Estimation Risk and Optimal Combined Portfolio Strategies

by

Zhenzhen Huang

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

Waterloo, Ontario, Canada, 2024

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: **Dr. Marcos Escobar-Anel**
Professor, Department of Statistical and Actuarial Sciences
Western University

Supervisor(s): **Dr. Pengyu Wei**
Assistant Professor, Division of Banking and Finance
Nanyang Technological University

Dr. Chengguo Weng
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal Member: **Dr. David Saunders**
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Dr. Tony Wirjanto
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal-External Member: **Dr. Yuying Li**
Professor, School of Computer Science
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

Chapter 2 is based on the following published paper:

Huang, Z., Wei, P., & Weng, C. (2024). Tail mean-variance portfolio selection with estimation risk. Insurance: Mathematics and Economics 116, 218-234.

Chapter 4 is based on the following submitted paper:

Huang, Z., Wei, P., Weng, C., & Wirjanto, T. (2023). Winning probability weighted combined portfolio. Submitted.

Abstract

The traditional Mean-Variance (MV) framework of Markowitz (1952) has been the foundation of numerous research works for many years, benefiting from its mathematical tractability and intuitive clarity for investors. However, a significant limitation of this framework is its dependence on the mean vector and covariance matrix of asset returns, which are generally unknown and have to be estimated using historical data. The resulting plug-in portfolio, which uses these estimates instead of the true parameter values, often exhibits poor out-of-sample performance due to estimation risk. A considerable amount of research proposes various sophisticated estimators for these two unknown parameters or introduces portfolio constraints and regularizations. In this thesis, however, we focus on an alternative approach to mitigate estimation risk by utilizing combined portfolios and directly optimizing the expected out-of-sample performance. We review the relevant literature and present essential preliminary discussions in Chapter 1. Building on this, we introduce three distinct perspectives in portfolio selection, each aimed at assessing the efficiency of combined portfolios in managing estimation risk. These perspectives guide the detailed examination of research projects presented in the subsequent three chapters of the thesis.

Chapter 2 discusses the Tail Mean-Variance (TMV) portfolio selection with estimation risk. The TMV risk measure has emerged from the actuarial community as a criterion for risk management and portfolio selection, with a focus on extreme losses. The existing literature on portfolio optimization under the TMV criterion relies on the plug-in approach, which introduces estimation risk and leads to significant deterioration in the out-of-sample portfolio performance. To address this issue, we propose a combination of the plug-in and 1/N rules and optimize its expected out-of-sample performance. Our study is based on the Mean-Variance-Standard-deviation (MVS) performance measure, which encompasses the TMV, classical MV, and Mean-Standard-Deviation (MStD) as special cases. The MStD criterion is particularly relevant to mean-risk portfolio selection when risk is assessed using quantile-based risk measures. Our proposed combined portfolio consistently outperforms the plug-in MVS and 1/N portfolios in both simulated and real-world datasets.

Chapter 3 focuses on Environmental, Social, and Governance (ESG) investing with estimation risk taken into account. Recently, there has been a significant increase in the commitment of institutional investors to responsible investment. We explore an ESG constrained framework that integrates the ESG criteria into decision-making processes, aiming

to enhance risk-adjusted returns by ensuring that the total ESG score of the portfolio meets a specified target. The optimal ESG portfolio satisfies a three-fund separation. However, similar to the traditional MV portfolio, the practical application of the optimal ESG portfolio often encounters estimation risk. To mitigate estimation risk, we introduce a combined three-fund portfolio comprising components corresponding to the plug-in ESG portfolio, and we derive the optimal combination coefficients under the expected out-of-sample MV utility optimization, incorporating either an inequality or equality constraint on the expected total ESG score of the portfolio. Both simulation and empirical studies indicate that the implementable combined portfolio outperforms the plug-in ESG portfolio.

Chapter 4 introduces a novel Winning Probability Weighted (WPW) framework for constructing combined portfolios from any pair of constituent portfolios. This framework is centered around the concept of winning probability, which evaluates the likelihood that one constituent portfolio will outperform another in terms of out-of-sample returns. To ensure comparability, the constituent portfolios are adjusted to align with their long-term risk profiles. We utilize machine learning techniques that incorporate financial market factors alongside historical asset returns to estimate the winning probabilities, which then taken as the combination coefficients for the combined portfolio. Additionally, we optimize the expected out-of-sample MV utility of the combined portfolio to enhance its performance. Extensive empirical studies demonstrate the superiority of the proposed WPW approach over existing analytical methods in terms of certainty equivalent return across various scenarios.

Finally, Chapter 5 summarizes the thesis and outlines potential directions for further research.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisors, Dr. Chengguo Weng and Dr. Pengyu Wei, for their sustained support and exceptional guidance throughout my entire Ph.D. study. They introduced me to the field of actuarial science, shared their extensive expertise, and broadened my vision with various intriguing problems in the field. Their efforts, guidance, and encouragement have been indispensable for me to complete this thesis. I am profoundly fortunate to have been their student, and will always cherish the journey with them.

I would like to extend my sincere gratitude to all the committee members: Dr. Marcos Escobar-Anel (Western University), Dr. Yuying Li, Dr. David Saunders, and Dr. Tony Wirjanto for dedicating their valuable time to review my thesis and provide insightful feedback. Special thanks to Dr. Tony Wirjanto for sharing his expertise in our collaboration and for the enlightening course on sustainable finance he taught, which greatly inspired my research. I am also grateful to Dr. David Saunders and Dr. Tony Wirjanto for their valuable contributions as members of my thesis proposal committee. Additionally, I appreciate Dr. Marcos Escobar-Anel and Dr. Yuying Li for agreeing to join the committee. I am truly honored by their involvement.

I am deeply grateful to Dr. Yue Kuen Kwok (The Hong Kong University of Science and Technology) for his extensive support on my first research project when I had no prior research experience. His encouragement and guidance have been crucial in initiating my journey into Ph.D. research and have significantly shaped my development as a researcher. I also appreciate Dr. Shuxun Wang (Southern University of Science and Technology) for the opportunity to be a visiting student, where I spent a valuable and enriching period.

Many thanks to the faculty and staff in the Department of Statistics and Actuarial Science, from whom I have received invaluable supports and guidance. I am grateful to have met Dr. Phelim Boyle, Dr. Ben Feng, Dr. Lisa Gao, Dr. Mario Ghossoub, Dr. Mary Hardy, Dr. David Landriault, Dr. Bin Li, Dr. Fangda Liu, Dr. Ruodu Wang, Dr. Fan Yang. Their collective efforts have been instrumental in fostering a vibrant academic community of the department. Furthermore, I extend my appreciation to Mary Lou Dufton, Carlos Mendes, and Greg Preston for their continuing assistance with administrative and technical matters.

Throughout my Ph.D. journey, I have been fortunate to receive tremendous support from numerous colleagues and friends. Many thanks to my fellow Ph.D. colleagues for the enriching study sessions together. I wish them all the best in their future endeavors. I am grateful to my dear friends for their unwavering encouragement and care. Their companionship have been profoundly meaningful to me, especially during challenging times. They have brightened my world like the sun on a rainy day, bringing joy and laughter when I needed it most.

Finally, I am indebted to my family for their unconditional love throughout life.

Dedication

This is dedicated to my family.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Portfolio selection is a universally relevant and important topic across various sectors, which generally encompasses a strategic approach to choosing and managing a collection of investment assets that aligns optimally with the financial goals of institutions and individual investors. In the field of actuarial science, portfolio selection is widely applied to ensure financial stability and manage risks, with applications including asset-liability and pension-fund management. Advanced portfolio strategies enable institutions such as insurance companies enhance their solvency, secure financial health in the foreseeable future, and improve the well-being of policyholders. For instance, offering attractive equity-linked variable annuities and defined benefit pension plans that utilize advanced portfolio strategies can expand the retirement savings options available to policyholders. Meanwhile, individual investors, such as policyholders with defined contribution pension plans, also seek optimal portfolio strategies to augment wealth while managing risks, aiming to achieve long-term financial security and growth. Therefore, the study of portfolio selection is an integral part of actuarial research, significantly contributing to the enhancement of overall social welfare.

There is a substantial body of literature dedicated to the study of portfolio selection. The seminal work of Markowitz (1952) laid the foundation for Modern Portfolio Theory (MPT), which revolutionized portfolio construction and selection by introducing a comprehensive mathematical framework. MPT emphasizes the balance between reward and risk, which determines the optimal asset allocation by utilizing the mean and variance of portfolio returns as key metrics to evaluate reward and risk, respectively. It leads to the concept of the Mean-Variance (MV) efficient frontier, depicting a series of efficient portfolios that offer the highest expected return for each level of risk (variance). Positioned along the efficient frontier, investors are equipped to make investment decisions that align with their risk tolerance and return objectives. Over the years, investment models have continuously evolved to incorporate advanced insights while adhering to the foundational principles of MPT (Fabozzi et al., 2002). For instance, sophisticated portfolio strategies now frequently include downside risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), which evaluate potential losses in extreme market scenarios (e.g., Alexander and Baptista, 2004; Bodnar et al., 2012). Moreover, recent trends have seen the integration of Environmental, Social, and Governance (ESG) criteria into the portfolio construction process to align investments with ethical standards and sustainability objectives (see Pástor et al., 2021; Pedersen et al., 2021, for example).

While theoretically sound, the practical implementation of many theoretical portfolio strategies often encounters challenges due to reliance on the true parameters (i.e., the mean vector and covariance matrix) of asset return distributions, which are rarely fully known to investors. A common approach is to estimate the unknown parameters from historical data over a specific period, and treat these estimates as if they were the true parameters. The so-called plug-in portfolio is then constructed by substituting the estimated values for the unknown parameters in the optimal portfolio. However, since these estimates rarely match the true parameter values exactly, it may cause the plug-in portfolio to deviate significantly from the truly optimal one. Furthermore, the plug-in portfolio weights often suffer from large fluctuations over time when different data periods are used for estimation, as noted by Green and Hollifield (1992) and others. These deviations and fluctuations can lead to a significant deterioration in the out-of-sample performance of the portfolio, which is essential for assessing how it might perform under future conditions. This issue is known as the estimation risk problem.

Earlier studies have devoted considerable effort to mitigating estimation errors and enhancing out-of-sample performance, including works by Kalymon (1971), Klein and Bawa (1976), and Klein and Bawa (1977), among others. One of the key research directions has been the development of more effective estimators compared to traditional sample estimators for unknown parameters. Since the seminal work of Stein (1956) which demonstrated that shrinking the sample mean toward a constant could yield a smaller expected quadratic loss compared to the sample mean itself, there has been extensive literature on various shrinkage and related Bayesian estimators. For example, Jorion (1986, 1991) introduces a Bayes-Stein estimator for the unknown mean vector of asset returns. Ledoit and Wolf (2004, 2017) propose linear and nonlinear shrinkage estimators for the unknown covariance matrix of asset returns, offering both tractability and interpretability. Berger (2013) provides a comprehensive review from a Bayesian perspective. Additionally, further research has explored additional constraints to enhance out-of-sample performance, such as the no-short-sale constraint examined by Jagannathan and Ma (2003), which shows a shrinkage effect on the covariance matrix. DeMiguel et al. (2009a) include constraints on the norm of the portfolio weight vector.

Despite extensive efforts to mitigate estimation risk, DeMiguel et al. (2009b) conduct a comparative analysis of various sophisticated extensions of the MV rule versus the naive 1/N rule, where the 1/N rule involves allocating wealth equally among available assets at each rebalancing date, without relying on historical data or theoretical optimization analysis. Surprisingly, they discover that the 1/N rule can outperform many other portfolio strategies in most scenarios, and the estimation window length required for other portfolios to surpass the 1/N rule is exceptionally long (e.g., approximately 3000 months for a portfolio of 25 assets). Even though Kirby and Ostdiek (2012) point out that the MV optimization is inherently disadvantaged compared to the 1/N rule in the research design of DeMiguel et al. (2009b), the 1/N rule remains a compelling research target and is widely used as a benchmark for comparing portfolio performance. It is also observed that investors continue to follow the 1/N rule in practice (Benartzi and Thaler, 2001). Furthermore, the literature has explored the reasons behind the superior out-of-sample performance of the 1/N rule. Hwang et al. (2018) discover that the outperformance of the 1/N rule can be attributed to the compensation for the increased tail risk and the reduced upside potential associated with a more concave payoff. Guo et al. (2019) propose the 1/N favorability index and validate that the 1/N portfolio is more challenging to beat in bull markets. Yuan and Zhou (2022) explore why the 1/N rule is difficult to outperform in practice and examine conditions under which it might be surpassed.

In view of the superior out-of-sample performance of the 1/N rule, Tu and Zhou (2011) propose an innovative combination of the plug-in MV portfolio with the 1/N rule. The combined portfolio functions as a shrinkage estimator, with the 1/N rule serving as the target to which the plug-in MV portfolio is shrunk. The optimal combination coefficients are determined using the analytical framework introduced by Kan and Zhou (2007), which directly optimizes the expected out-of-sample MV utility under the normality assumption of asset returns. Their findings suggest that the optimal combined portfolios can deliver superior out-of-sample performance compared to the individual constituent portfolios across various scenarios.

Central to the work of Tu and Zhou (2011) is the concept of combined portfolios, which can be generalized to a weighted combination of any two or more prespecified constituent portfolios. We focus on combining ex-ante (rather than ex-post) strategies, which utilize the information available at a given time to construct the constituent portfolios. Since the ex-post method relies on information available after the given time, the ex-ante approach is the one available and utilized in this thesis. Given that each constituent portfolio contributes unique information sets and analytical approaches, the rationale for combining distinct portfolios is both economically and statistically sound (Nardari and Schüssler, 2023). From an economic perspective, combining these portfolios facilitates the incorporation of diverse information sets, thereby allowing for a wider range of information to be utilized in decision-making compared to relying solely on individual portfolios. Additionally, since each portfolio carries its own specific risks, investors will seek to diversify the idiosyncratic risks of individual portfolios through such combination. Moreover, investors with concave utility functions can generally find desirable combination coefficients such that the utility of the combined portfolio is maximized, ensuring that adopting the combined portfolio achieves a utility no less than that of the constituent portfolios. Statistically, combined portfolios often trace back to methodologies such as shrinkage estimation, Bayesian approaches, and robust optimization, as studied in Jorion (1986), Frost and Savarino (1986), Garlappi et al. (2007), Kan and Zhou (2007), DeMiguel et al. (2013), among others. The performance of combined portfolios heavily depends on the selection of combination coefficients. Therefore, it is crucial to carefully determine the combination coefficients to ensure the effectiveness of the combined portfolios.

Most of the literature on combined portfolios focuses on analyzing the behavior of estimators and determines the combination coefficients by maximizing expected out-of-sample performance under the normality assumption of asset returns, as demonstrated in the studies by Kan and Zhou (2007) and Tu and Zhou (2011). For example, DeMiguel et al. (2015) examine the optimal multi-period combined portfolio subject to quadratic transaction costs. Lassance (2021) considers the maximization of the expected out-of-sample Sharpe ratio. Kan et al. (2022) study the MV portfolio selection with estimation risk when there is no risk-free asset included in the portfolio. Lassance et al. (2023) investigate the combination of the sample MV portfolio with the 1/N rule and allow the combination coefficients not to necessarily sum to one. Kan and Wang (2023) propose a combined portfolio that optimally balances the value of including test assets and the effect of estimation errors. Lassance et al. (2024) discuss the risk of expected utility under parameter uncertainty and study a robustness measure that balances the mean and volatility of out-of-sample utility.

Such analytical methods are typically constrained to specific pairs of constituent portfolios due to the inherent complexity in deriving explicit expressions for the performance measures. An alternative approach to studying combined portfolios without relying on the normality assumption involves employing high-dimensional asymptotic regimes, see, for example, Bodnar et al. (2018) and Bodnar et al. (2022) which require only the existence of the first four moments, and Kan and Lassance (2024) that account for the impact of fat tails of return distributions. Additionally, DeMiguel et al. (2013) employ a bootstrap approach to obtain the optimal combination coefficients for the combined portfolios under different criteria. Jiang et al. (2019) introduce a regression analysis to find the optimally combined coefficient and identify several factors that impact estimation errors. Other approaches for portfolio combination that require mild or no distribution assumptions include algorithms to minimize regret under uncertainty (Chakrabarti, 2021), the ensemble framework (Nardari and Schüssler, 2023), and the bagged pretested method (Kazak and Pohlmeier, 2023), among others.

Theoretically, if the optimal combination coefficients were known, combining multiple constituent portfolio strategies would dominate combining any subset of them. However, these optimal combination coefficients are generally unknown and must be estimated. When we increase the number of constituent portfolios in the combined portfolio, it introduces additional combination coefficients that need to be estimated. The estimation errors associated with these coefficients can outweigh the benefits of including more constituent portfolios. Therefore, most existing literature focuses on the combined portfolio with a limited number of constituent portfolios, typically two or three. This thesis similarly considers such a limited number of constituent portfolios.

1.2 Preliminaries

This section provides a concise overview of key concepts that are central to the thesis, specifically focusing on estimation risk in MV portfolios and the strategy of portfolio combination. These concepts will recur throughout the thesis.

1.2.1 Estimation Risk

Consider an investor who allocates wealth among N risky assets and a risk-free asset. Let $\boldsymbol{R} = [R_1, ..., R_N]^\top$ denote the N-dimensional return vector for the risky assets, where R_i is the rate of return for asset i. The rate of return for the risk-free asset is denoted by R_f. The excess return vector for the N risky assets, relative to the risk-free asset, is computed as $\boldsymbol{r} = \boldsymbol{R} - R_f\mathbf{1}$, where $\mathbf{1}$ is the N-dimensional vector of ones. Additionally, define $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as the N-dimensional mean vector and $N \times N$ covariance matrix of the full rank for the excess return vector $\boldsymbol{r}$. The elements of the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are respectively given by

$$\mu_i := \mathbb{E}[r_i] \quad \text{and} \quad \Sigma_{ij} := \mathbb{E}[(r_i - \mu_i)(r_j - \mu_j)], \quad i, j = 1, 2, ..., N,$$

where r_i is the excess return for asset i in the vector $\boldsymbol{r}$. A portfolio strategy refers to a vector of investment weights (or fractions) $\boldsymbol{w} = [w_1, ..., w_N]^\top$, where w_i represents the proportion of wealth invested in the risky asset i, $i = 1, \ldots, N$. The excess return of the portfolio $\boldsymbol{w}$ is calculated as $r_{\boldsymbol{w}} = \boldsymbol{w}^\top \boldsymbol{r}$, with its mean and variance given by

$$\mu_{\boldsymbol{w}} = \boldsymbol{w}^\top \boldsymbol{\mu} \quad \text{and} \quad \sigma^2_{\boldsymbol{w}} = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}.$$

In the traditional MV portfolio theory (Markowitz, 1952), the investor aims to find the optimal portfolio strategy $\boldsymbol{w}$, which balances the trade-off between the mean and variance of the portfolio according to a MV objective. We consider the following two distinct cases: the optimal MV portfolio that includes the risk-free asset, and the optimal constrained MV portfolio which excludes the risk-free asset.

Case 1: The Optimal MV Portfolio

In the scenario where the risk-free asset is included in the portfolio, the weight allocated to it is calculated as $1-\sum_{i=1}^{N} w_i$. The classical MV portfolio problem which aims to maximize the MV objective is formulated as follows:

$$\max_{\boldsymbol{w}} \quad \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \tag{1.1}$$

where γ is the risk aversion coefficient of the decision-maker. The theoretically optimal portfolio for this optimization problem (1.1) is determined by:

$$\boldsymbol{w}_{mv}^* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.$$

Case 2: The Optimal Constrained MV Portfolio

In the scenario where the risk-free asset is not included in the portfolio, we can set $R_f = 0$, which makes $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ the mean vector and covariance matrix of the asset returns $\boldsymbol{R}$, respectively. In this configuration, the sum of the portfolio weights equals one, leading to the following feasible set for the portfolio:

$$\mathcal{W} := \{\boldsymbol{w} \in \mathbb{R}^N | \boldsymbol{w}^\top \mathbf{1} = 1\}.$$

The classical MV portfolio problem that adheres to this feasible set with a constraint on the portfolio weights can be formulated as follows:

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \\ st. \quad & \boldsymbol{w}^\top \mathbf{1} = 1. \end{aligned} \tag{1.2}$$

The theoretically optimal solution for this constrained optimization problem (1.2) is obtained by

$$\boldsymbol{w}_{mr}^* = \boldsymbol{w}_{gmv} + \frac{1}{\gamma}\boldsymbol{w}_z,$$

where $\boldsymbol{w}_{gmv} = \boldsymbol{\Sigma}^{-1}\mathbf{1}/(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})$ is the global minimum variance portfolio with an expected return $\mu_{\boldsymbol{w}_{gmv}} = \boldsymbol{w}_{gmv}^\top\boldsymbol{\mu}$, and $\boldsymbol{w}_z = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_{\boldsymbol{w}_{gmv}}\mathbf{1})$ is the zero investment portfolio that ensures $\boldsymbol{w}_z^\top\mathbf{1} = 0$.

However, the theoretically optimal portfolios $\boldsymbol{w}_{mv}$ and $\boldsymbol{w}_{mr}$ are not directly implementable in practice because the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of asset returns are unknown to investors. A naive approach is to replace these parameters with their sample counterparts in the optimal portfolios, resulting in what are known as plug-in portfolios. Specifically, given the historical excess returns of the N risky assets over T periods, denoted as $\{\boldsymbol{r}_1, ...\boldsymbol{r}_T\}$, we compute the sample estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \qquad \text{and} \qquad \widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})^\top.$$

We then substitute the unknown parameters in the optimal portfolios with these estimators. The corresponding plug-in MV portfolios for $\boldsymbol{w}_{mv}$ and $\boldsymbol{w}_{mr}$ are respectively obtained by:

$$\begin{aligned}\hat{\boldsymbol{w}}_{mv} &= \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}},\\ \hat{\boldsymbol{w}}_{mr} &= \hat{\boldsymbol{w}}_{gmv} + \frac{1}{\gamma}\hat{\boldsymbol{w}}_{z},\end{aligned}$$

where $\hat{\boldsymbol{w}}_{gmv} = \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}/(\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1})$ and $\hat{\boldsymbol{w}}_z = \widehat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \mu_{\hat{\boldsymbol{w}}_{gmv}}\mathbf{1})$.

Even though the plug-in approach offers portfolio strategies that are implementable in practice, these plug-in portfolios often exhibit notoriously poor performance during the out-of-sample period. Here, the out-of-sample period is defined as the future time frame that is distinct from the period of historical returns used to develop the plug-in portfolios. In the context of this thesis that considers a one-period investment model rather than a multi-period investment portfolio, the out-of-sample period specifically refers to the single period following the portfolio construction. We use out-of-sample performance to evaluate the effectiveness of a portfolio strategy when applied to the out-of-sample period. Denote the asset excess return and the return of the risk-free asset in the subsequent time $T+1$ as $\boldsymbol{r}_{T+1}$ and $r_{f,T+1}$, respectively. The out-of-sample portfolio return for any plug-in portfolio strategy $\hat{\boldsymbol{w}}$, derived from historical information available up to time T and applied to the subsequent period $T+1$, is given by:

$$r_{\hat{\boldsymbol{w}}} = \hat{\boldsymbol{w}}^\top\boldsymbol{r}_{T+1} + r_{f,T+1}.$$

Assume that the asset returns in the out-of-sample period have the same mean vector $\boldsymbol{\mu}$

and covariance matrix $\boldsymbol{\Sigma}$ as those in the historical period. The mean and variance of the out-of-sample portfolio return $r_{\hat{\boldsymbol{w}}}$, referred to as the out-of-sample mean and out-of-sample variance of the portfolio $\hat{\boldsymbol{w}}$, are respectively calculated by:

$$\mu_{\hat{\boldsymbol{w}}} = \hat{\boldsymbol{w}}^\top \boldsymbol{\mu} \quad \text{and} \quad \sigma^2_{\hat{\boldsymbol{w}}} = \hat{\boldsymbol{w}}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}.$$

Since the portfolio $\hat{\boldsymbol{w}}$ is determined based on historical asset returns, $\hat{\boldsymbol{w}}$ is random rather than fixed. Consequently, both $\mu_{\hat{\boldsymbol{w}}}$ and $\sigma^2_{\hat{\boldsymbol{w}}}$ are also random variables. When we take the randomness of $\hat{\boldsymbol{w}}$ into account, we define the expected out-of-sample mean and expected out-of-sample variance as follows:

$$\mathbb{E}[\mu_{\hat{\boldsymbol{w}}}] = \mathbb{E}[\hat{\boldsymbol{w}}^\top \boldsymbol{\mu}] \quad \text{and} \quad \mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}}] = \mathbb{E}[\hat{\boldsymbol{w}}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}].$$

Additionally, in empirical studies where the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of asset returns are not directly obtainable, we commonly consider a rolling window approach to evaluate the out-of-sample performance. Assume there are M rolling windows in total. Each window yields an out-of-sample portfolio return as $r_{\hat{\boldsymbol{w}}_i} = \hat{\boldsymbol{w}}_i^\top \boldsymbol{r}_{i+1} + r_{f,i+1}, i = 1, 2, ..., M$, where $\hat{\boldsymbol{w}}_i$ is the portfolio derived based on information available up to time i and $\boldsymbol{r}_{i+1}, r_{f,i+1}$ are the asset excess return vector and the return of the risk-free asset in the subsequent time $i+1$. To estimate the expected out-of-sample mean and variance for the portfolio $\hat{\boldsymbol{w}}$, we calculate the sample mean and sample variance of these out-of-sample returns as follows:

$$\hat{\mu}_{r_{\hat{\boldsymbol{w}}}} = \frac{1}{M} \sum_{i=1}^{M} r_{\hat{\boldsymbol{w}}_i} \quad \text{and} \quad \hat{\sigma}^2_{r_{\hat{\boldsymbol{w}}}} = \frac{1}{M} \sum_{i=1}^{M} (r_{\hat{\boldsymbol{w}}_i} - \hat{\mu}_{\hat{\boldsymbol{w}}})^2.$$

The underperformance of plug-in portfolios is primarily attributed to estimation risks. In this thesis, estimation risks refer to the uncertainty that arises from estimating unknown parameters necessary for portfolio strategies, namely, the mean vector and covariance matrix of asset returns. Changes in the sample data used for estimating unknown parameters could lead to significant fluctuations in the estimates, thereby affecting the stability and reliability of portfolio strategies. Notably, the sample mean vector is a very noisy estimator of $\boldsymbol{\mu}$, which typically requires extensive historical asset returns to achieve an accurate estimate (see Merton, 1980, for example). Additionally, Best and Grauer (1991) demonstrate that the MV portfolio weights can be extremely sensitive to changes in the mean vector.

