorable to the $1/N$ rule.

We set out our analysis in this chapter by considering the following question: what features of a market make it most desirable to hold the $1/N$ portfolio? A natural answer would be "if the market parameters (i.e., the expected returns and the covariance matrix) are such that the actual Sharpe ratio maximizing portfolio in the market happens to be the $1/N$ portfolio". This condition, if translated into the mathematical language, will be shown to be $\boldsymbol{\mu} = c\boldsymbol{\Sigma}\mathbf{1}$ for some $c > 0$, where $\boldsymbol{\mu}$ is an $N \times 1$ vector of the expected asset returns in excess of the riskless rate, $\boldsymbol{\Sigma}$ is an $N \times N$ asset returns covariance matrix, and $\mathbf{1}$ is an $N \times 1$ vector of ones. (Hereafter, we use "return" to refer to "return in excess of the riskless rate" for convenience.) Similarly, we are also able to identity a condition which makes the $1/N$ portfolio the least desirable portfolio to hold, i.e., $\boldsymbol{\mu} = c\boldsymbol{\Sigma}\mathbf{1}$ for some $c < 0$.[2]

According to our analyses above, a market is more favorable to the $1/N$ portfolio if the positive proportionality between the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}\mathbf{1}$ holds more strongly. Thus, we devise a measure of how favorable a market is to the $1/N$ rule, which is given by $\cos(\boldsymbol{\mu}, \boldsymbol{\Sigma}\mathbf{1})$, and coin it as the "$1/N$ favorability index" of a given market. This index measures how close the $1/N$ portfolio is to optimality and thus is able to reflect how difficult it is for other portfolios to outperform the $1/N$ rule. Similarly defined cosine measures, i.e., $\cos(\boldsymbol{\Sigma}^{-\alpha}\boldsymbol{\mu}, \boldsymbol{\Sigma}^{1-\alpha}\mathbf{1})$, reflect the same idea. Importantly, our proposed index is the only one that does not entail inverting a covariance matrix, and as a result, we can easily calculate the observed index from asset returns even if the portfolio size is greater than the sample size.

We then study properties of this newly proposed index under the parsimonious yet popular single-factor model for cross-sectional asset returns. An interesting finding from our analysis is that, conditional on the factor returns during a certain period, the (theoretical) $1/N$ favorability index has upper and lower bounds which depend on the sign

[1] In [50], the value-weighted portfolio, instead of the $1/N$ portfolio, is viewed as the benchmark.

[2] In this case, the vertex of the efficient frontier lies below the x-axis, and the first order condition for finding the tangency portfolio will give rise to the minimum Sharpe ratio portfolio. We refer readers to Figure 1 of [68] for an illustration on this point.

of the factor return over the period. If the factor return is positive, the conditional $1/N$ favorability index falls in an interval with strictly positive bounds; otherwise, the index falls in an interval with strictly negative bounds. It is important to emphasize that the finding that bulls markets are more favorable to the $1/N$ portfolio is not as trivial as it may seem to be, because even in bull markets, a portfolio could do better than the $1/N$ rule by buying the top performers and short-selling the relatively bad ones. Further, we show the consistency of the sample-based estimator, which we call the observed $1/N$ favorability index, of the conditional $1/N$ favorability index under high-dimensional asymptotics. The implication is that we should be able to see a positive relationship between the observed $1/N$ favorability index over a certain period and the market average return during the same period. This implication of the single-factor model is tested and supported by both simulation and empirical studies.

The observed $1/N$ favorability index over a certain period is an indication of the difficulty for an optimized portfolio to outperform the $1/N$ portfolio over the same period. If we imagine the extreme case where there are no estimation errors whatsoever, so that the in-sample optimal portfolio coincides with the evaluation period, or out-of-sample, optimal portfolio, then the $1/N$ portfolio at best achieves the same performance as the optimized portfolio does, which happens when the observed $1/N$ favorability index is equal to 1. However, in the presence of estimation errors, there is a huge discount in the out-of-sample performance of an in-sample optimal portfolio. Therefore, the $1/N$ portfolio more likely outperforms the contaminated optimized portfolio if it is closer to out-of-sample optimality. This notion of closeness is measured by the observed $1/N$ favorability index.

We conduct another empirical experiment to study whether the model implication that "bull markets are usually accompanied by a high $1/N$ favorability index" can explain the varying degree of difficulty in outperforming the $1/N$ portfolio in different markets. Based on our proposed criterion for determining whether or not a market is conducive to outperformance of the $1/N$ portfolio, we identify two markets which are favorable to the $1/N$ portfolio, the US and Japanese markets. In addition, we find that it is not difficult to outperform the $1/N$ portfolio in the Australian, Canadian, and UK markets. A comparison of the study period (around 18 years) average returns in these markets shows that the US and Japanese markets are the top two among all markets in terms of the average return during the study period. This empirical evidence supports the proposition that it is more difficult to outperform the $1/N$ portfolio in bull markets. More importantly, it reveals an important fact that the perceived good performance of the $1/N$ portfolio is very likely due to researchers' sole focus on the US market and to the high favorability of the US market to the $1/N$ portfolio.

To complement the previous set of results which is to a large extent observational (because of the small number of markets included), we partition the long study period into sub-periods so that we obtain much more country-period records. We pool the records and perform a logistic regression to examine how the market average return over a period determines the probability that the $1/N$ portfolio outperforms the pre-specified comparison benchmark over the same period. We find a significant positive relationship between the average return and the performance of the $1/N$ portfolio. This result indicates that the single-factor model, despite of its simplicity, is able to capture the behavior of the $1/N$ portfolio in different market conditions.

This work contributes to the literature in several aspects. First, we provide a useful insight into the performance of the $1/N$ portfolio. In particular, we point out that the well-documented outperformance of the $1/N$ portfolio in the US market is not entirely due to the poor quality of the contaminated "optimal" portfolios, but more to the fact that the $1/N$ portfolio is innately closer to mean-variance optimality. In other words, the $1/N$ rule is not as naive as it has been perceived to be. Second, we propose a measure of how favorable a market is to the $1/N$ portfolio. This quantity is shown to be able to effectively measure whether a market is suitable for holding a $1/N$ portfolio, and its sample-based estimator is shown to be a consistent estimator under high-dimensional asymptotics. Lastly, we give an answer to the question of when the $1/N$ rule works. We use both a statistical model and a set of numerical studies to show that the $1/N$ rule works better in bull markets. This finding provides guidance to passive investors who pay extra attention to timing of investment and to international investment.

4.1.1 Literature Review

This chapter sits at the confluence of several strands of literature within the area of portfolio management.

Since the publication of the seminal study of [30], the $1/N$ portfolio has been elevated to be a benchmark for evaluating a portfolio construction method. A number of previous work on the $1/N$ portfolio has tried to explain the reported outperformance of this simple heuristic rule. [107] and [30] themselves attributed the outperformance of the $1/N$ portfolio to the contamination of "optimal" portfolios by the estimation errors in the input components. Some other researchers argued that the reported outperformance of the $1/N$ rule is due to specific research designs which are advantageous to the naive diversification. For instance, [70] pointed out that the research design of [30] makes no attempt to match the risk characteristics of the optimized portfolios with those of the $1/N$ portfolio and is therefore skewed in favor of the $1/N$ rule. In a related work, [65] attributed the inferior performance of the optimized portfolios to their exclusion of the riskless asset. [44] found that a longer investment horizon allows for optimizing strategies to exploit linear predictability in returns and therefore to observe satisfying performance of those strategies. An alternative strand of literature explained the outperformance of the $1/N$ rule from the perspective of the merits of the simple rule itself. For example, [91] claimed that an investor needs to buy the previous period losers and sell the previous period winners to maintain equal weights and that it is the contrarian nature of the strategy that ensures a good performance of the $1/N$ portfolio. In a more recent study, [59] proposed that the $1/N$ rule increases a portfolio's tail risk and results in more concave portfolio returns and the outperformance of the $1/N$ rule acts as a compensation for the tail risk and concavity.

The rest of this chapter is organized as follows. Section 4.2 introduces the $1/N$ favorability index and explains the rationale behind it. Section 4.3 discusses the properties of the $1/N$ favorability index under a single-factor model and the statistical properties of the observed index. The findings in Section 4.3 show that bull markets will give rise to a high $1/N$ favorability index. Section 4.4 assesses the implication of the earlier sections that the $1/N$ portfolio is more difficult to outperform in bull markets than in bear markets. Section 4.5 concludes the chapter.

4.2 $1/N$ favorability index

When explaining the reported failure of optimizing strategies to consistently outperform the simple $1/N$ rule, people are apt to blame those active strategies for various reasons, such as incorporating estimation errors, leading to extreme positions, over-fitting in-sample data, etc., as though it is unacceptable not to be able to beat such a simple rule. However, is it possible that such failure is due to the reason that, the markets where the optimizing strategies fail, are simply in favor of the $1/N$ rule? In this section, we explore such a possibility. In particular, we seek market-specific features which make a market a favorable environment to hold a $1/N$ portfolio.

We set out by considering two buy-and-hold portfolios, a mean-variance optimal one, which maximizes the in-sample Sharpe ratio, and a $1/N$ portfolio. We construct these portfolios on the same starting date and hold them for a common out-of-sample period. In the end, we calculate the holding period Sharpe ratios of both portfolios. Next we consider the following question: what features of the market[3] would make the $1/N$ portfolio hard to beat?

[3] We use the terminology "market-specific feature" to refer to a quantity that characterizes the market as a whole, rather then depicting the profile of individual assets.

In answering the above question, a quite natural thought is that, if the $1/N$ portfolio happens to achieve mean-variance efficiency in terms of the out-of-sample period market parameters, then it would be extremely difficult for the optimized portfolio to outperform the $1/N$ rule. This is, in the first place, due to a lack of stationarity in the time series of asset returns ([86]), which implies that there is often a discrepancy between the in-sample period optimal portfolio and the out-of-sample period one. Moreover, even if the time series of asset returns is stationary, the optimized portfolio will be to some extent, depending on the portfolio size, contaminated by the presence of estimation errors. Various error-mitigating techniques, e.g., [61] and [75], can help reduce but never eliminate those errors altogether.

Now we formally state the market-related condition which makes the $1/N$ portfolio favorable to hold. Let N represent the number of risky assets in the market, $\mathbf{1}$ an $N \times 1$ vector of ones, $\boldsymbol{\mu}$ the $N \times 1$ vector of expected rate of return over the out-of-sample period, and $\boldsymbol{\Sigma}$ the $N \times N$ covariance matrix of asset returns over the same period. Then, the $1/N$ rule produces the *Sharpe ratio maximizing portfolio* over the out-of-sample period if and only if $\boldsymbol{\mu} = c\boldsymbol{\Sigma}\mathbf{1}$ for some $c > 0$. This relationship is easily obtained by equating the Sharpe ratio maximizing portfolio with the equally weighted portfolio and checking the condition for the Sharpe ratio maximizing portfolio to exist. Similarly, we can show that the $1/N$ rule leads to the *Sharpe ratio minimizing portfolio* if and only if $\boldsymbol{\mu} = c\boldsymbol{\Sigma}\mathbf{1}$ for some $c < 0$. Note that the ith element in the column vector $\boldsymbol{\Sigma}\mathbf{1}$ represents the summation of covariances between asset i and each of the N risky assets. We hereafter refer to $\boldsymbol{\Sigma}\mathbf{1}$ as the "vector of aggregate covariances". The condition which makes the $1/N$ rule the most (least) favorable portfolio strategy over a specific period indicates that if a market is temporally featured with a positive proportionality between an asset's expected return and the asset's aggregate covariance with all risky assets, then this market is favorable for the $1/N$ portfolio; conversely, if there is a negative proportionality between the vector of expected returns and the vector of aggregate covariances, no other portfolios can be worse-off than the $1/N$ portfolio. An equivalent statement is that it is desirable to hold a $1/N$ portfolio in a market if taking risks (measured by the aggregate covariance) is rewarded in the market.

Actually, [30] hinted on the same arguments in Section 1.1 of their paper as those stated above by pointing out that the $1/N$ portfolio can be viewed as a strategy that does estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ but imposes the restriction that $\boldsymbol{\mu} \propto \boldsymbol{\Sigma}\mathbf{1}$. However, we recap and formalize the idea by including the other extreme case where the $1/N$ portfolio is the least desirable one. Later it will be seen that the two extreme cases correspond to the upper and lower bounds respectively of our devised measure of a market's favorability to the $1/N$ portfolio. Another aspect that distinguishes our discussion here from that in [30] is our focus on the variation in characteristics of different (either temporally or spatially) markets, instead of the estimation problem in a single market.

Following the same line of thought as in our earlier discussion, the stronger the positive (negative) proportional relationship between the vector of expected returns and the vector of aggregate covariances is, the more (less) favorable the market is to the $1/N$ portfolio. Therefore, the extent to which the proportional relationship holds can be viewed as a measure of how favorable a market is to holding a $1/N$ portfolio. This motivates us to construct a cosine measure expressed as:

$$\cos(\boldsymbol{\mu}, \boldsymbol{\Sigma 1}) = \frac{\langle \boldsymbol{\mu}, \boldsymbol{\Sigma 1} \rangle}{\|\boldsymbol{\mu}\| \|\boldsymbol{\Sigma 1}\|},$$

where $\|\cdot\|$ represents the Euclidean norm of a vector, to quantify how favorable a market is to the $1/N$ portfolio. Hereafter we refer to this cosine value as a market's "$1/N$ favorability index" over the time period which $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ correspond to. A higher value of this index indicates that the market is more favorable to the $1/N$ portfolio over the specific time period. If $\boldsymbol{\mu}$ is perfectly positively proportional to $\boldsymbol{\Sigma 1}$, the favorability index is valued at 1. According to our earlier discussion, in this case the $1/N$ portfolio maximizes the out-of-sample Sharpe ratio and can hardly be outperformed by any other portfolio. Otherwise if $\boldsymbol{\mu}$ is perfectly negatively proportional to $\boldsymbol{\Sigma 1}$, the favorability index becomes -1, and the $1/N$ portfolio would perform extremely poorly.

We need to point out that $\cos(\boldsymbol{\mu}, \boldsymbol{\Sigma 1})$ is not the only possible version of a $1/N$ favorability index. Actually, the relationship $\boldsymbol{\mu} \propto \boldsymbol{\Sigma 1}$ gives rise to a series of possible measures, for instance, $\cos(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \mathbf{1})$ and $\cos(\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}, \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{1})$. However, the version we adopt is the only one among the family of possible indices that does not involve inverting the covariance matrix. This advantage makes it convenient when it comes to calculating the $1/N$ favorability index from realized asset returns, because we do not need to worry about the issue of singularity caused by the number of assets being greater than the sample size. In addition, we will show in the next section that the sample-based version of $\cos(\boldsymbol{\mu}, \boldsymbol{\Sigma 1})$ has the important statistical property of consistency under the high-dimensional asymptotics when both the number of assets and the sample size go to infinity.

Readers are apt to mix up the relationship between $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma 1}$ that makes the $1/N$ portfolio favorable with implication of some models in the asset pricing literature, since most asset pricing models address the problem of how different types of risk are priced in terms of their risk premia. However, a key difference is that asset pricing models are usually derived with an equilibrium condition imposed, while in this chapter, we do not require a market to be in an equilibrium status. We believe that the market parameters, i.e., the expected returns and the covariance matrix, may change from time to time, making some periods more favorable to the $1/N$ portfolio and other time periods less so. Our goal is, as has been aptly indicated by the title of this chapter, to investigate the question of when the $1/N$ rule works well.

4.3 $1/N$ favorability index in bull and bear markets

In this section, we will show, by assuming a single-factor model for the cross-section of asset returns, that a market featured by a high $1/N$ favorability index is usually accompanied by a contemporaneous bullish trend. This model implication will be validated by using both a synthetic asset return dataset and a few real world asset return datasets. The finding of the relationship between the $1/N$ favorability index and the overall market performance will lead to an interesting proposition that the $1/N$ portfolio tends to outperform in bull markets.

4.3.1 Single-factor model

A statistical model in which asset returns are generated by one factor has often been employed in finance ([96, 82]). In this sub-section, we show that this simple model yields a pleasingly testable implication on the relationship between the level of cross-sectional average return and the $1/N$ favorability index in a market over a finite period of time. Since the $1/N$ favorability index reflects to what extent assets are compensated for taking risks, the relationship we discover has an alternative interpretation that the risk-return relationship varies in bull and bear markets.

We assume that the cross-section of asset returns are driven by a single factor whose return over the period from $t-1$ to t is denoted by R_{mt}. The $N \times 1$ vector $\mathbf{R}_t$ contains the return of the N risky assets over the period from $t-1$ to t. The individual asset returns are determined by the factor return through an $N \times 1$ vector of factor loadings $\boldsymbol{\beta}$, as well as by a residual term which is independent of the factor return, i.e.,

$$\mathbf{R}_t = \boldsymbol{\beta} R_{mt} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{\text{iid}}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{4.1}$$

where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. A general consensus about a single-factor model for cross-sectional asset returns is that the factor represents the market, thus the loading measures the extent to which each single asset co-moves with the market.

We are interested in exploring the decisive factors that affect the $1/N$ favorability index over a time period of length T. Define $\overline{\mathbf{R}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t$. Let

$$\boldsymbol{\mu}_T = E[\overline{\mathbf{R}}|R_{m1},\ldots,R_{mT}]$$

and

$$\boldsymbol{\Sigma}_T = E\left[\frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{R}_t-\overline{\mathbf{R}})(\mathbf{R}_t-\overline{\mathbf{R}})^\intercal|R_{m1},\ldots,R_{mT}\right]$$

denote the conditional expectation of the sample mean and that of the sample covariance matrix of T consecutive observations of asset returns respectively. Note that the expectation is taken conditionally on the factor returns, that is, taken over the randomness in the residuals only. Then, according to the single-factor model in eq. (4.1), we obtain:

$$\boldsymbol{\mu}_T = \boldsymbol{\beta}\overline{R}_m$$

and

$$\boldsymbol{\Sigma}_T = \frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2\boldsymbol{\beta}\boldsymbol{\beta}^\intercal + \sigma^2\mathbf{I},$$

where $\overline{R}_m = \frac{1}{T}\sum_{t=1}^{T}R_{mt}$. We name the cosine value of the angle between $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T\mathbf{1}$ the "conditional $1/N$ favorability index":

$$\cos(\boldsymbol{\mu}_T,\boldsymbol{\Sigma}_T\mathbf{1}) = \frac{\langle\boldsymbol{\mu}_T,\boldsymbol{\Sigma}_T\mathbf{1}\rangle}{\|\boldsymbol{\mu}_T\|\|\boldsymbol{\Sigma}_T\mathbf{1}\|}.$$

Next we expand the expression for this index by plugging in the expression for $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T$:

$$\cos(\boldsymbol{\mu}_T,\boldsymbol{\Sigma}_T\mathbf{1}) = \frac{(\boldsymbol{\beta}^\intercal\mathbf{1})\overline{R}_m\left[\|\boldsymbol{\beta}\|^2\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2+\sigma^2\right]}{\|\boldsymbol{\mu}_T\|\|\boldsymbol{\Sigma}_T\mathbf{1}\|}.$$

To analyze this equation, we first seek an explicit expression for its denominator components. According to the single-factor model, we have:

$$\|\boldsymbol{\mu}_T\| = \|\boldsymbol{\beta}\||\overline{R}_m|$$

and

$$\begin{aligned}\|\boldsymbol{\Sigma}_T\mathbf{1}\| &= \left\|(\boldsymbol{\beta}^\intercal\mathbf{1})\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2\boldsymbol{\beta}+\sigma^2\mathbf{1}\right\| \\ &\le (\boldsymbol{\beta}^\intercal\mathbf{1})\|\boldsymbol{\beta}\|\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2+\sigma^2\sqrt{N}.\end{aligned}$$

The inequality comes from expanding the norm and then applying the Cauchy-Schwartz inequality $\boldsymbol{\beta}^\intercal \mathbf{1} \leq \|\boldsymbol{\beta}\|\|\mathbf{1}\|$. Consequently, we obtain an upper bound for $\|\boldsymbol{\mu}_T\|\|\boldsymbol{\Sigma}_T\mathbf{1}\|$:

$$\begin{aligned}\|\boldsymbol{\mu}_T\|\|\boldsymbol{\Sigma}_T\mathbf{1}\| &\leq (\boldsymbol{\beta}^\intercal\mathbf{1})\|\boldsymbol{\beta}\|^2|\overline{R}_m|\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2+\sigma^2\sqrt{N}\|\boldsymbol{\beta}\||\overline{R}_m| \\ &\leq \sqrt{N}\|\boldsymbol{\beta}\||\overline{R}_m|\left[\|\boldsymbol{\beta}\|^2\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2+\sigma^2\right].\end{aligned}$$

The second inequality again comes from applying the Cauchy-Schwartz inequality. It follows that the conditional $1/N$ favorability index has a lower bound when $\overline{R}_m > 0$:

$$\cos(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T\mathbf{1}) \geq \frac{(\boldsymbol{\beta}^\intercal\mathbf{1})\overline{R}_m\left[\|\boldsymbol{\beta}\|^2\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2+\sigma^2\right]}{\sqrt{N}\|\boldsymbol{\beta}\||\overline{R}_m|\left[\|\boldsymbol{\beta}\|^2\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt}-\overline{R}_m)^2+\sigma^2\right]} = \cos(\boldsymbol{\beta}, \mathbf{1}).$$

Similarly, we can show that when $\overline{R}_m < 0$, there is an upper bound for the index, i.e., $\cos(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T\mathbf{1}) \leq -\cos(\boldsymbol{\beta}, \mathbf{1})$. Therefore, we obtain the following range for the conditional $1/N$ favorability index:

$$\cos(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T\mathbf{1}) \in \begin{cases}[\cos(\boldsymbol{\beta}, \mathbf{1}), 1], & \overline{R}_m > 0 \\ \{0\}, & \overline{R}_m = 0 \\ [-1, -\cos(\boldsymbol{\beta}, \mathbf{1})], & \overline{R}_m < 0\end{cases}. \tag{4.2}$$

The range for the conditional $1/N$ favorability index indicates that there is a clear distinction between markets which make the $1/N$ portfolio favorable with those which are not suitable for holding the $1/N$ portfolio. To be specific, if the cross-section of asset returns are driven by a single factor, whether the market is in a bullish or a bearish period is the critical factor of whether holding a $1/N$ portfolio is a wise choice.

Now, let us consider a $1/N$ portfolio strategy that rebalances, or restores to equal weight, after each T periods until the end of the Mth period of length T. In the extreme case where the single (market) factor earns a positive average return in all of these M sub-periods of length T, the M-period average $1/N$ favorability index will be an average of M values in the interval $[\cos(\boldsymbol{\beta}, \mathbf{1}), 1]$ and therefore be in the same interval. In the other extreme case where the single factor earns a negative average return in all the M sub-periods, the M-period average $1/N$ favorability index will be in the interval $[-1, -\cos(\boldsymbol{\beta}, \mathbf{1})]$. But the more probable case is that, some of the M sub-periods are featured with a positive average factor return, and others are accompanied with a negative one. If this is the case, the average $1/N$ favorability index is a result of averaging positive and negative sub period indices. The more of the M sub-periods are featured by a positive factor return, the more likely that the M-period average $1/N$ favorability index is positive. Thus, an important implication of the single-factor model is that the $1/N$ portfolio is more difficult to outperform in a bullish market.

Our analysis above covers two different $1/N$ portfolio strategies. The first is to build a buy-and-hold $1/N$ portfolio and track its performance for a relatively short investment horizon of length T. The second is to regularly rebalance the portfolio so that it is always restored to an equal weight portfolio on rebalancing dates. We assess the regularly rebalanced portfolio based on its performance over a longer investment horizon of length MT. Our analysis based on the single-factor model shows that both portfolio strategies work better in bullish markets. We never assess a buy-and-hold $1/N$ portfolio over a long investment horizon, because the weight of a buy-and-hold $1/N$ portfolio will evolve towards that of a value-weighted portfolio. Later, we will examine the empirical performance of both strategies to confirm the implication of the single-factor model.

If, hypothetically, $\boldsymbol{\beta}$ is positively proportional to the vector of ones, then the conditional $1/N$ favorability index is coerced to be 1 when the market factor earns a positive return. The implication of many asset pricing models that the market factor earns a positive risk premium may lead people to believe that a $\boldsymbol{\beta}$ vector closer to a positive multiple of $\mathbf{1}$ makes the market favorable to the $1/N$ portfolio. However, another side of the story is that the market factor, although is associated with a positive risk premium, usually oscillates between bullish and bearish periods. This means that the gain from the closeness of the $1/N$ portfolio and the Sharpe ratio maximizing portfolio during a bullish period is compromised by the loss from the closeness of the $1/N$ portfolio and the Sharpe ratio minimizing portfolio during a bearish period. Therefore, whether $\boldsymbol{\beta}$ is more proportional to $\mathbf{1}$ or less so is not a factor as important as the trend in the market, in answering the question of when the $1/N$ rule works.

4.3.2 Observed $1/N$ favorability index

Up to this point, we have only discussed properties of the "theoretical" version of the $1/N$ favorability index, in the sense that the two vectors used for calculating the cosine measure are conditional expectations of the vector of realized returns and the vector of realized aggregate covariances. The observed $1/N$ favorability index is calculated directly based on realized asset returns. If we let

$$\widehat{\boldsymbol{\mu}}_T = \overline{\mathbf{R}}$$

and

$$\widehat{\boldsymbol{\Sigma}}_T = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{R}_t - \overline{\mathbf{R}})(\mathbf{R}_t - \overline{\mathbf{R}})^\intercal$$

denote the sample mean and sample covariance matrix of T realized returns respectively, then, the observed $1/N$ favorability index is calculated as:

$$\cos(\widehat{\boldsymbol{\mu}}_T, \widehat{\boldsymbol{\Sigma}}_T\mathbf{1}) = \frac{\langle \widehat{\boldsymbol{\mu}}_T, \widehat{\boldsymbol{\Sigma}}_T\mathbf{1}\rangle}{\|\widehat{\boldsymbol{\mu}}_T\|\|\widehat{\boldsymbol{\Sigma}}_T\mathbf{1}\|}.$$

The following proposition shows that the observed, or sample-based, $1/N$ favorability index is a consistent estimator for the conditional $1/N$ favorability index.