1.2.2 Combined Portfolios

In this thesis, we mitigate the impact of estimation risk by concentrating on a widely recognized category of investment strategies known as combined portfolios. Each combined portfolio consists of a weighted combination of any two or more prespecified constituent portfolios. For illustrative purposes, a combined portfolio incorporating two distinct plug-in constituent portfolios can be formulated as:

$$\hat{\boldsymbol{w}}_c = \delta_1 \hat{\boldsymbol{w}}_1 + \delta_2 \hat{\boldsymbol{w}}_2, \tag{1.3}$$

where δ_i is the constant combination coefficient associated with the plug-in constituent portfolio $\hat{\boldsymbol{w}}_i, i = 1, 2$. Constituent portfolios can be flexibly selected from a range of strategies documented in both academic research and industry practices, which include but are not limited to the traditional MV portfolios, Bayesian portfolio rules (e.g., Jorion, 1986; Bodnar et al., 2017), portfolios that incorporate constraints or regularization (e.g., Jagannathan and Ma, 2003; DeMiguel et al., 2009a), and the 1/N rule which allocates wealth uniformly across all available assets (DeMiguel et al., 2009b).

As the plug-in constituent portfolios are derived from historical asset returns, the combined portfolio is similarly a function of the historical returns. When implementing the combined portfolio, the corresponding out-of-sample mean and variance are respectively defined as follows:

$$\mu_{\hat{\boldsymbol{w}}_c} = \hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu} = \delta_1 \hat{\boldsymbol{w}}_1^\top \boldsymbol{\mu} + \delta_2 \hat{\boldsymbol{w}}_2^\top \boldsymbol{\mu},$$
$$\sigma^2_{\hat{\boldsymbol{w}}_c} = \hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c = \delta_1^2 \hat{\boldsymbol{w}}_1^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_1 + \delta_2^2 \hat{\boldsymbol{w}}_2^\top \boldsymbol{\Sigma} \boldsymbol{w}_e + 2\delta_1\delta_2 \hat{\boldsymbol{w}}_1^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_2.$$

Note that due to the reliance on historical data within $\hat{\boldsymbol{w}}_c$, which inherently contains uncertainties, both $\mu_{\hat{\boldsymbol{w}}_c}$ and $\sigma^2_{\hat{\boldsymbol{w}}_c}$ are considered random variables. Kan and Zhou (2007) suggest the evaluation of the out-of-sample performance of the portfolio by aligning it with the MV objective function. For the combined portfolio, the out-of-sample MV utility is thus expressed as

$$U(\hat{\boldsymbol{w}}_c) = \hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu} - \frac{\gamma}{2} \hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c.$$

Given that the out-of-sample MV utility is a random variable, it is natural to assess the portfolio based on its expected out-of-sample performance, which is calculated as follows:

$$\mathbb{E}[U(\hat{\boldsymbol{w}}_c)] = \mathbb{E}\left[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}\right] - \frac{\gamma}{2}\mathbb{E}\left[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c\right].$$

The expected out-of-sample MV utility represents the average out-of-sample performance an investor can anticipate achieving under parameter uncertainty when following the portfolio rule $\hat{\boldsymbol{w}}_c$ repeatedly. Particularly, the expected out-of-sample mean and variance of the combined portfolio are calculated as $\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}]$ and $\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c]$, respectively.

Generally, the effectiveness of a combined portfolio depends the determination of the combination coefficients. A prevalent assumption in the literature for deriving the optimal combination coefficients is that asset returns are independent and follow a multivariate normal distribution. Under this assumption, many studies achieve the optimal combined portfolios for various pairs of constituent portfolios by maximizing the expected out-of-sample MV utility, including Kan and Zhou (2007), Tu and Zhou (2011), Kan et al. (2022), Lassance et al. (2023), among others.

1.3 Motivation and Outline

Given the growing interest in estimation risk and optimal combined portfolio strategies in recent academic literature, this thesis aims to explore these topics from the following three key perspectives:

- Tail Mean-Variance (TMV) portfolio selection;
- Environmental, Social, and Governance (ESG) investing;
- Development of a flexible framework for constructing combined portfolios.

The subsequent chapters of this thesis are dedicated to addressing these three critical aspects.

In Chapter 2 (Tail Mean-Variance Portfolio Selection with Estimation Risk), we focus on the TMV risk measure, which has emerged from the actuarial community as a critical criterion for risk management and portfolio selection, particularly emphasizing extreme losses or tail risks. Here, tail risks refer to the risks associated with tail events, which are characterized by their low probability of occurrence but potentially severe consequences on financial markets and institutions. For example, from the late 1980s to the early 2010s, there were at least seven episodes, such as the subprime mortgage crisis, that qualify as tail events. The existing literature on portfolio optimization under the TMV criterion relies on the plug-in approach which is known to introduce estimation risk. This issue has similarly discussed in Chapter 1.2.1.

We first introduce a Mean-Variance-Standard-deviation (MVS) performance measure, which encompasses the TMV, classical MV, and Mean-Standard-Deviation as special cases. A weight constraint is applied to ensure that the total allocation across all risky assets in the portfolio adds up to 100% in this chapter. Additionally, scenarios that include a risk-free asset can be similarly analyzed by optimizing the MVS measure without the weight constraint, which often simplifies the analysis. The optimal MVS portfolio satisfies a two-fund separation. However, similar to the traditional MV portfolio, the practical application of the optimal MVS portfolio encounters estimation risk. To mitigate it, we propose a combined three-fund portfolio with three constituent portfolios: the 1/N portfolio, and the other two constituent portfolios corresponding to the components found in the plug-in MVS portfolio. Importantly, this proposed combined portfolio satisfies the weight constraint, ensuring consistency with the MVS optimization that excludes the risk-free asset. Under the multivariate normal distribution assumption of asset returns, we analytically derive the optimal combination coefficients that maximize the expected out-of-sample MVS performance criterion. For practical implementation, we demonstrate how sophisticated estimators and a shrinkage covariance matrix can be used to estimate the optimal combination coefficients and achieve an implementable combined three-fund portfolio. Numerical studies demonstrate the superior performance of the proposed combined three-fund portfolio in terms of expected out-of-sample MVS utilities, which is robust to variations in the estimation window length, number of assets, and datasets.

In Chapter 3 (ESG Investing with Estimation Risk), we explore the rapidly expanding field of ESG investing, which has garnered significant attention as more investors commit to responsible investing practices over the past decade. ESG investing emphasizes the effective management of environmental, social, and governance risks to support long-term investment sustainability. The incorporation of ESG criteria has recently become prevalent in actuarial products, such as the Canada Post Corporation Registered Pension Plan as detailed in its 2022 Annual Report.[1] Despite the rapid evolution and growing importance of ESG investing, there remains a notable gap in the literature concerning the associated estimation risks. Consequently, this chapter aims to address this gap by exploring estimation risks within ESG investing frameworks. It is important to note that the estimation risk here refers to the uncertainty arising from estimating the mean vector and covariance matrix of asset returns, as similarly discussed in Chapter 1.2.1. Although ESG scores of assets are known to be ambiguous, with different agents providing varying scores for the same stock, this chapter focuses on the scenario where the ESG score is considered constant. Specifically, we assume that investors have chosen their preferred ESG rating agency and consistently utilize their scores.

We first explore an ESG-constrained optimization framework that incorporates ESG criteria into the decision-making process. Here, the objective is to maximize the MV utility while ensuring that the total ESG score of the portfolio meets a specified target. We exclude the risk-free asset from our discussion, as there are no commonly adopted ESG scores for risk-free assets. The resulting optimal ESG portfolio exhibits a three-fund separation, where each component meets specific weight and ESG criteria. To address the estimation risk associated with the practical application of the optimal ESG portfolio, we propose a combined portfolio in which the constituent portfolios correspond to the components identified in the plug-in ESG portfolio. Under the multivariate normal distribution assumption of asset returns, we utilize the expected out-of-sample MV utility as the optimization criterion, along with either an inequality or equality constraint on the expected total ESG score. We analytically derive the optimal combination coefficients, and explore sophisticated estimators to facilitate the estimation of these combination coefficients. Our numerical results demonstrate that the proposed combined portfolio consistently outperforms the plug-in ESG portfolio in terms of expected out-of-sample MV utility by a significant margin.

In Chapter 4 (Winning Probability Weighted Combined Portfolio), we pivot to explore a flexible statistical framework designed to construct desirable combination coefficients for any combined portfolios. Previous studies have often encountered challenges in determin-

[1] The 2022 Annual Report of Canada Post highlights that: "The Canada Post Corporation Registered Pension Plan integrates ESG principles into its investment strategy, while maintaining focus on returns for employees and pensioners". The full report is available at: https://www.canadapost-postescanada.ca/cpc/en/our-company/financial-and-sustainability-reports/2022-annual-report/social-and-environmental.page

ing combination coefficients analytically, and the analytical approach is typically applicable only to a limited range of combined portfolios. The difficulty generally arises from the complexity involved in deriving explicit expressions for the expected out-of-sample performance measures under the multivariate normal distribution assumption of asset returns. Consequently, this chapter focuses on developing a more flexible method to determine combination coefficients by moving beyond the normality assumption and employing statistical learning techniques.

In order to enhance out-of-sample performance and offer greater flexibility, we introduce a novel Winning Probability Weighted (WPW) framework for constructing combined portfolios. Unlike the aforementioned studies, we assume that each constituent portfolio exhibits weakly stationary out-of-sample returns, a premise commonly accepted in the literature (e.g., Györfi et al., 2008) and empirically validated in our numerical analysis. The WPW framework is founded on the innovative concept of characterizing combination coefficients as winning probabilities, which are the likelihoods of one constituent portfolio outperforming the other in the out-of-sample scenarios. This innovation lays the groundwork for employing various machine learning methods to determine the combination coefficients, further allowing the incorporation of exogenous factors, such as cross-sectional financial market factors, with traditional historical returns in the portfolio construction process. In the WPW framework, we first adjust the constituent portfolios to have the same long-term variance so that they are comparable in long-term risk profile, and then we apply classification models to estimate the winning probabilities which are taken as the combination coefficients. Finally, we further enhance the combined portfolio by scaling it when targeting the expected out-of-sample MV utility optimization. Extensive empirical studies demonstrate the superiority of the WPW approach over existing analytical methods and the corresponding constituent portfolios in terms of certainty equivalent return across different scenarios.

In Chapter 5, we conclude the thesis and raise several open questions to be addressed in future research. All proofs and supplementary materials for Chapters 2-4 are compiled in Appendices A-C.

Finally, it is important to note that Chapters 2-4 are structured to stand independently, as each chapter contains the main material of a distinct research paper. While we have made concerted efforts to maintain consistent notation throughout this thesis, some inconsistencies may remain. Therefore, we encourage readers to consider the notation used in each chapter independently.

Chapter 2

Tail Mean-Variance Portfolio Selection with Estimation Risk

2.1 Introduction

Because investment and insurance portfolios are sensitive to extreme losses, there has been a growing interest in risk measures that focus on the tails of distributions. In particular, the Tail Conditional Expectation (TCE) and Tail Variance (TV), as extensions of the classical mean and variance, have attracted the attention of actuaries. The TCE, also known as the Expected Shortfall (ES) or Conditional Value-at-Risk (CVaR), was initially proposed by Artzner et al. (1999) to measure the expectation of losses that exceed a particular Value-at-Risk (VaR), and has become a popular risk measure in actuarial science. There is extensive literature about the TCE measure for various loss distributions, including work by Panjer (2002) for multivariate normal distributions, Vernic (2006) for multivariate skew-normal distributions, and Landsman et al. (2016) for elliptical distributions, among others. The TV, proposed by Furman and Landsman (2006), measures the variability of losses along the tail of a loss distribution. It has been developed by a series of papers including Landsman et al. (2013) for log-elliptical distributions, Ignatieva and Landsman (2015) for the symmetric generalised hyperbolic family, and Eini and Khaloozadeh (2022*b*) for generalized skew-elliptical distributions, among others.

Landsman (2010) proposes the Tail Mean-Variance (TMV) risk measure, which is a weighted sum of the TCE and TV of portfolio losses, as a criterion for portfolio selection. Unlike the classical Mean-Variance (MV) model, the TMV criterion focuses on the tail of a loss distribution and is thus more sensitive to extreme losses. There is now a wealth of literature on TMV models. For instance, Owadally and Landsman (2013) derive the optimal portfolio under the TMV criterion when the asset return vector follows a multivariate elliptical distribution. Xu and Mao (2013) examine capital allocation problems based on the TMV principle. Landsman and Makov (2016) investigate the minimization of a function of a quadratic functional, motivated by a portfolio selection problem under the TMV model. Jiang et al. (2016) analyze the optimization of the TMV portfolio under the generalized Laplace distribution. Landsman et al. (2020) study the maximization of a general function involving the mean and variance of portfolio returns, for which the TMV model can be considered as a special example. More recently, Eini and Khaloozadeh (2021*a*) obtain the explicit solution to the TMV portfolio optimization problem under generalized skew-elliptical distributions.

The bulk of the literature on the TMV portfolio assumes that the true mean vector and covariance matrix of asset returns are known. However, in practice, these true parameters are rarely known and have to be estimated, which face estimation risk as we introduced in Chapter 1. This chapter aims to explore the optimal TMV portfolio selection problem while taking into account estimation risk. In line with the aforementioned literature, we assume that the return vector of the risky assets follows a multivariate normal distribution. We also assume that investors aim to maximize a proposed Mean-Variance-Standard-deviation (MVS) preference, which is a combination of the portfolio's mean, variance, and standard deviation. The proposed MVS objective nests the classical MV, TMV, and Mean-Standard-Deviation (MStD) as special cases. The MStD criterion is particularly relevant to mean-risk portfolio selection when risk is measured by quantile-based risk measures. As one of the primary contributions, given the true mean vector and covariance matrix of the asset returns, we derive the optimal MVS portfolio characterized by a two-fund separation including the global minimum variance portfolio and the zero investment portfolio.

We then mitigate estimation risk in the MVS portfolio selection problem. Given that the mean vector and the covariance matrix have to be estimated from historical data, the resulting plug-in MVS portfolio is a function of the samples and thus is random. Therefore, we focus on the expected out-of-sample MVS utility as the optimization criterion for the derivation of the portfolio weights. Inspired by Tu and Zhou (2011), we propose a combined three-fund rule to mitigate estimation risk. The three-fund rule consists of the 1/N portfolio and two other plug-in components from the optimal MVS portfolio derived without considering estimation risk: specifically, one from the global minimum variance portfolio and the other from the zero investment portfolio. We derive the optimal combination coefficients that maximize the approximated out-of-sample MVS utility, which is the first-order Taylor series expansion of the expected out-of-sample MVS utility. The true optimal combination coefficients involve unknown parameters that have to be estimated for practical use. We illustrate how sophisticated estimators and a shrinkage covariance matrix can be used to estimate these coefficients and obtain an implementable combined three-fund portfolio. We test the performance of our proposed portfolio strategies with both simulated and real-world datasets. Our results demonstrate that the proposed combined three-fund portfolio consistently outperforms both the plug-in MVS and 1/N portfolios by a significant margin in terms of the expected MVS utilities. Additionally, the out-of-sample performance of our proposed portfolio is robust to variations in the estimation window length, number of assets, and different datasets.

The rest of the chapter is organized as follows. Section 2.2 discusses the Tail Mean-Variance criterion and its connection with the Mean-Variance-Standard-deviation criterion. Section 2.3 introduces the estimation risk and provides the combination rules. Section 2.4 derives the optimal combined portfolios. Section 2.5 illustrates the performance of combined portfolios based on simulation and empirical studies. All proofs are placed in Appendix A.

2.2 Mean-Risk Portfolio Optimization

Consider an investor who allocates wealth among $N \geq 2$ risky assets. Let $\boldsymbol{r} = [r_1, ..., r_N]^\top$ denote the asset return vector, where r_i represents the return on asset i, $i = 1, \ldots, N$. Assume that the asset return vector follows a multivariate normal distribution with an N-dimensional mean vector $\boldsymbol{\mu} \in \mathbb{R}^N$ and a covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$. Assume that $\boldsymbol{\mu}$ is not collinear with $\mathbf{1}$, where $\mathbf{1}$ is an N-dimensional vector of ones, and that $\boldsymbol{\Sigma}$ is a symmetric positive definite and non-singular matrix. A portfolio strategy is a vector of investment weights (or fractions) $\boldsymbol{w} = [w_1, ..., w_N]^\top$, where w_i represents the proportion of wealth invested in asset i, $i = 1, \ldots, N$. The set of feasible portfolios considered in this chapter is defined as

$$\mathcal{P} = \{\boldsymbol{w} \in \mathbb{R}^N : \boldsymbol{w}^\top \mathbf{1} = 1\},$$

We denote the portfolio return as $r_{\boldsymbol{w}} = \boldsymbol{w}^\top \boldsymbol{r}$ and the corresponding portfolio loss as $L = -r_{\boldsymbol{w}}$ with a cumulative distribution function $F_L(\cdot)$.

Landsman (2010) proposes Tail Mean-Variance (TMV) as a risk measure for the portfolio loss L defined as follows:

$$\text{TMV}_q(L) = \mathbb{E}[L|L > \text{VaR}_q(L)] + \lambda \text{Var}[L|L > \text{VaR}_q(L)],$$

where $\lambda > 0$ is the weighting parameter, $q \in (0, 1)$ is the confidence level, and $\text{VaR}_q(L)$ is the Value-at-Risk of the portfolio loss L at the confidence level q, i.e.,

$$\text{VaR}_q(L) = \inf\{x \in \mathbb{R} : F_L(x) \geq q\}.$$

The TMV risk measure is a weighted combination of the Tail Conditional Expectation (TCE) and Tail Variance (TV). The TCE, also known as the Expected Shortfall or Conditional Value-at-Risk, was initially proposed by Artzner et al. (1999) and is calculated by $\mathbb{E}[L|L > \text{VaR}_q(L)]$. It can be interpreted as the expectation of the loss that exceeds a particular Value-at-Risk, providing information about the average on the tail part of the loss distribution. The TV, proposed by Furman and Landsman (2006) and calculated by $\text{Var}[L|L > \text{VaR}_q(L)]$, measures the deviation of the tail loss from the TCE. Under the TMV criterion, the investor focuses on the tail of the loss distribution and balances between the TCE and TV of portfolio losses.

Prior work studies the TMV portfolio selection under different distribution assumptions of asset return, such as the multivariate normal distribution (Landsman, 2010), the multivariate elliptical distribution (Owadally and Landsman, 2013), the generalized Laplace distribution (Jiang et al., 2016) and the generalized skew-elliptical distributions (Eini and Khaloozadeh, 2021*a*). From these studies, the TMV of the portfolio loss can be expressed as a weighted sum of the portfolio's expected loss, standard deviation, and variance. That is,

$$\text{TMV}_q(L) = -\boldsymbol{w}^\top \boldsymbol{\mu} + \xi_{1,q}\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} + \lambda \xi_{2,q} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w},$$

where $\xi_{1,q}$ and $\xi_{2,q}$ are two constants that depend on q and exhibit varied expressions under different distribution assumptions. Particularly, under the normality assumption, these constants are expressed as follows:

$$\xi_{1,q} = \frac{\phi(z_q)}{1-\Phi(z_q)} \text{ and } \xi_{2,q} = 1 - \xi_{1,q}(\xi_{1,q} - z_q). \tag{2.1}$$

Here, $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution functions of the standard normal random variable, and z_q denotes its q-quantile.

Motivated by the expression of $\text{TMV}_q(L)$ as given above, this chapter considers the following Mean-Variance-Standard-deviation (MVS) portfolio selection problem:

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \gamma_1 \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} - \gamma_2 \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}, \\ st. \quad & \boldsymbol{w}^\top \mathbf{1} = 1, \end{aligned} \tag{2.2}$$

where $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$ are two risk aversion coefficients such that at least one of them is non-zero. Assume that $\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} \neq 0$. When $\gamma_1 = 0$, we need the following additional technical assumption:

Assumption 2.1. *When $\gamma_1 = 0$, we assume it holds that $\gamma_2^2 > \psi^2$, where $\psi^2 = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - (\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})^2 / (\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1})$.*

Owadally and Landsman (2013) derive the optimal solution for the TMV portfolio optimization problem by leveraging its inherent connection to the classic Mean-Variance (MV) portfolio optimization problem. We adapt their method to obtain the optimal solution to the MVS portfolio problem and summarize the results in Proposition 2.1. As a formal proof, we verify the optimality of the solution directly by checking the first-order conditions arising in the method of Lagrangian multipliers as shown in Appendix A.1.1. Furthermore, we can plot the efficient frontier of the MVS optimization as similarly discussed in Owadally and Landsman (2013). Given that η_{mvs} (2.3) serves a similar role to the risk aversion coefficient in the MV optimization, the efficient frontiers derived from these two optimizations are expected to overlap.

Proposition 2.1. *Under Assumption 2.1, the solution to the MVS portfolio optimization problem (2.2) is given by*

$$\boldsymbol{w}^* = \boldsymbol{w}_{gmv} + \frac{1}{\eta_{mvs}} \boldsymbol{w}_z, \tag{2.3}$$

where $\boldsymbol{w}_{gmv} = \boldsymbol{\Sigma}^{-1}\mathbf{1}/(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})$ *is the global minimum variance portfolio with expected return* $\mu_{gmv} = \boldsymbol{w}_{gmv}^\top\boldsymbol{\mu}$ *and variance* $\sigma_{gmv}^2 = \boldsymbol{w}_{gmv}^\top\boldsymbol{\Sigma}\boldsymbol{w}_{gmv}$*, and* $\boldsymbol{w}_z = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_{gmv}\mathbf{1})$ *is the zero investment portfolio such that* $\boldsymbol{w}_z^\top\mathbf{1} = 0$*. The constant* η_{mvs} *is determined by:*

(a) When $\gamma_1 > 0$ *and* $\gamma_2 = 0$*,* $\eta_{mvs} = 2\gamma_1$*;*

(b) When $\gamma_1 = 0$ *and* $\gamma_2 > 0$*,* $\eta_{mvs} = \sqrt{(\gamma_2^2 - \psi^2)/\sigma_{gmv}^2}$*, where* $\psi^2 = \boldsymbol{w}_z^\top\boldsymbol{\mu} = \boldsymbol{w}_z^\top\boldsymbol{\Sigma}\boldsymbol{w}_z$ *is the expected return and also the variance of the zero investment portfolio;*

(c) When $\gamma_1 > 0$ *and* $\gamma_2 > 0$*,* η_{mvs} *is the unique solution in* $(2\gamma_1, \infty)$ *that satisfies*

$$(\eta_{mvs} - 2\gamma_1)^2 = \frac{\gamma_2^2\eta_{mvs}^2}{\eta_{mvs}^2\sigma_{gmv}^2 + \psi^2}. \tag{2.4}$$

Proof. See Appendix A.1.1. □

The above MVS portfolio selection framework encompasses many portfolio optimization problems as its special cases. There exists a one-to-one correspondence between the risk aversion coefficients γ_1 and γ_2 in the MVS objective function and the parameters q and λ in the TMV risk measure. Specifically, when $\gamma_1 = \lambda\xi_{2,q}$ and $\gamma_2 = \xi_{1,q}$ with ξ_1 and ξ_2 defined in (2.1), the MVS objective function in (2.2) becomes the negation of the TMV measure at a confidence level q and with a weighting parameter λ, under the normality assumption. For each commonly adopted confidence level $q \in \{0.9, 0.95, 0.99\}$ and each weighting parameter $\lambda \in \{3, 5\}$, we delineate the mapping between (q, λ) and (γ_1, γ_2) in Table 2.1.

Table 2.1: Transition between (q, λ) in TMV and (γ_1, γ_2) in MVS.

TMV		MVS
confidence level q	weighting parameter λ	$\gamma_1 = \lambda\xi_{2,q}, \gamma_2 = \xi_{1,q}$
0.9	3	0.5074, 1.7550
0.95	3	0.4142, 2.0627
0.99	3	0.2905, 2.6652
0.9	5	0.8457, 1.7550
0.95	5	0.6904, 2.0627
0.99	5	0.4842, 2.6652

The MVS optimization also includes the mean-TMV optimization as a special case, which uses TMV as the risk measure within the MV optimization framework. To see it, denote the risk aversion coefficient in the mean-TMV objective function as $\gamma_{mt} > 0$. Consequently, the MVS objective function aligns with that of the mean-TMV optimization when $\gamma_1 = \gamma_{mt}\lambda\xi_{2,q}/(1+\gamma_{mt})$ and $\gamma_2 = \gamma_{mt}\xi_{1,q}/(1+\gamma_{mt})$. The denominator $(1+\gamma_{mt})$, which is always positive, is to normalize the coefficient of the mean. Therefore, when $[\gamma_1, \gamma_2]$ are proportional to $[\lambda\xi_{2,q}, \xi_{1,q}]$, they aptly reflect the risk aversion coefficient γ in the mean-TMV optimization. Additionally, the MV portfolio optimization with multiple priors and ambiguity aversion studied in Garlappi et al. (2007) is equivalent to the above MVS optimization when the expected returns are estimated jointly for all assets.

Furthermore, as indicated by Proposition 2.1, the proposed MVS portfolio selection problem nests two special cases: the classical MV portfolio problem and the Mean-Standard-Deviation (MStD) portfolio problem as detailed below.

Case 1: Mean-Variance (MV) Criterion

The classical MV portfolio problem (Markowitz, 1952) can be recovered by setting $\gamma_1 = \gamma_{mv}/2$ and $\gamma_2 = 0$ in problem (2.2):

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\gamma_{mv}}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \\ st. \quad & \boldsymbol{w}^\top \mathbf{1} = 1. \end{aligned} \tag{2.5}$$

The optimal solution to (2.5) is given by

$$\boldsymbol{w}_{mv}^* = \boldsymbol{w}_{gmv} + \frac{1}{\gamma_{mv}}\boldsymbol{w}_z.$$

Case 2: Mean-Standard-Deviation (MStD) Criterion

The following MStD portfolio problem can be obtained by setting $\gamma_1 = 0$ and $\gamma_2 = \gamma_{ms}$ in problem (2.2):

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \gamma_{ms}\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}, \\ st. \quad & \boldsymbol{w}^\top \mathbf{1} = 1. \end{aligned} \tag{2.6}$$

When Assumption 2.1 holds, (2.6) has the optimal solution as follows:

$$\boldsymbol{w}_{ms}^{*} = \boldsymbol{w}_{gmv} + \frac{1}{\eta_{ms}}\boldsymbol{w}_{z},$$

where the constant η_{ms} is given by $\eta_{ms} = \sqrt{(\gamma_{ms}^2 - \psi^2)/\sigma_{gmv}^2}$.

Under the multivariate normal distribution assumption of asset returns, the MStD criterion encompasses the mean-risk portfolio selection problems where the risk is measured by a quantile-based risk measure such as VaR and CVaR. To be more specific, the return of a portfolio $\boldsymbol{w}$ follows a normal distribution with the mean $\mu_{\boldsymbol{w}} = \boldsymbol{w}^\top \boldsymbol{\mu}$ and variance $\sigma_{\boldsymbol{w}}^2 = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$. Therefore, at a confidence level $q \in (1/2, 1)$, the VaR and the CVaR of the portfolio loss L are, respectively, given by

$$\begin{aligned} \mathrm{VaR}_q(L) &= -\boldsymbol{w}^\top \boldsymbol{\mu} + \Phi^{-1}(q)\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}, \\ \mathrm{CVaR}_q(L) &= -\boldsymbol{w}^\top \boldsymbol{\mu} + \frac{\phi(\Phi^{-1}(q))}{1-q}\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}. \end{aligned}$$

Moreover, quantile-based risk measures such as weighted VaR (He et al., 2015; Wei, 2018) can be expressed as a weighted average of VaR at different confidence levels, and hence, the mean-risk model under these risk measures is equivalent to the MStD criterion with a special coefficient γ_{ms}; see, for example, Alexander and Baptista (2002, 2004) and Bodnar et al. (2012) for related studies. In addition, the mental account model of Das et al. (2010) can also be formulated as an MStD optimization problem.

2.3 Estimation Risk and Combination Rules

In practice, the portfolio rule (2.3) is not directly implementable because the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of asset returns are generally unknown. The standard plug-in approach entails estimating the mean vector and covariance matrix based on the observed historical data and then substituting the estimates into (2.3) to obtain an implementable portfolio rule.

Suppose we have historical asset returns of the N risky assets in the previous T periods, and denote them by $\{\boldsymbol{r}_1, \cdots, \boldsymbol{r}_T\}$, where $\boldsymbol{r}_t$ is the N-dimensional asset return vector for period $t, t = 1, ..., T$. Assume that $\{\boldsymbol{r}_1, \cdots, \boldsymbol{r}_T\}$ are independent and identically distributed, following the multivariate normal distribution assumption with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, the sample mean and sample covariance matrix are independent of each other and are respectively given by

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T),$$
$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})^\top \sim W_N(T-1, \boldsymbol{\Sigma})/T,$$

where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ represents a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}/T$, and $W_N(T-1, \boldsymbol{\Sigma})$ denotes a Wishart distribution with $T-1$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$. The plug-in MVS portfolio is constructed by replacing the true parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2.3) with their sample counterparts $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, respectively. Due to the inherent difference between these estimates and the true parameters, the plug-in MVS portfolio often exhibits notoriously poor performance during the out-of-sample period.

2.3.1 Combination Rules

We consider a combination of portfolio rules to address the issue of estimation risk. The combined portfolio strategy can be interpreted as a shrinkage technique to mitigate the adverse impact of estimation errors on the out-of-sample performance of portfolio rules. This approaches have been explored in the literature (see, for example, Tu and Zhou, 2011; DeMiguel et al., 2013; Kan et al., 2022; Lassance et al., 2023, among others). Specifically, Tu and Zhou (2011) demonstrate that combining the plug-in MV portfolio with the 1/N rule markedly enhances the out-of-sample performance of the plug-in MV portfolio. Here, the 1/N rule, which suggests that investors allocate their wealth equally among all risky assets, performs surprisingly well in practice (DeMiguel et al., 2009b). Due to its simplicity and efficiency, the 1/N has been widely used as a shrinkage target in the literature.

Inspired by Tu and Zhou (2011), we consider the following combined three-fund portfolio that is a combination of two components from the plug-in MVS portfolio and the 1/N rule:

$$\hat{\boldsymbol{w}}_c = \delta\hat{\boldsymbol{w}}_g + \beta\hat{\boldsymbol{w}}_m + (1 - \delta\hat{\boldsymbol{w}}_g^\top \mathbf{1} - \beta\hat{\boldsymbol{w}}_m^\top \mathbf{1})\boldsymbol{w}_e, \tag{2.7}$$

where $\boldsymbol{w}_e = \mathbf{1}/N$ represents the 1/N portfolio. Also, $\hat{\boldsymbol{w}}_g = \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$ and $\hat{\boldsymbol{w}}_m = \widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ correspond to the plug-in global minimum variance portfolio and the plug-in zero-investment portfolio, respectively, which originate from the plug-in version of the optimal MVS portfolio (2.3). In the above, δ and β are two constant combination coefficients. Unlike the combined portfolios in Tu and Zhou (2011) which allow an investment in the risk-free asset, the combined portfolio specified in (2.7) is fully invested in risky assets which aligns with the discussion in Section 2.2. To achieve it, the coefficient in front of the 1/N rule ensures that the portfolio satisfies the following constraint:

$$\hat{\boldsymbol{w}}_c^\top \mathbf{1} = 1.$$

As a relevant note, Lassance et al. (2024) derive an analytical expression for the risk, measured by the variance, of an investor's out-of-sample utility under parameter uncertainty. While their focus is placed on the analysis of the classic MV portfolio, the global minimum variance portfolio, and a shrinkage portfolio that combines both, they also study the risk of the out-of-sample utility when integrating the 1/N rule with the two sophisticated portfolios. In this special case, their combined portfolio is also a three-fund rule, including the plug-in global minimum variance portfolio, the plug-in zero investment portfolio, and the 1/N rule. Our three-fund rule differs from theirs in that we rearrange the two components in the plug-in MVS portfolio, which are then adopted as the constituent portfolios in our combined three-fund rule. This is out of consideration for mathematical convenience.

Given the combined three-fund portfolio $\hat{\boldsymbol{w}}_c$ in (2.7), the out-of-sample mean and variance are obtained by $\mu_{\hat{\boldsymbol{w}}_c} = \hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}$ and $\sigma^2_{\hat{\boldsymbol{w}}_c} = \hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma}\hat{\boldsymbol{w}}_c$, respectively, which represent the performance of the combined portfolio $\hat{\boldsymbol{w}}_c$ when it is implemented in an out-of-sample period. For ease of presentation, we denote

$$\begin{aligned}\mu_{\hat{\boldsymbol{w}}_g} &:= \hat{\boldsymbol{w}}_g^\top \boldsymbol{\mu} = \mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu},\\ \mu_{\hat{\boldsymbol{w}}_m} &:= \hat{\boldsymbol{w}}_m^\top \boldsymbol{\mu} = \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu},\\ \mu_N &:= \boldsymbol{w}_e^\top \boldsymbol{\mu} = \mathbf{1}^\top \boldsymbol{\mu}/N,\end{aligned}$$

and

$$\begin{aligned}\sigma^2_{\hat{\boldsymbol{w}}_g} &:= \hat{\boldsymbol{w}}_g^\top \boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g = \mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1},\\ \sigma^2_{\hat{\boldsymbol{w}}_m} &:= \hat{\boldsymbol{w}}_m^\top \boldsymbol{\Sigma}\hat{\boldsymbol{w}}_m = \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}},\\ \sigma^2_N &:= \boldsymbol{w}_e^\top \boldsymbol{\Sigma}\boldsymbol{w}_e = \mathbf{1}^\top \boldsymbol{\Sigma}\mathbf{1}/N^2.\end{aligned}$$

Furthermore, we let

$$\begin{aligned}\sigma^2_{\hat{\boldsymbol{w}}_{g,m}} &:= \hat{\boldsymbol{w}}_g^\top \boldsymbol{\Sigma}\hat{\boldsymbol{w}}_m = \mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}},\\ \sigma^2_{\hat{\boldsymbol{w}}_{g,e}} &:= \hat{\boldsymbol{w}}_g^\top \boldsymbol{\Sigma}\boldsymbol{w}_e = \mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}/N,\\ \sigma^2_{\hat{\boldsymbol{w}}_{m,e}} &:= \hat{\boldsymbol{w}}_m^\top \boldsymbol{\Sigma}\boldsymbol{w}_e = \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}/N.\end{aligned}$$

These are the conditional covariances of returns from any two of the three given components, $\hat{\boldsymbol{w}}_g$, $\hat{\boldsymbol{w}}_m$ and $\boldsymbol{w}_e$, included in our combined portfolio $\hat{\boldsymbol{w}}_c$. We also denote $\hat{s}_g = \hat{\boldsymbol{w}}_g^\top \mathbf{1}$ and $\hat{s}_m = \hat{\boldsymbol{w}}_m^\top \mathbf{1}$. Proposition 2.2 shows how one can use these notations, along with the combination coefficients δ and β to express the out-of-sample mean $\mu_{\hat{\boldsymbol{w}}_c}$ and variance $\sigma^2_{\hat{\boldsymbol{w}}_c}$ of the combined portfolio $\hat{\boldsymbol{w}}_c$.

Proposition 2.2. *The out-of-sample mean and variance of the combined three-fund portfolio $\hat{\boldsymbol{w}}_c$ are computed as follows:*

$$\begin{aligned}&\mu_{\hat{\boldsymbol{w}}_c} = \hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu} = \delta H_{11} + \beta H_{21} + \mu_N,\\ &\sigma^2_{\hat{\boldsymbol{w}}_c} = \hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma}\hat{\boldsymbol{w}}_c = \delta^2 A_1 + \beta^2 A_3 + 2\delta\beta A_2 + \delta H_{12} + \beta H_{22} + \sigma^2_N,\end{aligned} \tag{2.8}$$

where

$$\begin{aligned}
H_{11} &= \mu_{\hat{\boldsymbol{w}}_g} - \hat{s}_g\mu_N, \qquad H_{12} = 2(\sigma^2_{\hat{\boldsymbol{w}}_{g,e}} - \hat{s}_g\sigma^2_N),\\
H_{21} &= \mu_{\hat{\boldsymbol{w}}_m} - \hat{s}_m\mu_N, \qquad H_{22} = 2(\sigma^2_{\hat{\boldsymbol{w}}_{m,e}} - \hat{s}_m\sigma^2_N),\\
A_1 &= \sigma^2_{\hat{\boldsymbol{w}}_g} + \hat{s}_g^2\sigma^2_N - 2\hat{s}_g\sigma^2_{\hat{\boldsymbol{w}}_{g,e}},\\
A_2 &= \sigma^2_{\hat{\boldsymbol{w}}_{g,m}} + \hat{s}_g\hat{s}_m\sigma^2_N - \hat{s}_g\sigma^2_{\hat{\boldsymbol{w}}_{m,e}} - \hat{s}_m\sigma^2_{\hat{\boldsymbol{w}}_{g,e}},\\
A_3 &= \sigma^2_{\hat{\boldsymbol{w}}_m} + \hat{s}_m^2\sigma^2_N - 2\hat{s}_m\sigma^2_{\hat{\boldsymbol{w}}_{m,e}}.
\end{aligned}$$

Proof. See Appendix A.1.2. □

In addition, we can consider two special cases of the combined three-fund rule (2.7) when one of the combination coefficients is zero. If $\beta = 0$, the combined global minimum variance portfolio that includes the 1/N rule and $\hat{\boldsymbol{w}}_g = \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$ is given by

$$\hat{\boldsymbol{w}}_{cg} = \delta\hat{\boldsymbol{w}}_g + (1 - \delta\hat{\boldsymbol{w}}_g^\top\mathbf{1})\boldsymbol{w}_e. \tag{2.9}$$

If $\delta = 0$, the combined market portfolio that consists of the 1/N rule and $\hat{\boldsymbol{w}}_m = \widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ is given by

$$\hat{\boldsymbol{w}}_{cm} = \beta\hat{\boldsymbol{w}}_m + (1 - \beta\hat{\boldsymbol{w}}_m^\top\mathbf{1})\boldsymbol{w}_e. \tag{2.10}$$

The out-of-sample portfolio mean and variance for each of these two combined portfolio strategies can be obtained from (2.8) by setting either $\delta = 0$ or $\beta = 0$ in Proposition 2.2.

2.3.2 Expected Out-of-Sample Utility Criterion

We adapt the expected out-of-sample utility framework proposed in Kan and Zhou (2007) for evaluating the out-of-sample performance of a portfolio rule. To align with the MVS criterion (2.2), the proposed out-of-sample MVS utility for a plug-in portfolio $\hat{\boldsymbol{w}}$ is given by

$$U(\hat{\boldsymbol{w}}) = \hat{\boldsymbol{w}}^\top\boldsymbol{\mu} - \gamma_1\hat{\boldsymbol{w}}^\top\boldsymbol{\Sigma}\hat{\boldsymbol{w}} - \gamma_2\sqrt{\hat{\boldsymbol{w}}^\top\boldsymbol{\Sigma}\hat{\boldsymbol{w}}}.$$

Since the portfolio weight vector $\hat{\boldsymbol{w}}$ depends on historical realizations of asset returns, the out-of-sample MVS utility of $\hat{\boldsymbol{w}}$ is also a random variable. Hence, we evaluate its performance based on the expected out-of-sample utility

$$\mathbb{E}[U(\hat{\boldsymbol{w}})] = \mathbb{E}\left[\hat{\boldsymbol{w}}^\top \boldsymbol{\mu} - \gamma_1 \hat{\boldsymbol{w}}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}} - \gamma_2 \sqrt{\hat{\boldsymbol{w}}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}}\right].$$

The expected out-of-sample MVS utility represents the average utility level an investor can achieve by applying the portfolio rule $\hat{\boldsymbol{w}}$ repeatedly over time. It is worth mentioning that Lassance et al. (2024) further advocate considering the trade-off between the mean and variance of the out-of-sample utility, and propose a novel measure of portfolio robustness.

For the combined strategy $\hat{\boldsymbol{w}}_c$, its expected out-of-sample utility depends on the combination coefficients (δ, β) and the relationship can be expressed as:

$$G(\delta, \beta) = \mathbb{E}[U(\hat{\boldsymbol{w}}_c)] = \mathbb{E}[\mu_{\hat{\boldsymbol{w}}_c}] - \gamma_1 \mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}] - \gamma_2 \mathbb{E}[\sigma_{\hat{\boldsymbol{w}}_c}]. \tag{2.11}$$

To identify the optimal combination coefficients for the combined portfolio $\hat{\boldsymbol{w}}_c$, we seek to maximize its expected out-of-sample MVS utility (2.11).

However, analytical evaluation of the expected out-of-sample standard deviation in (2.11) is challenging, if not impossible. We consider its first-order Taylor series expansion[1]

$$\mathbb{E}[\sqrt{\sigma^2_{\hat{\boldsymbol{w}}_c}}] \approx \sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]},$$

which poses a simple form and is directly related to the expected out-of-sample variance. The first-order Taylor series expansion was also used in Lassance (2021) for analytical tractability, in the context of the expected out-of-sample Sharpe ratio maximization. We may also consider its second-order Taylor series expansion[2]

$$\mathbb{E}[\sqrt{\sigma^2_{\hat{\boldsymbol{w}}_c}}] \approx \sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]}\left[1 - \frac{\mathrm{Var}(\sigma^2_{\hat{\boldsymbol{w}}_c})}{8\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]^2}\right].$$

A detailed discussion of these two Taylor series expansions is provided in Appendix A.2. In view of the Taylor series expansion and Jensen's inequality, the true expected out-of-sample standard deviation lies within the first-order and second-order Taylor series ex-

[1] Denote that $x = \sigma^2_{\hat{\boldsymbol{w}}_c}$. We assume that $f(x) = \sqrt{x}$ is differentiable at $E[x]$ and use the first order Taylor expansion around $E[x]$ to estimate $\mathbb{E}[f(x)]$, which yields $E[\sqrt{x}] = \sqrt{E[x]} + (\sqrt{E[x]})' E[x - E[x]] = \sqrt{E[x]}$.

[2] Denote that $x = \sigma^2_{\hat{\boldsymbol{w}}_c}$. We consider the second order Taylor expansion of $f(x) = \sqrt{x}$ around $E[x]$ to estimate $\mathbb{E}[f(x)]$, which yields $E[\sqrt{x}] = \sqrt{E[x]} + \frac{1}{2}(\sqrt{E[x]})'' E[(x - E[x])^2] = \sqrt{E[x]} - Var(x)/(8E[x]^{3/2})$.

pansions. Appendix A.2 compares the performance of the first-order and second-order Taylor series expansions via numerical simulations. Even though the second-order Taylor series expansion is more accurate than the first-order Taylor series expansion, its complex expression hinders us from obtaining analytic optimal combination coefficients. Furthermore, the first-order Taylor series expansion remains highly proximate to the true value, as demonstrated in Appendix A.2.

Therefore, we utilize the first-order Taylor series expansion for the evaluation of the expected out-of-sample standard deviation. The optimization problem to find the optimal combination coefficients that maximize the approximated expected out-of-sample MVS utility is given by

$$\max_{\delta,\beta} \tilde{G}(\delta, \beta), \tag{2.12}$$

where the objective function is

$$\begin{aligned}
&\tilde{G}(\delta, \beta) = \mathbb{E}[\mu_{\hat{\boldsymbol{w}}_c}] - \gamma_1 \mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}] - \gamma_2 \sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]} \\
=&\mathbb{E}[H_{11}]\delta + \mathbb{E}[H_{21}]\beta + \mu_N - \gamma_1(\sigma_N^2 + 2\delta\beta\mathbb{E}[A_2] + \delta^2\mathbb{E}[A_1] + \beta^2\mathbb{E}[A_3] + \delta\mathbb{E}[H_{12}] + \beta\mathbb{E}[H_{22}]) \\
&-\gamma_2\sqrt{\sigma_N^2 + 2\delta\beta\mathbb{E}[A_2] + \delta^2\mathbb{E}[A_1] + \beta^2\mathbb{E}[A_3] + \delta\mathbb{E}[H_{12}] + \beta\mathbb{E}[H_{22}]}.
\end{aligned} \tag{2.13}$$

Parallel to the discussion in Section 2.2, the approximated expected out-of-sample MVS criterion also contains two special cases. When the risk aversion coefficients in (2.13) are set to $\gamma_1 = \gamma_{mv}/2$ and $\gamma_2 = 0$, it reduces to the expected out-of-sample MV utility function discussed in Kan and Zhou (2007). Furthermore, by setting $\gamma_1 = 0$ and $\gamma_2 = \gamma_{ms}$ in (2.13), we recover the approximated expected out-of-sample MStD utility function.

Given that the asset return vector follows the multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the following proposition presents the explicit expressions of those expectations involved in (2.13). We can then easily obtain the analytic expression of the approximated expected out-of-sample MVS utility (2.13) as a deterministic function of the combination coefficients.

Proposition 2.3. *Assume $T > N + 4$. For those quantities defined in Proposition 2.2, we* *have*

$$\mathbb{E}[H_{11}] = c_1(\theta_2^2 - \mu_N \theta_3^2), \qquad \mathbb{E}[H_{12}] = 2c_1(1 - \sigma_N^2 \theta_3^2),$$
$$\mathbb{E}[H_{21}] = c_1(\theta_1^2 - \mu_N \theta_2^2), \qquad \mathbb{E}[H_{22}] = 2c_1(\mu_N - \sigma_N^2 \theta_2^2),$$
$$\mathbb{E}[A_1] = c_1 c_2 \theta_3^2(2N - T) + c_2 \theta_3^4 \sigma_N^2 (c_1 + T),$$
$$\mathbb{E}[A_2] = c_1 c_2 \theta_2^2(N - 2 + \sigma_N^2 \theta_3^2) - c_2 T \theta_3^2(\mu_N - \sigma_N^2 \theta_2^2),$$
$$\mathbb{E}[A_3] = c_1 c_2 \left[\theta_1^2(T - 4 + \sigma_N^2 \theta_3^2) + \frac{(T-2)(N - 2 + \sigma_N^2 \theta_3^2)}{T}\right] + c_2 T \theta_2^2 \left(\sigma_N^2 \theta_2^2 - 2\mu_N\right),$$

where

$$\theta_1^2 = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad \theta_2^2 = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad \theta_3^2 = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1},^{3}$$
$$c_1 = T/(T - N - 2), \quad c_2 = T/(T - N - 1)(T - N - 4).$$

Proof. See Appendix A.1.3. □

2.4 Optimal Combined Portfolio

In this section, we derive the optimal combination coefficients that maximize the approximated expected out-of-sample MVS utility (2.12). We first prove the concavity of the objective function.

Lemma 2.1. *The approximated expected out-of-sample MVS objective function* $\tilde{G}(\delta, \beta)$ *is a concave function of* (δ, β)*.*

Proof. See Appendix A.1.4. □

We can then obtain the optimal combination coefficients for the combined three-fund rule $\hat{\boldsymbol{w}}_c$ by solving the optimization problem (2.12) using the first-order optimality condition.

[3] Note that for these three terms θ_1^2, θ_2^2 and θ_3^2, we can also express them as $\theta_1^2 = \mu_{gmv}^2/\sigma_{gmv}^2 + \psi^2$, $\theta_2^2 = \mu_{gmv}/\sigma_{gmv}^2$ and $\theta_3^2 = 1/\sigma_{gmv}^2$.

Theorem 2.1. *The optimal combination coefficients δ_c^* and β_c^* that solve the approximated expected out-of-sample MVS criterion optimization (2.12) for the combined three-fund rule $\hat{\boldsymbol{w}}_c$ are given by*

$$\delta_c^* = \frac{\mathbb{E}[H_1]\mathbb{E}[A_3] - \mathbb{E}[H_2]\mathbb{E}[A_2]}{2\eta(\mathbb{E}[A_1]\mathbb{E}[A_3] - \mathbb{E}[A_2]^2)} \quad \text{and} \quad \beta_c^* = \frac{\mathbb{E}[H_1] - 2\mathbb{E}[A_1]\eta\delta_c^*}{2\eta\mathbb{E}[A_2]}, \tag{2.14}$$

where $\mathbb{E}[H_1] = \mathbb{E}[H_{11}] - \eta\mathbb{E}[H_{12}]$, $\mathbb{E}[H_2] = \mathbb{E}[H_{21}] - \eta\mathbb{E}[H_{22}]$, and the positive constant $\eta \in (\gamma_1, \infty)$ solves

$$2(\eta - \gamma_1)\sqrt{\delta_c^*\mathbb{E}[H_{12}] + \beta_c^*\mathbb{E}[H_{22}] + \sigma_N^2 + 2\delta_c^*\beta_c^*\mathbb{E}[A_2] + (\delta_c^*)^2\mathbb{E}[A_1] + (\beta_c^*)^2\mathbb{E}[A_3]} = \gamma_2. \tag{2.15}$$

The optimal combined three-fund rule is given by

$$\hat{\boldsymbol{w}}_c^* = \delta_c^*\hat{\boldsymbol{w}}_g + \beta_c^*\hat{\boldsymbol{w}}_m + (1 - \delta_c^*\hat{\boldsymbol{w}}_g^\top\mathbf{1} - \beta_c^*\hat{\boldsymbol{w}}_m^\top\mathbf{1})\boldsymbol{w}_e. \tag{2.16}$$

Proof. See Appendix A.1.5. □

Besides the combined three-fund rule $\hat{\boldsymbol{w}}_c$, we also consider the combined global minimum variance portfolio $\hat{\boldsymbol{w}}_{cg}$ (2.9) and combined market portfolio $\hat{\boldsymbol{w}}_{cm}$ (2.10). The following theorem gives the analytic combination coefficients for the two special cases.