Proposition 4.3.1. Suppose that the parameter of the single-factor model in eq. (4.1) satisfies the following assumptions:

A1. $\|\boldsymbol{\beta}\| = O(\sqrt{N})$ as $N \to \infty$.

A2. There exists a fixed $\delta > 0$, such that $\cos(\boldsymbol{\beta}, \mathbf{1}) > \delta$ as $N \to \infty$.

Under the asymptotics that $N, T \to \infty$ with relative rate $\frac{N}{T} = O(1)$, the observed $1/N$ favorability index is a consistent estimator for the conditional $1/N$ favorability index, i.e.,

$$\cos(\widehat{\boldsymbol{\mu}}_T, \widehat{\boldsymbol{\Sigma}}_T\mathbf{1}) - \cos(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T\mathbf{1}) \xrightarrow{p} 0. \tag{4.3}$$

The consistency of the observed $1/N$ favorability index is a critical property. Without this property, the positive "$1/N$ favorability index - market condition" relationship implied by the single-factor model cannot be observed from realized asset returns. We show the consistency of the observed $1/N$ favorability index under so-called high-dimensional asymptotics that both N and T go to infinity to cater for the case of a relatively small T and a potentially large portfolio size. Note that the two assumptions we impose in Proposition 4.3.1 are relatively mild. A1 is satisfied as long as the elements in $\boldsymbol{\beta}$ are bounded in probability. A2 prevents the cosine of the angle between $\boldsymbol{\beta}$ and $\mathbf{1}$ from going to zero as the portfolio size increases.

4.3.3 Regime-switching in factor return

Up to this point, we have yet to discuss the dynamic of factor returns. However, the dynamic plays an important role in determining whether the relation between the favorability index and the market return implied by the single-factor model could be observed in the real world. If the market return factor remains to be positive across time, the conditional $1/N$ favorability index is always in its positive range and we will fail to observe the pattern described in the previous sub-section. [7] show the relevance of regime-switching in the market in portfolio optimization problems. In this sub-section, we introduce a relatively parsimonious regime-switching lognormal process (RSLN, [55]) for modeling the factor returns. In the next sub-section, we will calibrate model parameters using real world asset returns and then simulate a synthetic asset returns dataset based on these parameters to determine whether there is a positive relationship between the $1/N$ favorability index and the average return in the market.

We assume that the market oscillates between a bullish regime characterized by a high expected market factor return and a bearish regime featured by a low or even negative expected market factor return. Let ρ_t denote the regime applying over the period from $t-1$ to t, $\rho_t = 1, 2$. The transition matrix $P = \{p_{ij}\}_{i,j=1,2}$ contains the probabilities of switching regimes, i.e.,

$$p_{ij} = \Pr\{\rho_{t+1} = j | \rho_t = i\} \quad i = 1, 2,\ j = 1, 2.$$

The probabilities are all independent of t. We assume that the market factor return follows a normal distribution within regime, i.e.,

$$R_{mt} | \rho_t \overset{\text{iid}}{\sim} N(\mu_{\rho_t}, \sigma^2_{\rho_t}). \tag{4.4}$$

In the next sub-section, we will calibrate the model parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and the regime-switching probabilities from real data.

It is worth emphasizing that the positive relationship between the conditional $1/N$ favorability index and the cross-sectional average return is only manifested when we focus on a relatively short time horizon. This is because as $T \to \infty$, both $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T$ converge to their respective limits. In particular, it is easy to show that

$$\lim_{T\to\infty} \boldsymbol{\mu}_T \xrightarrow{p} \boldsymbol{\beta}(\pi_1\mu_1 + \pi_2\mu_2)$$

and

$$\lim_{T\to\infty} \boldsymbol{\Sigma}_T \xrightarrow{p} [\pi_1\sigma_1^2 + \pi_2\sigma_2^2 + \pi_1\pi_2(\mu_1 - \mu_2)^2]\boldsymbol{\beta}\boldsymbol{\beta}^\intercal + \sigma^2\mathbf{I},$$

where $\{\pi_1, \pi_2\}$ is the stationary distribution of the regime-switching model and can be solved from the equation $\pi P = \pi$. Therefore, if we consider an infinite horizon, both the conditional $1/N$ favorability index $\cos(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T\mathbf{1})$ and the average conditional expected return $\frac{1}{N}\mathbf{1}^\intercal\boldsymbol{\mu}_T$ converge to a single point, and thus it is not sensible to study the favorability-average return relationship any longer. When we consider a smaller value of T, there will be more variation in terms of the average market factor return across different samples. Consequently, the positive favorability-average return relationship can be better perceived. Another important reason for considering a small value of T is that, as we have discussed earlier, a $1/N$ portfolio which is held for a long period of time without being rebalanced is no longer a $1/N$ portfolio.

4.3.4 Favorability - market condition relationship: a simulation result

In this sub-section, we check whether the model implication that there is a positive relationship between a market's $1/N$ favorability index and its average return level can be observed from simulated asset returns. Before simulating the asset returns, we first calibrate the parameters in the regime-switching model and those in the single-factor model by using real-world market index returns and individual stock returns.

For the purpose of calibration, we assume a monthly regime-switching frequency, i.e., the time elapsed between $t-1$ and t stands for a month. The data used for calibrating the regime-switching model is monthly returns of the S&P 500 index from January 1960 to December 2016. We use this data to calibrate the regime-switching model because the S&P 500 index is a value-weighted index whose return, according to classical asset pricing theories, can be interpreted as the return of the market portfolio. Below are the parameters of the calibrated regime-switching model, including the annualized expected return and standard deviation in each of the two regimes and the transition probability matrix:

$$\begin{matrix} (12\mu_1, \sqrt{12}\sigma_1) = (11.17\%, 8.42\%) \\ (12\mu_2, \sqrt{12}\sigma_2) = (3.14\%, 17.13\%) \end{matrix}, P = \begin{bmatrix} 0.95 & 0.05 \\ 0.03 & 0.97 \end{bmatrix}.$$

In addition, we use monthly returns data of 100 stocks randomly picked from the S&P 500 index constituents to calibrate σ^2 and $\boldsymbol{\beta}$. According to the calibration results, the bound (other than ± 1) for the conditional $1/N$ favorability index (recall eq. (4.2)) is $\pm\cos(\boldsymbol{\beta}, \mathbf{1}) = \pm 0.96$.

Once all of the parameters have been obtained, we simulate a sequence of 1000 months of market factor returns based on the calibrated regime-switching model. Subsequently, we simulate 1000 monthly returns for 100 hypothetical stocks based on the calibrated single-factor model. In each year (every 12 consecutive hypothetical months, i.e., $T = 12$) we record the observed $1/N$ favorability index $\cos(\widehat{\boldsymbol{\mu}}_T, \widehat{\boldsymbol{\Sigma}}_T\mathbf{1})$ calculated using the simulated stock returns data within this particular year. In addition, we also record the annualized market factor return $\sum_t R_{mt}$ during the year. We collect in total 83 such yearly records based on the 1000 months' simulated data. Figure 4.1 panel (a) shows how the observed $1/N$ favorability index changes when the market factor experiences a different level of return. The pattern is consistent with our conjecture. In particular, when the market is bullish, the observed $1/N$ favorability index is high and holding a $1/N$ portfolio is highly favorable; when the market is in the state of depression, the observed $1/N$ favorability index is negative and the $1/N$ rule would lead to an abysmal performance.

According to our discussions in Section 4.3.1, the range of the conditional $1/N$ favorability index solely depends on the sign of the annualized market factor return. Therefore, if we create a similar plot as the panel (a) of Figure 4.1 but use the conditional $1/N$ favorability index to replace the observed index, we expect to see an S-shape, i.e., the points to the right of the vertical line $x = 0$ lie in the band $[0.96, 1]$, and those to the left of the line $x = 0$ lie in the band $[-1, -0.96]$. According to Proposition 4.3.1, as the number of assets N and the length of holding period T both increase to infinity at a relative rate of $\frac{N}{T} = O(1)$, the observed $1/N$ favorability index will converge to the conditional $1/N$ favorability index at all levels of the market factor return. Therefore, we expect to see an S-shape in the observed $1/N$ favorability index - market factor return plot when N and T are sufficiently large.

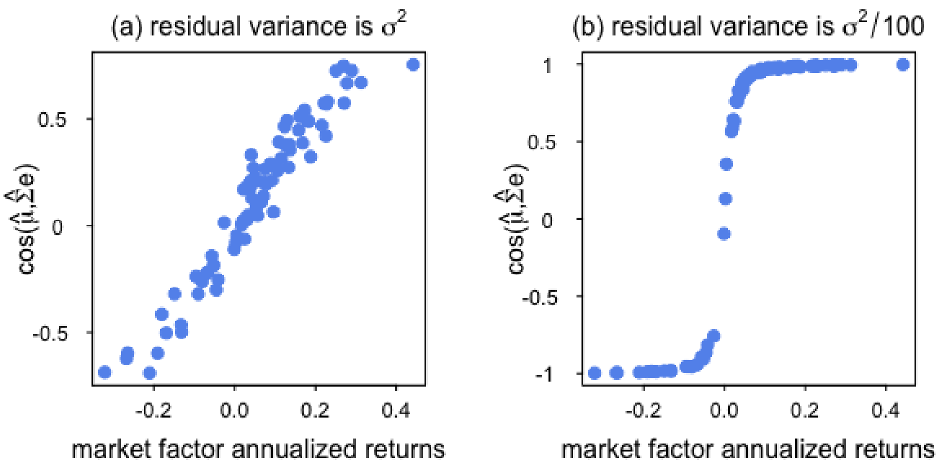
When plotting the panel (b) of Figure 4.1, we simulate another set of cross-sectional asset returns based on the same calibrated parameters but only replace the calibrated residual variance σ^2 with $\sigma^2/100$. Since the residual variance has been greatly reduced, the observed and theoretical $1/N$ favorability indices should be much closer[4], even though we still use $N = 100$ and $T = 12$. Therefore, the clear S-shape in panel (b) of Figure 4.1 provides a firm support to our finding in eq. (4.2).

4.3.5 Favorability - market condition relationship: an empirical result

In this sub-section, we use real-world asset returns data to assess whether a market's observed $1/N$ favorability index varies in bull and bear periods, as predicted jointly by the single-factor model introduced in Section 4.3.1 and the convergence property shown in Section 4.3.2.

[4]This is because the conditional $1/N$ favorability index is defined as the conditional expectation of the observed index, and the conditional expectation is taken only over the randomness in the assets' residual returns.

Figure 4.1: Observed $1/N$ favorability index vs. market factor annualized return: simulated data



We perform the same empirical experiment in seven different countries. For each country, we pick a representative equity market index and construct portfolios using all constituent stocks of the selected index. Names of the seven indices, as well as the countries, numbers of constituent stocks, and time periods of data, are listed in Table 4.1. Daily price data of all constituent stocks of each selected index is downloaded from the Compustat database. For each country, we record the observed $1/N$ favorability index, i.e., $\cos(\widehat{\boldsymbol{\mu}}_T, \widehat{\boldsymbol{\Sigma}}_T \mathbf{1})$, in each year within the data period ($T = 252$), as well as the average return of all assets, i.e. $\frac{1}{N}\widehat{\boldsymbol{\mu}}_T^\intercal \mathbf{1}$, during the same year. Note that this average return is equivalent to the return of the $1/N$ portfolio of all assets in the investment universe. As a result, we obtain 18 pairs of favorability - average return record for those indices whose corresponding data period is from January 1, 1999 to December 31, 2016.

Figure 4.2 presents the relationship between the observed $1/N$ favorability index and the market average return in each country. Each of the first seven panels shows the result obtained from a particular country. The last panel displays a mix of all points collected from the seven markets. In each of the first seven panels, we observe an apparently upward trend, which means that a bull market tends to be accompanied by a high observed $1/N$ favorability index. This is consistent with what our theoretical analysis suggests and with the simulation study in the previous sub-section. According to the last panel, when pooled together, the points from different countries roughly lie on an S-shape curve. Later, it will be seen that this S-shape pattern in pooled records from different countries is critical for us to be able to observe the cross-country phenomenon that the $1/N$ portfolio usually performs well in bullish markets, instead of just during bullish periods. Suppose that we

Figure 4.2: Observed $1/N$ favorability index vs. market average return: empirical data. Each point corresponds to a year of the data period. The x-coordinate represents the annualized return of the $1/N$ portfolio. The y-coordinate measures the observed $1/N$ favorability index.

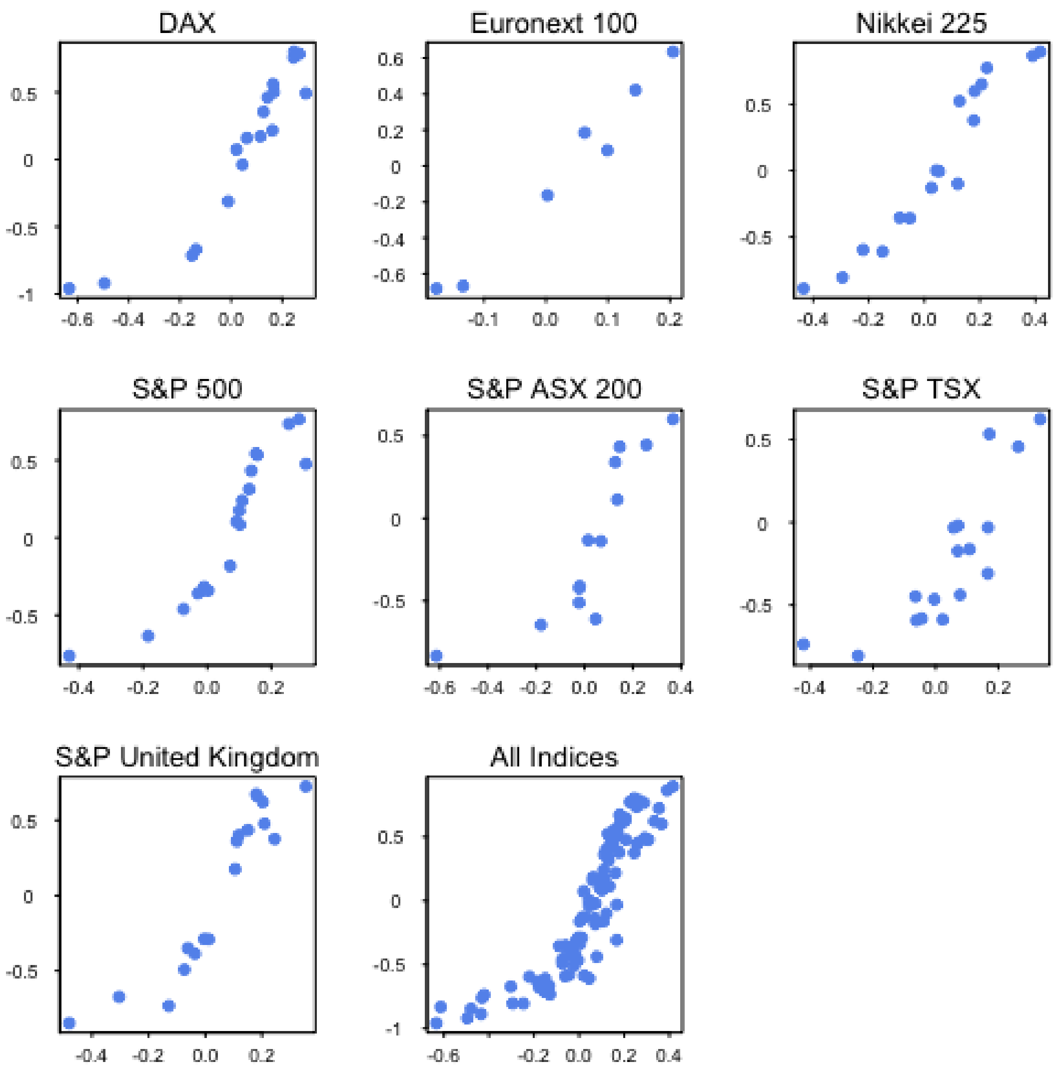


Table 4.1: List of equity market data used in the empirical study

Index name	Country	# of stocks	Data period	Description
DAX	Germany	30	Jan 1, 1999 - Dec 31, 2016	A blue chip index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange.
Euronext 100	EU	100	Oct 27, 2009 - Dec 31, 2016	A blue chip index of the pan-European exchange, Euronext NV.
Nikkei 225	Japan	225	Jan 1, 1999 - Dec 31, 2016	A stock market index for the Tokyo Stock Exchange (TSE).
S&P 500	US	500	Jan 1, 1999 - Dec 31, 2016	An American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ.
S&P ASX 200	Australia	200	Jan 23, 2004 - Dec 31, 2016	A market-capitalization weighted and float-adjusted stock market index of stocks listed on the Australian Securities Exchange.
S&P TSX	Canada	250	Jan 1, 1999 - Dec 31, 2016	A benchmark Canadian index representing roughly 70% of the total market capitalization on the Toronto Stock Exchange (TSX).
S&P United Kingdom	UK	86	Jan 1, 1999 - Dec 31, 2016	A sub-index of the S&P Europe 350 - includes all UK-domiciled stocks from the parent index.

see an upward trend in the panel corresponding to each market but cannot see a clear single upward trend when the points from each market are pooled (note that we may see a few S-shape curves lying next to each other). If this happens, a bull market (in terms of its average return in comparison with that in other countries) does not necessarily present a favorable condition for the $1/N$ portfolios because the $1/N$ favorability index is not necessarily high.

Ideally, we would like to appeal to an asset pricing model that predicts a positive relationship between the $1/N$ favorability index and the average return level. However, most classical asset pricing models, e.g., the CAPM, take a perspective of equilibrium and as a result do not admit a negative expected return of the market portfolio, thus making it difficult to justify the negative risk-return relationship in bear markets. Moreover, according to either the classical form of the CAPM or many of its extensions, there is some idiosyncratic-type risk that cannot be rewarded in the form of a higher expected return. Therefore, it is hardly possible to connect the risk measured by aggregate covariance to an extant asset pricing model. Since our focus is more on an ex-post (or realized), instead of an ex-ante (or expected), risk-return relationship in different market conditions, we resorted to a statistical model to account for our observation.

Fortunately, the single-factor model for cross-sectional asset returns is both realistic (given its degree of parsimony) and rich in implication. Since it is commonly agreed that the market return is the dominating factor that drives the cross-section of equity returns, the single-factor model seems to provide a reasonable explanation for the empirical fact that bull markets are usually accompanied by a high $1/N$ favorability index.

4.4 Empirical performance of the $1/N$ portfolio

In Section 4.3, we have shown by using both theoretical justification and numerical evidence that bull markets are usually accompanied by a high $1/N$ favorability index. Therefore, if the $1/N$ favorability index is a good measure of the difficulty for an optimized portfolio to outperform the $1/N$ portfolio in terms of the Sharpe ratio, the $1/N$ portfolio should be more difficult to beat in bull markets. In this section, we conduct an empirical study to check whether the difficulty for an optimized portfolio to outperform a $1/N$ portfolio indeed depends on the market average return. If the answer to this question is affirmative, it means that we have been adopting an overly demanding benchmark for evaluating portfolio optimization methods in bull markets and that the difficulty for an optimized portfolio to outperform the $1/N$ rule in such markets is partly due to the fact that the $1/N$ portfolio is close to being optimal.

Before proceeding, we first discuss how to quantify whether or not it is difficult for an optimized portfolio to outperform the $1/N$ portfolio in a particular market.

4.4.1 Assessment of the $1/N$ portfolio

Due to variation in market environment, it is not desirable to use an absolute measure, such as the Sharpe ratio, to evaluate the performance of the $1/N$ portfolio in different markets. [69] propose a novel portfolio performance measure which assesses the goodness of a portfolio based on the probability for a uniformly distributed random portfolio to outperform it. We use a similar portfolio comparison idea here. To quantify how well the $1/N$ portfolio performs in a particular market, we select a widely-recognized well-performing optimized portfolio as a benchmark. If the $1/N$ portfolio yields a higher Sharpe ratio than the benchmark portfolio does, we deem the $1/N$ portfolio as being hard to beat in the market of interest. Alternatively, we view this particular market as favorable to the $1/N$ portfolio. The choice of the benchmark portfolio will be discussed in the next section.

To ensure that the outperformance of one portfolio over another in terms of the holding period Sharpe ratio is statistically significant, instead of a result of purely random noise, we apply the Sharpe ratio difference test proposed in [73] to determine whether any two portfolios yield significantly different Sharpe ratios. The authors suggest to construct a studentized time-series bootstrap confidence interval for the difference of the Sharpe ratios and to conclude that the two ratios are statistically different if zero is not contained in the interval. This test is applicable to a wider range of testing scenarios, for example, when returns have heavier tail than the normal distribution or are serially dependent[5].

As has been stated in the introduction section, although we will carry out a systematic comparison between the $1/N$ portfolio and an optimized portfolio in the subsequent sections, we do not take a specific stand with regard to our preference for any of the portfolios; instead a comparison is made only for the purpose of providing a relative measure of how well the $1/N$ portfolio performs. That is, we are more interested in identifying the common features of the markets which favor the $1/N$ portfolio.

4.4.2 Selection of the portfolio benchmark

As mentioned earlier, an ideal benchmark for assessing the performance of the $1/N$ portfolio would be a widely-acknowledged well-performing optimized portfolio. This guideline directs us to consider a minimum-variance (MV) type portfolio as the benchmark. The reason for choosing an MV type portfolio is that it implicitly exploits risk-based pricing anomalies ([95]) and is widely recognized ([61], [27], [30]) for its surprisingly high returns and low realized volatility. Notably, a few techniques have been developed in the literature to refine the sample-based MV portfolio to make it less vulnerable to estimation errors. We equip our selected benchmark portfolio with such techniques, so that the benchmark portfolio is more likely to yield a good out-of-sample performance, making it more sensible to assess the difficulty of outperforming the $1/N$ portfolio based on a comparison with the benchmark portfolio.

We choose two variations of the sample-based MV portfolio as the benchmark for assessing the performance of the $1/N$ portfolio. The first one is the MV portfolio based on

[5]It is important to point out that this test statistics evaluates the performance between any two portfolios based on a single data set. A legitimate concern that can be raised for this type of tests, including the tests used in the empirical work reported, for instance, by [30], is whether the reported outperformance results could be attributed to the use of pre-selected datasets of stock portfolios, often characteristics based portfolios, inadvertently introducing a sample selection bias in the process. An alternative approach would assess the performance of portfolio rules based on multiple datasets and summarizes aggregate performance for any two portfolio rules.

the "shrinkage towards identity" covariance matrix estimator (see [75]). We hereafter refer to this portfolio as a "shrinkage-based MV portfolio". The weight vector for this portfolio is given by:

$$\widehat{\mathbf{w}}_{s-MV} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}^{\intercal}\widehat{\boldsymbol{\Sigma}}_s\mathbf{w} \quad \text{s.t. } \mathbf{w}^{\intercal}\mathbf{1} = 1,$$

where $\widehat{\boldsymbol{\Sigma}}_s$ is the shrinkage estimator for the covariance matrix. In our empirical study, this estimator is obtained from the past three years' daily price history. The second optimized portfolio we use as a benchmark is a short-sale-constrained MV portfolio, whose weight vector is given by:

$$\widehat{\mathbf{w}}_{c-MV} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}^{\intercal}\widehat{\boldsymbol{\Sigma}}\mathbf{w} \quad \text{s.t. } \mathbf{w}^{\intercal}\mathbf{1} = 1 \text{ and } \mathbf{w} \geq 0,$$

where $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix estimator. Again, in our analysis, the sample covariance matrix is estimated from three years' daily returns. Compared with the sample-based MV portfolio, the shrinkage-based MV portfolio is less vulnerable to estimation errors, because the shrinkage estimator for covariance matrix corrects the over-dispersed sample eigenvalues and thus is closer to the population covariance matrix in the sense of the Frobenius norm. Imposing the no-short-sale constraints is an alternative way of resolving the issue of estimation errors. [61] point out that constraining portfolio weights to be non-negative can reduce the risk in estimated optimal portfolios even when the constraints turn out to be wrong by showing that a short-sale constraint on the MV portfolio is equivalent to shrinking the elements of the sample covariance matrix. Another reason for using the short-sale-constrained MV as a benchmark is the finding reported by [30] that "of all the optimizing models studied here, the minimum-variance portfolio with constraints studied in [61] performs the best in terms of Sharpe ratio".

We also made an attempt to use some variations of the sample-based mean-variance portfolio as the performance benchmark for determining whether or not the $1/N$ portfolio is difficult to beat, but the MV type portfolios lead to a superior Sharpe ratio in most scenarios. So we give up selecting a mean-variance type portfolio as the benchmark.

4.4.3 Cross-country performance of $1/N$ portfolios

In this sub-section, we construct and evaluate the portfolios by using a number of datasets that include both equity returns in the US market and those in some non-US markets. In particular, we select the same equity market indices as those used in Section 4.3 and use their respective constituent stocks to construct a number of portfolios. A dynamic list of index constituents and daily price data are downloaded from the Compustat database. Daily price data of most index components cover the period from January 1, 1996 to December 31, 2016, except for the two indices (i.e., Euronext 100 and S&P ASX 200) which did not exist on January 1, 1996. For these three indices, their corresponding components' daily price data ranges from the date of their creation to December 31, 2016.

We are particularly interested in determining whether and how difficult it is for an optimized portfolio to outperform the $1/N$ portfolio, as measured by the relative performance of the latter compared with the benchmarks, in each of the seven markets we study. Below are the details about the portfolio construction procedure that we adopt in this chapter. All of the three portfolios are rebalanced at the same frequency. Later, we will report the performance measures of the portfolios constructed under an annual rebalancing scheme as well as a monthly rebalancing scheme. For a comparison purpose, at each portfolio rebalancing date, we use the same set of stocks to build the three updated portfolios. The stocks that enter the updated portfolios are those that 1) are an index constituent on the rebalancing date and that 2) have at least three years' price history before the rebalancing date. All portfolios are held for a fixed amount of time, depending on the rebalancing frequency, and then rebalanced again. This portfolio construction procedure leads to a lengthy holding period of around 18 years for the portfolios whose corresponding index was created before 1996.

Table 4.2 summarizes the holding period Sharpe ratios of the $1/N$ and the two benchmark portfolios, as well as the p-values for testing whether the $1/N$ portfolio yields a Sharpe ratio significantly different from that of the benchmark portfolio. The p-value for each Sharpe ratio difference test is reported below the Sharpe ratio of the benchmark portfolio and is shown in parentheses. The Sharpe ratio of an MV type portfolio is marked in a bold font if the p-value for the Sharpe ratio difference test is smaller than 0.1. The Sharpe ratios in panels (a) and (b) are calculated as if there is no transaction cost. Panels (c) and (d) contain the holding period Sharpe ratios which have been adjusted for transaction costs. We set the proportional transaction costs equal to 50 basis points per transaction as assumed in [10] and in [30]. If we denote by c the proportional transaction cost, then the evolution of wealth from time t to time $t+1$ for a portfolio strategy k is

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

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. In addition to the Sharpe ratio, we also record the holding period annualized returns (adjusted for transaction

Table 4.2: Holding period performance (adjusted for transaction costs) of the $1/N$, shrinkage-based MV, and shortsale-constrained MV portfolios. Portfolio "s-MV" stands for the shrinkage-based MV portfolio. Portfolio "c-MV" stands for the shortsale-constrained MV portfolio. In panels (a) and (b), the performance measures are reported as if any transaction cost is ignored. In panels (c) and (d), the performance measures are adjusted for a proportional transaction cost of 50 basis points. In panels (a) and (c), all of the three portfolios are rebalanced annually. In panels (b) and (d), they are rebalanced monthly. The "NA" indicates a negative Sharpe ratio.

	(a) Annual rebalance, no trans. cost				(b) Monthly rebalance, no trans. cost			
	$1/N$ return	$1/N$ Sharpe	s-MV Sharpe	c-MV Sharpe	$1/N$ return	$1/N$ Sharpe	s-MV Sharpe	c-MV Sharpe
DAX	3.43%	0.15	**0.44**	0.23	3.79%	0.17	0.37	0.26
			(0.05)	(0.54)			(0.27)	(0.52)
Euronext 100	2.90%	0.15	0.39	0.18	3.85%	0.21	0.19	0.16
			(0.42)	(0.90)			(0.96)	(0.85)
Nikkei 225	4.25%	0.18	0.09	NA	5.26%	0.22	0.03	NA
			(0.74)	(0.44)			(0.50)	(0.31)
S&P 500	6.54%	0.33	0.28	0.38	6.60%	0.31	0.36	0.35
			(0.80)	(0.72)			(0.87)	(0.81)
S&P ASX 200	2.30%	0.13	0.51	**0.39**	2.15%	0.12	**0.62**	**0.48**
			(0.12)	(0.06)			(0.08)	(0.04)
S&P TSX	3.91%	0.26	**0.68**	**0.74**	4.52%	0.27	0.69	**0.70**
			(0.08)	(0.02)			(0.13)	(0.01)
S&P United Kingdom	4.35%	0.24	0.37	**0.56**	4.01%	0.21	0.36	**0.54**
			(0.59)	(0.05)			(0.58)	(0.05)

	(c) Annual rebalance, with trans. costs				(d) Monthly rebalance, with trans. costs			
	$1/N$ return	$1/N$ Sharpe	s-MV Sharpe	c-MV Sharpe	$1/N$ return	$1/N$ Sharpe	s-MV Sharpe	c-MV Sharpe
DAX	3.38%	0.15	0.42	0.22	3.72%	0.16	0.31	0.23
			(0.12)	(0.60)			(0.43)	(0.62)
Euronext 100	2.83%	0.15	0.34	0.16	3.75%	0.20	0.02	0.10
			(0.51)	(0.98)			(0.62)	(0.63)
Nikkei 225	4.22%	0.18	NA	NA	5.22%	0.22	**NA**	NA
			(0.53)	(0.49)			(0.04)	(0.20)
S&P 500	6.49%	0.33	0.04	0.34	6.54%	0.31	**NA**	0.25
			(0.16)	(0.94)			(0.01)	(0.73)
S&P ASX 200	2.17%	0.13	0.44	0.36	1.96%	0.11	0.37	**0.39**
			(0.21)	(0.14)			(0.40)	(0.09)
S&P TSX	3.77%	0.25	0.58	**0.70**	4.33%	0.26	0.36	**0.60**
			(0.14)	(0.02)			(0.71)	(0.07)
S&P United Kingdom	4.29%	0.23	0.32	**0.53**	3.94%	0.20	0.18	0.48
			(0.77)	(0.06)			(0.93)	(0.11)

costs) of the $1/N$ portfolio, which is a reflection of the average return of each equity market during the holding period. This quantity is recorded because we are interested in whether or not the difficulty in outperforming a $1/N$ portfolio depends on the average return in the market.

According to Table 4.2, if we ignore the transaction costs and if the portfolios are rebalanced on an annual basis, it is not difficult to outperform the $1/N$ portfolio in Germany, Australia, Canada, and UK, in the sense that at least one of the benchmarks significantly outperforms the $1/N$ portfolio and no one significantly underperforms. If we switch to a monthly rebalancing frequency, it is still not difficult to outperform the $1/N$ portfolio in the same few countries except Germany. It is important to point out that all of the four countries mentioned above experience a relatively low average market return, evidenced by a return of the $1/N$ portfolio of not more than 4.52%. In panels (a) and (b) of Figure 4.2, no $1/N$ portfolio significantly outperforms any benchmark portfolio. The results lead us to conclude that Germany, Australia, Canada, and UK markets are not favorable to the $1/N$ portfolio but it is not possible to say definitely about which market is favorable to the $1/N$ portfolio.

If transaction costs are taken into account and if the portfolios are rebalanced on an annual basis, it is not difficult to outperform the $1/N$ portfolio in the Canadian and UK markets. When the portfolios are rebalanced on a monthly basis, the $1/N$ portfolio significantly outperforms a benchmark portfolio but is not significantly outperformed by any benchmark portfolio in the US and Japanese markets. In addition, it is not difficult to outperform the $1/N$ portfolio in the Australian and Canadian markets. Combining the results in panels (c) and (d), we conclude that the the $1/N$ portfolio is difficult to beat in US and Japan, but not so in Australia, Canada, and UK. A closer scrutiny at the return of the $1/N$ portfolio in each market reveals that the US and Japanese markets experienced the highest level of average market return during the portfolio holding period (5.22% and 6.54%, respectively). Therefore, the empirical evidence supports our proposition that the $1/N$ portfolio is more difficult to outperform in bull markets.

It is useful to clarify the rationale for presenting the results in panels (c) and (d), since in this chapter, we do not attribute the outperformance of a $1/N$ portfolio to practical reasons such as low transaction costs. The reason for including panels (c) and (d) is that, solely based on panels (a) and (b), we cannot find any market/country which is favorable to the $1/N$ portfolio in the sense that the $1/N$ portfolio significantly outperforms at least one benchmark portfolio and does not significantly underperform any. But this does not imply that the $1/N$ portfolio performs equally poorly across countries. When we take transaction costs into consideration, the performance of the benchmark portfolios deteriorates more than the $1/N$ portfolio does, especially when the monthly rebalancing frequency is adopted.

As a result, the variation across markets in terms of the difficulty to outperform the $1/N$ portfolio is manifested, and we are able to find two markets (i.e., S&P 500 and Nikkei 225) which are more favorable to the $1/N$ portfolio and then investigate whether those markets are featured by a high average return compared with those not favorable to the $1/N$ portfolio.

In a nutshell, our empirical investigation shows that when the monthly rebalancing scheme is adopted, the assorted performance of the $1/N$ portfolio is more prevalent: it significantly outperforms at least one benchmark portfolio and does not underperform any in the US and Japanese market; it significantly underperforms at least one benchmark portfolio and does not outperform any in the Australian and Canadian markets. Our findings are in accordance to those reported in [30], which focuses on the US market, and those reported in [42], which studies the UK market. The observation that the US and Japanese markets are featured by a high average return and that the Australian and Canadian markets are accompanied by a lower one supports our earlier analysis on the relationship between the $1/N$ favorability index and the market condition. Moreover, our results suggest that the perceived good performance of the $1/N$ portfolio is due to 1) many researchers' focus on the US market and 2) the bullish trend of the US market makes it a favorable environment for holding a $1/N$ portfolio.

Admittedly, the way in which we have quantified the superiority of the $1/N$ portfolio makes it credible to assert that the $1/N$ portfolio is not hard to beat in the Australian, Canadian, and UK markets but it less so to state that the $1/N$ portfolio is hard to beat in the US and Japanese markets, because the benchmark portfolios we have selected are not necessarily always the best-performing ones among all the optimized portfolios. However, since the MV type portfolios are known to have a good performance and we also have applied effective estimation error controlling techniques to improve the portfolios' out-of-sample performance, our conclusion on the cross-country performances of the $1/N$ portfolio cannot be easily dismissed.

4.4.4 Cross-sample performance of buy-and-hold $1/N$ portfolios

The empirical results obtained in Section 4.4.3 validate our earlier proposition that a $1/N$ portfolio is more difficult to be outperformed in bull markets. However, those results are mostly observational. This sub-section complements the previous set of results by including a systematic analysis on the relationship between the performance of a $1/N$ portfolio and the average market return. To accomplish this goal, we artificially enlarge the sample size of markets by switching our focus from a regularly rebalanced portfolio with a long investment horizon to a number of buy-and-hold portfolios which have a much shorter holding period.

To assess the conjecture that the $1/N$ portfolio more likely outperforms in a market with a high average return, we partition the 18-year holding period in the experiment in Section 4.4.3 into sub-periods and study the relationship between whether or not the $1/N$ portfolio outperforms during a sub-period and the market average return during the same sub-period. To ensure robustness of the conclusion to different lengths of each sub-period, we perform the study with a sub-period of one year and one month respectively. Since the sub-periods we consider in our study are relatively short, we let the portfolios be constructed at the beginning of each sub-period and held through the sub-period without rebalancing. Thus, we do not need to account for any transaction cost when calculating the sub-period performance measures. To ensure robustness of the results to different choices of the benchmark portfolio, we first treat the shrinkage-based MV portfolio as the benchmark when determining whether or not the $1/N$ portfolio outperforms and then repeat the experiment by treating the shortsale-constrained MV portfolio as the one to compare with. For clarity, we describe the data collection procedure below as if the length of the sub-period is one year. The other case where the sub-period is one month is readily understood.

For each year in the holding period, we record the return of the $1/N$ portfolio, which is also known as the market average return, as well as the Sharpe ratio of the two portfolios of interest, the $1/N$ portfolio and the benchmark portfolio (one of the shrinkage-based MV and the shortsale-constrained MV portfolios). As a result, for the markets with a representative index created before 1996, we obtain 18 such yearly records. Pooling the records from all of the seven markets together, we have 107 records in total. In some of the records, both the $1/N$ portfolio and the benchmark portfolio yield a negative Sharpe ratio. We discard those records because negative Sharpe ratios are generally more difficult to interpret and not amenable to a meaningful comparison.[6] We label the remaining 84 records with either 1 or 0, according to whether the $1/N$ portfolio or the benchmark portfolio yields a higher Sharpe ratio. We refer to the binary label as the "$1/N$ superiority indicator". Note that if one portfolio yields a positive Sharpe ratio and another yields a negative Sharpe ratio, we view the former as the more superior portfolio. We code the relative performance of the $1/N$ portfolio into a binary variable instead of simply taking the difference of the two Sharpe ratios because, on one hand, when any one of the Sharpe

[6] For example, if Portfolios A and B both achieve an annual return of -5% and standard deviations of 5% and 20% respectively, then Portfolio A has a more negative Sharpe ratio. However, Portfolio A may be the more desirable one, because it is much less volatile compared with Portfolio B. Therefore, a direct comparison between two negative Sharpe ratios can be misleading.

Table 4.3: Estimation and hypothesis testing results for logistic regression of the $1/N$ superiority indicator on the market average return. Significance codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1   1. In (a) and (b), the sub-period is one year. In (c) and (d), the sub-period is one month. In (a) and (c), the shrinkage-based MV portfolio is used as the benchmark for generating the $1/N$ superiority indicator. In (b) and (d), the shortsale-constrained MV portfolio is used as the benchmark.

(a) One year, shrinkage-based MV				(b) One year, shortsale-constrained MV			
	Estimate	Std. Error	p-value		Estimate	Std. Error	p-value
$\hat{\beta}_1$	-2.36	0.53	1.05e-05 ***	$\hat{\beta}_1$	-2.18	0.50	1.28e-05 ***
$\hat{\beta}_2$	14.14	3.31	1.98e-05 ***	$\hat{\beta}_2$	12.42	3.02	4.00e-05 ***
model	–	–	1.69e-08 ***	model	–	–	1.99e-07 ***

(c) One month, shrinkage-based MV				(d) One month, shortsale-constrained MV			
	Estimate	Std. Error	p-value		Estimate	Std. Error	p-value
$\hat{\beta}_1$	-0.94	0.10	<2e-16 ***	$\hat{\beta}_1$	-0.77	0.09	<2e-16 ***
$\hat{\beta}_2$	34.51	2.71	<2e-16 ***	$\hat{\beta}_2$	23.35	2.21	<2e-16 ***
model	–	–	<2.2e-16 ***	model	–	–	<2.2e-16 ***

ratios is negative, the difference becomes hard to interpret, and on the other hand, the maximum attainable Sharpe ratio in different markets and sub-periods varies, making the level of difference in Sharpe ratios quite volatile across samples.

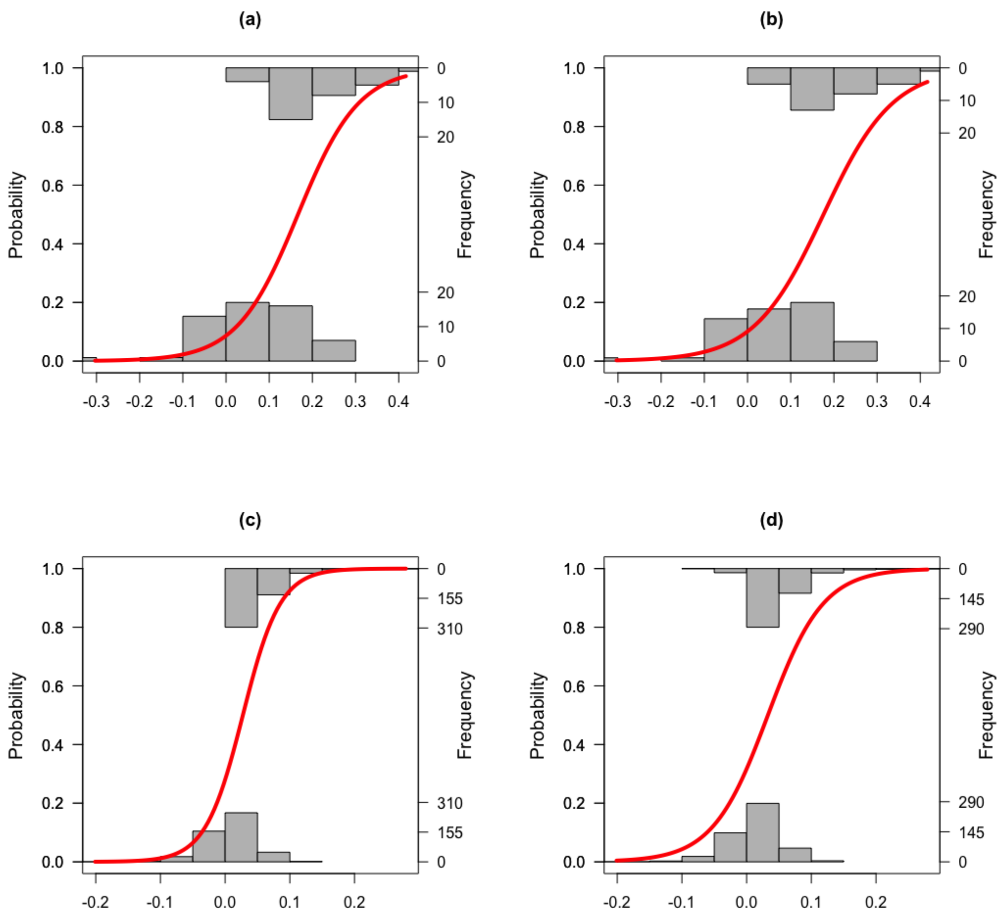
Then we perform a logistic regression, where the explanatory variable is the market average return during a sub-period (r_i) and the dependent variable is the $1/N$ superiority indicator assigned to each record (s_i). Recall that $s_i = 1$ corresponds to the case where the $1/N$ portfolio yields a higher Sharpe ratio, and $s_i = 0$ otherwise. Let p_i denote the probability that the $1/N$ portfolio outperforms in a market with an average return of r_i, i.e., $p_i = \Pr\{s_i = 1|r_i\}$. The logistic regression model is given by the following equation:

$$\log \frac{p_i}{1 - p_i} = \beta_1 + \beta_2 r_i.$$

The logistic regression model is estimated by using the maximum likelihood method. The estimates of the parameters and the estimated standard errors for $\hat{\beta}_1$ and $\hat{\beta}_2$, the respective p-values based on a Wald test, as well as the p-value for the joint significance of $\hat{\beta}_1$ and $\hat{\beta}_2$ based on a likelihood ratio test are reported in Table 4.3. The model is fitted in four different settings, namely all the possible combinations of the two choices of length of each sub-period and the two choices of benchmark portfolio.

Figure 4.3 shows the fitted probability that the $1/N$ portfolio outperforms the benchmark portfolio as well as the histograms of sub-period average returns grouped by the $1/N$

Figure 4.3: Logistic regression of the $1/N$ superiority indicator on the market average return. The x-axis measures average return (annualized) in a market during a sub-period. The grey barplots are histograms of the average returns grouped by the $1/N$ superiority indicator. The histogram at the top corresponds to the group in which the $1/N$ portfolio outperforms. The red curve represents the fitted probability that the $1/N$ portfolio is superior to the benchmark portfolio. In (a) and (b), the sub-period is one year. In (c) and (d), the sub-period is one month. In (a) and (c), the shrinkage-based MV portfolio is used as the benchmark for generating the $1/N$ superiority indicator. In (b) and (d), the shortsale-constrained MV portfolio is used as the benchmark.



superiority index. According to the figure, the higher the market average return is, the more probable that the $1/N$ portfolio outperforms the benchmark portfolio. According to the hypothesis testing results in Table 4.3, both β_1 and β_2 in the model are statistically significant, which reinforces our proposition that the $1/N$ portfolio performs better in bull markets. Furthermore, this conclusion is robust to the choice of the length of the sub-period and the choice of the benchmark portfolio.

4.5 Conclusion

We have devised a market-specific measure of how favorable a market is to holding a $1/N$ portfolio and name it the "$1/N$ favorability index". This index measures the deviance between the $1/N$ portfolio and the true optimal portfolio in a market. Consequently, the greater the $1/N$ favorability index is, the more likely that the $1/N$ portfolio outperforms an optimized portfolio.

We have analyzed the behavior of the $1/N$ favorability index under a single-factor model which explains the cross-section of asset returns using a market factor. The single-factor model predicts an S-shape of the conditional $1/N$ favorability index vs. average market factor return curve, which means that bull markets tend to be accompanied by a positive $1/N$ favorability index and vice versa. The conditional $1/N$ favorability index can be consistently estimated by its sample estimator, which is called the observed $1/N$ favorability index, even when both the portfolio size and the sample size are large. Therefore, bull markets are expected to be accompanied by a high observed $1/N$ favorability index as well. Since the observed $1/N$ favorability index reflects the difficulty for an optimized portfolio (containing estimation errors) to outperform the $1/N$ portfolio during the period that the observed index corresponds to, a further implication is that the $1/N$ portfolio is more difficult to beat in bull markets.

We have conducted a rather exhaustive empirical study to validate the model implications. By examining the cross-country performance of the $1/N$ portfolio relative to two carefully selected portfolio benchmarks, we find that over the study period of 18 years, the $1/N$ portfolio is difficult to beat in the US and Japanese markets, but not so difficult to beat in the Australian, Canadian, and UK markets. The US and Japanese markets experienced the highest market average return over the study period. Therefore, the empirical evidence supports the implication of the single-factor model together with our analyses. A more in-depth empirical study was then conducted to investigate the relationship between the superiority of the $1/N$ portfolio and the market average return by performing a logistic regression and doing statistical tests. The results provide a formal evidence in support of our finding that the $1/N$ rule works well in bull markets.

Chapter 5

Age Matters

5.1 Introduction

Financial economists have long been interested in the empirical distribution of individual stock returns. These returns provide the raw inputs for the evaluation of portfolio strategies as well as a testing ground for asset pricing theories. Indeed [84] in his classic paper on portfolio selection advocated the use of the empirical distribution of historical stock returns as the first step in providing parameter estimates for his optimization algorithms. More recently [14] has conducted an extensive analysis of individual stocks using returns from the Center for Research in Securities Prices (CRSP) database.

Portfolios of individual stocks are attractive to risk averse investors because of their potential diversification benefits. The equally weighted $1/N$ strategy has been widely studied in the finance literature. It is a very simple strategy since it involves no estimation, no optimization, no short positions and has relatively little turnover compared to other strategies. [13] document that it is widely used in practice by participants of defined contribution pension plans as a heuristic method for choosing asset classes. Despite its naive construction this strategy has been shown to outperform most alternative strategies. For example [30] examine the performance of the $1/N$ rule using a variety of datasets. They attribute the superior performance of the equally weighted strategy to the presence of estimation risk and its well known perverse interaction with optimization.

[20] also demonstrate the superior performance of an equally weighted $1/N$ rebalanced portfolio over the value weighted market portfolio. They use individual stock returns from the CRSP database for the period 1926-1997 to construct their equally weighted portfolio. They attribute this outperformance mainly to the small firm effect:

> “because of higher returns on small firms, an equally weighted portfolio of as few as five randomly chosen firms can provide the same level of expected utility as the value weighted market portfolio.”

[91] show that equally weighted portfolios outperform value weighted portfolio based on samples of individual stocks in the S&P indices. They show that the major source of the extra alpha in the equally weighted portfolio is due to the contrarian nature of the strategy.

This chapter examines the performance of equally weighted portfolios and we show that there is an additional reason for their superior returns. It is worth emphasizing that our portfolios are made up of individual stocks from the entire CRSP database. In contrast the datasets used by [30] where the N components of the equally weighted portfolios are themselves portfolios[1] or indices for seven of their datasets. Their eighth dataset is based on simulated stock returns for a single factor model. Our equally weighted portfolios are constructed as in [20]. We use the same comprehensive dataset as [14] since it facilitates comparisons with his results.

This approach, where the components of the equally weighted portfolios are individual securities rather than other portfolios, is better suited to our purpose. It permits us to keep track of the time series properties of the individual stocks and in particular their ages. Another difference between using individual stocks and portfolios is that when a stock is delisted it disappears from the equally weighted portfolio. If this happens it is replaced with another stock drawn at random from the available pool. This newly added stock will be representative of the market as a whole in particular in terms of age. The other stocks in the portfolio will age by one period so that the portfolio as a whole will grow older.

A simple and effective way to compare the performance of the equally weighted portfolio with that of the market is to use comparably sized portfolios which contain N equally weighted stocks at the start of each period. These portfolios are routinely liquidated at the end of each period and a new set of N stocks is selected at random from the available pool. By means of this construction these portfolios are representative of the market as a whole. [14] used the same type of construction except that his portfolios were value weighted instead of equally weighted. Following his convention we refer to these portfolios as equally weighted bootstrapped portfolios or just bootstrapped portfolios. We denote the traditional $1/N$ portfolios as equally weighted rebalanced portfolios or just rebalanced portfolios. One important property of the bootstrapped portfolios is that because of the

[1] For example their first data set consists of ten sector portfolios of the S&P plus the US equity market portfolio for the period 1981-2002.

periodic rebalancing they have the same[2] exposure to reversals as the traditional $1/N$ portfolios.

This chapter compares the returns on the rebalanced portfolio with the returns on the bootstrapped portfolio over the 1926-2016 period spanned by the CRSP data base. While there is some secular variation in the relative performance of the two types of portfolio over time, our most striking finding is that the rebalanced portfolio yields higher realized returns than the bootstrapped portfolio during the most recent forty-year period: 1977-2016. This finding is robust to the portfolio size and to the choice of different starting dates and to the investment horizon within this period. We contend that this difference is not due to the rebalanced portfolio benefitting from reversals since the bootstrapped portfolio will also benefit to the same extent. During the 1926-1976 period the returns on the rebalanced portfolio are very similar to the returns on the bootstrapped portfolio

There are two arguments for why one might not expect the rebalanced portfolio to outperform the market. The first has to do with delisting since holding a stock until it disappears from the market does not seem to be a smart strategy. Actually, the popular belief that being delisted is bad news is somewhat misleading. This is because a stock can disappear from the market for reasons other than bankruptcy. For example the most frequent reason for delisting is merger and acquisition, which often reflects the past success of a company. Even for the stocks that exit from the market due to unfavorable reasons investors rarely lose all of their investment. The second reason is that our results appear to run counter to the size[3] effect. If on average portfolios of small stocks outperform portfolios of large stocks and if it is true that older firms in general tend to be larger than younger firms and if age is positively correlated with firm size, the stocks in the rebalanced portfolios being older than average will be larger than the stocks in the bootstrapped portfolios. We explain in this chapter why the situation is much more nuanced than this and we disentangle the intertwined effects of size and age.

In this chapter we argue that the reason for the performance difference between the rebalanced portfolios and the bootstrapped portfolios stems from the relation between stock age and stock return. The age distribution of the stocks in the bootstrapped portfolio will be very similar to that of the stock universe whereas the age distribution of the stocks in the rebalanced portfolio will typically be older than those in the stock universe. Thus the age profile of the N stocks in the rebalanced portfolio will be older than those of the N stocks in the bootstrapped portfolio. If stock return is related to age this will impact the

[2]We show this in A.4.2.

[3]There is considerable evidence that the importance of the size effect has declined in recent years. See [58], [3].

relative performance of the two types of portfolios. We show in this chapter that there is a significant positive relation between stock return and stock age during the period 1977-2016 and that the relation is much weaker[4] during the first fifty years from 1926 to 1976.

This positive relation between age and return is consistent with the underperformance of IPO's documented by [92]. He finds that newly listed firms perform worse on average than a matched sample of older firms during the first five years after listing. Updated tables providing data on the long run performance of new issues are available from Jay Ritter's website[5]. During the period 1980-2016 IPO firms have underperformed matched (by size) firms by an average of 3.3% per annum during the first five years. [20] made this connection[6] between the poor returns on new listings and the composition of the equally weighted portfolios.