Theorem 2.2. *We have the following assertions:*

1. *The optimal combination coefficient δ_{cg}^* that solves the approximated expected out-of-sample MVS criterion optimization (2.12) for the combined global minimum variance portfolio $\hat{\boldsymbol{w}}_{cg}$ is*

$$\delta_{cg}^* = \frac{\mathbb{E}[H_{11}] - \eta_{cg}\mathbb{E}[H_{12}]}{2\eta_{cg}\mathbb{E}[A_1]}, \tag{2.17}$$

where $\eta_{cg} \in (\gamma_1, \infty)$ solves $2(\eta_{cg} - \gamma_1)\sqrt{\sigma_N^2 + (\delta_{cg}^*)^2\mathbb{E}[A_1] + \delta_{cg}^*\mathbb{E}[H_{12}]} = \gamma_2$. The optimal combined global minimum variance portfolio is given by

$$\hat{\boldsymbol{w}}_{cg}^* = \delta_{cg}^*\hat{\boldsymbol{w}}_g + (1 - \delta_{cg}^*\hat{\boldsymbol{w}}_g^\top\mathbf{1})\boldsymbol{w}_e. \tag{2.18}$$

2. *The optimal combination coefficient β_{cm}^* that solves the approximated expected out-of-sample MVS criterion optimization (2.12) for the combined market portfolio $\hat{\boldsymbol{w}}_{cm}$*

is

$$\beta^*_{cm} = \frac{\mathbb{E}[H_{21}] - \eta_{cm}\mathbb{E}[H_{22}]}{2\eta_{cm}\mathbb{E}[A_3]}, \tag{2.19}$$

where $\eta_{cm} \in (\gamma_1, \infty)$ *solves* $2(\eta_{cm} - \gamma_1)\sqrt{\sigma_N^2 + (\beta^*_{cm})^2\mathbb{E}[A_3] + \beta^*_{cm}\mathbb{E}[H_{22}]} = \gamma_2$. *The optimal combined market portfolio is given by*

$$\hat{\boldsymbol{w}}^*_{cm} = \beta^*_{cm}\hat{\boldsymbol{w}}_m + (1 - \beta^*_{cm}\hat{\boldsymbol{w}}_m^\top \mathbf{1})\boldsymbol{w}_e. \tag{2.20}$$

Proof. See Appendix A.1.6. □

Theorems 2.1 and 2.2 present the optimal combination coefficients of the combined portfolios that maximize the approximated expected out-of-sample MVS utility. The results also cover two special optimization problems: the expected out-of-sample MV utility optimization ($\gamma_1 = \gamma_{mv}/2, \gamma_2 = 0$) and the approximated expected out-of-sample MStD utility optimization ($\gamma_1 = 0, \gamma_2 = \gamma_{ms}$). Hence, our results are flexible to accommodate MV, TMV, and MStD portfolio optimization problems via a suitable choice of risk aversion coefficients γ_1 and γ_2.

In practice, the optimal combination coefficients are unknown, as they depend on θ_1^2, θ_2^2, θ_3^2, μ_N, and σ_N^2 (see Proposition 2.3) which further depend on the mean vector and covariance matrix of asset returns. Sample estimators of these five unknown terms, denoted by $\hat{\theta}_1^2, \hat{\theta}_2^2, \hat{\theta}_3^2, \hat{\mu}_N, \hat{\sigma}_N^2$, respectively, can be obtained by replacing the unknown mean and covariance matrix with corresponding sample counterparts. However, these sample estimators can be heavily biased when T is small (e.g. Kan and Zhou, 2007). We consider sophisticated estimators based on the distributional properties of these terms. Particularly, due to the difficulty in obtaining a sophisticated estimator for μ_N, we opt to keep the sample estimator for this parameter.

Lemma 2.2. *Let* $F_{a,b}(c)$ *represent a noncentral* F *distribution with* a *and* b *degrees of freedom and a noncentrality parameter* c. *Let Inv-*χ_x^2 *and* χ_y^2 *represent an inverse chi-square distribution with* x *degrees of freedom and a chi-square distribution with* y *degrees of freedom, respectively. We have the following assertions:*

1. $\hat{\theta}_1^2 = \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} \sim \left(\frac{N}{T-N}\right) F_{N,T-N}(T\theta_1^2)$.
2. $\hat{\theta}_3^2/T = \mathbf{1}^\top(T\widehat{\boldsymbol{\Sigma}})^{-1}\mathbf{1} \sim (\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})\textit{Inv-}\chi^2_{T-N}$.

3. $TN^2\hat{\sigma}_N^2 = T\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}\mathbf{1} \sim (\mathbf{1}^\top\boldsymbol{\Sigma}\mathbf{1})\chi^2_{T-1}$.

4. A stochastic representation of $\hat{\theta}_2^2 = \hat{\boldsymbol{\mu}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$ *is given by*

$$\hat{\theta}_2^2/T = \hat{\boldsymbol{\mu}}^\top(T\widehat{\boldsymbol{\Sigma}})^{-1}\mathbf{1} \overset{d}{=} \frac{1}{u_1}\left(\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1} + \sqrt{\left(\frac{1}{T} + \frac{N-1}{T(T-N+1)}u_3\right)(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})}u_2\right),$$

where $u_1 \sim \chi^2_{T-N}$*,* $u_2 \sim N(0,1)$*, and* $u_3 \sim F_{(N-1)/2,(T-N+1)/2}(T\psi^2)$*.*

Proof. See Appendix A.1.7. □

As suggested by Kan and Zhou (2007), estimating θ_1^2 is equivalent to estimating the noncentrality parameter of the noncentral F distribution. Hence, the sophisticated estimator of θ_1^2, which includes additional adjustments to the unbiased one, is given by

$$\tilde{\theta}_1^2 = \frac{(T-N-2)\hat{\theta}_1^2 - N}{T} + \frac{2(\hat{\theta}_1^2)^{N/2}(1+\hat{\theta}_1^2)^{-(T-2)/2}}{TB_{\hat{\theta}_1^2/(1+\hat{\theta}_1^2)}(N/2,(T-N)/2)}, \tag{2.21}$$

where $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\,\mathrm{d}y$ is the incomplete beta function. The first term on the right-hand side of (2.21) represents the unbiased estimator of θ_1^2, while the second term is the additional adjustment for small T. Note that $\tilde{\theta}_1^2$ is always positive, unlike the unbiased estimator of θ_1^2. In addition, using the distributional properties in Lemma 2.2, we obtain unbiased estimators of θ_2^2, θ_3^2 and σ_N^2 as sophisticated estimators, which are, respectively, given by

$$\tilde{\theta}_2^2 = \frac{T-N-2}{T}\hat{\theta}_2^2, \quad \tilde{\theta}_3^2 = \frac{T-N-2}{T}\hat{\theta}_3^2 \text{ and } \tilde{\sigma}_N^2 = \frac{T}{T-1}\hat{\sigma}_N^2. \tag{2.22}$$

Our discussion about the distributional properties of the sample estimators $\hat{\theta}_1^2$, $\hat{\theta}_2^2$, $\hat{\theta}_3^2$, and $\hat{\sigma}_N^2$ so far is based on the sample covariance matrix of asset returns. We also consider the shrinkage covariance matrix which is widely used in the literature (e.g., Ledoit and Wolf, 2004) and given by

$$\widehat{\boldsymbol{\Sigma}}_{LW} = (1-\rho)\widehat{\boldsymbol{\Sigma}} + \rho\upsilon\boldsymbol{I}_N,$$

where $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix and $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. The parameter υ is the average of the eigenvalues of the sample covariance matrix, and the parameter ρ satisfies

$$\rho = \frac{\min\left[\|\widehat{\boldsymbol{\Sigma}} - \upsilon \boldsymbol{I}_N\|^2, \frac{1}{T^2}\sum_{t=1}^{T}\|(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})^\top - \widehat{\boldsymbol{\Sigma}}\|^2\right]}{\|\widehat{\boldsymbol{\Sigma}} - \upsilon \boldsymbol{I}_N\|^2}.$$

Here, $\|A\| = \sqrt{\mathrm{tr}(AA')/N}$ for a matrix A represents the Frobenius norm. Therefore, we can also obtain sophisticated estimators for these four unknowns, θ_1^2, θ_2^2, θ_3^2, and σ_N^2, based on the shrinkage covariance matrix instead of the sample counterpart. To be more specific, when using the shrinkage covariance matrix, we first obtain the following shrinkage sample estimators:

$$\hat{\theta}_{1,LW}^2 = \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}_{LW}^{-1}\hat{\boldsymbol{\mu}},\ \ \hat{\theta}_{2,LW}^2 = \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}_{LW}^{-1}\mathbf{1},\ \ \hat{\theta}_{3,LW}^2 = \mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_{LW}^{-1}\mathbf{1},\ \ \hat{\sigma}_{N,LW}^2 = \mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_{LW}\mathbf{1}/N.$$

We then replace the sample estimators in (2.21) and (2.22) with the above shrinkage sample estimators to obtain a set of shrinkage sophisticated estimators $\tilde{\theta}_{1,LW}^2$, $\tilde{\theta}_{2,LW}^2$, $\tilde{\theta}_{3,LW}^2$, and $\tilde{\sigma}_{N,LW}^2$. It is worth noting that the last three shrinkage sophisticated estimators from this newly achieved set may become biased.

To summarize, after obtaining the estimates of θ_1^2, θ_2^2, θ_3^2, μ_N and σ_N^2, we replace the unknown terms in the optimal combination coefficients (2.14), (2.17), and (2.19) with those estimates and obtain the estimated combination coefficients. The resulting combined portfolios, known as implementable combined portfolios, are denoted as $\tilde{\boldsymbol{w}}_c$ (the implementable combined three-fund portfolio), $\tilde{\boldsymbol{w}}_{cg}$ (the implementable combined global minimum variance portfolio), and $\tilde{\boldsymbol{w}}_{cm}$ (the implementable combined market portfolio), respectively.

2.5 Numerical Studies

In this section, we evaluate the performance of the three proposed combined portfolios discussed in Section 2.4 and compare them with the plug-in MVS and 1/N portfolios, via both simulation and empirical studies.

The portfolio strategies we consider are summarized in Table 2.2. Panel A presents the oracle combined portfolios, which are based on optimal combination coefficients computed from the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, as given in Theorems 2.1 and 2.2; these portfolios are only available

Table 2.2: List of portfolio strategies.

Notation	Portfolio
Panel A: Oracle combined portfolios (only available in the simulation study)	
$\hat{\boldsymbol{w}}_c^*$	the oracle combined three-fund portfolio (2.16)
$\hat{\boldsymbol{w}}_{cg}^*$	the oracle combined global minimum variance portfolio (2.18)
$\hat{\boldsymbol{w}}_{cm}^*$	the oracle combined market portfolio (2.20)
Panel B: Implementable combined portfolios	
Set 1: apply the sample estimate $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and combination coefficients	
$\tilde{\boldsymbol{w}}_c$	the implementable combined three-fund portfolio
$\tilde{\boldsymbol{w}}_{cg}$	the implementable combined global minimum variance portfolio
$\tilde{\boldsymbol{w}}_{cm}$	the implementable combined market portfolio
Set 2: apply the shrinkage estimate $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and combination coefficients	
$\tilde{\boldsymbol{w}}_c^{LW}$	the implementable combined three-fund portfolio
$\tilde{\boldsymbol{w}}_{cg}^{LW}$	the implementable combined global minimum variance portfolio
$\tilde{\boldsymbol{w}}_{cm}^{LW}$	the implementable combined market portfolio
Set 3: apply $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the combination coefficients	
$\tilde{\boldsymbol{w}}_c^{LW1}$	the implementable combined three-fund portfolio
$\tilde{\boldsymbol{w}}_{cg}^{LW1}$	the implementable combined global minimum variance portfolio
$\tilde{\boldsymbol{w}}_{cm}^{LW1}$	the implementable combined market portfolio
Set 4: apply $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}$ in the combination coefficients	
$\tilde{\boldsymbol{w}}_c^{LW2}$	the implementable combined three-fund portfolio
$\tilde{\boldsymbol{w}}_{cg}^{LW2}$	the implementable combined global minimum variance portfolio
$\tilde{\boldsymbol{w}}_{cm}^{LW2}$	the implementable combined market portfolio
Panel C: Benchmark portfolios	
$\boldsymbol{w}_e$	the 1/N portfolio
$\hat{\boldsymbol{w}}_{mvs}$	the plug-in MVS portfolio that replaces $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2.3) with their sample counterparts
$\hat{\boldsymbol{w}}_{mvs}^{LW}$	the plug-in MVS portfolio that replaces $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2.3) with the sample mean and shrinkage covariance matrix

in the simulation study, and not implementable in practice. They provide (theoretically) tight upper bounds on the performance of combined portfolios. Panel B summarizes four sets of implementable combined portfolios that replace the unknown parameters in the optimal combined portfolios with corresponding estimates. All the four sets of portfolios estimate the mean vector $\boldsymbol{\mu}$ by its sample counterpart; however, they are distinguished from each other by the use of different estimates of the covariance matrix. For the portfolios in Sets 1 and 3, the constituent portfolios $\hat{\boldsymbol{w}}_g$ and $\hat{\boldsymbol{w}}_m$ are obtained with the sample covariance matrix, while the combination coefficients are estimated using the sophisticated estimators as discussed in Section 2.4 under two scenarios: one employing the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$, and the other utilizing the shrinkage covariance matrix $\widehat{\boldsymbol{\Sigma}}_{LW}$. To be more specific, the implementable combined three-fund portfolios for Sets 1 and 3 are respectively given by

$$\tilde{\boldsymbol{w}}_c = \tilde{\delta}_c \hat{\boldsymbol{w}}_g + \tilde{\beta}_c \hat{\boldsymbol{w}}_m + (1 - \tilde{\delta}_c \hat{\boldsymbol{w}}_g^\top \mathbf{1} - \tilde{\beta}_c \hat{\boldsymbol{w}}_m^\top \mathbf{1})\boldsymbol{w}_e,$$
$$\tilde{\boldsymbol{w}}_c^{LW1} = \tilde{\delta}_c^{LW} \hat{\boldsymbol{w}}_g + \tilde{\beta}_c^{LW} \hat{\boldsymbol{w}}_m + (1 - \tilde{\delta}_c^{LW} \hat{\boldsymbol{w}}_g^\top \mathbf{1} - \tilde{\beta}_c^{LW} \hat{\boldsymbol{w}}_m^\top \mathbf{1})\boldsymbol{w}_e,$$

where $\tilde{\delta}_c$ and $\tilde{\beta}_c$ are obtained using $\widehat{\boldsymbol{\Sigma}}$, while $\tilde{\delta}_c^{LW}$ and $\tilde{\beta}_c^{LW}$ are calculated using $\widehat{\boldsymbol{\Sigma}}_{LW}$. For portfolios in Sets 2 and 4, we plug in the shrinkage covariance matrix into the two constituent portfolios, and denote them by

$$\hat{\boldsymbol{w}}_g^{LW} = \widehat{\boldsymbol{\Sigma}}_{LW}^{-1} \mathbf{1} \quad \text{and} \quad \hat{\boldsymbol{w}}_m^{LW} = \widehat{\boldsymbol{\Sigma}}_{LW}^{-1} \hat{\boldsymbol{\mu}}.$$

Similarly, using two distinct sets of estimates for the combination coefficients, the implementable combined three-fund portfolios for Sets 2 and 4 are respectively given by

$$\tilde{\boldsymbol{w}}_c^{LW} = \tilde{\delta}_c^{LW} \hat{\boldsymbol{w}}_g^{LW} + \tilde{\beta}_c^{LW} \hat{\boldsymbol{w}}_m^{LW} + (1 - \tilde{\delta}_c^{LW} \mathbf{1}^\top \hat{\boldsymbol{w}}_g^{LW} - \tilde{\beta}_c^{LW} \mathbf{1}^\top \hat{\boldsymbol{w}}_m^{LW})\boldsymbol{w}_e,$$
$$\tilde{\boldsymbol{w}}_c^{LW2} = \tilde{\delta}_c \hat{\boldsymbol{w}}_g^{LW} + \tilde{\beta}_c \hat{\boldsymbol{w}}_m^{LW} + (1 - \tilde{\delta}_c \mathbf{1}^\top \hat{\boldsymbol{w}}_g^{LW} - \tilde{\beta}_c \mathbf{1}^\top \hat{\boldsymbol{w}}_m^{LW})\boldsymbol{w}_e.$$

The implementable combined global minimum variance portfolio and the implementable combined market portfolio listed in the four sets are defined similarly. Panel C shows the benchmark portfolios including the 1/N rule, the plug-in MVS portfolio $\hat{\boldsymbol{w}}_{mvs}$ that replaces $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2.3) with their sample counterparts, and the plug-in MVS portfolio $\hat{\boldsymbol{w}}_{mvs}^{LW}$ that replaces $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the sample mean and shrinkage covariance matrix, respectively.

2.5.1 Simulation Study

We consider a linear factor model for asset returns as in MacKinlay and Pástor (2000):

$$\boldsymbol{r} = \boldsymbol{\psi} + \boldsymbol{B}\boldsymbol{r}_f + \boldsymbol{\epsilon}, \tag{2.23}$$

where $\boldsymbol{\psi}$ is an $N \times 1$ mispricing coefficient vector with elements uniformly distributed between -0.002 and 0.002, $\boldsymbol{r}_f$ denotes the returns of factors that follow a multivariate normal distribution such that

$$\boldsymbol{r}_f \sim N(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b),$$

where we assume that $\boldsymbol{r}_f$ contains five factor variables. The mean return vector $\boldsymbol{\mu}_b$ is

$$\boldsymbol{\mu}_b = [0.018, 0.016, 0.014, 0.012, 0.01]^\top,$$

and $\boldsymbol{\Sigma}_b$ is 5×5 matrix with diagonal elements set at $0.006, 0.007, 0.008, 0.009, 0.01$, and off-diagonal elements equal to 0.001. The factor loading matrix $\boldsymbol{B}$ is an $N \times 5$ matrix with elements evenly spreading between 0.01 and 0.12. Specifically, we first create a vector ranging from 0.01 to 0.12 with $5N$ elements: $0.01 + i \times 0.11/(5N-1)$, $i = 0, 1, \ldots, 5N-1$. Then, for each $i = 1, \ldots, 5$, we take the i-th N elements from this vector as the i-th column of $\boldsymbol{B}$. The $N \times 1$ residual vector $\boldsymbol{\epsilon}$ follows a multivariate normal distribution with a zero mean vector and a covariance matrix $\boldsymbol{\Sigma}_\epsilon/12$. Here, $\boldsymbol{\Sigma}_\epsilon$ is a diagonal matrix whose diagonal elements are the squares of values randomly drawn from a uniform distribution with a range of $[0.02, 0.05]$. Given the above specifications, the asset return vector follows a multivariate normal distribution with the mean vector $\boldsymbol{\mu}_{\boldsymbol{r}} = \boldsymbol{\psi} + \boldsymbol{B}\boldsymbol{\mu}_b$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{r}} = \boldsymbol{B}\boldsymbol{\Sigma}_b\boldsymbol{B}^\top + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}/12$.

We evaluate the expected out-of-sample performance of the proposed implementable combined portfolios using Monte Carlo simulation. We simulate M sets of $T \times N$ asset returns from the five-factor model specified in (2.23). For each set of $T \times N$ simulated asset returns, we compute the sample mean $\hat{\boldsymbol{\mu}}$ and sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$ (or shrinkage covariance matrix $\widehat{\boldsymbol{\Sigma}}_{LW}$), and then get the estimates of θ_1^2, θ_2^2, θ_3^2, μ_N, and σ_N^2 as discussed in Section 2.4. These estimators are subsequently plugged into the formulas in Theorems 2.1 and 2.2 to obtain the implementable combined portfolios as listed in Panel B of Table 2.2. Let $\tilde{\boldsymbol{w}}_i$ be one such portfolio obtained based on the i-th set of simulated asset returns, $i = 1, \ldots, M$. We then take the average over M out-of-sample MVS utilities,

$$\frac{1}{M}\sum_{i=1}^{M}\left(\tilde{\boldsymbol{w}}_i^\top\boldsymbol{\mu} - \gamma_1\tilde{\boldsymbol{w}}_i^\top\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}_i - \gamma_2\sqrt{\tilde{\boldsymbol{w}}_i^\top\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}_i}\right),$$

as the expected out-of-sample MVS utility for performance comparison. For any benchmark portfolio listed in Panel C of Table 2.2, we estimate its expected out-of-sample MVS utility in the same way.

Table 2.3: Expected out-of-sample MVS utilities (in percentage) for $N = 15$.

(γ_1, γ_2)	$(1, 0.3735)$	$(1.25, 0.3112)$	$(2.5, 0)$	$(2.0984, 0.1)$	$(1.2952, 0.3)$	$(0, 0.6225)$
(q, λ)	(0.2159, 1.7642)	(0.1742, 2.0449)	-	(0.0454, 2.5563)	(0.1668, 2.0893)	-
Panel A: Oracle combined portfolio						
$\hat{\boldsymbol{w}}_c^*$	0.0788	0.3496	2.0213	1.4518	0.4019	-0.6351
$\hat{\boldsymbol{w}}_{cg}^*$	-0.0536	0.1673	1.6917	1.1534	0.2112	-0.6628
$\hat{\boldsymbol{w}}_{cm}^*$	-0.0426	0.1773	1.6961	1.1590	0.2209	-0.6533
Panel B: Implementable combined portfolios						
Set 1: apply the sample estimate $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and combination coefficients						
$\tilde{\boldsymbol{w}}_c$	-0.2730	0.0019	1.6903	1.1151	0.0547	-0.7783
$\tilde{\boldsymbol{w}}_{cg}$	-0.1830	0.0534	1.6261	1.0740	0.0996	-0.7359
$\tilde{\boldsymbol{w}}_{cm}$	-0.1705	0.0502	1.5828	1.0339	0.0935	-0.7048
Set 2: apply the shrinkage estimate $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW}$	-0.1925	0.0553	1.6813	1.1167	0.1038	-0.7543
$\tilde{\boldsymbol{w}}_{cg}^{LW}$	-0.1178	0.1024	1.6273	1.0840	0.1458	-0.6945
$\tilde{\boldsymbol{w}}_{cm}^{LW}$	-0.1390	0.0560	1.4995	0.9659	0.0946	-0.6934
Set 3: apply $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW1}$	-0.2611	-0.0159	1.5940	1.0371	0.0322	-0.7817
$\tilde{\boldsymbol{w}}_{cg}^{LW1}$	-0.1476	0.0727	1.5988	1.0577	0.1162	-0.7087
$\tilde{\boldsymbol{w}}_{cm}^{LW1}$	-0.1395	0.0611	1.5577	1.0057	0.1009	-0.6974
Set 4: apply $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}$ in the combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW2}$	-0.1827	0.0872	1.7343	1.1731	0.1388	-0.7497
$\tilde{\boldsymbol{w}}_{cg}^{LW2}$	-0.1386	0.0930	1.6174	1.0824	0.1380	-0.7174
$\tilde{\boldsymbol{w}}_{cm}^{LW2}$	-0.1467	0.0676	1.5395	1.0126	0.1096	-0.6964
Panel C: Other plug-in portfolios						
$\boldsymbol{w}_e$	-0.3015	-0.1965	0.3278	0.1593	-0.1777	-0.7211
$\hat{\boldsymbol{w}}_{mvs}$	-2.5279	-1.9802	0.2445	-0.4266	-1.8885	-1.3990
$\hat{\boldsymbol{w}}_{mvs}^{LW}$	-1.5094	-1.0850	0.9075	0.2770	-1.0104	-1.2205

Table 2.4: Expected out-of-sample MVS utilities (in percentage) for $N = 25$.

(γ_1, γ_2)	$(1, 0.3735)$	$(1.25, 0.3112)$	$(2.5, 0)$	$(2.0984, 0.1)$	$(1.2952, 0.3)$	$(0, 0.6225)$
(q, λ)	(0.2159, 1.7642)	(0.1742, 2.0449)	-	(0.0454, 2.5563)	(0.1668, 2.0893)	-
Panel A: Oracle combined portfolio						
$\hat{\boldsymbol{w}}_c^*$	0.4411	0.7879	2.6397	2.0340	0.8521	-0.5664
$\hat{\boldsymbol{w}}_{cg}^*$	0.2043	0.5092	2.2666	1.6797	0.5672	-0.6070
$\hat{\boldsymbol{w}}_{cm}^*$	0.1568	0.4478	2.1720	1.5918	0.5037	-0.6110
Panel B: Implementable combined portfolios						
Set 1: apply the sample estimate $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and combination coefficients						

$\tilde{\boldsymbol{w}}_c$	0.0880	0.4556	2.3663	1.7481	0.5231	-0.7038
$\tilde{\boldsymbol{w}}_{cg}$	0.0684	0.3913	2.2067	1.6057	0.4521	-0.6742
$\tilde{\boldsymbol{w}}_{cm}$	-0.0304	0.2666	2.0368	1.4399	0.3234	-0.6874
Set 2: apply the shrinkage estimate $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW}$	0.1852	0.5225	2.3846	1.7740	0.5859	-0.6674
$\tilde{\boldsymbol{w}}_{cg}^{LW}$	0.1460	0.4503	2.2341	1.6379	0.5085	-0.6415
$\tilde{\boldsymbol{w}}_{cm}^{LW}$	0.0057	0.2609	1.9529	1.3633	0.3111	-0.6608
Set 3: apply $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW1}$	0.0829	0.4197	2.2627	1.6616	0.4830	-0.6956
$\tilde{\boldsymbol{w}}_{cg}^{LW1}$	0.0905	0.3954	2.1704	1.5804	0.4537	-0.6587
$\tilde{\boldsymbol{w}}_{cm}^{LW1}$	-0.0030	0.2603	2.0075	1.4015	0.3122	-0.6654
Set 4: apply $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}$ in the combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW2}$	0.2279	0.5807	2.4249	1.8264	0.6451	-0.6706
$\tilde{\boldsymbol{w}}_{cg}^{LW2}$	0.1403	0.4522	2.2033	1.6226	0.5104	-0.6554
$\tilde{\boldsymbol{w}}_{cm}^{LW2}$	0.0308	0.3164	2.0124	1.4396	0.3705	-0.6700
Panel C: Other plug-in portfolios						
$\boldsymbol{w}_e$	-0.2939	-0.1900	0.3297	0.1627	-0.1712	-0.7096
$\hat{\boldsymbol{w}}_{mvs}$	-7.9030	-6.4797	-2.5593	-3.5428	-6.2666	-1.8972
$\hat{\boldsymbol{w}}_{mvs}^{LW}$	-4.4777	-3.5036	-0.4246	-1.2586	-3.3521	-1.7649

Table 2.5: Expected out-of-sample MVS utilities (in percentage) for $N = 50$.