The relation between firm age and stock return is also consistent with the recent model of [78] who analyze the conditions under which firms adopt new technology. In their setup firms differ in their capacity to adopt new (and costly) technology which will make them more efficient. The authors define the concept of capital age as the length of time since the last adoption of a new technology. Capital age is used to measure different levels of technical efficiency. Young capital age firms are closer to the technological frontier than old capital age firms. Young capital age firms are predicted to be more productive and less risky than old capital age firms. Hence they earn lower expected returns than old capital age firms and this is confirmed empirically. Our measure of calendar age bears a similar relation to expected return.

To better understand the age effect and the connection between the age effect and the size effect we construct 16 portfolios that are doubly sorted into four age groups and four size groups and compare their performance. We focus on the 1977-2016 period and report comparable results for the 1926-1976 period in A.4.1. The age effect is clearly observed in all size groups, and the size effect is evident in all age groups. When we divide stocks into four age groups we find that returns are increasing with age over the first three groups but

[4][12] examined the relation between stock return and stock age over a period roughly corresponding the first half of our sample period. They report a mildly negative relation between stock age and return. There are some differences between our approaches. Their data period is 1930-1980 whereas ours is 1926-1976. They only consider stocks that have been listed on the exchange for 60 months and so omit many stocks that we include. In addition their method of computing the delisting returns differs from ours.

[5]https://site.warrington.ufl.edu/ritter/ipo-data/

[6][20] note on page 138 that "The main difference between the randomly selected portfolios and the EW CRSP index portfolio is that the securities included in the former are all listed at the beginning of the 10- or 20- year period, whereas the constituents of the CRSP portfolio are continuously updated to reflect new listings. Therefore the superior performance of the randomly selected portfolios is consistent with the abnormal returns to new listings that have been documented by [80]."

are flat or drop a little for the oldest group. That is the age effect is not monotone. It holds over the bulk of a firm's life but may be reversed in the oldest age group. Hence our age effect is not inconsistent with the finding that firms are less profitable at older ages (see for instance [79]).

This leads us to conclude that the age and size are not spanned by a common underlying factor. Moreover the age effect seems to be in conflict with the size effect, since stock age and size are positively correlated but explain the stock returns in the opposite direction. To further resolve this puzzle we divide stocks into decile groups based on their age or size and calculate the return statistics within each decile group. The results suggest that the observed small firm effect is a result of the extremely positive return skewness in the smallest 10% of the stocks (This has been noted in [14], and the general return skewness problem is discussed in [57] for instance). If the within-group median return is used as the performance measure, the direction of how the two factors affect stock returns turns out to be the same.

This chapter makes the following contributions to the literature. First, we acquire deeper understanding of why the rebalanced portfolio outperforms the bootstrapped portfolio so impressively over the period from 1977 to 2016. We show that this is caused by a combination of the older age profile of the rebalanced portfolio and the relation between stock returns and firm age. Second, we empirically document an age effect: an asset pricing anomaly that is entangled with but quite distinct from the size effect. Third, our results provide a possible opportunity for investment management. An institution could in principle structure a portfolio to exploit the age effect.

The remaining part of this chapter is organized as follows. Section 5.2 analyzes the performance of the equally weighted bootstrapped portfolio and the equally weighted rebalanced portfolio and highlights the performance gap. Section 5.3 relates the performance gap to the difference in age distribution between the two portfolios and discusses some aspects of the age effect. We provide a detailed analysis of the these phenomena for the period 1977-2016 and give a summary of the results for the first 50 years of data in Appendix A. Section 5.4 discusses economic explanations for the age effect. Section 5.5 concludes the chapter.

5.2 Bootstrapped versus Rebalanced Portfolios

In this section we compare the realized returns on our two basic portfolio strategies. These are the conventional $1/N$ equally weighted strategy[7] that has been studied by [30] and the equally weighted bootstrapped strategy. We use the same data as [14]. The data is available from the Center for Research in Securities Prices (CRSP) monthly stock return database. As in [14] only common stocks with share codes 10, 11, and 12 are included in the study. The entire period runs from June 1926 to December 2016 and includes 26,051 distinct CRSP permanent numbers (PERMNOs). The monthly returns are inclusive of reinvested dividends.

We construct the bootstrapped portfolio by picking N stocks at the start of each month and investing equal amounts in each stock. We hold this portfolio for one month before liquidating the portfolio and starting this process all over again for the next month. By compounding all the monthly[8] returns we obtain the holding period return of the bootstrapped portfolio. The rebalanced portfolio is constructed by selecting N random stocks at inception and investing equal amounts in each stock. Each month the weights are adjusted to obtain equal investments in each stock. If a stock in this portfolio is delisted in a particular month it is replaced by another stock selected at random from the available pool of actively traded CRSP stocks at that time.

5.2.1 Relative Performance

We compare the performance of the bootstrapped and rebalanced N-stock portfolios in Table 5.1 and find that on average the returns on the rebalanced portfolios exceed those on the bootstrapped portfolios. These results are based on simulations of 20,000 portfolios of each type for $N = 5, 25, 50,$ or 100. As noted previously in Section 5.1 these results may appear counterintuitive. They provide the motivation for investigation of the age effect in the next section.

[7] It is denoted in this chapter as the rebalanced strategy.

[8] It is worth pointing out that some stocks enter the bootstrapped portfolio in their last trading month and are delisted during the month. These stocks are associated with a code of delisting reason and a delisting return. The delisting return is calculated by comparing the security's Amount After Delisting with its price on the last day of trading. In such a case we adjust the stock return by incorporating the delisting return to reflect the actual return an investor would obtain when holding the stock till it is delisted. There are a few occasions where the delisting reason is specified but the delisting return is missing. In such occasions we follow the method proposed in [99] to fill the delisting return according to the delisting reason .

Table 5.1: Summary of annualized returns of 20,000 bootstrapped and rebalanced N-stock portfolios. In a bootstrapped portfolio the indicated numbers of stocks are selected at random for each month. In a rebalanced portfolio the indicated numbers of stocks are selected at random at the beginning of investment horizon, the same stocks are adjusted to have equal weights each month unless one or more stocks are picked at random to make up for the delisted one(s). Equally weighted portfolio returns are computed each month and are linked over the horizon from July 1926 to December 2016. Annualized return is recorded for each of 20,000 simulations of each portfolio type. Mean, median, and skewness of the 20,000 annualized returns are reported, as well as the percentage out of the 20,000 returns that are positive, greater than the return on Treasury bill, and greater than the return on an equally weighted portfolio of the whole market.

	Bootstrap portfolios			Rebalanced portfolios		
	Mean	Median	Skew	Mean	Median	Skew
			N = 5			
Holding return	0.0978	0.0976	0.1131	0.1101	0.1108	-0.1365
% > 0	100.00%			100.00%		
% > T-bill	99.99%			100.00%		
% > EW mkt	14.89%			23.32%		
			N = 25			
Holding return	0.1179	0.1178	0.0725	0.1239	0.1239	-0.0037
% > 0	100.00%			100.00%		
% > T-bill	100.00%			100.00%		
% > EW mkt	31.76%			53.05%		
			N = 50			
Holding return	0.1205	0.1204	0.0703	0.1258	0.1258	-0.0090
% > 0	100.00%			100.00%		
% > T-bill	100.00%			100.00%		
% > EW mkt	36.32%			67.08%		
			N = 100			
Holding return	0.1220	0.1220	0.0404	0.1269	0.1269	0.0174
% > 0	100.00%			100.00%		
% > T-bill	100.00%			100.00%		
% > EW mkt	40.87%			82.25%		

Comparing the mean annualized returns in the same rows, we notice that the rebalanced portfolios outperform the bootstrapped portfolios for all four values of N. For $N = 5$ the performance gap is 1.23% per annum. This pattern becomes even more obvious when we look at the percentage out of the 20,000 portfolios that outperform the equally weighted portfolio of the whole market. For the bootstrapped portfolios, as the portfolio size increases, this percentage increases toward 50%[9]. However the proportion of rebalanced portfolios that outperform the equally weighted market gradually increases to be over 80%. Note we have not yet taken transaction costs into account when calculating the returns. That is, since the rebalanced portfolios have much less turnover compared with the bootstrapped ones, the former will be more favourable if transaction costs were included.

While Table 5.1 demonstrates that the returns on the rebalanced portfolios are consistently higher than those on the bootstrapped portfolios, the differences for $N = 50$ and $N = 100$ do not seem large at around fifty basis points. However recall that these results are based on the entire 90 year period from 1926 to 2016 and that there were relatively few stocks at the start of this period. We obtain more interesting and more dramatic results when we divide the period up into smaller subperiods. We redo the same calculations as in Table 5.1 but based on shorter investment horizons. The first period is from July 1926 to December 1976 which leads to a holding period of about 50 years. The second period from January 1977 to December 2016 coincides with a typical time period that would be currently used for asset pricing empirical tests. The third period is from January 2007 to December 2016, leading to a 10-year holding period. In addition we set $N = 100$.

Table 5.2 reports the performance of the bootstrapped and rebalanced portfolios over these subperiods. There is a substantial difference in the relative performance of the two portfolios in the first fifty years and in the last forty years. During the earlier period the returns are very close with the bootstrapped portfolio being marginally better by 0.08% per annum. However during the most recent forty years the returns on the rebalanced portfolio are on average 2.32% per annum higher than those on the bootstrapped portfolio. For the most recent decade (2007-2016) the rebalanced portfolio return is 1.49% per annum higher than the return on the bootstrapped portfolio. We recall from Table 5.1 that over the entire 90 year period with $N = 100$ that the mean return on the rebalanced portfolio exceeds the mean return on the bootstrapped portfolio by 0.49% per annum. This suggests something quite different is happening in the last four decades as compared to the first five decades.

[9]Actually the limit of this percentage as N increases is not exactly 50%, because the return of the equally weighted market is only the expected value instead of the 50% quantile of the return distribution of the bootstrapped portfolios

Table 5.2: Summary of annualized returns of 20,000 bootstrapped and rebalanced 100-stock portfolios over three shorter holding periods: July 1926 - December 1976, January 1977 - December 2016, and January 2007 - December 2016. Construction of bootstrapped and rebalanced portfolios is the same as described in Table 5.1 except that the monthly returns of equally weighted portfolios are linked over indicated investment horizons and that the portfolio size is fixed at $N = 100$. Annualized return is recorded for each of 20,000 simulations of each portfolio type. Mean, median, and skewness of the 20,000 annualized returns are reported, as well as the percentage out of the 20,000 returns that are positive, greater than the return on Treasury bill, and greater than the return on an equally weighted portfolio of the whole market.

July 1926 - December 1976						
	Bootstrapped portfolios			Rebalanced portfolios		
	Mean	Median	Skew	Mean	Median	Skew
Holding return	0.1167	0.1167	-0.0058	0.1159	0.1159	-0.0050
% > 0	100.00%			100.00%		
% > T-bill	100.00%			100.00%		
% > EW mkt	45.01%			35.07%		
January 1977 - December 2016						
	Bootstrapped portfolios			Rebalanced portfolios		
	Mean	Median	Skew	Mean	Median	Skew
Holding return	0.1286	0.1284	0.0913	0.1518	0.1517	0.0605
% > 0	100.00%			100.00%		
% > T-bill	100.00%			100.00%		
% > EW mkt	41.36%			99.36%		
January 2007 - December 2016						
	Bootstrapped portfolios			Rebalanced portfolios		
	Mean	Median	Skew	Mean	Median	Skew
Holding return	0.0579	0.0575	0.1110	0.0728	0.0728	-0.0302
% > 0	99.91%			100.00%		
% > T-bill	99.69%			100.00%		
% > EW mkt	45.44%			78.28%		

5.2.2 How Bad is Being Delisted?

In Section 5.1 we mentioned that some observers tend to think that the rebalanced portfolio would perform poorly because it holds a stock until it disappears from the market. However it is sometimes overlooked that being delisted is not necessarily bad news. We refer readers to Table 2B in [14] for a detailed summary of lifetime buy-and-hold returns by final delisting status. The results suggest that the majority of stocks that are finally delisted due to Merger, Exchange, or Liquidation yield a lifetime buy-and-hold return exceeding that of the one-month Treasury bill. Even for the stocks that are delisted by the exchange, the mean lifetime buy-and-hold return is -0.8%, which is far from a devastating outcome. However it should be noted that this is thanks to the diversification effect - the median lifetime buy-and-hold return is much more negative. In addition more stocks were delisted due to Merger, Exchange, or Liquidation than any other reasons. These results together explain why delisting does not unduly penalize the returns on the rebalanced portfolios.

5.2.3 Comparison with Value Weighted Bootstrapped Portfolio

We can gain additional insight by comparing the returns on equally weighted strategies with the returns on value weighted strategies. Specifically we compare the performance of equally weighted bootstrapped portfolios with similar value weighted bootstrapped portfolios. [14] has already computed the returns on value weighted bootstrapped portfolios and we compared his results with our equally weighted bootstrapped portfolios. In our comparison we use the same set of stocks in each comparison pair so that the portfolios differ only by their respective weights. We find[10] that the returns on the equally weighted bootstrapped portfolios are on average 2.24% per annum higher than the returns on the value weighted bootstrapped portfolios. Since both portfolios have the same age distribution this performance cannot be explained by an age effect. It is due to the contrarian nature of the equally weighted portfolio and the small firm effect. As we will see in the next section the equally weighted rebalanced portfolio and the equally weighted bootstrapped portfolio have quite different age distributions and this can impact their relative performance.

5.3 Stock Age and Cross-sectional Returns

In this section we demonstrate that stock age is an important determinant of returns. In particular we show that portfolio age is a key difference between the bootstrapped

[10] The results are available on request.

portfolios and the rebalanced ones and that this difference leads to the performance gap between these two portfolio types. The numerical analysis presented in this section is based on the period 1977-2016. This is because the recent 40-year period is more relevant to the current financial market. For completeness we report the corresponding results for the period 1926 to 1976 in A.4.1. The age effect is observable but much weaker during this earlier period.

5.3.1 A Probabilistic View on Age Distribution

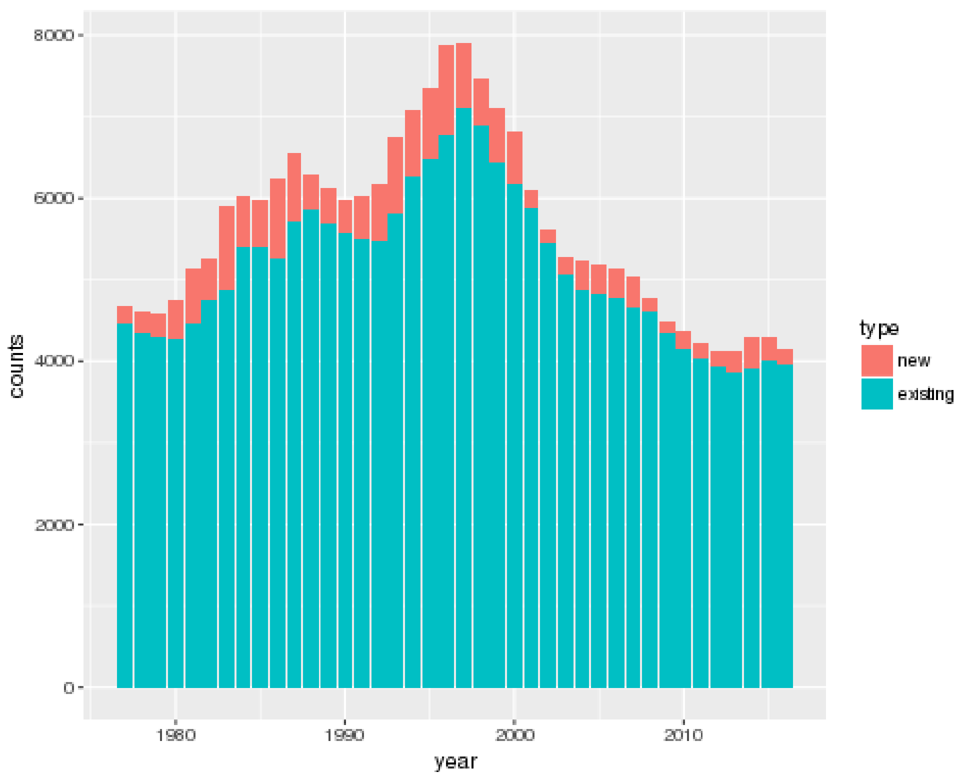
In this subsection we explain using a probabilistic argument why the rebalanced portfolio will have an older age distribution than the bootstrapped portfolio. Consider a rebalancing date when there are M stocks available in the stock universe. Then each of the M stocks has a probability of N/M of being included in the N-stock bootstrapped portfolio. If K stocks that were in the rebalanced portfolio in the previous period leave the portfolio because of delisting, then the $N - K$ stocks that already exist in the rebalanced portfolio will remain in the portfolio with a probability of one. Moreover each of the remaining $M - (N - K)$ stocks in the pool will be selected into the rebalanced portfolio with a probability of $K/(M - N + K)$ (¡N/M). From this perspective an important difference between the two portfolio types rests squarely on the rebalanced portfolio favouring *seasoned* stocks by assigning them a much higher probability of staying in the portfolio. In other words the component stocks in the rebalanced portfolio become mature in terms of age as time elapses. In contrast, the bootstrapped portfolio does not take into account the age of the stocks. This means that the average age of the rebalanced portfolio will increase over time, whereas the average age of the bootstrapped portfolio will be similar to that of the stock universe.

Figure 5.1 shows the profile of the stock population in the CRSP data from 1977 to 2016. The red portion in each bar represents the number of stocks that entered the universe in the current calendar year. The blue portion represents the number of stocks that have existed in the universe at the beginning of each calendar year. In demographic parlance the new listings correspond to births and the delistings correspond to deaths. The crude birth rate is the total number of births in a given year divided by the size of the population. The crude death rate is the total number of deaths in a given year divided by the size of the population. For this data the average crude birth rate for the period 1977-2016 is 10% while the crude death rate is 9%. These numbers are similar[11] to those obtained by [79]

[11] Based on the period 1978-2004, [79] obtained 10.3% for the crude birth rate and 9.9% for the crude death rate.

but higher[12] than those obtained by [31] who focus only on domestic US stocks. A delisting rate of 10% implies that in the rebalanced portfolio about 90% of the constituent stocks remain in place each year and as a result they age by one year. The other ten percent that are added to the portfolio will have an age distribution similar to that of the stock universe. The age distribution of the bootstrapped portfolio reflects the age distribution of the stock universe. Hence the average age of the stocks in the rebalanced portfolio increases[13].

Figure 5.1: Population of stocks in the CRSP database: existing stocks and new listings. The figure shows the change in the stock population in the CRSP database from 1977 to 2016. The red portion in each bar represents the number of stocks that entered the universe in the indicated calendar year. The blue portion represents the number of stocks that have existed in the universe at the beginning of the indicated calendar year.



[12][31] estimate an average crude birth rate of 7.5% and an average crude death rate of 8.2% based on the period 1975-2012. They just focus on US stocks whereas we follow [14] and retain securities with share codes 10, 11 and 12. Hence our rates are higher.

[13]We can show that if the stock universe is stationary over time, then the average age in the rebalanced portfolio keeps increasing until it reaches an asymptotic limit.

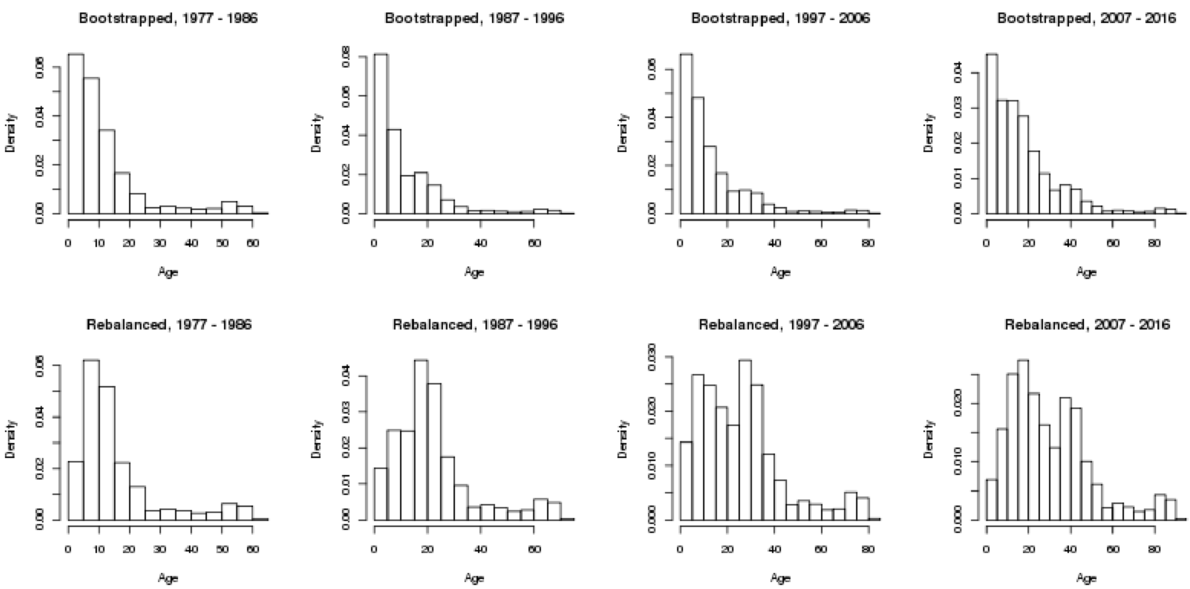
5.3.2 Age Distribution in Bootstrapped and Rebalanced Portfolios

Before presenting the age distributions for different portfolios, it is useful to clarify the calculation of age in our study. On any given month the age of a stock is the number of months that have elapsed since the first month the stock appeared in the CRSP database divided by twelve. Our empirical data confirm our predictions about the difference in age distribution between the bootstrapped and rebalanced portfolios. We consider the holding period from January 1977 to December 2016 and simulate 1000 bootstrapped portfolios as well as 1000 rebalanced portfolios, each containing 100 component stocks. In each month we record the age of all component stocks in each simulated portfolio. We report the empirical age distribution in both portfolios over each decade in the 40-year horizon according to the aggregate result across 1000 simulations. The aim of this decade-by-decade breakdown is to highlight the evolution of the age distribution in the two portfolios as time passes. Figure 5.2 presents the age distribution of the two representative portfolios over time. Furthermore, based on the 1000 simulations, we report the average age of component stocks in each portfolio type over the four non-overlapping decades. The result is shown in Table 5.3.

Some institutional background is helpful in interpreting these graphs. The average age of the entire stock universe is increasing over these four decades for two main reasons. The first is due to the aging of the large influx of Nasdaq stocks that entered the database as a group in 1972. Their entry is shown clearly in Figure A.1 of Appendix A. The second is due to the decline[14] in new listings since the mid 1990's as evidenced by Figure 5.1. From Figure 5.2 we see that the histograms of the age distribution of both the bootstrapped portfolio and the rebalanced portfolio move to the right over the four decades. However if we compare the histograms of the two portfolios across each decade we see that the age histogram of the rebalanced portfolio is consistently further to the right of the corresponding histogram of the bootstrapped portfolio. Table 5.3 confirms this observation. The age difference is 4.3 years for the period 1977-1986 and averages 12 years over the next three decades. The fact that the stock universe is replenished each year with newly listed stocks and the way in which the rebalanced portfolio is constructed leads to the difference in the age distributions between the two portfolio types.

[14] See also [31].

Figure 5.2: Age distribution in 100-stock bootstrapped and rebalanced portfolios over different time periods. The distributions are based on age of components of one thousand 100-stock bootstrapped and rebalanced portfolios held over the period from January 1977 to December 2016. We report the empirical age distribution in both portfolios over each decade in the 40-year horizon.



5.3.3 Age Effect

We have identified the age distribution as a key difference between the bootstrapped and rebalanced portfolios. The next step is to investigate whether age can explain the cross-section of stock returns. If age is a significant predictor of stock returns, then the performance gap between the rebalanced and bootstrapped portfolios can be explained by the difference in the age distributions.

To study the age effect we include each stock that has ever appeared in the stock universe from January 1977 to December 2016 in our analysis. Each month we record the age and monthly return (annualized by multiplying by 12) of each stock in the entire universe. Let R_{it} denote the annualized return[15] of the ith stock in the tth month and A_{it} denote the age of the ith stock in the tth month. Now that the data has both time-series and cross-section dimensions, panel data modeling techniques are used to estimate the

[15]The purpose of annualizing the monthly returns is to bring the estimate of the model coefficient to a more visible scale.

Table 5.3: Average age of 100-stock bootstrapped and rebalanced portfolios over different time periods. The average ages are calculated based on age of components of one thousand 100-stock bootstrapped and rebalanced portfolios held over the period from January 1977 to December 2016. We report the average age in both portfolios over each decade in the 40-year horizon.

Portfolio Type	1977 - 1986	1987 - 1996	1997 - 2006	2007 - 2016
Bootstrapped	11.60	11.24	12.90	17.35
Rebalanced	15.90	22.11	25.61	29.89

parameter of interest, which is the effect of age on asset returns. Equation (5.1) is a simple (time effects) model relating the stock's return to its age[16].

$$R_{it} = \beta A_{it} + \gamma_t + \epsilon_{it} \tag{5.1}$$

The parameter β quantifies the change in the cross-sectional stock return when the stock age increases by one unit. The parameter γ_t characterizes the level of average cross-sectional returns in the tth month. The time effect term is included because the market movement from month to month could make it problematic to pool samples across time and thus affect the estimation of the age effect. In addition it can be argued that the age effect in the panel-data regression in equation (5.1) is potentially confounded by vintage years. A vintage year is the year in which a company receives its first influx of investment capital. It is the year when capital is contributed by a venture capital, a private equity fund or a partnership drawing down from its investors. A vintage year at the peak or bottom of a business cycle can potentially affect subsequent returns on initial investment as the company undergoes over or under-valuation at the time. The introduction of the time effect in the panel-data regression in equation (5.1) can be viewed as a rough measure to control for this potentially omitted confounding influence. Lastly, ϵ_{it} is the error term.

Subtracting the cross-sectional average $\bar{R}_t = \frac{1}{N_t}\sum_i R_{it}$, where N_t is the number of existing stocks in the tth month, from the initial model equation (5.1) becomes

$$\ddot{R}_{it} = \beta \ddot{A}_{it} + \ddot{\epsilon}_{it}, \tag{5.2}$$

where $\ddot{R}_{it} = R_{it} - \bar{R}_t$, $\ddot{A}_{it} = A_{it} - \bar{A}_t$ and $\ddot{\epsilon}_{it} = \epsilon_{it} - \bar{\epsilon}_t$. Note that $\bar{A}_t$ and $\bar{\epsilon}_t$ are defined in the same way as $\bar{R}_t$. The first row in Table 5.4 summarizes the estimation and hypothesis testing results of the time effects model. The sign of the $\hat{\beta}$ and the highly significant p-value for the t-test confirm age has a significantly positive impact on the cross-sectional return of a stock. In other words, stocks that are older tend to outperform younger stocks.

[16]The estimate of β will be biased if size and possibly other firm characteristics affect asset returns and are correlated with age.

Table 5.4: Empirical results for time effects model: January 1977 - December 2016. In the first row, the estimate, standard error, t-statistics, and the associated p-value for the overall time effects model are reported. In each month, the age group each stock belongs to is determined based on the cross-sectional ranking of the stock's current age. The breakpoints between age groups are the first quartile, median, and third quartile of the cross-sectional age distribution. All stock-month observations are divided into four age groups in this way. In each of the second to fifth rows, model fitting results for the indicated age group are reported.

Age Group	Estimate	Std. Error	t-value	p-value
All ages	0.0007	0.0001	7.7808	0.0000
Infant	0.0061	0.0026	2.3621	0.0182
Youth	0.0068	0.0019	3.6444	0.0003
Adult	0.0008	0.0010	0.7864	0.4316
Senior	-0.0006	0.0001	-4.8058	0.0000

We fit the same regression model with sub-groups of the data to provide additional robustness to our result. At the beginning of each month in the investment horizon, each stock in the universe is labeled with one of the four age groups, Infant, Youth, Adult, and Senior, according to their current age. The breakpoints between adjacent age groups are the first quartile, median, and third quartile of the cross-sectional age distribution[17]. In this way we add an additional categorical feature to each stock-month observation. Then we divide the data into four sub-groups according to the age group label and fit the model in equation (5.1) using each of the four subsets. The estimation and hypothesis testing results are also presented in Table 5.4. Within each of the youngest two age groups there is a significant and positive relationship between stock age and stock return. The age-return relation in the second oldest age group is insignificant. A significant and negative age effect is observed in the oldest age group which represents a downturn in performance when a stock gets really old. However the magnitude of the coefficient in the two younger age groups is nine times larger than that in the two older age groups. The sub-group analysis allows us to acquire a deeper understanding of the nature of the age effect at different stages of a firm's life cycle.

We claim that the significant relationship between a stock's age and its return, together with the difference in age distribution between the bootstrapped and rebalanced portfolios, explains the performance gap between these two types of portfolios. We show in Appendix B that the returns on the rebalanced and bootstrapped portfolios have the same exposure to mean reversion so that any performance difference cannot be attributed to mean reversion. This does not contradict the results of [91] which show that the performance of

[17]This grouping method leads to a dynamic group membership. Size of different groups may be different because there may be multiple stocks at the breakpoint ages.

the rebalanced $1/N$ portfolio does benefit from mean reversion. The performance of the bootstrapped $1/N$ portfolio benefits from mean reversion to the same extent. When we compare returns on the two portfolios, the impact of mean reversion cancels out so that the difference in returns on the two portfolios cannot be accounted for by mean reversion.

5.3.4 Age Effect vs. Size Effect

The small firm effect is a well-known pricing anomaly in finance which holds that smaller firms, or those companies with a small market capitalization (as a product of price and number of outstanding shares), outperform larger companies. This effect has been documented by many researchers (see [106] for a review) but the consensus is (cf [3]) that it has become less important in more recent years. In results not reported here we confirmed its presence in our data as well.

In the previous subsection we showed that senior firms generally outperform junior firms. The "senior firm effect" and the "small firm effect" seem to be complementary (and not substituting) effects, because a stock's age and its market capitalization are positively correlated measures of scale of the issuing company. However these two measures explain the cross-sectional stock returns in opposite directions since age and size are positively correlated. We confirm the presence of this correlation in our data. We record the age and market capitalization of all available stocks in each current month as well as the ranking (according to age and size respectively) of each stock within the current stock universe. Pooling the records across months we obtain vectors of stock age, size, rank by age, as well as rank by size. The correlation coefficient between the raw values of age and size is 0.23, and the correlation coefficient between the rank of age and rank of size is 0.29.

The finding that two positively correlated stock characteristics explain the cross-sectional stock returns in opposite directions is puzzling at first. To give a more detailed picture of how the age and size factors affect stock returns, we construct quartile portfolios which are doubly sorted according to both the age and size factors. At the beginning of each month all stocks in the universe are divided into four roughly equal-size age groups, i.e., Infant, Youth, Adult, and Senior, according to their current age. The breakpoints between adjacent groups are the first quartile, median, and third quartile of the stock age distribution in the particular month[18]. Within each of the four age groups, the stocks are further divided into four size groups, i.e., Tiny, Small, Medium, and Big, according to their current market capitalization. The doubly sorting procedure yields 16 roughly equal-size

[18] This grouping method leads to a dynamic group membership. Size of different groups may be different because there may be multiple stocks at the breakpoint ages.

stock groups. For each of the 16 groups we construct an equally weighted portfolio. At the beginning of each month all these doubly sorted factor portfolios are liquidated and reconstructed to reflect the change in group members. The portfolio construction date in our study is the beginning of January 1977. All of the portfolios formed on age and size are rebuilt each month until the end of December 2016. Table 5.5 summarizes three performance measures of these 16 portfolios, namely the annualized return, the standard deviation, and the Sharpe ratio. The riskless rate used in the calculation of Sharpe ratios is downloaded from the Kenneth French website[19]. A comparison among these doubly sorted quartile portfolios reveals how each factor affects cross-sectional stock returns.

Table 5.5: Performance of sixteen doubly sorted equally weighted portfolios formed on age and size. Starting from January 1977, at the beginning of each month each available stock is assigned to one of sixteen factor portfolios based on its cross-sectional ranking of age and size. The breakpoints between adjacent age/size groups are the first quartile, median, and third quartile of the age/size distribution. the returns on equally weighted portfolio of all stocks in each factor portfolio are calculated. All factor portfolios are held until December 2016. The average annualized return, standard deviation, and Sharpe ratio of all sixteen portfolios are reported.

Age Group	Size Group	Return	Std Dev	Sharpe
Infant	Tiny	13.61%	26.75%	0.34
	Small	8.89%	23.52%	0.18
	Medium	10.66%	24.80%	0.25
	Big	12.66%	23.97%	0.34
Youth	Tiny	17.58%	24.08%	0.54
	Small	11.05%	21.21%	0.31
	Medium	12.24%	21.21%	0.36
	Big	12.17%	20.22%	0.38
Adult	Tiny	21.20%	23.32%	0.71
	Small	16.47%	21.32%	0.56
	Medium	15.81%	20.27%	0.55
	Big	14.23%	17.97%	0.54
Senior	Tiny	20.43%	22.92%	0.69
	Small	14.62%	19.62%	0.51
	Medium	16.12%	18.80%	0.61
	Big	13.79%	15.38%	0.60

Table 5.5 displays two main features in the returns. It confirms the existence of both an age effect and a size effect. We focus initially on the age effect since the size effect is already well documented in the literature. The age effect is quite pronounced but it is not uniformly monotonic across all age groups. The average returns generally increase as the age group moves through the first three age groups. There is a slight decrease in returns as we move from the third age group to the oldest age group for three of the four size

[19] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

groups. An exception occurs for the third largest size group which we have denoted as the Medium sized group. However we notice that with the size group fixed, the returns of the two oldest age groups (Adult and Senior) are very close to each other, and each is much higher than the returns of the two youngest age groups. The closeness in performance between the oldest two age groups explains why the rebalanced portfolios outperform the bootstrapped ones notwithstanding the apparent non-monotonicity of the age effect. It is also worth pointing out that the size factor is not monotone either. Within all of the age groups, the size group Tiny always outperforms the size group Big, yet the performance does not deteriorate monotonically with size. The important implication of our findings is that stocks that are both mature and small tend to outperform the market. These two features, although seemingly having opposite effects, when appearing together can lead to profitable returns.

Table 5.6: Returns of sixteen doubly sorted equally weighted portfolios: by decade. Starting from January 1977, at the beginning of each month each available stock is assigned to one of sixteen factor portfolios based on its cross-sectional ranking of age and size. The breakpoints between adjacent age/size groups are the first quartile, median, and third quartile of the age/size distribution. The average return on an equally weighted portfolio of all stocks in each factor portfolio is calculated. All factor portfolios are held until December 2016. Annualized returns over the four non-overlapping decades of each factor portfolio are reported.

	January 1977 - December 1986					January 1987 - December 1996			
	Tiny	Small	Medium	Big		Tiny	Small	Medium	Big
Infant	23.40%	17.71%	16.86%	17.64%	Infant	11.68%	4.22%	10.59%	16.50%
Youth	22.38%	13.66%	14.52%	14.35%	Youth	21.38%	7.79%	10.96%	13.54%
Adult	26.00%	25.84%	22.95%	30.12%	Adult	26.17%	11.04%	12.29%	15.64%
Senior	30.12%	21.80%	22.31%	17.23%	Senior	20.72%	10.35%	13.50%	14.94%
	January 1997 - December 2006					January 2007 - December 2016			
	Tiny	Small	Medium	Big		Tiny	Small	Medium	Big
Infant	18.70%	10.13%	8.70%	8.90%	Infant	0.65%	3.48%	6.49%	7.61%
Youth	21.88%	15.75%	14.89%	11.47%	Youth	4.68%	7.00%	8.58%	9.30%
Adult	23.56%	18.92%	16.35%	12.52%	Adult	9.08%	10.09%	11.65%	9.11%
Senior	20.41%	15.14%	17.01%	13.32%	Senior	10.45%	11.16%	11.65%	9.67%

The Sharpe ratio results in Table 5.5 provide an even more striking demonstration of the age effect. For each size group the Sharpe ratio of the oldest age group is typically double the Sharpe ratio of the youngest age group. For the first three size groups the Sharpe ratio of the most senior age group is at least twice the Sharpe ratio of the youngest age group. For the largest size group the Sharpe ratio of the oldest age group is 1.8 times the Sharpe ratio of the youngest age group. On the other hand if we hold age fixed the Sharpe ratios of the different size portfolios are much less disperse.

Since the age effect is the key finding in this chapter, we explore its robustness across different periods. Table 5.6 contains a more detailed decade-by-decade breakdown of the doubly sorted portfolios. This breakdown shows clearly that the age effect is both strong and persistent across all four decades. For all decades the returns are generally increasing in age. The age group Youth has in general higher returns than age group Infant with the average difference being 1.80% over all size groups and decades. In turn age group Adult has in general higher returns than age group Youth with the average difference being 4.33% over all size groups and decades. Age groups Adult and Senior represent the two oldest groups. The difference between the two oldest age groups is somewhat lower and negative. Over all the 16 combinations, the average return for age group Senior is lower than the average return for age group Adult. The average difference is 1.35% per annum which is small relative to the other differences. These results are consistent with our earlier regression results in Table 5.4.

We obtain a more compelling demonstration of the impact of age when we combine the two youngest age groups by taking their average and the two oldest age groups in the same way. The group containing the two youngest age groups is labelled Junior and the group containing the two oldest age groups is labelled Senior. The left hand side of Table 5.7 compares the returns on these age sorted portfolios over four size groups for each of the four decades. Differences between the return of age group Senior and return of age group Junior are reported as SMJ which is a short notation for "Senior minus Junior". The most striking result from Table 5.7 is that the return on the Senior portfolios exceeds the return on the Junior portfolios for each size group within each decade. Furthermore the average difference in these portfolio returns over all sixteen combinations is 4.55% which is very significant. According to the same set of results presented in A.4.1 for the period before 1977, the SMJ is also positive for each size group within each decade, but the average return difference is only 1.79%.

It is instructive to conduct a similar grouping based on size to compare the relative importance of the age effect and the size effect. We group the two smallest size groups together (Tiny+Small) and the two largest size groups together (Medium+Big) and calculate the returns on these size sorted portfolios over all four age groups for each of the four decades. Differences between return of these two coarser size groups are reported as SMB which is a short notation for "Small minus Big" in the right half of Table 5.7. It should be clarified that SMB represents the return difference between the smaller half (Tiny+Small) and the bigger half (Medium+Big) rather than that between the size groups Small and Big. It turns out that the average SMB return over the sixteen age-decade combinations is 1.69% which is much lower then the average SMJ return. In fact the magnitude of the age effect in this framework based on these calculations is 2.7 times as large as the size

Table 5.7: Analysis of portfolios formed on age and size: for each decade. The sixteen factor portfolios are re-organized into eight by combining the youngest (smallest) two age (size) groups and the oldest (biggest) two age (size) groups with equal weights for each of the four size (age) groups. Annualized returns of the merged portfolios over each decade between January 1977 and December 2016 are reported. Confounding the size group, the average return differences between the two combined age groups are reported as SMJ. Confounding the age group, the return differences between the two merged size groups are reported as SMB.

January 1977 - December 1986											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Infant	Youth	Adult	Senior	Average
Infant +Youth	22.89%	15.69%	15.69%	15.99%		Tiny+Small	20.56%	18.02%	25.92%	25.96%	
Adult+Senior	28.06%	23.82%	22.63%	23.68%		Medium+Big	17.25%	14.44%	26.54%	19.77%	
SMJ	5.17%	8.13%	6.94%	7.68%	6.98%	SMB	3.31%	3.58%	-0.62%	6.19%	3.12%
January 1987 - December 1996											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Infant	Youth	Adult	Senior	Average
Infant +Youth	16.53%	6.01%	10.78%	15.02%		Tiny+Small	7.95%	14.59%	18.61%	15.54%	
Adult+Senior	23.45%	10.70%	12.90%	15.29%		Medium+Big	13.55%	12.25%	13.97%	14.22%	
SMJ	6.92%	4.69%	2.12%	0.27%	3.50%	SMB	-5.59%	2.34%	4.64%	1.32%	0.67%
January 1997 - December 2006											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Infant	Youth	Adult	Senior	Average
Infant +Youth	20.29%	12.94%	11.80%	10.18%		Tiny+Small	14.42%	18.82%	21.24%	17.78%	
Adult+Senior	21.99%	17.03%	16.68%	12.92%		Medium+Big	8.80%	13.18%	14.44%	15.16%	
SMJ	1.69%	4.09%	4.88%	2.74%	3.35%	SMB	5.62%	5.63%	6.80%	2.62%	5.17%
January 2007 - December 2016											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Infant	Youth	Adult	Senior	Average
infant +Youth	2.67%	5.24%	7.54%	8.46%		Tiny+Small	2.07%	5.84%	9.59%	10.81%	
Adult+Senior	9.76%	10.63%	11.65%	9.39%		Medium+Big	7.05%	8.94%	10.38%	10.66%	
SMJ	7.10%	5.39%	4.11%	0.94%	4.38%	SMB	-4.99%	-3.10%	-0.79%	0.14%	-2.19%
Average					4.55%						1.69%

effect. Over the early years before 1977 the average SMB return is 5.10% (see Table 7(b) in A.4.1). Furthermore, over the most recent decade, the average SMB across the four age groups is negative. The disappearance of the size effect over more recent periods has been documented in the literature (see [58] and [3]) and our findings are consistent with this evidence.

5.3.5 Effect of Aging on Size Distribution

In this subsection we examine the relation between age and size. We use longitudinal data techniques to study this problem. Specifically we identify cohorts of stocks and follow their evolution over time. The use of longitudinal data has the advantage of reducing heterogeneity from two sources typically associated with the use of cross-sectional data, i.e., they include stocks that were created at different times and subject to different selection processes. We track three cohorts of stocks over a seven-year period in order to study the effect of aging on stock size distribution over time. The first cohort of stocks were issued in 1984 and were followed until 1991; the second cohort issued in 1994 and tracked until 2001; and the last one issued in 2004 and tracked until 2011. The effect of aging is evaluated by comparing the market capitalization distributions of the set of stocks in each issuance year and the corresponding end-of-tracking year. In the first cohort, for example, from the 632 stocks identified as new in 1984, only 240 were still active in 1991. This leads to three different distributions of interest. The first one is the distribution of all entrants in 1984; the second one, the distribution of survivors in 1991; and the third one, the size distribution in 1984 of those stocks that survived until 1991.

Table 5.8 reports the median, mean, standard deviation, and quartile coefficient of dispersion of nine distributions of interest (each of the three cohorts is associated with three distributions) mentioned in the previous paragraph. The quartile coefficient of dispersion is a robust measure of dispersion and is defined as $(Q_3 - Q_1)/(Q_3 + Q_1)$, where Q_1 and Q_3 are the first and the third quartiles of each dataset. Figure 5.3 presents the kernel density functions of the nine sets of log-transformed market capitalization, each panel corresponding to one of the three cohorts. Comparing the solid curves (All, 1984/1994/2004) with the dashed curves (Survivors, 1984/1994/2004), we find that stocks that survived through the seven-year tracking period tend to have larger market capitalization going back in the year of issuance. Comparing the dashed curves with the dotted curves (Survivors, 1991/2001/2011), we observe that the market capitalization distribution moves to the right and becomes more dispersed as stocks become mature. The increasing dispersion in market capitalization distribution is also evident according to the last column in Table 5.8. Note that the graphs in Figure 5.3 use the log of size and the actual dispersion in dollar terms is considerably greater. Therefore we conclude that the effect of aging on market capitalization distribution is two-fold. The market capitalization on average becomes larger as time passes. However the market capitalization distribution also becomes significantly more disperse over time as well. Therefore it is possible for a stock to be both senior in terms of age and small in terms of market capitalization.

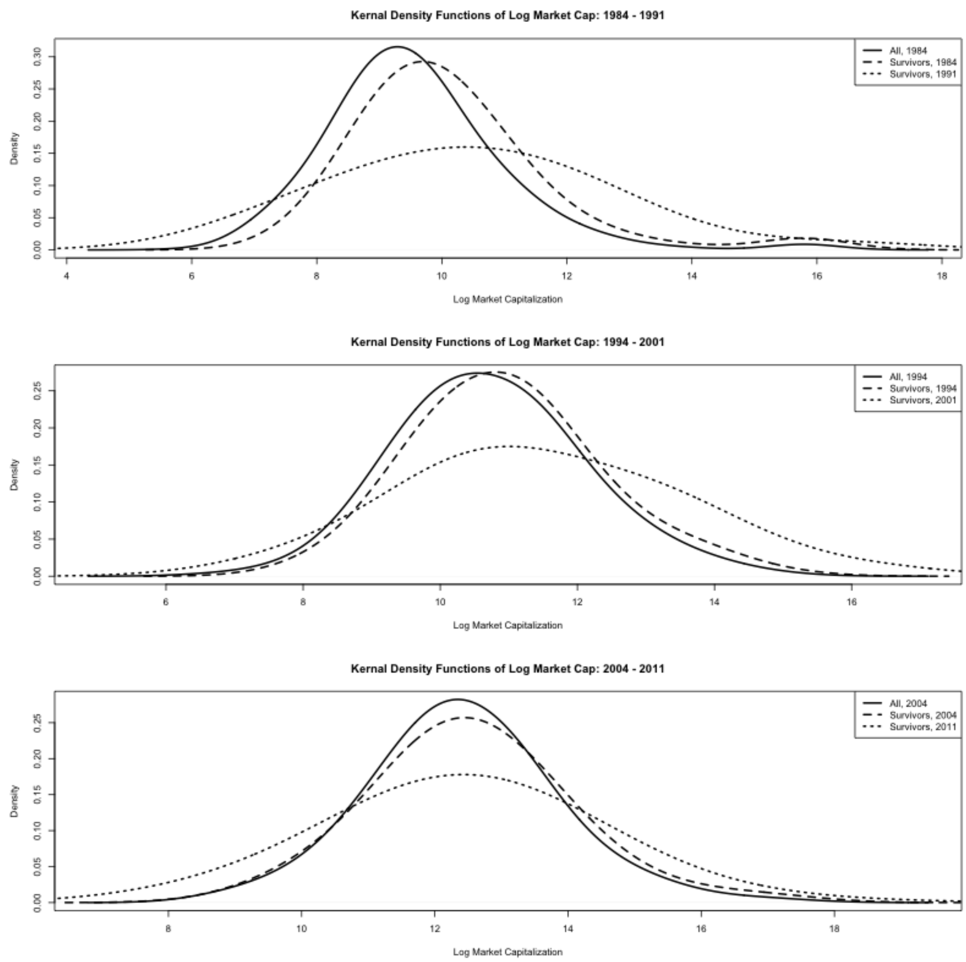
Table 5.8: Summary of three market capitalization distributions for three cohorts of stocks. Market capitalization of three cohorts of stocks, namely those entered the CRSP database in 1984, 1994, and 2004, are tracked over a seven-year period after their entrance. Each cohort is associated with three market capitalization distributions of interest. The first one is the distribution among all entrants at the beginning of the tracking period; the second one is the distribution among survivors at the end of the tracking period; the third one is the distribution at the beginning of the tracking period of those stocks that survived until the end of the tracking period. We report the mean, median, standard deviation, and quartile coefficient of dispersion of each distribution for each cohort.

	Median	Mean	Std Dev.	Dispersion
New entrants in 1984, in 1984	12,159	125,383	834,804	0.63
Survivors through 1991, in 1984	18,755	291,655	1,338,064	0.66
Survivors through 1991, in 1991	31,149	717,302	3,200,902	0.89
New entrants in 1994, in 1994	42,513	124,181	292,084	0.71
Survivors through 2001, in 1994	55,147	163,698	385,767	0.69
Survivors through 2001, in 2001	78,300	726,510	2,529,198	0.88
New entrants in 2004, in 2004	232,365	799,082	2,648,895	0.67
Survivors through 2011, in 2004	257,991	1,104,531	3,678,639	0.70
Survivors through 2011, in 2011	231,092	2,263,304	13,310,190	0.86

5.3.6 Return Skewness within Age and Size Decile Groups

In this subsection we examine the skewness of the stock return distribution for different age and size deciles and discuss the implications for our results. [14] studied the distributions of monthly buy-and-hold stock returns in different size decile groups. According to his Table 3A the *median* return in each size group increases (non-strictly) monotonically as we move from small to big groups. However the *mean* return in each size group does not show a clear pattern except that the smallest group yields a mean return much higher than any other size group. This observation implies that the observed "small firm effect" is to a large extent a result of the extreme positive skewness in the smallest 10% of firms. Another important implication of this finding is that heterogeneity in the smallest 10% of firms in terms of return is unmatchable by that in any other size group. Since stock age is a key feature in our study, it is also of interest to explore the pattern in the within-group skewness when stocks are grouped by age.

Figure 5.3: Kernal Density Functions of Log Market Capitalization. Market capitalization of three cohorts of stocks, namely those entered the CRSP database in 1984, 1994, and 2004, are tracked over a seven-year period after their entrance. Each cohort is associated with three market capitalization distributions of interest. The first one is the distribution among all entrants at the beginning of the tracking period; the second one is the distribution among survivors at the end of the tracking period; the third one is the distribution at the beginning of the tracking period of those stocks that survived until the end of the tracking period. Each panel in the figure corresponds to an indicated cohort, and the kernel density functions of the three distributions of log market capitalization are shown.



The leftmost columns of Table 5.9 report the mean, median, and skewness of monthly buy-and-hold stock returns grouped by size and the rightmost columns report the the set of statistics when the stocks are grouped by size. For each month during the period from January 1977 to December 2016 each available stock is assigned to a size (age) decile group based on its market capitalization (age since issuance) at the end of the last month. A group number closer to 10 means a more senior group or a larger-cap group. In this way each stock-month combination is tagged with an age group number and a size group number. Each decile group contains roughly 10% of the stock-month observations. Each stock-month observation is associated with a buy-and-hold return[20] of the particular stock over the particular month. The reported statistics are calculated based on all annualized monthly returns that belong to each decile group. The first four columns in Table 5.9 report similar information to that reported in Table 3A - Panel A of [14]. The minor discrepancy between the two sets of results appears to have originated from the difference in the data used[21].

Despite the difference in source data the pattern in return skewness across different decile groups shown in Table 5.9 is similar to that reported in Bessembinder's study: extremely positive skewness is observed in the smallest decile group, and there is a decreasing trend in the within-group skewness as we move from small to big size groups. The last four columns in Table 5.9 show the return statistics in different age decile groups. Compared with what we can observe from the within-group mean returns in different size groups, here we can see a more distinctly increasing trend in the within-group mean return as we move from young to senior age groups. However it is worth noting that the mean returns in the youngest and in the second youngest decile groups are reversely ordered compared with the general trend; this is also true in the comparison between the oldest and second oldest decile groups. A potential explanation for both observations can be found from the industrial organization literature. [41] explain why firms face an initial "honeymoon" period in which they are buffered from a sudden exit by their initial stock of resources. [11] argue that old firms are prone to suffer from a "liability of obsolescence" and also a "liability of senescence". This effect will be discussed in more detail in Section 5.4.