(γ_1, γ_2)	$(1, 0.3735)$	$(1.25, 0.3112)$	$(2.5, 0)$	$(2.0984, 0.1)$	$(1.2952, 0.3)$	$(0, 0.6225)$
(q, λ)	(0.2159, 1.7642)	(0.1742, 2.0449)	-	(0.0454, 2.5563)	(0.1668, 2.0893)	-
Panel A: Oracle combined portfolio						
$\hat{\boldsymbol{w}}_c^*$	0.8763	1.2580	3.1819	2.5620	1.3266	-0.4763
$\hat{\boldsymbol{w}}_{cg}^*$	0.5359	0.8945	2.7702	2.1597	0.9599	-0.5489
$\hat{\boldsymbol{w}}_{cm}^*$	0.4011	0.7428	2.5821	1.9787	0.8057	-0.5703
Panel B: Implementable combined portfolios						
Set 1: apply the sample estimate $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and combination coefficients						
$\tilde{\boldsymbol{w}}_c$	0.5209	0.9383	2.9443	2.3087	1.0123	-0.5903
$\tilde{\boldsymbol{w}}_{cg}$	0.4219	0.8015	2.7384	2.1149	0.8703	-0.6043
$\tilde{\boldsymbol{w}}_{cm}$	0.1622	0.5211	2.4363	1.8136	0.5871	-0.6429
Set 2: apply the shrinkage estimate $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW}$	0.6776	1.0672	3.0360	2.4072	1.1378	-0.5615
$\tilde{\boldsymbol{w}}_{cg}^{LW}$	0.5944	0.9654	2.8999	2.2769	1.0334	-0.5363
$\tilde{\boldsymbol{w}}_{cm}^{LW}$	0.1354	0.4268	2.2793	1.6508	0.4835	-0.6366
Set 3: apply $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW1}$	0.4443	0.8246	2.7195	2.1183	0.8934	-0.6207
$\tilde{\boldsymbol{w}}_{cg}^{LW1}$	0.4250	0.7858	2.6486	2.0520	0.8518	-0.5726
$\tilde{\boldsymbol{w}}_{cm}^{LW1}$	0.1349	0.4391	2.3759	1.7233	0.4985	-0.6402
Set 4: apply $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}$ in the combination coefficients						
$\tilde{\boldsymbol{w}}_c^{LW2}$	0.8383	1.2126	3.0533	2.4661	1.2794	-0.5354

$\tilde{\boldsymbol{w}}_{cg}^{LW2}$	0.6024	0.9539	2.7514	2.1725	1.0176	-0.5658
$\tilde{\boldsymbol{w}}_{cm}^{LW2}$	0.3552	0.6871	2.4574	1.8817	0.7482	-0.6089
Panel C: Other plug-in portfolios						
$\boldsymbol{w}_e$	-0.2891	-0.1857	0.3309	0.1649	-0.1671	-0.7025
$\hat{\boldsymbol{w}}_{mvs}$	-60.5886	-51.1542	-31.7297	-35.5664	-49.8273	-
$\hat{\boldsymbol{w}}_{mvs}^{LW}$	-26.2911	-21.8778	-12.1049	-14.1801	-21.2482	-

Tables 2.3-2.5 provide the resulting expected out-of-sample MVS utilities (in percentage) of various portfolios based on the $M = 50,000$ simulated sets of $T \times N$ asset returns when the estimation window length is set to be $T = 120$ and the number of assets is set to be $N = 15, 25, 50$, respectively. In these tables, the six sets of parameter values for γ_1 and γ_2 are set so that the constant η_{mvs} in (2.4) equals 5 for the case with $N = 25$. Meanwhile, "$(\gamma_1, \gamma_2) = (2.5, 0)$" corresponds to the MV criterion, and "$(\gamma_1, \gamma_2) = (0, 0.6225)$" is a case for the MStD criterion. Notably, for $N = 50$ in Table 2.5, the plug-in MVS portfolio is not achievable when $(\gamma_1, \gamma_2) = (0, 0.6225)$, since the required condition $\gamma_2^2 > \hat{\psi}^2$ in Assumption 2.1 is not satisfied for some simulated sets, where $\hat{\psi}^2$ is the estimate of ψ^2 with the sample mean and sample covariance matrix (or shrinkage covariance matrix) plugged in. For the other four sets of values of (γ_1, γ_2), the equivalent confidence level q and weighting parameter λ in the definition of the TMV risk measure are also provided in the three tables. Panel A of the three tables lists the expected out-of-sample MVS utilities of the oracle combined portfolios, and they represent the highest expected out-of-sample MVS utilities that the combined portfolios in the forms of (2.7), (2.9), (2.10) can respectively achieve. Panels B and C tabulate the expected out-of-sample MVS utilities for the implementable combined portfolios and the benchmark portfolios, respectively.

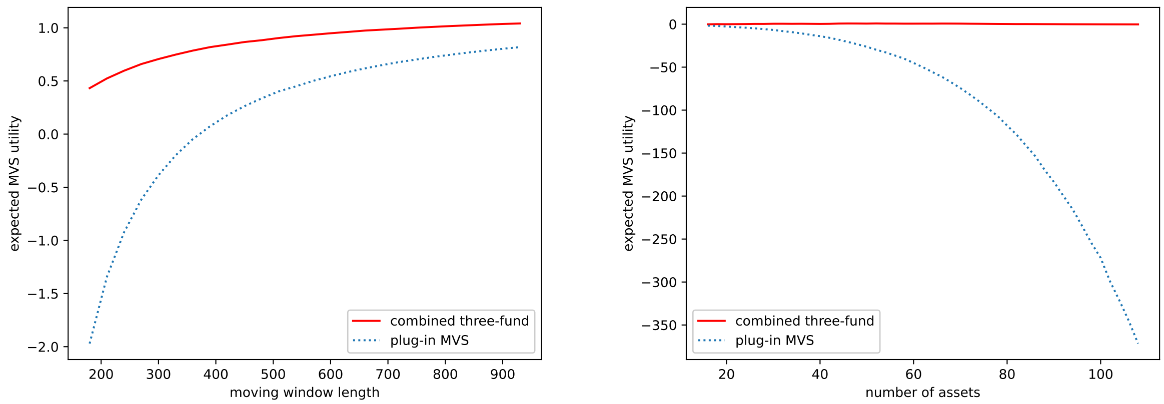
A comparison of Panels B and C in Tables 2.3-2.5 shows that our proposed implementable combined portfolios consistently outperform the plug-in MVS portfolios (with either sample estimate or shrinkage estimate of the covariance matrix) by a considerable margin, irrespective of the number of assets. The outperformance is evident not only in the combined three-fund portfolios but also in the combined global minimum variance portfolio and the combined market portfolio. Furthermore, all the proposed implementable combined portfolios generally yield higher expected out-of-sample MVS utilities compared to the 1/N portfolio. The only exception happens with the combined portfolios in Table 2.3 when $(\gamma_1, \gamma_2) = (0, 0.6225)$. Even in this exceptional case, some combined portfolios still outperform the 1/N portfolio and the underperformance is minimal when it occurs. These results imply that there is indeed value added through the combination of constituent portfolios.

Furthermore, note that there are four sets of implementable combined portfolios in Tables 2.3-2.5. These sets of portfolios are distinguished by their different ways of utilizing either the sample or shrinkage estimate of the covariance matrix for the constituent portfolios and combination coefficients. It is insightful to compare the four portfolio sets to assess the impact of using different estimates of the covariance matrix. When utilizing the same combination coefficients, the combined three-fund portfolios and the combined global minimum variance portfolios that employ the shrinkage covariance matrix $\widehat{\mathbf{\Sigma}}_{LW}$ for the constituent portfolios generally outperform those using the sample covariance matrix $\widehat{\mathbf{\Sigma}}$ (compare Set 1 with Set 4, and Set 3 with Set 2). However, there is no significant performance advantage in favor of using $\widehat{\mathbf{\Sigma}}_{LW}$ over $\widehat{\mathbf{\Sigma}}$ for the combined market portfolios. In many cases, both estimates used in the constituent portfolios lead to comparable expected out-of-sample MVS utilities, when the combination coefficients are the same. Therefore, it might be generally wise to adopt the shrinkage covariance matrix instead of the sample covariance matrix in the constituent portfolios when we construct implementable portfolios. Notably, Kan et al. (2022) also consider the shrinkage covariance matrix in their constituent portfolios and observe improved expected out-of-sample MV utilities.

Lastly, comparing portfolios within each set across Tables 2.3-2.5 offers intriguing insights into the relative performance of the three combined portfolios. When $N = 15$, as shown in Table 2.3, there is no consistent outperformance among the combined portfolios. The combined three-fund portfolio occasionally performs the best, while in other cases, it is the combined global minimum variance portfolio that outperforms. The combined market portfolio typically lags behind at least one of the other two. When the number of assets is large, as seen in Tables 2.4 and 2.5 (for $N = 25$ and $N = 50$ respectively), the combined three-fund portfolio usually exhibits the highest expected out-of-sample MVS utilities compared to the other two combined portfolios in most cases. Exceptions are primarily observed in the last column of Table 2.4. Even in these exceptional cases, the underperformance of the combined three-fund portfolio is slight. This observation suggests that the combined three-fund portfolio tends to outperform the other two when the portfolio includes a large number of assets.

Figure 2.1 depicts the effect of the estimation window length T and the number of assets N on the expected out-of-sample utilities for both the implementable combined three-fund

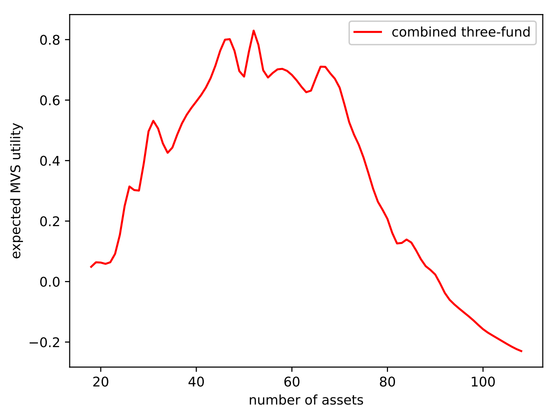
Figure 2.1: Expected out-of-sample MVS utility with the change of T and N.



(a) $N = 25$ **(b)** $T = 120$

This figure presents the expected out-of-sample MVS utilities (in percentage) of the implementable combined three-fund portfolio ($\tilde{\boldsymbol{w}}_c^{LW}$) and the plug-in MVS portfolio ($\hat{\boldsymbol{w}}_{mvs}^{LW}$). The risk aversion coefficients are set at $(\gamma_1, \gamma_2) = (1, 0.3735)$. Panel (a) shows the performance against the estimation window length with a fixed number of assets $N = 25$, and Panel (b) illustrates the performance against the number of assets with a fixed estimation window length $T = 120$.

Figure 2.2: Expected out-of-sample MVS utility for the combined three-fund portfolio.



This figure presents the expected out-of-sample MVS utilities (in percentage) of the implementable combined three-fund portfolio ($\tilde{\boldsymbol{w}}_c^{LW}$) against the number of assets. The risk aversion coefficients are set at $(\gamma_1, \gamma_2) = (1, 0.3735)$, and the estimation window is fixed at $T = 120$.

portfolio and the plug-in MVS portfolio. In accordance with expectations, Figure 2.1a confirms that performance improves as the estimation window length increases, highlighting the benefits of using more historical data for accurate estimation. Significantly, our implementable combined three-fund portfolio consistently outperforms the plug-in MVS portfolio across all estimation window lengths. Additionally, Figure 2.1b demonstrates that as the number of assets increases, the performance of the plug-in MVS portfolio deteriorates, whereas the performance of the implementable combined three-fund portfolio remains robust. This finding underscores the superiority and stability of combining portfolio rules. To offer a clearer view of the expected out-of-sample utility in relation to the number of assets for the combined three-fund portfolio, we present an enlarged curve in Figure 2.2.

2.5.2 Empirical Study

In this section, we perform empirical studies using monthly returns data from seven datasets, as described in Table 2.6, with the first five datasets sourced from the Kenneth French data library.[4] The last two datasets consist of portfolio returns derived from the long and short legs of specific anomalies studied in Novy-Marx and Velikov (2016), which are downloaded from the Robert Novy-Marx's website.[5] The first column of Table 2.6 provides an abbreviation for each dataset, while the remaining columns provide a description, the number of assets, and the data period, respectively.

Table 2.6: List of empirical datasets.

Abbreviation	Dataset	N	period
SM6	6 Portfolios Formed on Size and Momentum	6	1970/01-2020/01
M10	10 Portfolios Formed on Momentum	10	1970/01-2020/01
OPI25	25 Portfolios Formed on Operating Profitability and Investment	25	1970/01-2020/01
SOI32	32 Portfolios Formed on Size, Operating Profitability, and Investment	32	1970/01-2020/01
SBM100	100 Portfolios Formed on Size and Book-to-Market	100	1970/01-2020/01
L16	The long and short legs of eight low-turnover anomalies	16	1963/07-2013/12
A46	The long and short legs of all 23 anomalies	46	1973/07-2013/12

[4] The first five datasets in Table 2.6 are downloaded from the following link: `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[5] The last two datasets in Table 2.6 are downloaded from the following link: `https://simon.rochester.edu/digital-measures-faculty/robert-novy-marx`.

Given that the true mean and covariance matrix are unknown for empirical datasets, we adopt a rolling window approach to evaluate the out-of-sample performances of portfolios. Specifically, starting from the $(T+1)$-th month and for each subsequent month, we utilize data from the preceding T months to calculate the sample mean vector, sample covariance matrix, and shrinkage covariance matrix for asset returns. We then construct and implement the portfolios for the corresponding month. If a dataset spans X months in total, this rolling window approach yields $(X-T)$ out-of-sample returns for each portfolio $\hat{\boldsymbol{w}}$. We calculate the sample mean (denoted by $\hat{\mu}_{\hat{\boldsymbol{w}}}$) and sample variance (denoted by $\hat{\sigma}_{\hat{\boldsymbol{w}}}$) of these $(X-T)$ out-of-sample portfolio returns as the expected out-of-sample mean and variance, respectively. The empirical MVS utility for any portfolio $\hat{\boldsymbol{w}}$ then is computed as $\hat{\mu}_{\hat{\boldsymbol{w}}} - \gamma_1 \hat{\sigma}^2_{\hat{\boldsymbol{w}}} - \gamma_2 \hat{\sigma}_{\hat{\boldsymbol{w}}}$, and its negation is referred to as the empirical TMV risk value.

Tables 2.7 and 2.8 present the empirical TMV risk values of various portfolios listed in the Panels B and C of Table 2.2. We set the length of the rolling window at $T = 120$

Table 2.7: Empirical TMV risk values (in percentage) when $T = 120$.

	Dataset						
	SM6	M10	OPI25	SOI32	SBM100	L16	A46
Panel A: Implementable combined portfolios							
Set 1: apply the sample estimate $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and combination coefficients							
$\tilde{\boldsymbol{w}}_c$	5.4122	6.0310	5.4433	5.1751	7.1359	5.7968	5.2151
$\tilde{\boldsymbol{w}}_{cg}$	5.5160	6.0345	5.4100	5.1685	7.2389	5.8407	5.5495
$\tilde{\boldsymbol{w}}_{cm}$	6.6773	6.7596	6.2940	6.2208	7.5300	6.8716	6.9571
Set 2: apply the shrinkage estimate $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and combination coefficients							
$\tilde{\boldsymbol{w}}_c^{LW}$	5.7518	5.9002	5.2155	4.7989	8.3276	5.7079	4.8749
$\tilde{\boldsymbol{w}}_{cg}^{LW}$	5.8161	5.9453	5.2054	4.8378	8.3493	5.7160	5.2987
$\tilde{\boldsymbol{w}}_{cm}^{LW}$	6.9501	6.6024	6.1436	6.0163	8.0345	6.8963	6.6492
Set 3: apply $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the combination coefficients							
$\tilde{\boldsymbol{w}}_c^{LW1}$	5.8481	6.0599	5.4971	5.4324	11.0509	6.0279	5.7400
$\tilde{\boldsymbol{w}}_{cg}^{LW1}$	5.8928	6.0630	5.4658	5.4594	11.2275	6.0887	6.1740
$\tilde{\boldsymbol{w}}_{cm}^{LW1}$	6.8047	6.8142	6.2613	6.2302	7.9957	6.8931	6.6555
Set 4: apply $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}$ in the combination coefficients							
$\tilde{\boldsymbol{w}}_c^{LW2}$	5.8338	5.8706	5.2011	4.9348	7.6956	5.8217	5.2350
$\tilde{\boldsymbol{w}}_{cg}^{LW2}$	5.9264	5.9191	5.2043	4.9967	7.7096	5.8491	5.7214
$\tilde{\boldsymbol{w}}_{cm}^{LW2}$	6.9129	6.5525	6.1109	5.9331	7.8514	6.8858	6.4390
Panel B: Other plug-in portfolios							
$\boldsymbol{w}_e$	7.7882	7.3369	6.9278	7.2910	8.0397	8.2629	8.5412
$\hat{\boldsymbol{w}}_{mvs}$	5.5608	6.1932	6.0931	5.9048	2812.2717	6.1215	53.2100
$\hat{\boldsymbol{w}}_{mvs}^{LW}$	6.0319	6.6156	6.4703	6.5851	11.1154	6.1961	7.2991

Table 2.8: Empirical TMV risk values (in percentage) when $T = 240$.

	Dataset						
	SM6	M10	OPI25	SOI32	SBM100	L16	A46
Panel A: Implementable combined portfolios							
Set 1: apply the sample estimate $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and combination coefficients							
$\tilde{\boldsymbol{w}}_c$	5.6998	5.8100	5.4881	5.1609	5.3653	5.2696	5.4345
$\tilde{\boldsymbol{w}}_{cg}$	5.7637	5.8685	5.5070	5.1969	5.5720	5.3586	5.0920
$\tilde{\boldsymbol{w}}_{cm}$	6.4312	6.3660	5.9550	5.8627	6.1796	5.8822	7.1295
Set 2: apply the shrinkage estimate $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and combination coefficients							
$\tilde{\boldsymbol{w}}_c^{LW}$	5.1995	5.6480	5.3552	4.9481	5.1219	5.1802	4.3141
$\tilde{\boldsymbol{w}}_{cg}^{LW}$	5.3923	5.7163	5.3789	4.9750	5.2700	5.2342	4.4935
$\tilde{\boldsymbol{w}}_{cm}^{LW}$	6.2786	6.1474	5.7999	5.6553	6.2005	5.7057	5.8567
Set 3: apply $\widehat{\boldsymbol{\Sigma}}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the combination coefficients							
$\tilde{\boldsymbol{w}}_c^{LW1}$	6.0991	5.7939	5.5085	5.3346	5.7515	5.3565	6.2090
$\tilde{\boldsymbol{w}}_{cg}^{LW1}$	5.9723	5.8556	5.5283	5.3711	5.9455	5.4521	5.7337
$\tilde{\boldsymbol{w}}_{cm}^{LW1}$	6.3874	6.3753	5.9613	5.9057	6.1193	5.9138	6.9117
Set 4: apply $\widehat{\boldsymbol{\Sigma}}_{LW}$ in the constituent portfolios and $\widehat{\boldsymbol{\Sigma}}$ in the combination coefficients							
$\tilde{\boldsymbol{w}}_c^{LW2}$	5.2485	5.6579	5.3423	4.9083	5.4393	5.2090	4.2790
$\tilde{\boldsymbol{w}}_{cg}^{LW2}$	5.4427	5.7269	5.3664	4.9541	5.6223	5.2843	4.6518
$\tilde{\boldsymbol{w}}_{cm}^{LW2}$	6.2730	6.1506	5.7948	5.6322	6.1055	5.7216	5.9002
Panel B: Other plug-in portfolios							
$\boldsymbol{w}_e$	7.8666	7.3521	6.7918	7.1757	8.1691	8.0054	8.7756
$\hat{\boldsymbol{w}}_{mvs}$	5.8751	6.0034	5.8208	5.4785	7.6620	5.3280	8.1222
$\hat{\boldsymbol{w}}_{mvs}^{LW}$	5.6099	6.4800	6.5847	6.8156	6.6703	5.6119	7.5460

and $T = 240$ months, respectively. The parameters in the MVS criterion is $(\gamma_1, \gamma_2) = (0.8457, 1.7550)$, which is equivalent to setting $(q, \lambda) = (0.9, 5)$ in the TMV risk measure as studied in Landsman (2010). Notably, a smaller value in the table indicates better portfolio performance, since the empirical TMV risk value is the negation of the empirical MVS utility.

From the two tables, we have the following observations regarding the relative performance of the various portfolios. Firstly, the implementable combined three-fund portfolio consistently outperforms the plug-in MVS and 1/N portfolios across all the datasets, with the exception of the dataset SM6, which has the smallest number of assets compared to the other datasets. Even for the dataset SM6, the combined three-fund portfolio outperforms the 1/N portfolio in all the cases considered in Tables 2.7 and 2.8, and it outperforms the plug-in MVS portfolios for most cases when the estimation window length is $T = 240$. Secondly, the combined three-fund portfolio performs better than the combined two-fund portfolios ($\hat{\boldsymbol{w}}_{cg}$ and $\hat{\boldsymbol{w}}_{cm}$) for most cases. It consistently outperforms the combined market portfolio over almost all the scenarios. The only exception happens in the dataset SBM100 at Table 2.7. The combined three-fund portfolio only underperforms the combined global minimum variance portfolio in a few cases and when the underperformance happens, the margin is rather small. Thirdly, the combined three-fund portfolio that uses the shrinkage covariance matrix in the constituent portfolios (in either Set 2 or 4) generally provides the smallest TMV risk value. The only exception occurs when we apply the shrinkage covariance matrix with a smaller estimation window length $T = 120$ for the dataset SM6.

2.6 Conclusion

This chapter studies the TMV portfolio optimization under parameter uncertainty. We work on the proposed MVS optimization framework and derive the optimal MVS portfolio using its inherent connection with the traditional MV framework. To matigate estimation in the plug-in MVS portfolio, we introduce a combined three-fund portfolio strategy $\hat{\boldsymbol{w}}_c$, which incorporates the 1/N rule, along with the two other constituent portfolios corresponding to the components from the plug-in MVS portfolio. We obtain the optimal combination coefficients by maximizing the expected out-of-sample MVS utility under the multivariate normal distribution assumption of asset returns. Through the simulation study, we find that the proposed combined three-fund portfolio consistently outperforms both the plug-in MVS and 1/N portfolios by a significant margin in terms of the expected MVS utility. Meanwhile, in empirical studies where the normality assumption of asset returns typically no longer holds, the proposed combined three-fund portfolio can still outperform the plug-in MVS and 1/N portfolios in terms of the empirical TMV risk.

Chapter 3

ESG Investing with Estimation Risk

3.1 Introduction

In recent years, there has been a significant increase in the commitment of institutional investors to responsible investment. ESG investing emphasizes the effective management of environmental, social, and governance risks to support long-term investment sustainability. Recent research works have broadened the incorporation of ESG criteria into investment analysis. For example, Pedersen et al. (2021) introduce the concept of an ESG-efficient frontier, which identifies the highest attainable Sharpe ratio for each ESG level. Pástor et al. (2021) offer a tractable equilibrium model that incorporates ESG criteria and provide a number of empirical implications. Avramov et al. (2022) analyze the portfolio implications that account for the uncertainty about the corporate ESG profile. Furthermore, De Spiegeleer et al. (2023) examine the impact of including ESG criterion on the allocation of equity portfolios, with a focus on the risk and return characteristics of the portfolios.

While theoretically sound, the practical application of the theoretical ESG portfolio strategies from the bulk of the literature often encounters challenges due to the reliance on the true parameters (i.e., the mean vector and covariance matrix) of asset returns. These parameters are rarely fully known to investors and must be estimated, similar to the discussion in Chapter 1. This inherent estimation risk could significantly affect the effectiveness of these investment strategies. Given that there is a lack of studies addressing parameter uncertainty in ESG investing, this chapter is to explore the optimal ESG portfolio selection problem while accounting for estimation risk. Note that ESG scores are generally used to assess a company's sustainability and ethical performance, and are obtainable from various rating agencies. Although ESG scores may vary among different agencies, in this chapter, we assume that investors select their preferred ESG rating agency and consistently apply its scores. Given that ESG scores are specifically linked to individual firms, there is no universally accepted ESG score for risk-free assets, which typically include U.S. Treasury securities (such as Treasury bonds, notes, and bills). Consequently, we exclude the risk-free asset from the portfolio, thereby making the sum of the portfolio weights on risky assets equals one.

As one of the primary contributions of this study, we develop an ESG-constrained Mean-Variance (MV) optimization framework that ensures the total ESG score of the portfolio meets a prespecified target ESG score. The resulting optimal ESG portfolio adheres to a three-fund separation, where each constituent portfolio meets specific weight and ESG criteria. Additionally, the utility derived from holding the optimal ESG portfolio can be expressed as a quadratic function of the target ESG score. This enable us to identify a specific ESG score that maximizes the utility. When the target ESG score deviates from this specific value, the utility is reduced. Setting the ESG score to this specific value ensures that the optimal ESG portfolio aligns with the one achieved by the traditional MV optimization without the ESG constraint.