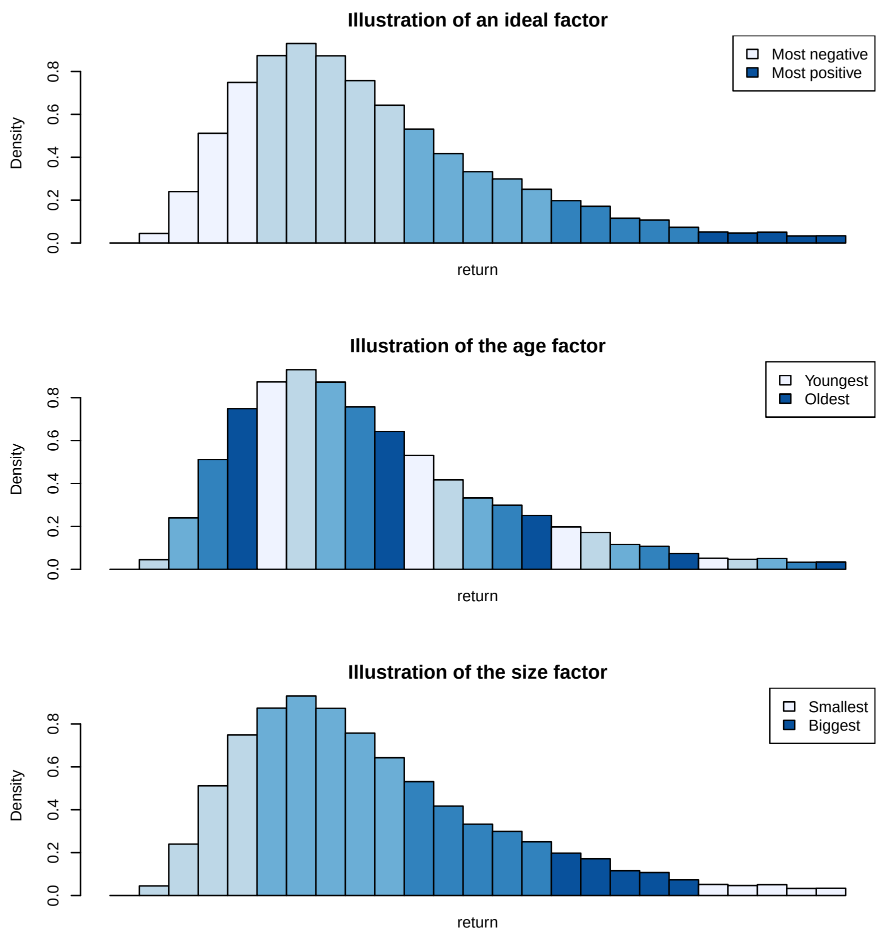
There is a marked difference in the skewness patterns in the size decile and in the age decile. For the size decile the return skewness decreases as we move in the direction of increasing firm size with the smallest decile having a skewness of 6.08 and the largest decile having a skewness of just 0.41. For the age decile the return skewness is remarkably stable with an average value of 5.07 across all age groups.

If we assign stocks to decile groups based on an ideal (hypothetical) factor, then the conditional return distribution in different groups should be clearly distinct. The overall

[20] Any delisting return is included into the calculation.

[21] It is important to point out that [14] uses the period 1926-2016, while our calculations are based on the period 1977-2016.

Figure 5.4: Illustration of portfolios sorted based on different factors. All three panels show a same histogram of hypothetical stock returns. We intend to illustrate the effect of factor-based grouping on within-group mean and skewness. In each panel, different bins are colored differently to reflect the level of factor. A darker blue color represents a more favorable factor group in the top panel, a more senior age group in the middle panel, and a bigger size group in the bottom panel.



return distribution is a mixture of ten distinct conditional distributions. In such a scenario we should expect a monotonic trend in the within-group mean and reduced (compared with skewness of unconditional return distribution) within-group skewness. Figure 5.4 shows a hypothetical unconditional return distribution and illustrates the effect of grouping according to different factors on the within-group (conditional) mean and skewness. The effect of grouping according to an ideal factor on within-group mean and skewness is illustrated in the top panel of Figure 5.4. Here a darker color stands for a more favorable factor group. According to Table 5.9 neither factor appears to be that ideal. For the age factor monotonicity is roughly observed but the conditional distributions are quite skewed. The middle panel of Figure 5.4 illustrates such a possibility. Here a darker color represents a more senior age group. For the size factor the monotonicity is violated. However most within-group skewness is reduced compared with skewness of the unconditional distribution of 5.13. This effect is illustrated in the bottom panel of Figure 5.4. Our key results depend more on the monotonicity feature and are not affected much by the skewness feature since skewness can be diversified away in portfolios.

Table 5.9: Statistics of one-month buy-and-hold returns in different size and age decile groups. For each month during the period from January 1977 to December 2016, each available stock is assigned to a size (age) decile group based on its market capitalization (age since issuance) at the end of the last month. A group number closer to 10 means a more senior group or a larger-cap group. Each stock-month observation is associated with a buy-and-hold return of the particular stock over the particular month. The reported statistics are calculated based on all annualized monthly returns that belong to each decile group.

Size Group	Mean	Median	Skew	Age Group	Mean	Median	Skew
1	0.2883	0.0000	6.0821	1	0.0988	0.0000	2.8856
2	0.1008	0.0000	3.3851	2	0.0825	0.0000	5.5759
3	0.0981	0.0000	3.1559	3	0.1131	0.0000	5.8414
4	0.1109	0.0000	3.5532	4	0.1543	0.0000	4.4354
5	0.1208	0.0000	2.4979	5	0.1254	0.0000	5.2705
6	0.1259	0.0000	1.8118	6	0.1766	0.0000	5.8495
7	0.1313	0.0593	1.3073	7	0.1589	0.0000	4.2460
8	0.1384	0.0945	1.2735	8	0.1624	0.0027	5.0487
9	0.1344	0.1126	0.7520	9	0.1575	0.0550	4.5711
10	0.1249	0.1173	0.4136	10	0.1400	0.0938	6.9448

5.4 Understanding Stock Age Effects

Many economic theories tend to treat firm size and firm age as capturing the same fundamental information. For example [49] presents his "stages of growth" model of organizational change in growing firms, in which size is linearly related to age. Other scholars have nonetheless made specific predictions about how firm performance changes with age. In this section we review these theoretical predictions in terms of three categories: selection effects, learning-by-doing effects, and inertia effects.

It is worth pointing out that the stock age used in our study is calculated based on the date on which a stock appears in the CRSP database for the first time. The three effects to be reviewed are from the industrial organization literature and characterize firms' earning ability in different stages. There is a gap between this stock age and the actual firm age. However we find the three effects capable of explaining the stock age effects we have discovered and therefore briefly discuss them since they may help explain some of the key findings in this chapter.

5.4.1 Selection Effects

Selection effects arise when selection pressures progressively eliminate the weakest firms, and result in an increase in the average productivity level of surviving firms, even if the productivity levels of individual firms do not change with age. This situation corresponds to an influential model in Industrial Organization in Economics proposed by [64]. According to this model, firms are born with fixed productivity levels, and learn about their productivity levels as time passes. In this model, low productivity firms are observed to exit, while high productivity firms remain in business. As a result the average productivity of the cohort increases with the cohort ages, even if the productivity levels of individual firms remain constant over time. Therefore selection effects provide a potential explanation for the age effect uncovered in our study.

5.4.2 Learning-by-doing Effects

Learning-by-doing effects occur when firms increase their productivity as they learn about more productive production techniques and incorporate these improvements in their production routines. (See [105] for a survey of the learning-by-doing concept.) Learning-by-doing effects can be expected to be particularly relevant for young firms. According to [46]: "New firms are hampered by their need to make search processes a prelude to every new problem they encounter. As learning occurs benefits can be obtained from the introduction of a repertoire of problem-solving procedures ... eliminating open search from the problem-solving response greatly reduces the labor and time required to address recurrent problems."

Furthermore older firms may benefit from their greater business experience, established contacts with customers, and easier access to resources. For example, [100] point out that entrepreneurs often lack detailed information about their jobs, firms and even the environments until they are active in the market. After a firm's creation, an intense learning process starts and contributes to the firm's growth and survival in the long-term. Also [24] provide evidence on the existence of microeconomic "learning-by-doing" effects with positive effects on the aggregate output. The learning-by-doing effects may provide an explanation to the positive association between stock age and expected returns reported in this chapter.

5.4.3 Inertia Effects

As firms get older they might become less productive if they become increasingly inert and inflexible. [11] argue that old firms are prone to suffer from a "liability of obsolescence" (because they do not fit in well to the changing business environment) and also a "liability of senescence" (according to which they become ossified by accumulated rules, routines, and organizational structures). At a theoretical level [54] justify inertia effects as "an outcome of an ecological evolutionary process". The idea is that firms are not able to change as fast as their environment would dictate. Firms with inertia effects can survive by adopting strategies such as the creation of new firms designed specifically to take advantage of new opportunities. However, if firms are not able to adapt to the changes in the business environment, new entrants will enter the industry. Accordingly it is the changes in the business environment, which favor some bundles of firm resources over others, that lead to differences in firm performance. Inertia effects provide a potential explanation to the non-monotonicity in the age effect particularly in the most senior age groups documented in our study.

5.5 Conclusion

We have documented a persistent performance gap between the equally weighted rebalanced portfolio and the equally weighted bootstrapped portfolio during the last forty years. Both portfolios involve randomly picked stocks. In the bootstrapped portfolio N stocks are randomly selected and assigned equal weight at the beginning of each month. In the rebalanced portfolio N stocks are randomly chosen at inception and the portfolio is rebalanced to be equally weighted at the beginning of each month. If a stock is delisted in the rebalanced portfolio it is replaced with a randomly picked stock from the available universe.

The key feature that differentiates these two types of portfolios is the age distribution of the portfolio constituents. There is considerable turnover in the composition of the stock universe. During the last forty years, the average annual rate of new listings was around ten percent and the average annual rate of withdrawals was about the same magnitude. This means that the bootstrapped portfolio will in general include a sizable proportion of younger stocks. In contrast any stock that is included in the rebalanced remains there until it is delisted. As a consequence the rebalanced portfolio contains more senior stocks.

We analyzed the effect of stock age on the cross-sectional stock return and showed that the difference in age distribution is the primary reason behind the performance gap between the bootstrapped and rebalanced portfolios. Our regression results show that stock age has a significantly positive effect on stock return during the last four decades. This finding supports our conjecture that age explains the performance gap. We doubly sort portfolios by age and by size to assess the effect of age particularly when the size factor is controlled for. The age effect is observed in all of the four size groups. In addition a decade-by-decade breakdown of the returns on the sixteen factor portfolios indicates that the age effect dominates the size effect over the period from 1977 to 2016. It is robust when we consider the decade by decade results. Prior to 1977 the age effect was found to be weaker than it was after 1977. In addition the age effect while present in the earlier time period was dominated by the size effect.

The size effect is a well-known asset pricing anomaly. However the age effect that we have uncovered here appears to be in conflict with it, since age and size are generally positively correlated measures. To resolve this puzzle we monitored the size of three cohorts of stocks and analyzed the effect of aging on the market capitalization distribution within each cohort. The results suggest that the average market capitalization of a cohort of stocks increases as the stocks are aging. However the dispersion in market capitalization rises much more dramatically. So it is not uncommon for a stock to remain or to become a small cap as it ages. In an experiment where stocks are divided into decile groups according to age or size, we find that when the within-group median return is used as the performance measure, the direction of the size effect and that of the age effect are consistent with our intuition. Larger-size groups and more senior groups outperform; however the extremely positive skewness in the returns of the smallest 10% stocks causes the mean return of the smallest stocks to be much higher than the median return.

Chapter 6

Conclusion

This thesis studies two critical problems in the portfolio management area. First, the presence of estimation errors in a large covariance matrix estimator renders the portfolio optimization theory of Markowitz less desirable in practical applications. Second, the outperformance of some passive portfolio investment strategies, especially the $1/N$ rule, is puzzling to say the least. In this thesis, we provide a theoretical justification for the use of a few existing methods in portfolio optimization applications, propose a novel approach to constructing an optimal portfolio robust to estimation errors, and provide two reasons behind the documented superior performance of the $1/N$ portfolio. The line of inquiry leads to discovery of an equity market pricing anomaly named the age effect.

One stream of literature on mitigating estimation errors in the covariance matrix obtains an optimized portfolio more robust to estimation errors by proposing a better covariance estimator and replacing the sample covariance matrix with the newly proposed estimator in the two-step approach. We investigate that, if the new covariance estimator is constructed via an adjustment of sample eigenvalues, what adjustment patterns would ensure an improvement in the out-of-sample Sharpe ratio. Our results show that simultaneously amplifying a collection of tail sample eigenvalues according to a certain nonlinear pattern yields a positive effect for the out-of-sample Sharpe ratio improvement. This result provides additional justification for using the “shrinkage towards identity” method and the spectral cut-off method in portfolio optimization applications.

We propose an “expected returns approximation” approach to constructing a Sharpe ratio maximizing portfolio robust to estimation errors in the covariance matrix. In this approach, the exogenously given expected returns vector is approximated by a collection of sample eigenvectors that relatively well estimate their population counterparts. The optimized portfolio constructed with the approximation plugged into the two-step approach is more robust to estimation errors. To ensure the proximity between the original returns vector and its approximation we impose an upper bound on the relative approximation error introduced. Two concrete approximation methods are discussed. The first one has an equivalent mathematical form to the spectral cut-off method in the literature; however the number of principle components to drop depends on the imposed upper bound on the approximation error and the geometric relationship between the expected returns vector and the eigenvectors. The second approximation method leads to a completely novel "spectral selection method". Both methods reduce the exposure of the optimized portfolio to estimation errors as much as possible within the region implied by the constraint on the approximation error. The spectral selection method is shown by simulation and empirical studies to outperform competing methods in terms of their out-of-sample performance.

Although the $1/N$ portfolio has achieved a good performance in the US equity market in the past decades, the outperformance of $1/N$, naturally, does not persist during each sample period or in each equity market. We identify the reason behind the assorted performance of the $1/N$ portfolio in different markets and during different sample periods. Although the $1/N$ portfolio does not involve any explicit optimization, it can be the solution to an optimization problem. A market favorable to holding a $1/N$ portfolio is featured by the condition that the mean-variance optimal portfolio is close to the $1/N$ portfolio. We use a novel cosine measure named the "$1/N$ favorability index" to quantify the proximity between the mean-variance optimal portfolio and the $1/N$ portfolio. We show that a single (market) factor model predicts a positive relationship between the market factor return and the $1/N$ favorability index. In other words, a $1/N$ portfolio is more difficult to outperform in bullish markets. As a long-only strategy, the $1/N$ portfolio is, unsurprisingly, favored more in bullish markets than in bearish markets in terms of its expected return. However in our context the notation of favorability is represented by how difficult it is for other portfolios to outperform the $1/N$ portfolio. In bullish markets, all portfolios are exposed to high asset returns and low volatilities yet optimized portfolios have chances to buy relative winners (in risk-adjusted sense) and sell relative losers to outperform the $1/N$. Therefore our main implication is not as intuitive as it may be perceived to be. Further, the positive relationship between the market factor return and the difficulty to outperform a $1/N$ portfolio at least partially explains the lasting outperformance of the $1/N$ portfolio in the US equity market.

We discover important facts about portfolio rebalancing through digging into two types of equally weighted random portfolios, the rebalanced portfolio and the bootstrapped portfolio. We observe a consistent outperformance of the rebalanced portfolio relative to the bootstrapped portfolio over the past 40 years. In addition, the portfolio construction mechanism for the two portfolios together with the frequent listings and delistings in the stock universe gives rise to different age distribution of components in the two portfolios. More specifically, stocks in the rebalanced portfolio are on average older, in terms of age since listing, than those in the bootstrapped portfolio. A panel-data regression analysis and a comparison among factor-sorted portfolio confirm a generally positive relation between stock age and return. This age effect is able to explain the outperformance of the rebalanced portfolio compared with the bootstrapped portfolio. Therefore, another reason behind the superior performance of the conventional $1/N$ portfolio is that it implicitly exploits the age effect. With regard to the panel-data regression analysis, it is possible to perform additional robustness analysis by including fixed effects, which is a vector of sectoral dummy variables to control for time-invariant sector characteristics. In addition, explicit firm characteristics, in particular size and other relevant ones, which are time-varying, can also be included in the regression in addition to the time effect variable.

References

[1] Jeffery S Abarbanell and Brian J Bushee. Abnormal returns to a fundamental analysis strategy. *Accounting Review*, pages 19–45, 1998.

[2] Tobias Adrian, Erkko Etula, and Tyler Muir. Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, 69(6):2557–2596, 2014.

[3] Ron Alquist, Ronen Israel, and Tobias J. Moskowitz. Fact, fiction, and the size effect. *The Journal of Portfolio Management*, 2018(1):082, 2018.

[4] Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006.

[5] 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.

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

[7] Geum Il Bae, Woo Chang Kim, and John M Mulvey. Dynamic asset allocation for varied financial markets under regime switching framework. *European Journal of Operational Research*, 234(2):450–458, 2014.

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

[9] Malcolm Baker and Jeffrey Wurgler. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680, 2006.

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

[11] David N Barron, Elizabeth West, and Michael T Hannan. A time to grow and a time to die: Growth and mortality of credit unions in new york city, 1914-1990. *American Journal of Sociology*, 100(2):381–421, 1994.

[12] Christopher B Barry and Stephen J Brown. Differential information and the small firm effect. *Journal of Financial Economics*, 13(2):283–294, 1984.

[13] Shlomo Benartzi and Richard H Thaler. Naive diversification strategies in defined contribution saving plans. *American Economic Review*, 91(1):79–98, 2001.

[14] Hendrik Bessembinder. Do stocks outperform treasury bills? *Journal of Financial Economics*, 129(3):440–457, 2018.

[15] Fischer Black and Robert Litterman. Asset allocation: combining investor views with market equilibrium. Technical report, Discussion paper, Goldman, Sachs & Co, 1990.

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

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

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

[19] Phelim P Boyle and Thierno B N'Diaye. Correlation matrices with the perron-frobenius property. *Electronic Journal of Linear Algebra*, 34:240–268, 2018.

[20] Michael J Brennan and Walter N Torous. Individual decision making and investor welfare. *Economic Notes*, 28(2):119–143, 1999.

[21] 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.

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

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

[24] Yongsung Chang, Joao F Gomes, and Frank Schorfheide. Learning-by-doing as a propagation mechanism. *American Economic Review*, 92(5):1498–1520, 2002.

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

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

[27] Roger G Clarke, Harindra De Silva, and Steven Thorley. Minimum-variance portfolios in the us equity market. *The Journal of Portfolio Management*, 33(1):10–24, 2006.

[28] Gregory Connor and Robert A Korajczyk. A test for the number of factors in an approximate factor model. *The Journal of Finance*, 48(4):1263–1291, 1993.

[29] 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, 2009.

[30] 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):1915–1953, 2009.

[31] Craig Doidge, G Andrew Karolyi, and René M Stulz. The us listing gap. *Journal of Financial Economics*, 123(3):464–487, 2017.

[32] David L Donoho, Matan Gavish, and Iain M Johnstone. Optimal shrinkage of eigenvalues in the spiked covariance model. *arXiv preprint arXiv:1311.0851*, 2013.

[33] Edwin J Elton, Martin J Gruber, and Thomas J Urich. Are betas best? *The Journal of Finance*, 33(5):1375–1384, 1978.

[34] Robert F Engle, Olivier Ledoit, and Michael Wolf. Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, pages 1–13, 2017.

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

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

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

[38] Jianqing Fan, Yingying Fan, and Jinchi Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197, 2008.

[39] 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.

[40] 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.

[41] Mark Fichman and Daniel A Levinthal. Honeymoons and the liability of adolescence: A new perspective on duration dependence in social and organizational relationships. *Academy of Management Review*, 16(2):442–468, 1991.

[42] Jonathan Fletcher. Do optimal diversification strategies outperform the 1/n strategy in uk stock returns? *International Review of Financial Analysis*, 20(5):375–385, 2011.

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

[44] Carolina Fugazza, Massimo Guidolin, and Giovanna Nicodano. Equally weighted vs. long-run optimal portfolios. *European Financial Management*, 21(4):742–789, 2015.

[45] Lorenzo Garlappi, Raman Uppal, and Tan Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1):41–81, 2006.

[46] Elizabeth Garnsey. A theory of the early growth of the firm. *Industrial and Corporate Change*, 7(3):523–556, 1998.

[47] Paul A Gompers and Andrew Metrick. Institutional investors and equity prices. *The Quarterly Journal of Economics*, 116(1):229–259, 2001.

[48] Parameswaran Gopikrishnan, Bernd Rosenow, Vasiliki Plerou, and H Eugene Stanley. Quantifying and interpreting collective behavior in financial markets. *Physical Review E*, 64(3):035106, 2001.

[49] Larry E Greiner. Evolution and revolution as organizations grow. In *Readings in Strategic Management*, pages 373–387. Springer, 1989.

[50] Richard C Grinold. Are benchmark portfolios efficient? *Journal of Portfolio Management*, 19:34–34, 1992.

[51] Danqiao Guo, Phelim Boyle, Chengguo Weng, and Tony Wirjanto. Eigen portfolio selection: A robust approach to sharpe ratio maximization, 2019.

[52] Danqiao Guo, Chengguo Weng, and Tony S Wirjanto. Sample eigenvalues adjustment for portfolio performance improvement under factor models. *Working paper, SSRN 2959808*, 2019.

[53] James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, pages 357–384, 1989.

[54] Michael T Hannan and John Freeman. Structural inertia and organizational change. *American Sociological Review*, pages 149–164, 1984.

[55] Mary R Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53, 2001.

[56] Campbell R Harvey, Yan Liu, and Heqing Zhu. and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68, 2016.

[57] JB Heaton, NG Polson, and Jan Hendrik Witte. Why indexing works. *Applied Stochastic Models in Business and Industry*, 33(6):690–693, 2017.

[58] Joel L Horowitz, Tim Loughran, and NE Savin. The disappearing size effect. *Research in Economics*, 54(1):83–100, 2000.

[59] Inchang Hwang, Simon Xu, and Francis In. Naive versus optimal diversification: Tail risk and performance. *European Journal of Operational Research*, 265(1):372–388, 2018.

[60] Heiko Jacobs, Sebastian Müller, and Martin Weber. How should individual investors diversify? an empirical evaluation of alternative asset allocation policies. *Journal of Financial Markets*, 19:62–85, 2014.

[61] 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.

[62] Philippe Jorion. International portfolio diversification with estimation risk. *Journal of Business*, pages 259–278, 1985.

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

[64] Boyan Jovanovic. Selection and the evolution of industry. *Econometrica: Journal of the Econometric Society*, pages 649–670, 1982.

[65] Raymond Kan, Xiaolu Wang, and Guofu Zhou. On the value of portfolio optimization in the presence of estimation risk: the case with and without risk-free asset. *Working paper, SSRN 2819254*, 2016.

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

[67] Alexander Kempf and Christoph Memmel. Estimating the global minimum variance portfolio. *Schmalenbach Business Review*, 58(4):332–348, 2006.

[68] Min Jeong Kim, Yongjae Lee, Jang Ho Kim, and Woo Chang Kim. Sparse tangent portfolio selection via semi-definite relaxation. *Operations Research Letters*, 44(4):540–543, 2016.

[69] Woo Chang Kim and Yongjae Lee. A uniformly distributed random portfolio. *Quantitative Finance*, 16(2):297–307, 2016.

[70] Chris Kirby and Barbara Ostdiek. Its all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, 47(2):437–467, 2012.

[71] 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.

[72] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467, 1999.

[73] Oliver Ledoit and Michael Wolf. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859, 2008.

[74] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621, 2003.

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

[76] 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.

[77] Xiao Li. Aggregating information for optimal portfolio weights. *Working paper, SSRN 3248680*, 2018.

[78] Xiaoji Lin, Berardino Palazzo, and Fan Yang. The risks of old capital age: Asset pricing implications of technology adoption. *Working paper, SSRN 2535409*, 2018.

[79] Claudio F Loderer and Urs Waelchli. Firm age and performance. *Working paper, SSRN 1342248*, 2010.

[80] Tim Loughran and Jay R Ritter. The new issues puzzle. *The Journal of Finance*, 50(1):23–51, 1995.

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

[82] 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.

[83] Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

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

[85] Lionel Martellini. Toward the design of better equity benchmarks: Rehabilitating the tangency portfolio from modern portfolio theory. *The Journal of Portfolio Management*, 34(4):34–41, 2008.

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

[87] Alexei Onatski. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479, 2009.

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

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

[90] Georg Ch Pflug, Alois Pichler, and David Wozabal. The 1/n investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.

[91] Yuliya Plyakha, Raman Uppal, and Grigory Vilkov. Why do equal-weighted portfolios outperform value-weighted portfolios? *Working paper, SSRN 2724535*, 2015.

[92] Jay R Ritter. The long-run performance of initial public offerings. *The Journal of Finance*, 46(1):3–27, 1991.

[93] Richard Roll. Orthogonal portfolios. *Journal of Financial and Quantitative Analysis*, 15(5):1005–1023, 1980.

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

[95] Bernd Scherer. A new look at minimum variance investing. *Working paper, SSRN 1681306*, 2010.

[96] William F Sharpe. A simplified model for portfolio analysis. *Management Science*, 9(2):277–293, 1963.

[97] William F Sharpe. Imputing expected security returns from portfolio composition. *Journal of Financial and Quantitative Analysis*, 9(3):463–472, 1974.

[98] 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.

[99] Tyler Shumway. The delisting bias in CRSP data. *The Journal of Finance*, 52(1):327–340, 1997.

[100] Jesper B Sørensen and Toby E Stuart. Aging, obsolescence, and organizational innovation. *Administrative Science Quarterly*, 45(1):81–112, 2000.

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

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

[103] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[104] 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.

[105] Thomas S Ulen and Peter Newman. *The New Palgrave Dictionary of Economics and the Law*. Palgrave Macmillan UK, 1998.

[106] Mathijs A Van Dijk. Is size dead? a review of the size effect in equity returns. *Journal of Banking & Finance*, 35(12):3263–3274, 2011.

[107] Heath Windcliff and Phelim P Boyle. The 1/N pension investment puzzle. *North American Actuarial Journal*, 8(3):32–45, 2004.