Given that the mean vector and the covariance matrix have to be estimated from historical data, the resulting plug-in ESG portfolio is a function of the samples and thus is random. To mitigate the associated estimation risk, we propose a combined rule which consists of three constituent portfolios corresponding to the components in the plug-in ESG portfolio. These constituent portfolios are specifically tailored to facilitate computations and ensure compliance with the weight constraint. Inspired by Kan and Zhou (2007), we utilize the expected out-of-sample MV utility as the optimization criterion to determine the optimal combination coefficients. Due to the structure of our proposed combined portfolio and the assumption of constant combination coefficients, it is infeasible to meet the pointwise ESG constraint in the same manner as the aforementioned ESG-constrained MV framework. Instead, we adopt an inequality expected ESG constraint, which requires the expected out-of-sample ESG score of the combined portfolio to meet or exceed a specified target. We analytically derive the optimal combination coefficients under the multivariate normal distribution assumption of asset returns. Additionally, we can readily extend the results from this inequality constraint to accommodate an equality expected ESG constraint, which is appealing to investors who prefer the expected total ESG score of the portfolio to precisely meet the target.

The optimal combination coefficients involve several unknown parameters that have to be estimated for practical application. We further illustrate how sophisticated estimators and a shrinkage covariance matrix can be used to estimate the combination coefficients and facilitate the construction of an implementable combined portfolio. Interestingly, when these sophisticated estimators are used to calculate the estimated combination coefficients, the resulting implementable version of the optimal combined portfolio, derived from the equality expected ESG constraint, coincidentally satisfies the pointwise ESG constraint in a similar manner as the previously described ESG-constrained MV framework.

We test the performance of our proposed portfolio strategies with both simulated and real-world datasets. Our results demonstrate that the combined portfolios employing sophisticated estimators generally perform better than those using sample estimators in estimating the optimal combination coefficients. Also, the implementable combined portfolios consistently outperform the plug-in ESG portfolios in terms of the MV utility and Sharpe ratio.

The rest of the chapter is organized as follows. Section 3.2 discusses the ESG constrained MV framework and some of its properties. Section 3.3 introduces the estimation risk and provides the combination rule. Section 3.4 derives the optimal combined portfolios using the expected out-of-sample MV utility criterion with the expected ESG constraints. Section 3.5 illustrates the performance of combined portfolios based on simulation and empirical studies. All proofs are placed in Appendix B.

3.2 ESG Investing

Consider an investor who allocates wealth among $N \geq 2$ risky assets over a single trading period. That is, the investor trades the risky assets at time 0, with the objective of optimizing utility under certain constraints by time 1. Let $\boldsymbol{r} = [r_1, ..., r_N]^\top$ be the asset return vector, where r_i is the return of risky asset i, $i = 1, \ldots, N$. Assume that the asset return vector follows a multivariate normal distribution with an N-dimensional mean vector $\boldsymbol{\mu} \in \mathbb{R}^N$ and a covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ of full rank. Additionally, denote the ESG scores of the risky assets as $\boldsymbol{S} = (S_1, .., S_N)^\top$. Assume that $\boldsymbol{\mu}$ is not collinear with $\mathbf{1}$, where $\mathbf{1}$ is an N-dimensional vector of ones, and that $\boldsymbol{\Sigma}$ is a symmetric positive definite and non-singular matrix. A portfolio strategy is represented by a vector of investment weights (or fractions) $\boldsymbol{w} = [w_1, ..., w_N]^\top$, where w_i is the proportion of wealth allocated to the risky asset i, $i = 1, \ldots, N$.

For the scope of this chapter, our focus is on investments excluding the risk-free asset, as there is no commonly accepted ESG scores for risk-free assets. Consequently, the portfolio $\boldsymbol{w} \in \boldsymbol{R}^N$ considered in this chapter satisfy the weight constraint:

$$\boldsymbol{w}^\top \mathbf{1} = 1.$$

Furthermore, even though there is a wealth of literature discussing the ambiguity of the ESG scores due to variations across different rating agencies (see, for example, Berg et al., 2022; Avramov et al., 2022; Luo et al., 2023, among others), we assume that the investor has selected a reliable rating agency and consistently uses the ESG scores provided by this agency. As the portfolio satisfying the weight constraint is overall long, we evaluate the ESG score of the portfolio by assessing the collective ESG performance of all the assets within the portfolio. Specifically, we compute a weighted average of these ESG scores based on the proportion of each asset within the portfolio, represented as $\boldsymbol{w}^\top \boldsymbol{S}$. This calculation yields an ESG score that reflects the overall ESG performance of the portfolio. For assets held in long positions, their ESG scores positively contribute to the total ESG score of the portfolio, with higher scores indicating a better alignment with sustainability practices. Conversely, when assets are held in short positions, we assume that holding a short position is to bet against the company's sustainability, and hence, we subtract the ESG scores of shorted assets from the total ESG score of the portfolio. Notably, Pedersen et al. (2021) considers the average ESG score of the portfolio as $\boldsymbol{w}^\top \boldsymbol{S} / \boldsymbol{w}^\top \mathbf{1}$. In our analysis, with the weight constraint that $\boldsymbol{w}^\top \mathbf{1} = 1$, this average ESG score simplifies to the total ESG score.

3.2.1 ESG-Constrained MV Framework

We assume that ESG investors integrate ESG scores of risky assets into their investment analysis. Specifically, we assume that the investors are concerned with the overall ESG performance of the portfolio and require that the total ESG score meets a specified target, denoted as

$$\boldsymbol{w}^\top \boldsymbol{S} = \bar{s},$$

where $\bar{s}$ is the target ESG score. This target value reflects the investor's ESG preferences, with a higher $\bar{s}$ indicating stronger ESG preferences. Including this ESG constraint facilitates backtesting by enabling investors to efficiently verify that their portfolios align with the predefined ESG score over time. Consequently, we incorporate this ESG criterion into portfolio construction, and we adapt the Mean-Variance (MV) optimization framework with both weight and ESG constraints as follows:[1]

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \\ \text{s.t.} \quad & \boldsymbol{w}^\top \boldsymbol{S} = \bar{s}, \quad \boldsymbol{w}^\top \mathbf{1} = 1, \end{aligned} \tag{3.1}$$

where $\gamma > 0$ is a risk aversion coefficient and $\bar{s}$ is the target ESG score. The set of feasible portfolios is defined as $\mathcal{F}_0 = \{\boldsymbol{w} \in \mathbb{R}^N : \boldsymbol{w}^\top \boldsymbol{S} = \bar{s}, \boldsymbol{w}^\top \mathbf{1} = 1\}$.

The ESG-constrained MV framework, detailed in (3.1), presents distinct difference over existing models. Firstly, Pedersen et al. (2021) extend the MV objective by incorporating an ESG preference function that evaluates the average ESG score of the portfolio, and limit to long-biased portfolios. In contrast, our framework aims to meet a specified total ESG score for the portfolio. Secondly, Pástor et al. (2021) introduce an exponential utility that accounts for the ESG attributes of individual assets and require the construction of ESG characteristics to assess firms' social impacts. Our formulation, however, efficiently utilizes available ESG scores of risky assets. Thirdly, these studies integrate the risk-free asset in their portfolios and assume it contributes no ESG utility to investors, while our framework intentionally excludes the risk-free asset due to the lack of universally accepted ESG scores for such assets in the market.

Assume that $\mathbf{1}^\top \boldsymbol{\Omega} \boldsymbol{S} \neq 0$, where $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{S}\mathbf{1}^\top - \mathbf{1}\boldsymbol{S}^\top)\boldsymbol{\Sigma}^{-1}$. We employ the Lagrangian

[1] Given that we focus on a single-period optimization, we do not consider modeling the investor's preferences as being state nonseparable, time nonseparable, or both.

method for the optimization problem (3.1) and analytically determine the optimal portfolio weights, as presented in Theorem 3.1.

Theorem 3.1. *The optimal ESG portfolio that satisfies the optimization problem* (3.1) *is*

$$\boldsymbol{w}_{esg}^{*} = \boldsymbol{w}_a + \bar{s}\boldsymbol{w}_e + \frac{1}{\gamma}\boldsymbol{w}_o, \tag{3.2}$$

where the constituent portfolios are the weight-constrained portfolio $\boldsymbol{w}_a$, the ESG-constrained portfolio $\boldsymbol{w}_e$, and the zero-weight-ESG portfolio $\boldsymbol{w}_o$, respectively, defined as follows:

$$\boldsymbol{w}_a = \frac{\boldsymbol{\Omega S}}{\mathbf{1}^\top \boldsymbol{\Omega S}}, \qquad \boldsymbol{w}_e = -\frac{\boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega S}}, \tag{3.3a}$$

$$\boldsymbol{w}_o = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega S}}, \tag{3.3b}$$

with $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{S}\mathbf{1}^\top - \mathbf{1}\boldsymbol{S}^\top)\boldsymbol{\Sigma}^{-1}$.

Proof. See Appendix B.1.1. □

Theorem 3.1 implies a three-fund separation in the optimal portfolio, which comprises three distinct constituent portfolios: the weight-constrained portfolio $\boldsymbol{w}_a$, the ESG-constrained portfolio $\boldsymbol{w}_e$, and the zero-weight-ESG portfolio $\boldsymbol{w}_o$. According to the formulation in (3.2), investors are advised to hold 100% of $\boldsymbol{w}_a$, while the exposures to $\boldsymbol{w}_e$ and $\boldsymbol{w}_o$ depend on the targeted ESG score $\bar{s}$ and risk aversion parameter γ.

3.2.2 Properties of the ESG-Constrained MV Framework

In this subsection, we provide three properties regarding the ESG-constrained MV framework, including the utility from holding the optimal ESG portfolio, the specific target ESG score such that the utility is maximized, and the unique features regarding the three constituent portfolios in (3.3a) and (3.3b).

Denote the return of any portfolio $\boldsymbol{w}$ as $R = \boldsymbol{w}^\top \boldsymbol{r}$. The mean and variance of the portfolio return are respectively denoted by $\mu_{\boldsymbol{w}} = \boldsymbol{w}^\top \boldsymbol{\mu}$ and $\sigma_{\boldsymbol{w}}^2 = \boldsymbol{w}^\top \boldsymbol{\Sigma}\boldsymbol{w}$. Consequently, the mean and variance of the three constituent portfolios in (3.3a) and (3.3b) can be expressed as follows:

$$\mu_{\boldsymbol{w}_e} = \boldsymbol{w}_e^\top \boldsymbol{\mu}, \qquad \mu_{\boldsymbol{w}_a} = \boldsymbol{w}_a^\top \boldsymbol{\mu}, \qquad \mu_{\boldsymbol{w}_o} = \boldsymbol{w}_o^\top \boldsymbol{\mu},$$
$$\sigma^2_{\boldsymbol{w}_e} = \boldsymbol{w}_e^\top \boldsymbol{\Sigma} \boldsymbol{w}_e, \qquad \sigma^2_{\boldsymbol{w}_a} = \boldsymbol{w}_a^\top \boldsymbol{\Sigma} \boldsymbol{w}_a, \qquad \sigma^2_{\boldsymbol{w}_o} = \boldsymbol{w}_o^\top \boldsymbol{\Sigma} \boldsymbol{w}_o,$$

where we can easily use the expression of $\boldsymbol{w}_o$ to verify that $\mu_{\boldsymbol{w}_o} = \sigma^2_{\boldsymbol{w}_o}$. Additionally, we define the covariances as follows: $\sigma^2_{\boldsymbol{w}_{a,e}} = \boldsymbol{w}_a^\top \boldsymbol{\Sigma} \boldsymbol{w}_e$, $\sigma^2_{\boldsymbol{w}_{a,o}} = \boldsymbol{w}_a^\top \boldsymbol{\Sigma} \boldsymbol{w}_o$ and $\sigma^2_{\boldsymbol{w}_{e,o}} = \boldsymbol{w}_e^\top \boldsymbol{\Sigma} \boldsymbol{w}_o$. Note that we can easily verify $\sigma^2_{\boldsymbol{w}_{a,o}} = \sigma^2_{\boldsymbol{w}_{e,o}} = 0$. Using these notations, Corollary 3.1 presents the utility from holding the optimal ESG portfolio (3.2).

Corollary 3.1. *The mean-variance utility from holding the optimal ESG portfolio* (3.2) *is*

$$U^*_{esg} = \bar{s}\mu_{\boldsymbol{w}_e} + \mu_{\boldsymbol{w}_a} - \frac{\gamma}{2}\left(\bar{s}^2\sigma^2_{\boldsymbol{w}_e} + 2\bar{s}\sigma^2_{\boldsymbol{w}_{a,e}} + \sigma^2_{\boldsymbol{w}_a}\right) + \frac{\mu_{\boldsymbol{w}_o}}{2\gamma}. \tag{3.4}$$

Proof. See Appendix B.1.2. □

Note that holding the combination of the two constituent portfolios $\boldsymbol{w}_a + \bar{s}\boldsymbol{w}_e$ lead to an utility:

$$\bar{s}\mu_{\boldsymbol{w}_e} + \mu_{\boldsymbol{w}_a} - \frac{\gamma}{2}\left(\bar{s}^2\sigma^2_{\boldsymbol{w}_e} + 2\bar{s}\sigma^2_{\boldsymbol{w}_{a,e}} + \sigma^2_{\boldsymbol{w}_a}\right).$$

Hence, holding the optimal ESG portfolio $\boldsymbol{w}^*_{esg}$ yields higher utility than holding $\boldsymbol{w}_a + \bar{s}\boldsymbol{w}_e$ by the amount $\mu_{\boldsymbol{w}_o}/2\gamma$, which comes from the exposure to $\boldsymbol{w}_o$. Additionally, the utility U^*_{esg} is influenced by both the risk aversion coefficient γ and the total ESG score $\bar{s}$, and can be expressed as a quadratic function of $\bar{s}$. In Proposition 3.1, we provide a specific target ESG score that investors may adopt to achieve a maximum amount of U^*_{esg}.

Proposition 3.1. *We have the following assertions:*

1. *The utility U^*_{esg} given in* (3.4) *achieves the largest value at*

$$\bar{s}^* = \frac{\mu_{\boldsymbol{w}_e} - \gamma\sigma^2_{\boldsymbol{w}_{a,e}}}{\gamma\sigma^2_{\boldsymbol{w}_e}}. \tag{3.5}$$

2. *When $\bar{s} = \bar{s}^*$ in the optimization problem (3.1), the optimal ESG portfolio coincides with the one from the following MV optimization problem:*

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \\ s.t. \quad & \boldsymbol{w}^\top \mathbf{1} = 1. \end{aligned} \tag{3.6}$$

Proof. See Appendix B.1.3. □

As shown in Proposition 3.1, when the investor sets the target ESG score as (3.5), i.e., $\bar{s} = \bar{s}^*$, they achieve the maximum value of the utility U^*_{esg}. Notably, in this setting, the optimal ESG portfolio corresponds exactly with the one derived from the traditional MV optimization with only the weight constraint. This alignment indicates that the optimal portfolio obtained from (3.6) inherently achieves a total ESG score $\bar{s}^*$. Meanwhile, the MV frontiers of the optimization problems detailed in (3.1) when $\bar{s} = \bar{s}^*$ and (3.6) overlap. In contrast, when the target ESG score is not $\bar{s}^*$, i.e., $\bar{s} \neq \bar{s}^*$, the resulting utility U^*_{esg} is lower than that achieved by holding the optimal portfolio from (3.6). This underscores the principle that including additional constraints typically reduces utility.

Finally, Proposition 3.2 describes the specific properties of the three constituent portfolios identified in Theorem 3.1. It highlights how $\boldsymbol{w}_e$ and $\boldsymbol{w}_a$ can be derived from specific optimization problems, which minimize portfolio variance under certain constraints related to the total ESG score and portfolio weights. The names for these two constituent portfolios originates from their respective constraints: the ESG-constrained portfolio ($\boldsymbol{w}_e$) achieves a total ESG score of one while maintains a zero total portfolio weight, whereas the weight-constrained portfolio ($\boldsymbol{w}_a$) requires the total portfolio weight to be one with a zero ESG score. Concurrently, the zero-weight-ESG portfolio ($\boldsymbol{w}_o$) is characterized by its unique attributes regarding the total ESG score and portfolio weights, which has both a zero total ESG score and zero total portfolio weight.

Proposition 3.2. *We have the following properties for the three constituent portfolios in (3.3a) and (3.3b). :*

1. *The ESG-constrained portfolio $\boldsymbol{w}_e$ is the optimal solution to the following optimization problem:*

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \\ s.t. \quad & \boldsymbol{w}^\top \boldsymbol{S} = 1, \quad \boldsymbol{w}^\top \mathbf{1} = 0. \end{aligned} \tag{3.7}$$

2. *The weight-constrained portfolio* $\boldsymbol{w}_a$ *is the optimal solution to the following optimization problem:*

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \\ s.t. \quad & \boldsymbol{w}^\top \boldsymbol{S} = 0, \quad \boldsymbol{w}^\top \mathbf{1} = 1. \end{aligned} \tag{3.8}$$

3. *The zero-weight-ESG portfolio* $\boldsymbol{w}_o$ *satisfies* $\boldsymbol{w}_o^\top \boldsymbol{S} = 0$ *and* $\boldsymbol{w}_o^\top \mathbf{1} = 0$.

Proof. See Appendix B.1.4. □

3.3 Estimation Risk and Combination Rules

In real-world markets, the implementation of the optimal ESG portfolio (3.2) faces challenges due to the unknown mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of asset returns. The plug-in method provides a practical approach to address this issue. It involves first calculating the sample mean vector and sample covariance matrix based on historical data, and then these estimates are used to replace the unknown parameters in the optimal ESG portfolio to obtain an implementable one.

Suppose we have historical asset returns of the N risky assets in the previous T periods, and denote them by $\{\boldsymbol{r}_1, \cdots, \boldsymbol{r}_T\}$, where $\boldsymbol{r}_t$ is the N-dimensional asset return vector for period $t, t = 1, ..., T$. Assume that $\{\boldsymbol{r}_1, \cdots, \boldsymbol{r}_T\}$ are independent and identically distributed, following the multivariate normal distribution assumption with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, the sample mean and sample covariance matrix are independent of each other and are respectively given by

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{r}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T),$$
$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} (\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})^\top \sim W_N(T-1, \boldsymbol{\Sigma})/T,$$

where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ represents a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}/T$, and $W_N(T-1, \boldsymbol{\Sigma})$ denotes a Wishart distribution with $T-1$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$. The plug-in ESG portfolio $\hat{\boldsymbol{w}}_{esg}$ is obtained by replacing the true parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (3.2) with their sample counterparts $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, respectively.

However, the plug-in ESG portfolio completely ignores the detrimental effect of estimation errors of the unknown parameters and usually underperforms the true optimal one $\boldsymbol{w}_{esg}^*$ due to the estimation risk.

To mitigate the estimation risk, we focus on a popular category of investment strategies known as combined portfolios, as discussed in Chapter 1. Directly considering a weighted combination of the constituent portfolios involved in the plug-in ESG portfolio poses calculation challenges, because the complex structure of these constituent portfolios complicates the analysis of the analytic expression of the expected out-of-sample performance. Note that we can rearrange the plug-in ESG portfolio as:

$$\hat{\boldsymbol{w}}_{esg} = C_1\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1} + C_2\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S} + C_3\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}, \tag{3.9}$$

where C_1, C_2 and C_3 are the coefficients of the vectors $\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$, $\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}$ and $\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ after the rearrangement, respectively. This motivates us to consider a combined portfolio containing these three vectors. As pointed out in Section 3.2, we focus on the portfolio that satisfies the weight constraint. To achieve it, we may directly consider the combination of the following three constituent portfolios that correspond to the components specified in (3.9): $\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}/\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$, $\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}/\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}$, $\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}/\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$, or we may express one of the combination coefficients as a function of the others as discussed in Section 2.3.1. However, deriving the analytic expressions for the expected out-of-sample performances of such potential combined portfolios proves to be very challenging, if not impossible.

In view of the challenges in structuring a combined portfolio to meet the weight constraint and facilitate the analysis of the expected out-of-sample performances, we propose the following practical configuration for the combined portfolio:

$$\hat{\boldsymbol{w}}_c = \hat{\boldsymbol{w}}_g + \delta\hat{\boldsymbol{w}}_z + \beta\hat{\boldsymbol{w}}_s, \tag{3.10}$$

where δ and β are two constant combination coefficients, and the constituent portfolios are

$$\hat{\boldsymbol{w}}_g = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}, \quad \hat{\boldsymbol{w}}_z = \hat{\boldsymbol{\Upsilon}}\hat{\boldsymbol{\mu}} \quad \text{and} \quad \hat{\boldsymbol{w}}_s = \hat{\boldsymbol{\Upsilon}}\boldsymbol{S}.$$

where $\hat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\Sigma}}^{-1} - \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}^{-1}/\mathbf{1}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N$. This combined portfolio $\hat{\boldsymbol{w}}_c$ can be expressed as a special linear combination of the three components in (3.9). Specifically, the first two constituent portfolios ($\hat{\boldsymbol{w}}_g$ and $\hat{\boldsymbol{w}}_z$) are the plug-in version of the global minimum variance (GMV) portfolio and the zero-investment portfolio derived from the optimization problem (3.6), which correspond to the components $\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$ and $\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$. Additionally, we introduce $\hat{\boldsymbol{w}}_s$ to correspond to the component $\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}$. Meanwhile, the structures of these three constituent portfolios ensure that the combined portfolio meets the weight constraint as

$$\hat{\boldsymbol{w}}_c^\top \mathbf{1} = 1.$$

When implementing the combined portfolio (3.10), the investor always holds 100% of $\hat{\boldsymbol{w}}_g$. The exposures on $\hat{\boldsymbol{w}}_z$ and $\hat{\boldsymbol{w}}_s$ are determined by the combination coefficients δ and β. Note that $\hat{\boldsymbol{w}}_g$ and $\hat{\boldsymbol{w}}_s$ are primarily affected by the estimation error from the covariance matrix, while $\hat{\boldsymbol{w}}_z$ encounters additional estimation error from the mean vector. Under the multivariate normal distribution assumption for asset returns, Proposition 3.3 gives the analytic form of the following three expectations: the expected out-of-sample mean $\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}]$, the expected out-of-sample ESG score $\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}]$ and expected out-of-sample variance $\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c]$.

Proposition 3.3. *For $T > N + 3$, we have*

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}] &= \mu_g + k_1 \delta \theta_1^2 + k_1 \beta \theta_2^2, \\ \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}] &= \mu_s + k_1 \delta \theta_2^2 + k_1 \beta \theta_3^2, \\ \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c] &= k_1 k_3 [\sigma_g^2 + T k_2 (\delta^2 \theta_1^2 + \beta^2 \theta_3^2 + 2\delta\beta\theta_2^2) + k_2 (N-1)\delta^2], \end{aligned} \tag{3.11}$$

where $k_1 = T/(T-N-1)$*,* $k_2 = T/(T-N)(T-N-3)$*,* $k_3 = (T-2)/T$*, and*

$$\begin{aligned} &\theta_1^2 = \boldsymbol{\mu}^\top \boldsymbol{\Upsilon} \boldsymbol{\mu}, \qquad \theta_2^2 = \boldsymbol{\mu}^\top \boldsymbol{\Upsilon} \boldsymbol{S}, \qquad \theta_3^2 = \boldsymbol{S}^\top \boldsymbol{\Upsilon} \boldsymbol{S}, \\ &\sigma_g^2 = \frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}, \qquad \mu_g = \frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}, \qquad \mu_s = \frac{\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}, \end{aligned} \tag{3.12}$$

with $\boldsymbol{\Upsilon} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} / \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}$.

Proof. See Appendix B.1.5. □

3.4 Optimization with Expected ESG Constraints

We determine the optimal combination coefficients for the combined portfolio $\hat{\boldsymbol{w}}_c$ that maximize the expected out-of-sample MV utility as proposed by Kan and Zhou (2007). For the ESG score of the combined portfolio, we may include a pointwise ESG constraint in a similar manner as the aforementioned ESG-constrained MV framework in Section 3.2, i.e., $\hat{\boldsymbol{w}}_c^\top \boldsymbol{S} = \bar{s}$. However, due to the structure and randomness of the constituent portfolios and the assumption of constant combination coefficients in $\hat{\boldsymbol{w}}_c$, such a stringent pointwise ESG constraint leads to an infeasible set for the optimization problem. Consequently, we introduce a more flexible constraint by focusing on the expectation of the total ESG score of the combined portfolio.

3.4.1 Optimal Combined Portfolios

We consider an expected out-of-sample MV optimization with an inequality expected ESG constraint, where the expectation of the total ESG score meets and exceeds a certain level. It is formulated as follows:

$$\begin{aligned} \max_{\delta,\beta} \quad & \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}] - \frac{\gamma}{2}\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c], \\ s.t. \quad & \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}] \geq \bar{s}. \end{aligned} \tag{3.13}$$

The constraint in this optimization problem (3.13) focuses on the expected ESG score of the portfolio, which offers considerable flexibility, especially in dynamic market environments when strict adherence to a specific ESG score might limit investment opportunities. The use of an inequality constraint is advantageous for investors seeking to achieve a higher average total ESG score of the portfolio than the target $\bar{s}$.

We find the optimal combination coefficients using the Karush-Kuhn-Tucker (KKT) conditions and present the result in Theorem 3.2.