APPENDICES

Appendix A

Proofs and Auxiliary Results

A.1 Proof of Propositions in Chapter 1

Proof of Theorem 2.2.1. Let $g(\mathbf{V},\lambda) := \frac{\boldsymbol{\mu}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\Sigma}\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}{(\boldsymbol{\mu}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu})^2}$. We first calculate $g'_{\mathbf{V}}(0) \equiv \frac{\partial g(\mathbf{V},\lambda)}{\partial\lambda}\big|_{\lambda=0}$. Taking a partial derivative of $g(\mathbf{V},\lambda)$ with respect to λ, we obtain

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

The derivative in eq. (A.1) can be simplified as follows:

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

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

$$\frac{\partial\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}{\partial\lambda}\bigg|_{\lambda=0} = \widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu} \quad \text{and} \quad \frac{\partial\boldsymbol{\mu}^T\mathbf{S}_{\mathbf{V},\lambda}^{-1}\boldsymbol{\mu}}{\partial\lambda}\bigg|_{\lambda=0} = \boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu}. \tag{A.3}$$

Now, we can readily evaluate $g'_{\mathbf{V}}(0)$ with the assistance of eq. (A.2) and eq. (A.3). After a few straightforward steps of calculation, we obtain:

$$g'_{\mathbf{V}}(0) = \frac{2}{(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu})^2}\left(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu} - \boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu}\frac{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}\right). \tag{A.4}$$

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

$$SR'_{\mathbf{V}}(0) = -\frac{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}{(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu})^{\frac{3}{2}}}\left(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu} - \boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2}\mathbf{V}\widehat{\mathbf{U}}^T\boldsymbol{\mu}\frac{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}\right). \tag{A.5}$$

□

Proof of Theorem 2.2.2. Before proceeding to the proof, we introduce a few necessary notations. Let $\mathbf{U}_F = (\mathbf{u}_1, \ldots, \mathbf{u}_K)$ and $\mathbf{U}_I = (\mathbf{u}_{K+1}, \ldots, \mathbf{u}_N)$ denote the eigenvectors that correspond to the factors and the idiosyncratic components respectively. Further, let $\boldsymbol{\Lambda}_F = \text{diag}\{\lambda_1, \ldots, \lambda_K\}$ and $\boldsymbol{\Lambda}_I = \text{diag}\{\lambda_{K+1}, \ldots, \lambda_N\}$. The sample quantities can be conformably partitioned into a factor part and an idiosyncratic part, i.e. $\widehat{\mathbf{U}}_F = (\widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_K)$, $\widehat{\mathbf{U}}_I = (\widehat{\mathbf{u}}_{K+1}, \ldots, \widehat{\mathbf{u}}_N)$, $\widehat{\boldsymbol{\Lambda}}_F = \text{diag}\{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_K\}$, and $\widehat{\boldsymbol{\Lambda}}_I = \text{diag}\{\widehat{\lambda}_{K+1}, \ldots, \widehat{\lambda}_N\}$. Let $\mathbf{a} = \mathbf{U}^T\boldsymbol{\mu} = (\mathbf{a}_F^T, \mathbf{a}_I^T)^T = (a_1, \ldots, a_N)^T$ denote the loading of the expected returns vector on the linear space spanned by the population eigenvectors. Similarly, let $\widehat{\mathbf{a}} = \widehat{\mathbf{U}}^T\boldsymbol{\mu} = (\widehat{\mathbf{a}}_F^T, \widehat{\mathbf{a}}_I^T)^T = (\widehat{a}_1, \ldots, \widehat{a}_N)^T$ denote the loading on the sample eigenvectors. Here, $\mathbf{a}_F$ and $\widehat{\mathbf{a}}_F$ are $K \times 1$ vectors; $\mathbf{a}_I$ and $\widehat{\mathbf{a}}_I$ are $(N-K) \times 1$ vectors. Let $\mathbf{e}_k$ denote a conformable column vector with its kth element being 1 and everywhere else being 0. Denote $a_\tau = o_{a.s.}(b_\tau)$ if $\lim_{\tau\to\infty} \frac{a_\tau}{b_\tau} = 0$ almost surely. Denote $a_\tau = O_{a.s.}(b_\tau)$ if $\overline{\lim}_{\tau\to\infty}|\frac{a_\tau}{b_\tau}| \leq M$, where M is a positive constant.

According to Theorem 1 case (a) in [98], as both N and T go to infinity with the relative rate being $T \asymp N^{1+c}$ for some $c > 0$, we obtain the following results:

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

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

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

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

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

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

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

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

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

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

$$\widehat{\mathbf{u}}_j^T \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{u}}_j = (1 - 0.5(o_{a.s.}(N^{-1/2}))^2)^2 = 1 + o_{a.s.}(N^{-1}), \quad j = K+1, \ldots, N. \tag{A.6}$$

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

$$\begin{aligned} & \widehat{\mathbf{u}}_j^T \left(\sum_{i=1}^{K} \mathbf{u}_i \mathbf{u}_i^T + \mathbf{U}_I \mathbf{U}_I^T \right) \widehat{\mathbf{u}}_j = 1 \\ \rightarrow \quad & \sum_{i=1}^{K} (\mathbf{u}_i^T \widehat{\mathbf{u}}_j)^2 = o_{a.s.}(N^{-1}) \\ \rightarrow \quad & \mathbf{u}_i^T \widehat{\mathbf{u}}_j = o_{a.s.}(N^{-1/2}), \quad i = 1, \ldots, K, \; j = K+1, \ldots, N. \end{aligned}$$

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

$$\widehat{\mathbf{u}}_i^T \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{u}}_j = -\widehat{\mathbf{u}}_i^T \left(\sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^T \right) \widehat{\mathbf{u}}_j = - \sum_{k=1}^{K} \widehat{\mathbf{u}}_i^T \mathbf{u}_k \mathbf{u}_k^T \widehat{\mathbf{u}}_j = o_{a.s.}(N^{-1}). \tag{A.7}$$

According to eq. (A.6) and eq. (A.7), $\widehat{\mathbf{U}}_I^T \mathbf{U}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_I = \mathbf{I}_{(N-K)\times(N-K)} + \mathbf{E}$, where $\mathbf{E}$ is a noise matrix whose elements have rate $o_{a.s.}(N^{-1})$.

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

$$\begin{aligned} \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} &= \begin{bmatrix} \widehat{\mathbf{U}}_F^T \\ \widehat{\mathbf{U}}_I^T \end{bmatrix} \begin{bmatrix} \mathbf{U}_F & \mathbf{U}_I \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_F & \\ & \mathbf{\Lambda}_I \end{bmatrix} \begin{bmatrix} \mathbf{U}_F^T \\ \mathbf{U}_I^T \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{U}}_F & \widehat{\mathbf{U}}_I \end{bmatrix} \\ &= \begin{bmatrix} \widehat{\mathbf{U}}_F^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_F + \widehat{\mathbf{U}}_F^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_F & \widehat{\mathbf{U}}_F^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_I + \widehat{\mathbf{U}}_F^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_I \\ \widehat{\mathbf{U}}_I^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_F + \widehat{\mathbf{U}}_I^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_F & \widehat{\mathbf{U}}_I^T \mathbf{U}_F \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_I + \widehat{\mathbf{U}}_I^T \mathbf{U}_I \mathbf{\Lambda}_I \mathbf{U}_I^T \widehat{\mathbf{U}}_I \end{bmatrix}. \end{aligned} \tag{A.8}$$

Let $\mathbf{b} = (\mathbf{b}_F^T, \mathbf{b}_I^T)^T$ be a $p \times 1$ random vector. According to the convergence results derived above, one of the two terms in the summation in each element of eq. (A.8) is negligible compared with the other term. The quadratic form $\mathbf{b}^T \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} \mathbf{b}$, under the large dimensional asymptotics and Assumption 2.2.4, can be rewritten as:

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

For example, when $\mathbf{b} = \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \boldsymbol{\mu} = \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{a}}$ and under the Assumption 2.2.5 that $\|\boldsymbol{\mu}\| = O(N^{1/2})$ and $\|\mathbf{a}_F\|_\infty = \left\|\mathbf{U}_F^T \boldsymbol{\mu}\right\|_\infty = o(N^{1/2})$,

$$\begin{aligned}
&\mathbf{b}^T \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} \mathbf{b} \\
=&\boldsymbol{\mu}^T \mathbf{S}^{-1} \mathbf{\Sigma} \mathbf{S}^{-1} \boldsymbol{\mu} \\
=&\widehat{\mathbf{a}}_F^T \widehat{\mathbf{\Lambda}}_F^{-1} \mathbf{\Lambda}_F \widehat{\mathbf{\Lambda}}_F^{-1} \widehat{\mathbf{a}}_F (1 + o_{a.s.}(1)) + 2\widehat{\mathbf{a}}_F^T \widehat{\mathbf{\Lambda}}_F^{-1} \mathbf{\Lambda}_F \mathbf{U}_F^T \widehat{\mathbf{U}}_I \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)) \\
&+ \sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)) \\
=&\widehat{\mathbf{a}}_F^T \widehat{\mathbf{\Lambda}}_F^{-1} \widehat{\mathbf{a}}_F (1 + o_{a.s.}(1)) + 2\widehat{\mathbf{a}}_F^T \mathbf{U}_F^T \widehat{\mathbf{U}}_I \widehat{\mathbf{\Lambda}}_I^{-1} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)) + \sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)) \\
=&\sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-2} \widehat{\mathbf{a}}_I (1 + o_{a.s.}(1)).
\end{aligned} \tag{A.9}$$

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

(a) When $\mathbf{V}$ takes the form of $\mathbf{E}_k$ and as N, T go to infinity with the relative rate being $T \asymp N^{1+c}$ $(c > 0)$, for any $k \in \{1, \ldots, N\}$, the first term inside the bracket in eq. (2.6) can be expanded as follows:

$$\begin{aligned}
\boldsymbol{\mu}^T \mathbf{S}^{-1} \mathbf{\Sigma} \widehat{\mathbf{U}} \widehat{\mathbf{\Lambda}}^{-2} \mathbf{E}_k \widehat{\mathbf{U}}^T \boldsymbol{\mu} &= \widehat{\mathbf{a}}^T \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} \widehat{\mathbf{\Lambda}}^{-2} \mathbf{E}_k \widehat{\mathbf{a}} \\
&= \frac{\widehat{a}_k}{\widehat{\lambda}_k^2} \widehat{\mathbf{a}}^T \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{U}}^T \mathbf{\Sigma} \widehat{\mathbf{U}} \mathbf{e}_k \\
&= \begin{cases} \frac{\widehat{a}_k}{\widehat{\lambda}_k^2} \widehat{\mathbf{a}}_F^T \mathbf{e}_k (1 + o_{a.s.}(1)) & k \leq K \\ \frac{\widehat{a}_k}{\widehat{\lambda}_k^2} \sigma^2 \widehat{\mathbf{a}}_I^T \widehat{\mathbf{\Lambda}}_I^{-1} \mathbf{e}_{k-K} (1 + o_{a.s.}(1)) & k > K \end{cases} \\
&= \begin{cases} \frac{\widehat{a}_k^2}{\widehat{\lambda}_k^2} (1 + o_{a.s.}(1)) & k \leq K \\ \frac{\sigma^2 \widehat{a}_k^2}{\widehat{\lambda}_k^3} (1 + o_{a.s.}(1)) & k > K \end{cases}.
\end{aligned}$$

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

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

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

$$SR'_{\mathbf{E}_k}(0) = -\frac{\widehat{a}_k^2\widehat{\mathbf{a}}_I^T\widehat{\boldsymbol{\Lambda}}_I^{-1}\widehat{\mathbf{a}}_I}{\widehat{\lambda}_k^2(\sigma^2\widehat{\mathbf{a}}_I^T\widehat{\boldsymbol{\Lambda}}_I^{-2}\widehat{\mathbf{a}}_I)^{\frac{3}{2}}}(1+o_{a.s.}(1)) + \frac{\widehat{a}_k^2}{\widehat{\lambda}_k^2(\sigma^2\widehat{\mathbf{a}}_I^T\widehat{\boldsymbol{\Lambda}}_I^{-2}\widehat{\mathbf{a}}_I)^{\frac{1}{2}}}(1+o_{a.s.}(1)) \quad \text{(A.10)}$$

Since $\frac{\widehat{\mathbf{a}}_I^F\widehat{\boldsymbol{\Lambda}}_I^{-1}\widehat{\mathbf{a}}_I}{\widehat{\mathbf{a}}_I^F\widehat{\boldsymbol{\Lambda}}_I^{-2}\widehat{\mathbf{a}}_I} \in [\widehat{\lambda}_N, \widehat{\lambda}_{K+1}]$ and $\frac{1}{(\widehat{\mathbf{a}}_I^T\widehat{\boldsymbol{\Lambda}}_I^{-2}\widehat{\mathbf{a}}_I)^{\frac{1}{2}}} \le \frac{\widehat{\lambda}_{K+1}}{(\widehat{\mathbf{a}}_I^T\widehat{\mathbf{a}}_I)^{\frac{1}{2}}}$, the absolute value of the first term (without the $(1+o_{a.s.}(1))$ part) on the RHS of eq. (A.10) satisfies

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

Similarly, the absolute value of the second term satisfies

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

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

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

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

(b) When $\mathbf{V}$ takes the form of $\mathbf{E}_{k^+}$ and as N, T go to infinity with the relative rate being $T \asymp N^{1+c}$ $(c > 0)$, for any $k \in \{1, 2, \ldots, N\}$, the first term inside the bracket in eq.

(2.6) can be expanded as follows:

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

The second term inside the bracket in eq. (2.6) can be written as:

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

Let

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

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

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

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

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

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

□

Proof of Theorem 2.2.3. With the additional assumption that $\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^T\frac{1}{s_h^2} + \sigma^2\mathbf{I}$, the terms inside the bracket in eq. (2.6) become:

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

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

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

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

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

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

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

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

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

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

$$SR'_{-\widehat{\boldsymbol{\Lambda}}^a\mathbf{E}_{k+}}(0) = \frac{\sigma^2\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2+a}\mathbf{E}_{k+}\widehat{\mathbf{U}}^T\boldsymbol{\mu}}{(\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}\boldsymbol{\mu})^{\frac{3}{2}}}\left(\frac{\boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-3+a}\mathbf{E}_{k+}\widehat{\mathbf{U}}^T\boldsymbol{\mu}}{\boldsymbol{\mu}^T\widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}^{-2+a}\mathbf{E}_{k+}\widehat{\mathbf{U}}^T\boldsymbol{\mu}} - \frac{\boldsymbol{\mu}^T\mathbf{S}^{-2}\boldsymbol{\mu}}{\boldsymbol{\mu}^T\mathbf{S}^{-1}\boldsymbol{\mu}}\right). \tag{A.11}$$

The terms inside the bracket in eq. (A.11), denoted as Δ, can be further simplified as:

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

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

□

A.2 Proof of Propositions in Chapter 3

Proof of Proposition 3.2.1. 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. (3.2) 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

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

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. (3.2):

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

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

Proof of Proposition 3.2.2. We first verify the existence of the MSR portfolio. The expected return of the minimum-variance portfolio is:

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

Therefore, the weight of the MSR portfolio is:

$$\mathbf{w}_{msr} = \frac{\boldsymbol{\Sigma}^{-1}\sum_{i=1}^N a_i\mathbf{u}_i}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\sum_{i=1}^N a_i\mathbf{u}_i} = \frac{\sum_{i=1}^N \frac{a_i}{\lambda_i}\mathbf{u}_i}{\sum_{i=1}^N \frac{a_i}{\lambda_i}\mathbf{1}^T\mathbf{u}_i} = \sum_{i=1}^N \frac{\frac{a_i\mathbf{1}^T\mathbf{u}_i}{\lambda_i}}{\sum_{i=1}^N \frac{a_i\mathbf{1}^T\mathbf{u}_i}{\lambda_i}}\mathbf{z}_i.$$

□

Proof of Proposition 3.3.2. 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 [98], under Assumptions 3.3.1 and 3.3.2, the first K sample eigenvalues and eigenvectors have such a limiting behavior that $\frac{\widehat{\lambda}_j}{\lambda_j^{(T)}} \xrightarrow{a.s.} 1$ and $\langle \widehat{\mathbf{u}}_j, \mathbf{u}_j \rangle \xrightarrow{a.s.} 1$, $j = 1, 2, \ldots, K$, as $T \to \infty$ and $N(T)/T \to c \in (0, 1)$. 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 $T \to \infty$ and $N(T)/T \to c \in (0, 1)$. □

Proof of Proposition 3.3.3. It is straightforward to check that the solution to the least square problem in eq. (3.7) is $\widehat{\mathbf{a}}^{ls} = (\widehat{a}_1^{ls}, \ldots, \widehat{a}_N^{ls})^T = (\widehat{\mathbf{U}}^T\widehat{\mathbf{U}})^{-1}\widehat{\mathbf{U}}^T\boldsymbol{\mu} = \widehat{\mathbf{U}}^T\boldsymbol{\mu}$. Next, we solve the optimization problem in eq. (3.14). Let $\mathbf{a} = (a_1, \ldots, a_N)^T$ denote the decision vector and define another vector $\mathbf{b} = (b_1, \ldots, b_N)^T := \widehat{\mathbf{\Lambda}}^{-c}\mathbf{a}$. 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:

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

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. □

A.3 Proof of Propositions in Chapter 4

Proof of Proposition 4.3.1. Let $\bar{\boldsymbol{\epsilon}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\epsilon}_t$ and $\widehat{\boldsymbol{\Omega}}_T = \frac{1}{T-1} \sum_{t=1}^{T} (\boldsymbol{\epsilon}_T - \bar{\boldsymbol{\epsilon}})(\boldsymbol{\epsilon}_T - \bar{\boldsymbol{\epsilon}})^\intercal$. We have

$$\widehat{\boldsymbol{\mu}}_T = \overline{\mathbf{R}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{R}_t = \boldsymbol{\beta} \overline{\mathbf{R}}_m + \bar{\boldsymbol{\epsilon}} = \boldsymbol{\mu}_T + \bar{\boldsymbol{\epsilon}}$$

and

$$\widehat{\mathbf{\Sigma}}_T = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{R}_t - \overline{\mathbf{R}})(\mathbf{R}_t - \overline{\mathbf{R}})^\intercal = \mathbf{\Sigma}_T + 2\boldsymbol{\beta}\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)(\boldsymbol{\epsilon}_t - \overline{\boldsymbol{\epsilon}})^\intercal + \widehat{\mathbf{\Omega}}_T - \sigma^2\mathbf{I}.$$

It follows that

$$\|\widehat{\mathbf{\Sigma}}_T\mathbf{1}\| \le \|\mathbf{\Sigma}_T\mathbf{1}\| + 2\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)(\boldsymbol{\epsilon}_t^\intercal\mathbf{1} - \overline{\boldsymbol{\epsilon}}^\intercal\mathbf{1})\|\boldsymbol{\beta}\| + \|(\widehat{\mathbf{\Omega}}_T - \sigma^2\mathbf{I})\mathbf{1}\|.$$

We further let the elements in $\boldsymbol{\epsilon}_t$ be $\boldsymbol{\epsilon}_t = (\epsilon_{t1}, \epsilon_{t2}, \ldots, \epsilon_{tN})^\intercal$ and those in $\overline{\boldsymbol{\epsilon}}$ be $\overline{\boldsymbol{\epsilon}} = (\overline{\epsilon}_1, \overline{\epsilon}_2, \ldots, \overline{\epsilon}_N)^\intercal$. To derive the order of $\|\widehat{\mathbf{\Sigma}}_T\mathbf{1}\|$, we examine the terms on the RHS of the equation above.

$$\begin{aligned}\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)(\boldsymbol{\epsilon}_t^\intercal\mathbf{1} - \overline{\boldsymbol{\epsilon}}^\intercal\mathbf{1})\|\boldsymbol{\beta}\| &= \sqrt{N}\|\boldsymbol{\beta}\|\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)\left(\frac{\boldsymbol{\epsilon}_t^\intercal\mathbf{1}}{\sqrt{N}} - \frac{\overline{\boldsymbol{\epsilon}}^\intercal\mathbf{1}}{\sqrt{N}}\right) \\ &= \sqrt{N}\|\boldsymbol{\beta}\|O_p\left(\frac{1}{\sqrt{T}}\right) \\ &= O_p\left(\frac{N}{\sqrt{T}}\right).\end{aligned}$$

The second last equality in the equation above is a direct application of the result on the convergence rate of sample covariance. The last equality holds because $\|\boldsymbol{\beta}\| = O(\sqrt{N})$.

Since $\boldsymbol{\epsilon}_t \overset{\text{iid}}{\sim} N(\mathbf{0}, \sigma^2\mathbf{I})$, for each element $\overline{\epsilon}_j$, $j = 1, 2, \ldots, N$, we have $\sqrt{T}\overline{\epsilon}_j = O_p(1)$, i.e., $\overline{\epsilon}_j = O_p(T^{-\frac{1}{2}})$. Therefore,

$$\|\overline{\boldsymbol{\epsilon}}\| = \sqrt{\sum_{j=1}^{N}\overline{\epsilon}_j^2} = O_p\left(\sqrt{\frac{N}{T}}\right).$$

$$\begin{aligned}\|(\widehat{\mathbf{\Omega}}_T - \sigma^2\mathbf{I})\mathbf{1}\| &= \sqrt{\sum_{k=1}^{N}\left[\sum_{j=1}^{N}\frac{1}{T-1}\sum_{t=1}^{T}(\epsilon_{tk} - \overline{\epsilon}_k)(\epsilon_{tj} - \overline{\epsilon}_j) - \sigma^2\right]^2} \\ &= \sqrt{\sum_{k=1}^{N}\left[\frac{1}{T-1}\sum_{t=1}^{T}(\epsilon_{tk} - \overline{\epsilon}_k)\left(\sum_{j=1}^{N}\epsilon_{tj} - \sum_{j=1}^{N}\overline{\epsilon}_j\right) - \sigma^2\right]^2}.\end{aligned}$$

Let independent random variables X_t and W_t denote $X_t = \epsilon_{tk}$ and $W_t = \sum_{j \neq k} \epsilon_{tj}$. Let $Y_t = X_t + W_t = \sum_{j=1}^{N} \epsilon_{tj}$. It is straightforward to show that $E[X_1 Y_1] = \sigma^2$. The variance of $X_1 Y_1$ is:

$$\begin{aligned}
\text{Var}(X_1 Y_1) &= E[X_1^2 Y_1^2] - \{E[X_1 Y_1]\}^2 \\
&= E[X_1^2 (X_1 + W_1)^2] - \sigma^4 \\
&= E[X_1^4 + 2X_1^3 W_1 + X_1^2 W_1^2] - \sigma^4 \\
&= 3\sigma^4 + (N-1)\sigma^4 - \sigma^4 \\
&= (N+1)\sigma^4.
\end{aligned}$$

According to the Central Limit Theorem, we have:

$$\frac{\frac{1}{T}\sum_{t=1}^{T} X_t Y_t - \sigma^2}{\sqrt{\frac{(N+1)\sigma^4}{T}}} \xrightarrow{D} Z \sim N(0,1).$$

Therefore, we obtain:

$$\frac{1}{T}\sum_{t=1}^{T} X_t Y_t = \sigma^2 + O_p\left(\sqrt{\frac{N}{T}}\right).$$

In addition, $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ and $\overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$ have rate $O_p(\frac{1}{\sqrt{T}})$ and $O_p(\sqrt{\frac{N}{T}})$ respectively. According to these results, we have:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}(\epsilon_{tk} - \bar{\epsilon}_k)\left(\sum_{j=1}^{N}\epsilon_{tj} - \sum_{j=1}^{N}\bar{\epsilon}_j\right) &= \frac{1}{T}\sum_{t=1}^{T}\epsilon_{tk}\sum_{j=1}^{N}\epsilon_{tj} - \overline{\epsilon_k}\sum_{j=1}^{N}\bar{\epsilon}_j \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t Y_t - \overline{XY} \\
&= \sigma^2 + O_p\left(\sqrt{\frac{N}{T}}\right).
\end{aligned}$$

Plugging this result into the expression for $\|(\widehat{\mathbf{\Omega}}_T - \sigma^2 \mathbf{I})\mathbf{1}\|$, we obtain:

$$\|(\widehat{\mathbf{\Omega}}_T - \sigma^2 \mathbf{I})\mathbf{1}\| = O_p\left(\frac{N}{\sqrt{T}}\right).$$

Therefore, the denominator of the observed $1/N$ favorability index is:

$$\begin{aligned}
\|\widehat{\boldsymbol{\mu}}_T\|\|\widehat{\boldsymbol{\Sigma}}_T\mathbf{1}\| &= \left(\|\boldsymbol{\mu}_T\| + O_p\left(\sqrt{\frac{N}{T}}\right)\right)\left(\|\boldsymbol{\Sigma}_T\mathbf{1}\| + O_p\left(\frac{N}{\sqrt{T}}\right)\right) \\
&= \|\boldsymbol{\mu}_T\|\|\boldsymbol{\Sigma}_T\mathbf{1}\| + \|\boldsymbol{\mu}_T\|O_p\left(\frac{N}{\sqrt{T}}\right) + \|\boldsymbol{\Sigma}_T\mathbf{1}\|O_p\left(\sqrt{\frac{N}{T}}\right) + O_p\left(\frac{N^{\frac{3}{2}}}{T}\right).
\end{aligned} \tag{A.12}$$

Now, we look at the numerator of the observed $1/N$ favorability index. Since $\|\boldsymbol{\beta}\| = O(\sqrt{N})$ and $\cos(\boldsymbol{\beta}, \mathbf{1})$ is bounded away from 0 as $N \to \infty$, we have $\|\boldsymbol{\mu}_T\| = O(\sqrt{N})$ and $\|\boldsymbol{\Sigma}_T\mathbf{1}\| = O(N^{\frac{3}{2}})$. The second conclusion comes from the relationship

$$(\boldsymbol{\beta}^\intercal\mathbf{1})\|\boldsymbol{\beta}\|\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)^2 \leq \|\boldsymbol{\Sigma}_T\mathbf{1}\| \leq (\boldsymbol{\beta}^\intercal\mathbf{1})\|\boldsymbol{\beta}\|\frac{1}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)^2 + \sigma^2\sqrt{N}$$

and the rate

$$\boldsymbol{\beta}^\intercal\mathbf{1} = \cos(\boldsymbol{\beta}, \mathbf{1})\|\boldsymbol{\beta}\|\|\mathbf{1}\| = O(N).$$

These results are used in analyzing the rate of the numerator of the observed $1/N$ favorability index.

$$\begin{aligned}
\langle\widehat{\boldsymbol{\mu}}_T, \widehat{\boldsymbol{\Sigma}}_T\mathbf{1}\rangle &= (\boldsymbol{\mu}_T + \overline{\boldsymbol{\epsilon}})^\intercal\left[\boldsymbol{\Sigma}_T\mathbf{1} + \frac{2\sqrt{N}}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)\left(\frac{\boldsymbol{\epsilon}_t^\intercal\mathbf{1}}{\sqrt{N}} - \frac{\overline{\boldsymbol{\epsilon}}^\intercal\mathbf{1}}{\sqrt{N}}\right)\boldsymbol{\beta} + (\widehat{\boldsymbol{\Omega}}_T - \sigma^2\mathbf{I})\mathbf{1}\right] \\
&= \boldsymbol{\mu}_T^\intercal\boldsymbol{\Sigma}_T\mathbf{1} + \frac{2\overline{R}_m\sqrt{N}\|\boldsymbol{\beta}\|^2}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)\left(\frac{\boldsymbol{\epsilon}_t^\intercal\mathbf{1}}{\sqrt{N}} - \frac{\overline{\boldsymbol{\epsilon}}^\intercal\mathbf{1}}{\sqrt{N}}\right) + \boldsymbol{\mu}_T^\intercal(\widehat{\boldsymbol{\Omega}}_T - \sigma^2\mathbf{I})\mathbf{1} \\
&\quad + \overline{\boldsymbol{\epsilon}}^\intercal\boldsymbol{\Sigma}_T\mathbf{1} + \frac{2\sqrt{N}\overline{\boldsymbol{\epsilon}}^\intercal\boldsymbol{\beta}}{T-1}\sum_{t=1}^{T}(R_{mt} - \overline{R}_m)\left(\frac{\boldsymbol{\epsilon}_t^\intercal\mathbf{1}}{\sqrt{N}} - \frac{\overline{\boldsymbol{\epsilon}}^\intercal\mathbf{1}}{\sqrt{N}}\right) + \overline{\boldsymbol{\epsilon}}^\intercal(\widehat{\boldsymbol{\Omega}}_T - \sigma^2\mathbf{I})\mathbf{1} \\
&= \boldsymbol{\mu}_T^\intercal\boldsymbol{\Sigma}_T\mathbf{1} + O_p\left(\frac{N^{\frac{3}{2}}}{\sqrt{T}}\right) + \|\boldsymbol{\mu}_T\|O_p\left(\frac{N}{\sqrt{T}}\right) \\
&\quad + \|\boldsymbol{\Sigma}_T\mathbf{1}\|O_p\left(\sqrt{\frac{N}{T}}\right) + O_p\left(\frac{N^{\frac{3}{2}}}{T}\right) + O_p\left(\frac{N^{\frac{3}{2}}}{T}\right) \\
&= \boldsymbol{\mu}_T^\intercal\boldsymbol{\Sigma}_T\mathbf{1} + O_p\left(\frac{N^2}{\sqrt{T}}\right).
\end{aligned} \tag{A.13}$$

Therefore, under the asymptotic that $N, T \to \infty$ with a relative rate of $\frac{N}{T} = O(1)$, the last three terms in eq. (A.12) are all of a lower order compared with $\|\boldsymbol{\mu}_T\|\|\boldsymbol{\Sigma}_T\mathbf{1}\|$. The absolute value of the first term in the expression for the numerator is $|\boldsymbol{\mu}_T^\intercal\boldsymbol{\Sigma}_T\mathbf{1}| \geq \cos(\boldsymbol{\beta}, \mathbf{1})\|\boldsymbol{\mu}_T\|\|\boldsymbol{\Sigma}_T\mathbf{1}\| = O(N^2)$. As a result, $\boldsymbol{\mu}_T^\intercal\boldsymbol{\Sigma}_T\mathbf{1}$ is the dominating term in eq. (A.13) as T increases to infinity. Thus, we have:

$$\cos(\widehat{\boldsymbol{\mu}}_T, \widehat{\boldsymbol{\Sigma}}_T\mathbf{1}) - \cos(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T\mathbf{1}) \xrightarrow{p} 0$$

as $N, T \to \infty$ with the relative rate of $\frac{N}{T} = O(1)$. □

A.4 Auxiliary Results for Chapter 5

A.4.1 Main Results Based on Data from July 1926 to December 1976

Figure A.1 presents the profile of stock population over the period from 1927 to 1976. A comparison between Figure A.1 and Figure 5.1 further justifies the need of analyzing the earlier 50-year period and the more recent 40-year period separately. There is a marked difference in the activity level of the market, in terms of number of listed stocks and frequency of enters and exits, between the two periods. Further the CRSP database went through drastic changes in 1962 due to the inclusion of AMEX stocks and in 1972 due to the inclusion of Nasdaq stocks. The stock population is much more stable in the more recent 40 years.