Theorem 3.2. *The optimal combination coefficients δ_I^* and β_I^* that solve the optimization* *problem* (3.13) *are given by*

$$\delta_I^* = \frac{\theta_1^2\theta_3^2 - \theta_2^4}{\gamma k_2 k_3 \left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]},$$
$$\beta_I^* = \begin{cases} \frac{(N-1)\theta_2^2}{\gamma k_2 k_3 T\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]}, & \zeta \leq 0, \\ \frac{\theta_2^2(\theta_1^2\theta_3^2 - \theta_2^4)}{\gamma k_2 k_3 \theta_3^2\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]} - \frac{\mu_s - \bar{s}}{k_1\theta_3^2}, & \zeta > 0, \end{cases} \tag{3.14}$$

where $\zeta = -[k_1\theta_2^2 + \gamma k_2 k_3 T(\mu_s - \bar{s})]/k_1\theta_3^2$*. The corresponding optimal combined portfolio is*

$$\hat{\boldsymbol{w}}_{c\text{-}I}^* = \hat{\boldsymbol{w}}_g + \delta_I^*\hat{\boldsymbol{w}}_z + \beta_I^*\hat{\boldsymbol{w}}_s. \tag{3.15}$$

Proof. See Appendix B.1.6. □

Theorem 3.2 shows the impact of the parameter ζ on the optimal combination coefficient β_I^*, particularly through its role in the activation of the inequality constraint. Specifically, when $\zeta \leq 0$, the inequality constraint becomes inactive, making the solutions identical to those obtained by maximizing the expected out-of-sample MV utility without any constraints. The corresponding optimal combined portfolio satisfies that $\mathbb{E}[(\hat{\boldsymbol{w}}_{c\text{-}I}^*)^\top \boldsymbol{S}] > \bar{s}$. Conversely, when $\zeta > 0$, the inequality constraint impacts the outcomes at its boundary, which can then be converted into an equality constraint. The corresponding optimal combined portfolio satisfies that $\mathbb{E}[(\hat{\boldsymbol{w}}_{c\text{-}I}^*)^\top \boldsymbol{S}] = \bar{s}$. We can identify the threshold for the target ESG score at which the parameter ζ equals 0, which is presented in Corollary 3.2.

Corollary 3.2. *Define the threshold target ESG score such that* $\zeta = 0$ *as:*

$$\bar{s}^{**} = \mu_s + \frac{k_1\theta_2^2}{\gamma k_2 k_3 T}.$$

We have the following assertions:

1. *If* $\bar{s} < \bar{s}^{**}$*, the optimal combined portfolio* $\hat{\boldsymbol{w}}_{c\text{-}I}^*$ (3.15) *satisfies* $\mathbb{E}[(\hat{\boldsymbol{w}}_{c\text{-}I}^*)^\top \boldsymbol{S}] > \bar{s}$*.*

2. *If* $\bar{s} \geq \bar{s}^{**}$*, the optimal combined portfolio* $\hat{\boldsymbol{w}}_{c\text{-}I}^*$ (3.15) *satisfies* $\mathbb{E}[(\hat{\boldsymbol{w}}_{c\text{-}I}^*)^\top \boldsymbol{S}] = \bar{s}$*.*

Particularly, for ease of discussion regarding the second scenario outlined in Corollary 3.2, we introduce an optimization framework that specifically addresses the equality expected ESG constraint. It is suitable for investors who prefer the expectation of the total ESG score to precisely meet a predetermined target level. The framework is formulated as follows:

$$\begin{aligned} \max_{\delta,\beta} \quad & \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}] - \frac{\gamma}{2}\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c], \\ s.t. \quad & \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}] = \bar{s}. \end{aligned} \tag{3.16}$$

The optimal combination coefficients for this optimization problem (3.16) can be readily achieved using the results from Theorem 3.2. For clarity and to facilitate discussion, we summarize the related results in Theorem 3.3.

Theorem 3.3. *The optimal combination coefficients δ_E^* and β_E^* that solve the optimization problem* (3.16) *are given by*

$$\begin{aligned} \delta_E^* &= \frac{\theta_1^2\theta_3^2 - \theta_2^4}{\gamma k_2 k_3 \left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]}, \\ \beta_E^* &= \frac{\theta_2^2(\theta_1^2\theta_3^2 - \theta_2^4)}{\gamma k_2 k_3 \theta_3^2 \left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]} - \frac{\mu_s - \bar{s}}{k_1\theta_3^2}. \end{aligned} \tag{3.17}$$

The corresponding optimal combined portfolio is given by

$$\hat{\boldsymbol{w}}_{c\text{-}E}^* = \hat{\boldsymbol{w}}_g + \delta_E^* \hat{\boldsymbol{w}}_z + \beta_E^* \hat{\boldsymbol{w}}_s. \tag{3.18}$$

Proof. See Appendix B.1.7. □

The constraints in the optimization problems (3.13) and (3.16) focus primarily on the expected ESG score of the combined portfolio, while understanding the variability of the ESG score is also essential for a comprehensive assessment of the stability and fluctuations associated with the investment. To address it, we extend our analysis to the variance of the ESG score of the combined portfolio $\hat{\boldsymbol{w}}_c$, which is presented in Proposition 3.4. When applying the optimal combination coefficients derived from either Theorem 3.2 or Theorem 3.3, we can obtain the variance of the total ESG score for the corresponding optimal combined portfolio using (3.19).

Proposition 3.4. *The variance of $\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}$ is obtained as*

$$Var[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}] = k_1\theta_3^2 \left[\frac{\sigma_g^2}{T} + \delta^2 k_2 k_3 + k_2(\delta^2\theta_1^2 + \beta^2\theta_3^2 + 2\delta\beta\theta_2^2)\right] + (Tk_2 - k_1^2)\left(\delta\theta_2^2 + \beta\theta_3^2\right)^2 \tag{3.19}$$

Proof. See Appendix B.1.8. □

3.4.2 Implementable Combined Portfolios

In practice, the optimal combination coefficients presented in Theorems 3.2 and 3.3 are not directly implementable because they depend on the four unknown parameters θ_1^2, θ_2^2, θ_3^2 and μ_s. These parameters in turn rely on the mean vector and covariance matrix of asset returns. To address this issue, it is necessary to estimate these four parameters to derive the corresponding combination coefficients. The implementable combined portfolios are thus formed by replacing the optimal combination coefficients with these estimated values in the optimal combined portfolios, making them practically applicable.

To estimate these four unknown parameters, a simple method involves replacing the unknown mean vector and covariance matrix with their sample counterparts. The sample estimators are denoted by

$$\hat{\theta}_1^2 = \hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Upsilon}} \hat{\boldsymbol{\mu}}, \;\; \hat{\theta}_2^2 = \hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Upsilon}} \boldsymbol{S}, \;\; \hat{\theta}_3^2 = \boldsymbol{S}^\top \hat{\boldsymbol{\Upsilon}} \boldsymbol{S} \text{ and } \hat{\mu}_s = \frac{\boldsymbol{S}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}, \tag{3.20}$$

However, these sample estimators can exhibit significant bias, especially when T is small (e.g., Kan and Zhou, 2007). Therefore, we further consider sophisticated estimators for θ_1^2, θ_2^2 and θ_3^2 based on the distributional properties of their sample estimators. Additionally, due to the difficulty in obtaining a sophisticated estimator for μ_s, we opt to keep the sample estimator for this parameter in the calculations.

As suggested by Kan et al. (2022), the sophisticated estimator for θ_1^2 is obtained as

$$\tilde{\theta}_1^2 = \frac{(T-N-1)\hat{\theta}_1^2 - (N-1)}{T} + \frac{2(\hat{\theta}_1^2)^{\frac{N-1}{2}}(1+\hat{\theta}_1^2)^{-\frac{T-2}{2}}}{TB_{\hat{\theta}_1^2/(1+\hat{\theta}_1^2)}((N-1)/2, (T-N+1)/2)}, \tag{3.21a}$$

where $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\,\mathrm{d}y$ is the incomplete beta function. The first term involved in $\tilde{\theta}_1^2$ represents the unbiased estimator for θ_1^2, while the second term is the additional adjustment for small sample size T. Note that $\tilde{\theta}_1^2$ is always positive, unlike the unbiased estimator of θ_1^2. We also derive the sophisticated (unbiased) estimators for θ_2^2 and θ_3^2 as

$$\tilde{\theta}_2^2 = \frac{T-N-1}{T}\hat{\theta}_2^2 \quad \text{and} \quad \tilde{\theta}_3^2 = \frac{T-N-1}{T}\hat{\theta}_3^2. \tag{3.21b}$$

A detailed proof of these two unbiased estimators is provided in Appendix B.1.9.

Therefore, after obtaining the estimates of the four unknown parameters, θ_1^2, θ_2^2, θ_3^2, and μ_s, we replace the unknown terms in the optimal combination coefficients with those estimates and obtain the estimated combination coefficients. Particularly, for the optimal combined portfolio $\hat{\boldsymbol{w}}_{c\text{-}E}^*$ given in (3.18), we denote the estimated combination coefficients as $\hat{\delta}_E$ and $\hat{\beta}_E$ when using the sample estimators (3.20). The corresponding implementable combined portfolio is then given by

$$\hat{\boldsymbol{w}}_{c\text{-}E} = \hat{\boldsymbol{w}}_g + \hat{\delta}_E \hat{\boldsymbol{w}}_z + \hat{\beta}_E \hat{\boldsymbol{w}}_s.$$

Similarly, when using the sophisticated estimators (3.21) to obtain the estimated combination coefficients, denoted by $\tilde{\delta}_E$ and $\tilde{\beta}_E$, the corresponding implementable combined portfolio is then calculated by

$$\tilde{\boldsymbol{w}}_{c\text{-}E} = \hat{\boldsymbol{w}}_g + \tilde{\delta}_E \hat{\boldsymbol{w}}_z + \tilde{\beta}_E \hat{\boldsymbol{w}}_s.$$

We have the following proposition for these two implementable combined portfolios.

Proposition 3.5. *For the implementable version of the optimal combined portfolio $\hat{\boldsymbol{w}}_{c\text{-}E}^*$ given in* (3.18)*, we have the following assertions:*

1. *When adopting sample estimators* (3.20) *to estimate the optimal combination coefficients, the variance of $\hat{\boldsymbol{w}}_{c\text{-}E}^\top \boldsymbol{S}$ will not be affected by the change of $\bar{s}$.*

2. *When adopting the proposed sophisticated estimators* (3.21) *to estimate the optimal combination coefficients, the implementable combined portfolio satisfies*

$$\tilde{\boldsymbol{w}}_{c\text{-}E}^\top \boldsymbol{S} = \bar{s}.$$

Proof. See Appendix B.1.10. □

To be more specific, for the implementable combined portfolio $\hat{\boldsymbol{w}}_{c\text{-}E}$ obtained using sample estimators, the total ESG score $\hat{\boldsymbol{w}}_{c\text{-}E}^{\top}\boldsymbol{S}$ can be written as a linear function of the target $\bar{s}$ linked with a constant coefficient. It indicates that the variance of this total ESG score remains constant regardless of changes in $\bar{s}$. When we adopt the proposed sophisticated estimators, the corresponding implementable combined portfolio $\tilde{\boldsymbol{w}}_{c\text{-}E}$ satisfies the pointwise ESG constraint, in a similar manner as the previously described ESG-constrained MV framework.

Our discussion about the sophisticated estimators so far is based on the sample covariance matrix of asset returns. We can also consider the shrinkage covariance matrix which is widely used in the literature (e.g., Ledoit and Wolf, 2004) and given by

$$\widehat{\boldsymbol{\Sigma}}_{LW} = (1-\rho)\widehat{\boldsymbol{\Sigma}} + \rho\upsilon \boldsymbol{I}_N,$$

where $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix and $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. The parameter υ is equal to the average of the eigenvalues of the sample covariance matrix, and the parameter ρ is set at

$$\rho = \frac{\min\left[\|\widehat{\boldsymbol{\Sigma}} - \upsilon\boldsymbol{I}_N\|^2, \frac{1}{T^2}\sum_{t=1}^{T}\|(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})^{\top} - \widehat{\boldsymbol{\Sigma}}\|^2\right]}{\|\widehat{\boldsymbol{\Sigma}} - \upsilon\boldsymbol{I}_N\|^2}.$$

Here, $\|A\| = \sqrt{\mathrm{tr}(AA')/N}$ is the Frobenius norm for a matrix A. Upon the use of the shrinkage covariance matrix $\widehat{\boldsymbol{\Sigma}}_{LW}$, we also have two versions of estimators for $\theta_1^2, \theta_2^2, \theta_3^2$. Specifically, we insert the sample mean vector $\hat{\boldsymbol{\mu}}$ and the shrinkage covariance matrix into the corresponding expressions in (3.12) to obtain the shrinkage sample estimators denoted by

$$\hat{\theta}_{1,LW}^2,\ \ \hat{\theta}_{2,LW}^2 \ \text{ and } \ \hat{\theta}_{3,LW}^2. \tag{3.22}$$

We then replace the sample estimators involved in (3.21a) and (3.21b) with the above shrinkage sample estimators to obtain a set of shrinkage sophisticated estimators

$$\tilde{\theta}_{1,LW}^2,\ \ \tilde{\theta}_{2,LW}^2 \ \text{ and } \ \tilde{\theta}_{3,LW}^2. \tag{3.23}$$

In the meantime, we may also replace the sample covariance matrix in the constituent portfolios with the shrinkage one. However, such a replacement makes it hard to find the explicit expressions of expected out-of-sample utilities and analytical combination coefficients. As an alternative, we directly employ the combination coefficients achieved from Theorems 3.2 and 3.3, in the same manner as advocated by Kan and Zhou (2007).

3.5 Numerical Studies

In this section, we evaluate the performance of the proposed portfolios via both simulation and empirical studies. The simulation study evaluates the performance under the normal distribution assumption for asset returns, whereas the empirical study examines the portfolio performances in real-market conditions.

Table 3.1: List of portfolio strategies.

Notation	Portfolio
Panel A: Oracle combined portfolios (only available in the simulation study)	
$\hat{\boldsymbol{w}}^*_{c\text{-}I}$	the optimal combined portfolio given in (3.15)
$\hat{\boldsymbol{w}}^*_{c\text{-}E}$	the optimal combined portfolio given in (3.18)
Panel B: Implementable combined portfolios	
Set 1: apply the sample estimators (3.20)	
$\hat{\boldsymbol{w}}_{c\text{-}I}$	the implementable combined portfolio for (3.15)
$\hat{\boldsymbol{w}}_{c\text{-}E}$	the implementable combined portfolio for (3.18)
Set 2: apply the sophisticated estimators (3.21)	
$\tilde{\boldsymbol{w}}_{c\text{-}I}$	the implementable combined portfolio for (3.15)
$\tilde{\boldsymbol{w}}_{c\text{-}E}$	the implementable combined portfolio for (3.18)
Set 3: apply the shrinkage sample estimators (3.22)	
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	the implementable combined portfolio for (3.15)
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	the implementable combined portfolio for (3.18)
Set 4: apply the shrinkage sophisticated estimators (3.23)	
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	the implementable combined portfolio for (3.15)
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	the implementable combined portfolio for (3.18)
Panel C: Plug-in portfolios	
$\hat{\boldsymbol{w}}_{esg}$	the plug-in ESG portfolio for (3.2) using the sample covariance matrix
$\hat{\boldsymbol{w}}^{LW}_{esg}$	the plug-in ESG portfolio for (3.2) using the shrinkage covariance matrix

The portfolio strategies we consider are summarized in Table 3.1. Panel A presents the oracle combined portfolios, which are based on true optimal combination coefficients computed from the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, as given in Theorems 3.2 and 3.3. These portfolios are only available in the simulation study, and not implementable in practice. They provide (theoretically) tight upper bounds on the performance of combined portfolios. Panel B summarizes implementable combined portfolios that utilize different estimators as discussed in Section 3.4.2. Specifically, Sets 1-4 in Panel B utilize various estimators, which are respectively the sample estimators (3.20), the sophisticated estimators (3.21), the shrinkage sample estimators (3.22), and the shrinkage sophisticated estimators (3.23). Particularly, the adoption of the shrinkage covariance matrix in the constituent portfolios generally results in better out-of-sample portfolio performance compared to the use of the sample covariance matrix. This observation is consistent with the findings in Section 2.5 and Kan et al. (2022). Consequently, we apply the shrinkage covariance matrix in the constituent portfolios across Sets 1-4. Panel C shows the plug-in ESG portfolio that replaces $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the sample mean and sample (or shrinkage) covariance matrix, respectively.

3.5.1 Simulation Study

In the simulation study, we focus on constructing investment portfolios which are weighted combinations of the latest components from the Dow Jones Industrial Average (DJ30) as of the year 2024. To conduct our analysis, we first collect historical ESG scores and asset returns for these components from the real market. Specifically, we gather yearly updated ESG scores from ASSET4 (now Refinitiv), covering the period from 2010 to 2021. Notably, the ESG scores for Dow Inc., one of the DJ30 components, are only available starting from 2018. Due to this limitation, we exclude Dow Inc. from our subsequent analyses, reducing the total number of assets in our portfolios to 29. That is, $N = 29$. Based on the available data, we compute the average ESG scores for these stocks and summarize the statistical characteristics of these ESG scores in Table 3.2. These average ESG scores are used as the definitive ESG scores for these assets in the construction of portfolios.

Table 3.2: Descriptive statistics of the ESG scores of components in DJ30.

average	smallest	25^{th} quantile	median	75^{th} quantile	largest
0.8645	0.5391	0.8226	0.9071	0.9224	0.9656

Additionally, we download the monthly adjusted close prices of these stocks from Yahoo Finance, starting from "2008-07-01" to "2023-07-01". The starting date "2008-07-01" is around the earliest date that all 29 stocks have available data. We calculate the monthly returns by using the ratio of price differences to previous prices, and then calculate the corresponding mean vector and covariance matrix of these monthly returns. These computed mean vector and covariance matrix are treated as the true statistical measures of the asset return distribution. Consequently, we generate samples from a multivariate normal distribution with these parameters to simulate asset returns.

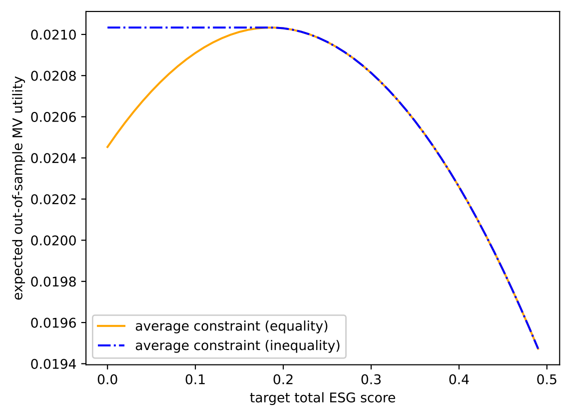
We begin by examining the optimal combined portfolios, $\hat{\boldsymbol{w}}_{c\text{-}I}$ and $\hat{\boldsymbol{w}}_{c\text{-}E}$, derived from Theorems 3.2 and 3.3. Using Proposition 3.3, the expected out-of-sample MV utility for the combined portfolio $\hat{\boldsymbol{w}}_c$ is calculated by

$$\begin{aligned}&\mathbb{E}[U_{\hat{\boldsymbol{w}}_c}] = \mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu}] - \frac{\gamma}{2}\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c] \\ =&\mu_g - \frac{\gamma k_1 k_3 \sigma_g^2}{2} + k_1\theta_1^2\delta + k_1\theta_2^2\beta - \frac{\gamma k_1 k_2 k_3}{2}[(T\theta_1^2 + N - 1)\delta^2 + T\theta_3^2\beta^2 + 2T\theta_2^2\delta\beta].\end{aligned} \tag{3.24}$$

We substitute the combination coefficients δ and β with the optimal values specified by either Theorem 3.2 (δ_I^*, β_I^*) or 3.3 (δ_E^*, β_E^*). Given that the true mean vector and covariance matrix of asset returns are known in advance in the simulation study, we use these two parameter values to calculate the expected out-of-sample utility for the corresponding combined portfolio. Additionally, we set the risk aversion coefficient at $\gamma = 3$ and the moving window length at $T = 120$.

Figure 3.1 shows the expected out-of-sample MV utility of the two optimal combined portfolios against the target total ESG scores $\bar{s}$, which range from 0 to 0.5. Notably, at the threshold target ESG score of $\bar{s}^{**} = 0.1856$ as outlined in Corollary 3.2, these two optimal combined portfolios achieve the highest expected out-of-sample utility. When the target ESG score is below this threshold (i.e., $\bar{s} < \bar{s}^{**}$), the inequality constraint in the optimization problem (3.13) is inactive, and consequently, the utility of $\hat{\boldsymbol{w}}_{c\text{-}I}$ remains at its maximum and does not fluctuate with changes in the target ESG score. In contrast, the optimal combined portfolio $\hat{\boldsymbol{w}}_{c\text{-}E}$ under the optimization problem (3.16) is required to adhere to the equality constraint, resulting in lower utilities. Moreover, when the target ESG score exceeds the threshold ($\bar{s} > \bar{s}^{**}$), the inequality constraint in the optimization problem (3.13) activates at the boundary, which makes it align with the equality constraint. Consequently, the utilities of the two optimal combined portfolios, $\hat{\boldsymbol{w}}_{c\text{-}I}$ and $\hat{\boldsymbol{w}}_{c\text{-}E}$, converge,

Figure 3.1: Expected out-of-sample MV utility for the optimal combined portfolios.



This figure presents the expected out-of-sample MV utilities (in percentage) of the optimal combined portfolios from different optimization problems against the target total ESG scores $\bar{s}$. The risk aversion coefficients are set at $\gamma = 3$ and the estimation window is fixed at $T = 120$.

which explains the observed overlap shown in the figure. Furthermore, the cost associated with increasing the target total ESG scores beyond the threshold can be quantified by the reduction in expected out-of-sample MV utility. If investors prioritize a higher total ESG score, they will correspondingly experience lower utilities.

Next, we evaluate the performance of implementable combined portfolios using Monte Carlo simulation. We simulate $M = 10,000$ sets of $T \times N$ asset returns, each generated from the multivariate normal distribution using the mean vector and covariance matrix of historical asset returns as discussed before. For each simulated set of $T \times N$ returns, we compute the sample mean and sample (or shrinkage) covariance matrix, and then plug them into the optimal combined portfolios to obtain the implementable ones. Let $\tilde{\boldsymbol{w}}$ be any implementable portfolio and $\tilde{\boldsymbol{w}}_i$ be the specific realization obtained based on the i-th set of simulated asset returns, $i = 1, \ldots, M$. Given that the constraints in the optimization problems (3.13) and (3.16) are tied to the expectation of the total ESG scores, we explore the total ESG scores of the implementable combined portfolios from these simulations by focusing on assessing the average and standard deviation. The calculations are formulated as follows:

$$\widehat{ES}_{\tilde{\boldsymbol{w}}} = \frac{1}{M}\sum_{i=1}^{M} \tilde{\boldsymbol{w}}_i^\top \boldsymbol{S} \qquad \text{and} \qquad \widehat{VS}_{\tilde{\boldsymbol{w}}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\tilde{\boldsymbol{w}}_i^\top \boldsymbol{S} - \widehat{ES}_{\tilde{\boldsymbol{w}}_i})^2}. \tag{3.25}$$

These two equations serve as the estimators for the expectation $\mathbb{E}[\tilde{\boldsymbol{w}}^\top \boldsymbol{S}]$ and the standard deviation $\sqrt{Var[\tilde{\boldsymbol{w}}^\top \boldsymbol{S}]}$ of the total ESG scores for $\tilde{\boldsymbol{w}}$, respectively, which provide insights into the consistency and variability of the ESG scores across the implementable portfolios.

Table 3.3: The average total ESG scores for portfolios in the simulation study.

$\bar{s}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
Panel A: Oracle combined portfolio								
$\hat{\boldsymbol{w}}^*_{c\text{-}E}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
	(0.1777)	(0.1727)	(0.1679)	(0.1633)	(0.1589)	(0.1548)	(0.1473)	(0.1411)
$\hat{\boldsymbol{w}}^*_{c\text{-}I}$	0.1856	0.1856	0.1856	0.2	0.25	0.3	0.4	0.5
	(0.1646)	(0.1646)	(0.1646)	(0.1633)	(0.1589)	(0.1548)	(0.1473)	(0.1411)
Panel B: Implementable combined portfolio								
Set 1: apply the sample estimates and sample covariance matrix								
$\hat{\boldsymbol{w}}_{c\text{-}E}$	0.2489	0.2864	0.3239	0.3614	0.3989	0.4364	0.5114	0.5864
	(0.0091)	(0.0091)	(0.0091)	(0.0091)	(0.0091)	(0.0091)	(0.0091)	(0.0091)
$\hat{\boldsymbol{w}}_{c\text{-}I}$	0.3459	0.368	0.3919	0.4173	0.4444	0.473	0.5343	0.6001
	(0.1584)	(0.1454)	(0.1325)	(0.1199)	(0.1078)	(0.0962)	(0.075)	(0.0569)
Set 2: apply the sophisticated estimates and sample covariance matrix								
$\tilde{\boldsymbol{w}}_{c\text{-}E}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
$\tilde{\boldsymbol{w}}_{c\text{-}I}$	0.2577	0.2754	0.2961	0.3199	0.3468	0.3768	0.4456	0.525
	(0.2245)	(0.2088)	(0.1923)	(0.1752)	(0.1577)	(0.1403)	(0.107)	(0.0775)
Set 3: apply the sample estimates and shrinkage covariance matrix								
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.2507	0.2882	0.3257	0.3632	0.4007	0.4382	0.5132	0.5882
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.3856	0.401	0.4188	0.4388	0.4612	0.4857	0.5408	0.603
	(0.1612)	(0.1489)	(0.136)	(0.1229)	(0.1099)	(0.097)	(0.0731)	(0.0525)
Set 4: apply the sophisticated estimates and shrinkage covariance matrix								
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.3366	0.3444	0.3548	0.3685	0.3856	0.4066	0.461	0.5308
	(0.2121)	(0.2021)	(0.1903)	(0.1765)	(0.1611)	(0.1446)	(0.1098)	(0.0768)
Panel C: Plug-in portfolios								
$\hat{\boldsymbol{w}}_{esg}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
$\hat{\boldsymbol{w}}^{LW}_{esg}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Table 3.4: Proportion of simulations with inactive inequality constraints.

$\bar{s}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
$\hat{\zeta} \leq 0$	0.8661	0.8197	0.7616	0.6909	0.621	0.5376	0.3752	0.2322

This table presents the ratio of simulations where the constraint is inactive relative to the total number of simulations, $M = 10,000$, when constructing $\tilde{\boldsymbol{w}}_{c\text{-}I}$. The parameter $\hat{\zeta}$, which estimates ζ in Theorem 3.2, is calculated using the sample mean vector and covariance matrix.