Figure A.2 and Table 3(b) contain similar information as Figure 5.2 and Table 5.3 except that the time range of the data used here is from July 1926 to December 1976. As in Table 5.3, we can see a clear increasing trend in the average age in the rebalanced portfolio from decade to decade. In addition, there is an increasing age gap between the bootstrapped and rebalanced portfolios. This is unsurprising because the age gap comes from the mechanism of portfolio construction and should not depend on the data period. As long as there are listings and delistings in the market, we should be able to see such an age gap. The average age in bootstrapped portfolios can be viewed as a proxy for the average age of all available stocks. According to Table 3(b), the average age of bootstrapped portfolios varies substantially from decade to decade, compared with the much more stable average age reported in Table 5.3. The volatile average age across time again implies that the stock universe prior to 1977 is highly unstable.

Figure A.1: Population of stocks in the CRSP database: existing stocks and new listings. The figure shows the change in the stock population in the CRSP database from 1927 to 1976. The red portion in each bar represents the number of stocks that enter the universe in the indicated calendar year. The blue portion represents the number of stocks that have existed in the universe by the beginning of the indicated calendar year.

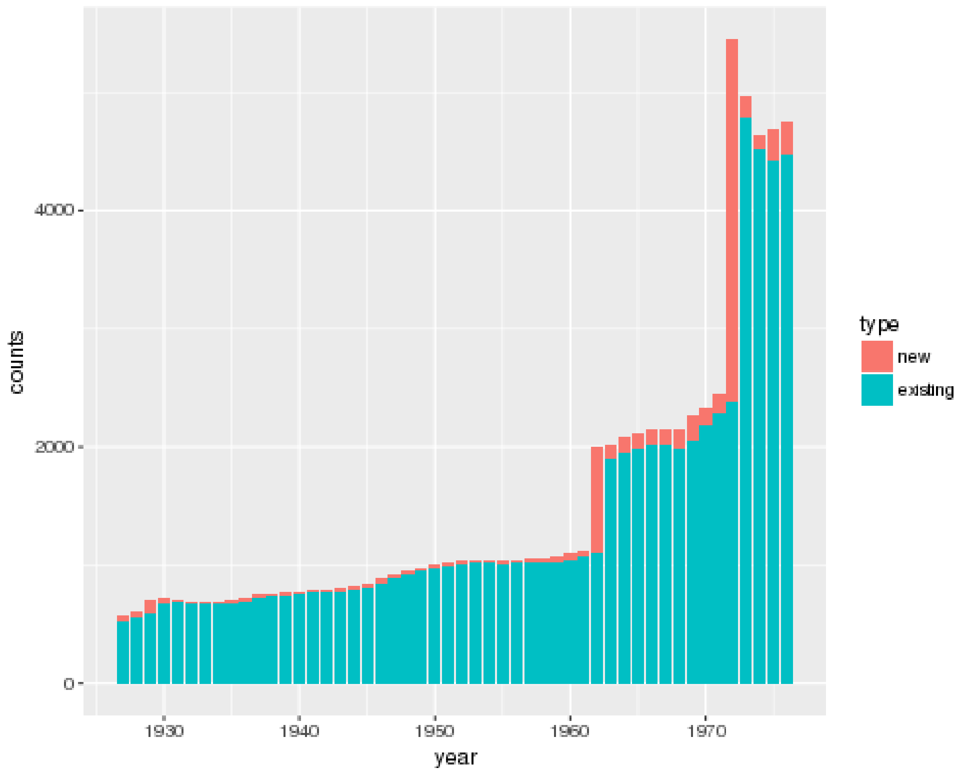


Table 4(b) reports the estimation and hypothesis testing results of the time effects model in equation (5.1). The model is fitted with stock-month data from July 1939 to December 1976. We use July 1939 as the starting time because prior to this month, the first quartile, median, and third quartile of the cross-sectional age distribution are not three distinct numbers so that we have difficulty in assigning stocks to the four age groups.

Regression over the full dataset yields a significant and positive coefficient that represents the age effect. However, compared with the coefficient estimate of 0.001 in Table 5.4, the magnitude of the age effect over the earlier years is much smaller. In addition, the t-test is much less significant according to the p-value. As for model fitting within each age group, the age effect is significant and positive in three out of the four age groups, but the p-values are relatively high compared with those in Table 5.4.

Table 3(b): Average age of 100-stock bootstrapped and rebalanced portfolios over different time periods. The average ages are calculated based on age of components of one thousand 100-stock bootstrapped and rebalanced portfolios held over the period from July 1926 to December 1976. We report the average age in both portfolios over each decade in the horizon of around 50 years.

Portfolio Type	1926 - 1936	1937 - 1946	1947 - 1956	1957 - 1966	1967 - 1976
Bootstrapped	4.49	11.98	17.20	17.54	11.98
Rebalanced	5.16	15.25	24.80	32.31	31.38

Table 4(b): Estimation and testing result of time effects model: July 1939 - December 1976. In the first row, the estimate, standard error, t-statistics, and the associated p-value for the overall time effects model are reported. In each month, the age group each stock belongs to is determined based on the cross-sectional ranking of the stock's current age. The breakpoints between age groups are the first quartile, median, and third quartile of the cross-sectional age distribution. All stock-month observations are divided into four age groups in this way. In each of the second to fifth rows, model fitting results for the indicated age group are reported.

Age Group	Estimate	Std. Error	t-value	p-value
All ages	0.0003	0.0001	2.5544	0.0106
Child	0.0036	0.0017	2.1480	0.0317
Youth	-0.0025	0.0018	-1.3851	0.1660
Adult	0.0020	0.0009	2.3186	0.0204
Senior	0.0007	0.0003	2.1954	0.0281

Figure A.2: Age distribution in 100-stock bootstrapped and rebalanced portfolios over different time periods. The distributions are based on age of components of one thousand 100-stock bootstrapped and rebalanced portfolios held over the period from July 1926 to December 1976. We report the empirical age distribution in both portfolios over each decade in the horizon of about 50 years.

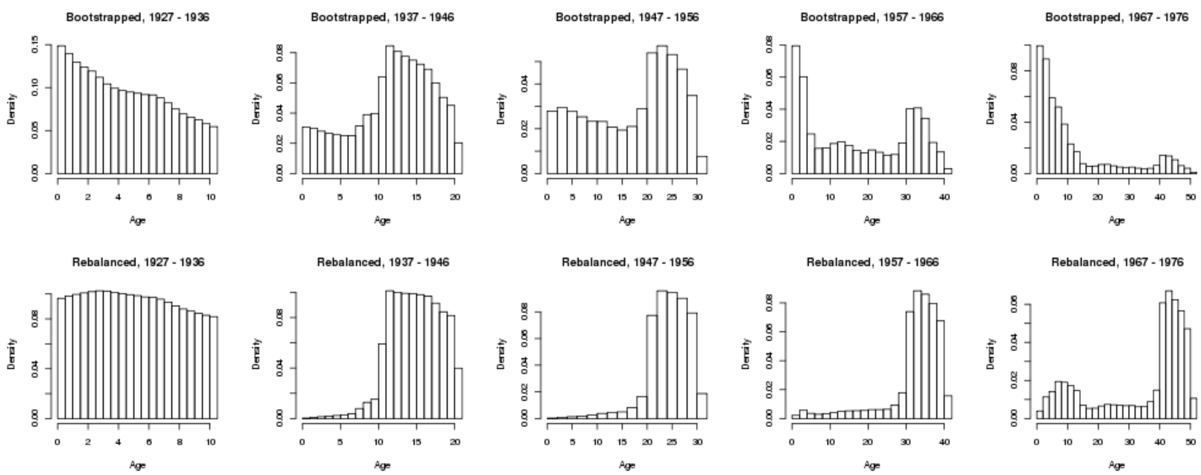


Table 5(b) is a replicate of Table 5.5 with data over the earlier years. Again, the starting time of the period over which portfolio returns are calculated is July 1939 because there are overlappings in the first quartile, median, and third quartile of the cross-sectional age distribution prior to this month. We can see how strong the small firm effect was during the earlier years in this table. Regardless of the age group, the Tiny group always outperforms the other three size groups in terms of return. Compared with the notable size effect the age effect is less visible. Among the four Tiny portfolios, the Youth-Tiny portfolio achieves the highest return, closely followed by Adult and Senior. Among the four Small portfolios or the four Medium portfolios, the Senior age group always has the best return. Among the four Big portfolios, the Adult group performs the best. Although the more senior age groups in general perform better the return difference between a more senior group and a less senior one is quite small compared with the return different between the Tiny portfolios and the rest ones.

Table 6(b) provides a decade-by-decade breakdown of the performance of the sixteen doubly sorted factor portfolios discussed earlier. To better visualize the magnitude of the age and size effects as well as the change in the magnitude across time, we perform the same merging procedure as what we did in Section 5.3. The decade-by-decade returns on these merged groups, as well as return differences between merged groups, are summarized

Table 5(b): Performance of sixteen doubly sorted equally weighted portfolios formed on age and size. Starting from July 1939, at the beginning of each month each available stock is assigned to one of sixteen factor portfolios based on its cross-sectional ranking of age and size. The breakpoints between adjacent age/size groups are the first quartile, median, and third quartile of the age/size distribution. Return on equally weighted portfolio of all stocks in each factor portfolio is calculated. All factor portfolios are held until December 1976. The annualized return, standard deviation, and Sharpe ratio of all sixteen portfolios are reported.

Age Group	Size Group	Return	Std Dev	Sharpe
Child	Tiny	18.60%	25.15%	0.74
	Small	12.42%	20.46%	0.61
	Medium	11.78%	18.34%	0.64
	Big	11.64%	15.56%	0.75
Youth	Tiny	20.82%	27.10%	0.77
	Small	15.10%	21.86%	0.69
	Medium	13.83%	18.77%	0.74
	Big	11.51%	15.92%	0.72
Adult	Tiny	20.60%	29.02%	0.71
	Small	13.93%	21.47%	0.65
	Medium	14.17%	19.23%	0.74
	Big	12.33%	15.53%	0.79
Senior	Tiny	20.31%	29.51%	0.69
	Small	16.02%	22.28%	0.72
	Medium	14.39%	19.48%	0.74
	Big	12.29%	15.47%	0.79

Table 6(b): Returns of sixteen doubly sorted equally weighted portfolios: by decade. Starting from July 1939, at the beginning of each month each available stock is assigned to one of sixteen factor portfolios based on its cross-sectional ranking of age and size. The breakpoints between adjacent age/size groups are the first quartile, median, and third quartile of the age/size distribution. Return on equally weighted portfolio of all stocks in each factor portfolio is calculated. All factor portfolios are held until December 1976. Annualized returns over the four non-overlapping decades of each factor portfolio are reported.

Jul 1939 - Dec 1946					Jan 1947 - Dec 1956				
	Tiny	Small	Medium	Big		Tiny	Small	Medium	Big
Child	37.66%	22.47%	18.76%	14.08%	Child	14.74%	13.74%	14.63%	14.53%
Youth	40.90%	25.02%	19.69%	14.69%	Youth	13.74%	15.53%	15.59%	15.34%
Adult	36.30%	20.61%	18.99%	42.97%	Adult	14.16%	13.01%	15.12%	18.10%
Senior	42.97%	26.13%	20.06%	14.13%	Senior	14.28%	15.00%	14.62%	15.15%
Jan 1957 - Dec 1966					Jan 1967 - Dec 1976				
	Tiny	Small	Medium	Big		Tiny	Small	Medium	Big
Child	11.44%	9.70%	8.59%	9.79%	Child	16.78%	6.42%	6.92%	7.35%
Youth	18.45%	12.16%	12.82%	10.37%	Youth	18.94%	10.15%	8.83%	6.36%
Adult	15.97%	13.11%	12.30%	10.56%	Adult	18.69%	11.65%	11.96%	7.04%
Senior	12.16%	12.39%	11.26%	10.24%	Senior	18.63%	13.08%	12.78%	10.11%

in Table 7(b). The SMJ return characterizes the magnitude of age effect, and the SMB return represents the size effect.

The average SMJ return over sixteen combinations of four size groups and four decades is 1.02%, much lower than the average SMJ return over the recent 40 years. The positive age effect despite its magnitude provides additional robustness to our main findings in Chapter 5. In contrast, the average SMB return over sixteen combinations of four age groups and four decades is 4.44%, much greater than that over the recent 40 years. The size effect clearly dominates the age effect in the earlier years.

It is comforting to see that in the decade from 1967 to 1976, the average SMJ across the four size groups is higher than that in the earlier three decades. This decade can be viewed as a period of transition after which the age effect becomes dominant. In addition, even in the earlier decades when the age effect is weak, the sign of SMJ is always positive. In contrast, the size effect peaks with an average SMB of 11.09% in the earliest period from 1939 to 1946 and becomes much weaker or even negative in the subsequent decades. Combining the results in Table 5.7 and Table 7(b), we conclude that the size effect is vanishing and highly variable, while the age effect is gradually become more dominating and robust with respect to time period.

Table 7(b): Return difference between portfolios formed on age and size: by decade. The sixteen factor portfolios are re-organized into eight by merging the youngest (smallest) two age (size) groups and the oldest (biggest) two age (size) groups with equal weights for each of the four size (age) groups. Annualized returns of the merged portfolios over each decade between July 1939 and December 1976 are reported. Confounding the size group, the return differences between the two merged age groups are reported as SMJ. Confounding the age group, the return differences between the two merged size groups are reported as SMB.

July 1939 - December 1946											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Child	Youth	Adult	Senior	Average
Child+Youth	39.28%	23.75%	19.22%	14.38%		Tiny+Small	30.06%	32.96%	28.45%	34.55%	
Adult+Senior	39.63%	23.37%	19.53%	28.55%		Medium+Big	16.42%	17.19%	30.98%	17.10%	
SMJ	0.35%	-0.38%	0.30%	14.17%	3.61%	SMB	13.64%	15.78%	-2.53%	17.45%	11.09%
January 1947 - December 1956											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Child	Youth	Adult	Senior	Average
Child+Youth	14.24%	14.63%	15.11%	14.93%		Tiny+Small	14.24%	14.64%	13.58%	14.64%	
Adult+Senior	14.22%	14.00%	14.87%	16.63%		Medium+Big	14.58%	15.46%	16.61%	14.89%	
SMJ	-0.02%	-0.63%	-0.24%	1.70%	0.20%	SMB	-0.34%	-0.83%	-3.03%	-0.24%	-1.11%
January 1957 - December 1966											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Child	Youth	Adult	Senior	Average
Child+Youth	14.94%	10.93%	10.71%	10.08%		Tiny+Small	10.57%	15.30%	14.54%	12.28%	
Adult+Senior	14.07%	12.75%	11.78%	10.40%		Medium+Big	9.19%	11.60%	11.43%	10.75%	
SMJ	-0.88%	1.82%	1.07%	0.32%	0.58%	SMB	1.38%	3.71%	3.11%	1.53%	2.43%
January 1967 - December 1976											
Age group	Tiny	Small	Medium	Big	Average	Size Group	Child	Youth	Adult	Senior	Average
Child+Youth	17.86%	8.28%	7.88%	6.85%		Tiny+Small	11.60%	14.54%	15.17%	15.85%	
Adult+Senior	18.66%	12.36%	12.37%	8.58%		Medium+Big	7.14%	7.60%	9.50%	11.45%	
SMJ	0.80%	4.08%	4.49%	1.72%	2.77%	SMB	4.46%	6.95%	5.67%	4.41%	5.37%
Average					1.79%						4.44%

A.4.2 Portfolio Performance in Presence of Negative Serial Correlation

In this section we suppose that monthly stock returns have negative serial correlation and examine the effects of this return dynamic on the expected return of the bootstrapped and rebalanced portfolios. We assume an AR(1) model to reflect the negative serial correlation in stock returns. Let r_{it} denote the simple rate of return of stock i over the tth month. The following AR(1) model with negative values of ϕ_i is able to capture the negative serial correlation.

$$(r_{it} - \mu_i) = \phi_i(r_{i(t-1)} - \mu_i) + \epsilon_{it}, \tag{A.14}$$

where $|\phi_i| < 1$ and ϵ_{it}'s are i.i.d. normal residuals with mean 0 and variance σ_i^2. Next we consider a fixed stock universe case where there are no new listings and delistings at all since the portfolio construction time as well as a dynamic stock universe case where in each month there are an equal number of new listings and delistings. We show that in both cases the negative serial correlation does not lead to outperformance of rebalanced portfolios (RP) over bootstrapped portfolios (BP).

Case 1: Fixed universe

Consider a fixed stock universe (no listings and delistings) of M stocks and a holding period of T months. The log holding period returns of our N-stock bootstrapped and rebalanced portfolios are:

$$\sum_{t=1}^{T} \log\left\{1 + \frac{1}{N}\sum_{i=1}^{M} \mathbb{I}_{\{i \text{ in BP in } t\text{th month}\}} r_{it}\right\} \tag{A.15}$$

and

$$\sum_{t=1}^{T} \log\left\{1 + \frac{1}{N}\sum_{i=1}^{M} \mathbb{I}_{\{i \text{ in RP in } t\text{th month}\}} r_{it}\right\} \tag{A.16}$$

respectively, where $\mathbb{I}_{\{A\}}$ is an indicator function valued at 1 if event A occurs and 0 otherwise. To show that the dynamic of r_{it} does not lead to a difference in expected portfolio returns, it suffices to show that for each $t = 1, 2, \ldots, T$,

$$E\log\left\{1 + \frac{1}{N}\sum_{i=1}^{M} \mathbb{I}_{\{i \text{ in BP in } t\text{th month}\}} r_{it}\right\} = E\log\left\{1 + \frac{1}{N}\sum_{i=1}^{M} \mathbb{I}_{\{i \text{ in RP in } t\text{th month}\}} r_{it}\right\}. \tag{A.17}$$

Since the stock picking is completely random, the random variables $\sum_{i=1}^{M} \mathbb{I}_{\{i \text{ in BP in } t\text{th month}\}} r_{it}$ and $\sum_{i=1}^{M} \mathbb{I}_{\{i \text{ in RP in } t\text{th month}\}} r_{it}$ are identically distributed for $t = 1, 2, \ldots, T$. Therefore equation (A.17) must hold, which means that negative serial correlation in stock return time series does not lead to any difference in the expected portfolio returns. Actually what is affected by the stock return dynamic is the variance of the portfolios returns.

Case 2: Dynamic universe

Now we move to the dynamic stock universe case. We assume for the sake of simplicity that there are always M available stocks in the investment universe. We further assume that at the beginning of each month $K(> N)$ out the M stocks are delisted from the market and the same number of new stocks are immediately added to the stock universe. In addition in each month all available stocks have an equal probability of being delisted. Let $\mathcal{M}_t$ denote the set of available stocks at the beginning of the tth month. The log holding period returns of the two portfolios are:

$$\sum_{t=1}^{T} \log \left\{ 1 + \frac{1}{N} \sum_{i \in \mathcal{M}_t} \mathbb{I}_{\{i \text{ in BP in } t\text{th month}\}} r_{it} \right\} \tag{A.18}$$

and

$$\sum_{t=1}^{T} \log \left\{ 1 + \frac{1}{N} \sum_{i \in \mathcal{M}_t} \mathbb{I}_{\{i \text{ in RP in } t\text{th month}\}} r_{it} \right\} \tag{A.19}$$

respectively. If the rebalanced portfolio seeks to exploit the negative serial correlation in stock return time series, it should pick stocks in a way such that $E[\mathbb{I}_{\{i \text{ in RP in } t\text{th month}\}}]$ is higher for a lower $r_{i(t-1)}$ and vice versa.

We examine the probability that a stock from the available stock pool at the beginning of the tth month is a component of the rebalanced portfolio over the tth month, i.e. $E[\mathbb{I}_{\{i \text{ in RP in } t\text{th month}\}}]$ for some $i \in \mathcal{M}_t$. If the stock i has been in the market before the portfolio construction date, i.e. $i \in \bigcap_{k=1}^{t} \mathcal{M}_k$,

$$
\begin{aligned}
&E\left[\mathbb{1}_{\{i \text{ in RP in } t\text{th month}\}} \middle| i \in \bigcap_{k=1}^{t} \mathcal{M}_k\right] \\
&= 1 - \Pr\{i \text{ not in RP in } t\text{th month}\} \\
&= 1 - \Pr\{i \text{ not in RP in 1st month}\} \\
&\quad \times \Pr\{i \text{ not in RP in 2nd month} | i \text{ not in RP in 1st month}\} \times \dots \\
&\quad \times \Pr\{i \text{ not in RP in } t\text{th month} | i \text{ not in RP in } (t-1)\text{th month}\} \\
&= 1 - \frac{M-N}{M} p^{t-1},
\end{aligned}
$$

where

$$
p = \sum_{k=0}^{N} \frac{\binom{K}{k}\binom{M-1-K}{N-k}}{\binom{M-1}{N}} \frac{M-N}{M-N+k}. \tag{A.20}
$$

Note that p is the conditional probability that a stock is not included in the rebalanced portfolio in the current month given that it was not in the rebalanced portfolio in the previous month. The k involved in the expression for p represents the number of delisted stocks within the rebalanced portfolio.

If the stock i first appears in the investment universe at the beginning of the sth month ($s > 1$), i.e., $i \notin \bigcup_{k=1}^{s-1} \mathcal{M}_k$ and $i \in \bigcap_{k=s}^{t} \mathcal{M}_k$,

$$
\begin{aligned}
&E\left[\mathbb{1}_{\{i \text{ in RP in } t\text{th month}\}} \middle| i \notin \bigcup_{k=1}^{s-1} \mathcal{M}_k, i \in \bigcap_{k=s}^{t} \mathcal{M}_k\right] \\
&= 1 - \Pr\{i \text{ not in RP in } t\text{th month}\} \\
&= 1 - \Pr\{i \text{ not in RP in } s\text{th month} | i \text{ not in RP in } (s-1)\text{st month}\} \times \dots \\
&\quad \times \Pr\{i \text{ not in RP in } t\text{th month} | i \text{ not in RP in } (t-1)\text{th month}\} \\
&= 1 - p^{t-s+1}
\end{aligned}
$$

where p is given in equation (A.20).

In summary we have

$$
E[\mathbb{1}_{\{i \text{ in RP in } t\text{th month}\}} | i \text{ was listed at } s] = \begin{cases} 1 - \frac{M-N}{M} p^{t-1} & \text{, if } s = 1 \\ 1 - p^{t-s+1} & \text{, if } 1 < s \le t \end{cases}. \tag{A.21}
$$

It is easy to check that $\frac{M-N}{M} < p$. Therefore $1-\frac{M-N}{M}p^{t-1} > 1-p^t$ and $E[\mathbb{1}_{\{i \text{ in RP in } t\text{th month}\}}]$ is decreasing in s (increasing in age of stock i). In contrast $E[\mathbb{1}_{\{i \text{ in BP in } t\text{th month}\}}]$ is equal to $\frac{N}{M}$ for $\forall i \in \mathcal{M}_t$. This leads us to conclude that the difference in stock picking scheme between the bootstrapped and rebalanced portfolios only comes through the age factor.