Table 3.3 provides the average and standard deviation (in brackets) of the total ESG scores across different portfolios, with the target ESG score $\bar{s}$ varying from 0.05 to 0.5. Panel A presents the expected values and standard deviations of the total ESG scores for the optimal combined portfolios, which are calculated using the analytic formulas from Propositions 3.3 and 3.4, based on the true mean vector and covariance matrix of asset returns. Panels B and C show the averages and standard deviations of the total ESG scores for the implementable portfolios from simulations, computed using (3.25). Particularly, as established in Proposition 3.5, the standard deviations of the total ESG scores for $\hat{\boldsymbol{w}}_{c\text{-}E}$ and $\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$ (see Sets 1 and 3) do not vary with changes in $\bar{s}$. Additionally, the total ESG scores for $\tilde{\boldsymbol{w}}_{c\text{-}E}$ and $\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$ (see Sets 2 and 4) shows zero variability in the total ESG scores across different target ESG values, similarly to the plug-in ESG portfolios in Panel C. It highlights the ability to precisely meet the target ESG score for each simulated set.

We draw the following observations about the total ESG score for different portfolios based on the results in Table 3.3. Firstly, the expectation of the total ESG scores for $\hat{\boldsymbol{w}}^{*}_{c\text{-}I}$ in Panel A confirms a threshold target ESG score of 0.1856, which aligns with the findings from Figure 3.1. The standard deviations of the total ESG scores for $\hat{\boldsymbol{w}}^{*}_{c\text{-}E}$ decrease as the target ESG score increases, similar to $\hat{\boldsymbol{w}}^{*}_{c\text{-}I}$ when $\bar{s} > 0.1856$. Secondly, the differences between Panels A and B come from the estimation risks associated with obtaining combination coefficients for combined portfolios. Generally, the implementable combined portfolios for $\hat{\boldsymbol{w}}^{*}_{c\text{-}E}$ has a smaller standard deviation than the ones for $\hat{\boldsymbol{w}}^{*}_{c\text{-}I}$, which indicates that the implementable combined portfolios from the equality constraint are less volatile in terms of the total ESG scores than the ones from the inequality constraint. Moreover, denote $\hat{\zeta}$ as the estimated value for ζ in Theorem 3.2. When the inequality constraint is inactive (i.e., $\hat{\zeta} \leq 0$), the implementable combined portfolios for $\hat{\boldsymbol{w}}^{*}_{c\text{-}E}$ differs from those for $\hat{\boldsymbol{w}}^{*}_{c\text{-}I}$. Conversely, they are identical if inequality constraint actives at the boundary (i.e., $\hat{\zeta} > 0$). Table 3.4 lists the proportion of simulation sets, out of a total of $M = 10,000$, in which the inequality constraint is inactive. It suggests that the inequality constraint is less likely to be inactive as the target ESG score $\bar{s}$ increases.

We then compare the out-of-sample performances of various portfolios by assessing the expected out-of-sample portfolio returns, MV utilities, and Sharpe ratios. For any implementable portfolio $\tilde{\boldsymbol{w}}$ and given its realizations $\tilde{\boldsymbol{w}}_i$ derived from the i-th set of simulated asset returns, $i = 1, 2, ..., M$, we estimate the expected out-of-sample mean $\mathbb{E}[\tilde{\boldsymbol{w}}^\top \boldsymbol{\mu}]$ and the expected out-of-sample variance $\mathbb{E}[\tilde{\boldsymbol{w}}^\top \boldsymbol{\Sigma} \tilde{\boldsymbol{w}}]$ of this portfolio respectively by

$$\widehat{ER}_{\tilde{\boldsymbol{w}}} = \frac{1}{M} \sum_{i=1}^{M} \tilde{\boldsymbol{w}}_i^\top \boldsymbol{\mu} \quad \text{and} \quad \widehat{VR}_{\tilde{\boldsymbol{w}}} = \frac{1}{M} \sum_{i=1}^{M} \tilde{\boldsymbol{w}}_i^\top \boldsymbol{\Sigma} \tilde{\boldsymbol{w}}_i. \tag{3.26}$$

Accordingly, the expected out-of-sample MV utility and the zero-order Taylor-series expansion of the expected out-of-sample Sharpe ratio [2] (see Lassance, 2021) are respectively estimated by

$$\widehat{EU}_{\tilde{\boldsymbol{w}}} = \widehat{ER}_{\tilde{\boldsymbol{w}}} - \frac{\gamma}{2} \widehat{VR}_{\tilde{\boldsymbol{w}}} \quad \text{and} \quad \widehat{SR}_{\tilde{\boldsymbol{w}}} = \frac{\widehat{ER}_{\tilde{\boldsymbol{w}}}}{\sqrt{\widehat{VR}_{\tilde{\boldsymbol{w}}}}}. \tag{3.27}$$

Here, the expected out-of-sample MV utility represents the trade-off between expected returns and the associated risk, adjusted by the risk aversion coefficient γ. It tells investors about the overall satisfaction expected from the portfolio considering both potential profits and losses. Meanwhile, the expected out-of-sample Sharpe ratio quantifies the risk-adjusted return, which tells investors how much return they are gaining for each unit of risk undertaken.

Tables 3.5-3.7 presents the expected out-of-sample mean, MV utilities, and Sharpe ratio for various portfolios. The standard deviations, which are calculated by the square root of the expected out-of-sample variance, are presented in the brackets in Table 3.5. For the oracle combined portfolios listed in the Panel A of these three tables, the expected out-of-sample mean and variance are calculated using the analytic formulas provided in Proposition 3.3, from which the expected out-of-sample MV utilities and Sharpe ratios are subsequently calculated. In contrast, for the other implementable portfolios, these values are computed using (3.26) and (3.27).

From Table 3.5, the combined portfolios from the equality constraint typically exhibit

[2] Note that in this chapter, we do not consider the risk-free asset. Therefore, we utilize the mean and variance of the portfolio's total return rather than its excess return to evaluate the Sharpe ratio.

Table 3.5: Expected out-of-sample portfolio returns in the simulation study.

$\bar{s}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
Panel A: Oracle combined portfolio								
$\hat{\boldsymbol{w}}^*_{c\text{-}E}$	0.0416	0.0405	0.0394	0.0382	0.0371	0.0360	0.0338	0.0316
	(0.0139)	(0.0130)	(0.0122)	(0.0115)	(0.0108)	(0.0101)	(0.0090)	(0.0081)
$\hat{\boldsymbol{w}}^*_{c\text{-}I}$	0.0386	0.0386	0.0386	0.0382	0.0371	0.0360	0.0338	0.0316
	(0.0117)	(0.0117)	(0.0117)	(0.0115)	(0.0108)	(0.0101)	(0.0090)	(0.0081)
Panel B: Implementable combined portfolio								
Set 1: apply the sample estimates and sample covariance matrix								
$\hat{\boldsymbol{w}}_{c\text{-}E}$	0.0480	0.0471	0.0463	0.0455	0.0446	0.0438	0.0421	0.0405
	(0.1427)	(0.1410)	(0.1394)	(0.1379)	(0.1365)	(0.1352)	(0.1329)	(0.1311)
$\hat{\boldsymbol{w}}_{c\text{-}I}$	0.0460	0.0455	0.0449	0.0443	0.0437	0.0431	0.0417	0.0402
	(0.1402)	(0.1391)	(0.1380)	(0.1369)	(0.1359)	(0.1348)	(0.1328)	(0.1311)
Set 2: apply the sophisticated estimates and sample covariance matrix								
$\tilde{\boldsymbol{w}}_{c\text{-}E}$	0.0408	0.0397	0.0385	0.0374	0.0363	0.0352	0.0330	0.0308
	(0.1196)	(0.1160)	(0.1125)	(0.1092)	(0.1061)	(0.1031)	(0.0976)	(0.0931)
$\tilde{\boldsymbol{w}}_{c\text{-}I}$	0.0364	0.0360	0.0355	0.0350	0.0343	0.0337	0.0321	0.0303
	(0.1090)	(0.1076)	(0.1060)	(0.1042)	(0.1024)	(0.1004)	(0.0965)	(0.0927)
Set 3: apply the sample estimates and shrinkage covariance matrix								
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.0438	0.0429	0.0421	0.0413	0.0404	0.0396	0.0379	0.0363
	(0.1200)	(0.1181)	(0.1164)	(0.1147)	(0.1131)	(0.1117)	(0.1091)	(0.1070)
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.0409	0.0406	0.0402	0.0397	0.0392	0.0386	0.0374	0.0360
	(0.1154)	(0.1146)	(0.1137)	(0.1127)	(0.1117)	(0.1107)	(0.1087)	(0.1069)
Set 4: apply the sophisticated estimates and shrinkage covariance matrix								
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.0364	0.0353	0.0342	0.0331	0.0320	0.0309	0.0286	0.0264
	(0.1030)	(0.0991)	(0.0953)	(0.0916)	(0.0881)	(0.0847)	(0.0785)	(0.0732)
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.0303	0.0301	0.0298	0.0295	0.0291	0.0286	0.0274	0.0258
	(0.0856)	(0.0849)	(0.0839)	(0.0828)	(0.0814)	(0.0798)	(0.0762)	(0.0723)
Panel C: Plug-in portfolios								
$\hat{\boldsymbol{w}}_{esg}$	0.0952	0.0941	0.0930	0.0919	0.0908	0.0897	0.0874	0.0852
	(0.3249)	(0.3236)	(0.3223)	(0.3212)	(0.3201)	(0.3191)	(0.3174)	(0.3160)
$\hat{\boldsymbol{w}}^{LW}_{esg}$	0.0871	0.0860	0.0849	0.0838	0.0826	0.0815	0.0793	0.0771
	(0.2739)	(0.2725)	(0.2711)	(0.2699)	(0.2687)	(0.2676)	(0.2657)	(0.2641)

higher average out-of-sample returns and standard deviations compared to those from the inequality constraint. Furthermore, the plug-in ESG portfolios in Panel C display the highest average returns and standard deviations among the all listed portfolios, indicating significant fluctuations in the out-of-sample performance when we directly implement the portfolio obtained without accounting for estimation risk.

From Table 3.6, the utilities of the orcale combined portfolios show a consistent pattern of change with the target ESG $\bar{s}$ as observed in Figure 3.1. Meanwhile, due to the estimation

Table 3.6: Expected out-of-sample MV utilities for portfolios in the simulation study.

$\bar{s}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
Panel A: Oracle combined portfolio								
$\hat{\boldsymbol{w}}^*_{c\text{-}E}$	0.0207	0.0209	0.0210	0.0210	0.0210	0.0208	0.0203	0.0194
$\hat{\boldsymbol{w}}^*_{c\text{-}I}$	0.0210	0.0210	0.0210	0.0210	0.0210	0.0208	0.0203	0.0194
Panel B: Implementable combined portfolio								
Set 1: apply the sample estimates and sample covariance matrix								
$\hat{\boldsymbol{w}}_{c\text{-}E}$	0.0174	0.0173	0.0171	0.0169	0.0167	0.0164	0.0156	0.0147
$\hat{\boldsymbol{w}}_{c\text{-}I}$	0.0165	0.0164	0.0163	0.0162	0.0160	0.0158	0.0152	0.0144
Set 2: apply the sophisticated estimates and sample covariance matrix								
$\tilde{\boldsymbol{w}}_{c\text{-}E}$	0.0193	0.0195	0.0195	0.0195	0.0194	0.0193	0.0187	0.0178
$\tilde{\boldsymbol{w}}_{c\text{-}I}$	0.0186	0.0187	0.0187	0.0187	0.0186	0.0185	0.0181	0.0174
Set 3: apply the sample estimates and shrinkage covariance matrix								
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.0222	0.0220	0.0218	0.0215	0.0212	0.0209	0.0201	0.0191
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.0209	0.0209	0.0208	0.0206	0.0205	0.0202	0.0196	0.0188
Set 4: apply the sophisticated estimates and shrinkage covariance matrix								
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.0205	0.0206	0.0206	0.0205	0.0203	0.0201	0.0194	0.0184
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.0193	0.0193	0.0193	0.0192	0.0192	0.0191	0.0187	0.0179
Panel C: Plug-in portfolios								
$\hat{\boldsymbol{w}}_{esg}$	-0.0631	-0.0629	-0.0629	-0.0629	-0.0629	-0.0631	-0.0637	-0.0646
$\hat{\boldsymbol{w}}^{LW}_{esg}$	-0.0254	-0.0254	-0.0254	-0.0255	-0.0256	-0.0259	-0.0266	-0.0276

Table 3.7: Expected out-of-sample Sharpe ratio for portfolios in the simulation study.

$\bar{s}$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
Panel A: Oracle combined portfolio								
$\hat{\boldsymbol{w}}^*_{c\text{-}E}$	0.3526	0.3544	0.3559	0.3570	0.3577	0.3578	0.3558	0.3500
$\hat{\boldsymbol{w}}^*_{c\text{-}I}$	0.3567	0.3567	0.3567	0.3570	0.3577	0.3578	0.3558	0.3500
Panel B: Implementable combined portfolio								
Set 1: apply the sample estimates and sample covariance matrix								
$\hat{\boldsymbol{w}}_{c\text{-}E}$	0.3362	0.3343	0.3321	0.3297	0.3269	0.3239	0.3169	0.3087
$\hat{\boldsymbol{w}}_{c\text{-}I}$	0.3280	0.3269	0.3255	0.3238	0.3218	0.3195	0.3138	0.3066
Set 2: apply the sophisticated estimates and sample covariance matrix								
$\tilde{\boldsymbol{w}}_{c\text{-}E}$	0.3409	0.3419	0.3425	0.3427	0.3425	0.3416	0.3378	0.3304
$\tilde{\boldsymbol{w}}_{c\text{-}I}$	0.3342	0.3348	0.3353	0.3356	0.3356	0.3351	0.3325	0.3265
Set 3: apply the sample estimates and shrinkage covariance matrix								
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.3649	0.3636	0.3619	0.3599	0.3575	0.3547	0.3478	0.3390
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.3546	0.3541	0.3533	0.3522	0.3508	0.3489	0.3438	0.3365
Set 4: apply the sophisticated estimates and shrinkage covariance matrix								
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.3538	0.3565	0.3590	0.3612	0.3630	0.3644	0.3648	0.3610
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.3536	0.3545	0.3555	0.3567	0.3577	0.3586	0.3592	0.3566
Panel C: Plug-in portfolios								
$\hat{\boldsymbol{w}}_{esg}$	0.2931	0.2908	0.2885	0.2861	0.2835	0.2809	0.2754	0.2696
$\hat{\boldsymbol{w}}^{LW}_{esg}$	0.3180	0.3156	0.3131	0.3104	0.3076	0.3047	0.2985	0.2918

risks associated with estimating combination coefficients, the implementable combined portfolios have lower utilities compared to their orcale counterparts. Notably, the impact of estimation risks is more pronounced for the implementable versions of $\hat{\boldsymbol{w}}^*_{c\text{-}I}$ since it is additionally affected by the estimation of ζ (see Theorem 3.2), a parameter determining the activity of the inequality constraint. It introduces an additional layer of complexity and risk, and contributes to lower utilities compared to those for $\hat{\boldsymbol{w}}^*_{c\text{-}E}$. Moreover, both implementable combined portfolios still significantly outperform the plug-in ESG portfolios.

Finally, although maximizing the expected out-of-sample utilities does not necessarily maximize the expected out-of-sample Sharpe ratio, the results in Table 3.7 demonstrate that the implementable combined portfolios still perform well. Particularly, the implementable combined portfolios from the equality constraint have higher Sharpe ratios than the ones from the inequality constraint. Moreover, they outperform the plug-in ESG portfolios.

3.5.2 Empirical Study

In the empirical study, we investigate the performance of various portfolios on up to 100 major stocks selected from S&P 500, as listed in Table 3.8. To select these stocks, we rank the S&P 500 components by market capitalization based on the most recent data as of the year 2024. Then we select the top 100 stocks that have complete price data and available ESG scores from our data sources. Specifically, these stocks have yearly updated ESG scores ranging from 2010 to 2021 from our available source of the ASSET4 dataset. The ESG scores of stocks are updated at the end of each year. Accordingly, we focus the investment analysis on the period from 2011-01-01 to 2022-12-31 and utilize the ESG scores from the previous year for each new investment year. For instance, for the investment year 2011, we utilize the ESG score from the year 2010. We download weekly adjusted close prices for these stocks from Yahoo Finance and calculate weekly returns using the ratios of the price changes to previous prices. Note that while selecting stocks with complete data facilitates our computation, it may introduce survivorship bias or overlook the potential effects of delisting. However, since our primary goal is to compare the performance of different portfolios using the same dataset, the impact of these potential biases on our findings is likely minimal.

Table 3.8: List of the top 100 selected components from the S&P500.

1	2	3	4	5	6	7	8	9	10
MSFT	AAPL	NVDA	GOOG	GOOGL	AMZN	BRK-B	LLY	JPM	V
WMT	XOM	UNH	MA	PG	ORCL	COST	JNJ	HD	MRK
BAC	NFLX	CVX	KO	AMD	QCOM	ADBE	PEP	CRM	TMO
LIN	TMUS	WFC	AMAT	DHR	CSCO	MCD	DIS	ABT	ACN
TXN	GE	VZ	INTU	AXP	AMGN	PM	CAT	MU	PFE
IBM	ISRG	NEE	CMCSA	NKE	SPGI	HON	UNP	LRCX	SCHW
SYK	BKNG	INTC	ETN	COP	LOW	T	VRTX	TJX	PGR
UPS	ADI	BLK	REGN	C	BSX	KLAC	LMT	BA	PLD
MDT	DE	MMC	ADP	CI	AMT	SNPS	SBUX	CMG	MDLZ
FI	SO	CDNS	BMY	APH	WM	GILD	GD	DUK	ICE

We employ a rolling window approach with a window length $T = 120$. To be more specific, starting from the $(T+1)$-th day and for each subsequent day, we utilize data from the preceding T days to compute the sample mean vector and shrinkage covariance matrix for asset returns. These matrices are then used to construct and implement portfolios for the corresponding day. To estimate the estimation window for the first investment day in year 2011, additional historical data prior to this date are required. In particular, when the rolling window length is set at $T = 120$, the entire data period for the historical stock prices extends from 2008-09-08 to 2022-12-31, where data before 2011-01-01 are utilized to generate the first rolling window. Furthermore, if we let X denote the number of weeks the whole dataset spans, the rolling window approach yields $(X - T)$ realized out-of-sample weekly returns for each portfolio.

In constructing various portfolios, we examine several cases with different numbers of stocks, namely, $N = 6, 10, 15, 25, 35, 55, 75, 100$. These cases involve using the top N stocks as detailed in Table 3.8. Similar to the previous simulation study, we first assess the performance by analyzing the total ESG scores of the implementable combined portfolios. We calculate the average and standard deviation of the total ESG scores across all rolling windows for each implementable combined portfolio. Note that unlike in the simulation study, the ESG scores for these stocks are updated annually. Table 3.9 provides results for three different scenarios, $\bar{s} = 0.7, 0.8, 0.9$, allowing us to analyze how changes in target ESG scores might impact performance.

The results presented in Table 3.9 provide the following observations. Firstly, the average total ESG scores for $\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$ and $\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$ are consistently lower than those for $\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$

Table 3.9: The average total ESG scores for portfolios in the empirical study.

N	6	10	15	25	35	55	75	100
Panel A: $\bar{s} = 0.7$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.7061	0.6935	0.7097	0.7274	0.7441	0.7605	0.8127	0.8241
	(0.0048)	(0.0088)	(0.0120)	(0.0189)	(0.0240)	(0.0311)	(0.0383)	(0.0533)
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.8478	1.0445	1.0785	0.8927	0.8566	0.8561	0.8630	0.8364
	(0.1773)	(0.4323)	(0.5465)	(0.3250)	(0.2496)	(0.1664)	(0.1204)	(0.0621)
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.8462	1.0482	1.0753	0.8731	0.8226	0.8130	0.7910	0.7964
	(0.1786)	(0.4293)	(0.5486)	(0.3347)	(0.2643)	(0.1895)	(0.1546)	(0.0798)
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Panel B: $\bar{s} = 0.8$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.8002	0.7843	0.7964	0.8058	0.8141	0.8180	0.8452	0.8400
	(0.0048)	(0.0088)	(0.0120)	(0.0189)	(0.0240)	(0.0311)	(0.0383)	(0.0533)
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.8898	1.0858	1.1250	0.9475	0.9101	0.8960	0.8874	0.8478
	(0.1477)	(0.3999)	(0.5157)	(0.2983)	(0.2272)	(0.1457)	(0.1099)	(0.0595)
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.8903	1.0945	1.1286	0.9422	0.8962	0.8797	0.8570	0.8329
	(0.1473)	(0.3933)	(0.5134)	(0.3005)	(0.2321)	(0.1522)	(0.1200)	(0.0498)
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Panel C: $\bar{s} = 0.9$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.8944	0.8751	0.8830	0.8841	0.8841	0.8755	0.8777	0.8558
	(0.0048)	(0.0088)	(0.0120)	(0.0189)	(0.0240)	(0.0311)	(0.0383)	(0.0533)
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.9478	1.1324	1.1735	1.0048	0.9662	0.9383	0.9124	0.8606
	(0.1171)	(0.3657)	(0.4855)	(0.2724)	(0.2053)	(0.1257)	(0.1002)	(0.0574)
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.9526	1.1469	1.1845	1.0153	0.9756	0.9522	0.9331	0.9050
	(0.1148)	(0.3555)	(0.4788)	(0.2675)	(0.2008)	(0.1171)	(0.0877)	(0.0150)
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

and $\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$. This is anticipated since $\tilde{\boldsymbol{w}}_c^I$ is structured under an inequality constraint, which allows the total ESG score for the realized portfolio to exceed the target one $\bar{s}$. Secondly, estimation errors associated with the combination coefficients are more pronounced as the number of assets N increases. Specifically, when N is relatively low (e.g., $N = 6, 10$), the average total ESG scores for $\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$ closely align with the target one $\bar{s}$. Conversely, at higher values of N (e.g., $N = 75, 100$), the significant estimation risk leads to wider divergence

Table 3.10: The annualized average out-of-sample returns for portfolios.

N	6	10	15	25	35	55	75	100
Panel A: $\bar{s} = 0.7$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.2955	0.2402	0.2198	0.1437	0.1224	0.1268	0.1144	0.0934
	(0.3622)	(0.3081)	(0.3648)	(0.3677)	(0.3474)	(0.3365)	(0.1932)	(0.1380)
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.2709	0.2139	0.1418	0.1195	0.1122	0.1119	0.1312	0.1044
	(0.3255)	(0.2057)	(0.1951)	(0.1540)	(0.1403)	(0.1377)	(0.1354)	(0.1328)
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.3064	0.2831	0.2638	0.1504	0.1238	0.1193	0.1182	0.0938
	(0.3399)	(0.3532)	(0.4089)	(0.3788)	(0.3550)	(0.3394)	(0.1958)	(0.1381)
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.2833	0.2570	0.1867	0.1266	0.1146	0.1039	0.1382	0.0979
	(0.2965)	(0.2581)	(0.2466)	(0.1700)	(0.1523)	(0.1408)	(0.1354)	(0.1309)
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.3390	0.2743	0.4246	0.2497	0.1727	0.2528	-0.0981	0.0630
	(0.5479)	(0.6712)	(0.9193)	(1.2206)	(1.5108)	(2.0467)	(2.8845)	(3.7646)
Panel B: $\bar{s} = 0.8$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.3061	0.2483	0.2174	0.1392	0.1180	0.1219	0.1117	0.0920
	(0.3326)	(0.3116)	(0.3650)	(0.3667)	(0.3468)	(0.3361)	(0.1928)	(0.1380)
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.2822	0.2228	0.1390	0.1137	0.1058	0.1033	0.1231	0.0958
	(0.2867)	(0.2100)	(0.1947)	(0.1517)	(0.1375)	(0.1354)	(0.1326)	(0.1314)
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.3094	0.2863	0.2604	0.1449	0.1177	0.1148	0.1148	0.0926
	(0.3399)	(0.3539)	(0.4084)	(0.3772)	(0.3539)	(0.3388)	(0.1952)	(0.1381)
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.2865	0.2605	0.1825	0.1196	0.1058	0.0957	0.1302	0.0959
	(0.2949)	(0.2594)	(0.2464)	(0.1678)	(0.1497)	(0.1393)	(0.1336)	(0.1309)
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.3504	0.2832	0.4218	0.2439	0.1663	0.2442	-0.1062	0.0544
	(0.5286)	(0.6745)	(0.9197)	(1.2200)	(1.5107)	(2.0466)	(2.8839)	(3.7650)
Panel C: $\bar{s} = 0.9$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.3168	0.2564	0.2150	0.1346	0.1135	0.1169	0.1091	0.0907
	(0.3230)	(0.3163)	(0.3661)	(0.3663)	(0.3466)	(0.3360)	(0.1925)	(0.1380)
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.2936	0.2317	0.1362	0.1079	0.0994	0.0947	0.1150	0.0873
	(0.2729)	(0.2166)	(0.1965)	(0.1513)	(0.1370)	(0.1351)	(0.1316)	(0.1313)
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.3243	0.2902	0.2572	0.1390	0.1124	0.1098	0.1114	0.0914
	(0.3454)	(0.3554)	(0.4082)	(0.3761)	(0.3532)	(0.3385)	(0.1948)	(0.1381)
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.3023	0.2648	0.1787	0.1122	0.0977	0.0869	0.1197	0.0881
	(0.2994)	(0.2618)	(0.2471)	(0.1667)	(0.1486)	(0.1390)	(0.1328)	(0.1313)
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.3617	0.2921	0.4190	0.2381	0.1599	0.2356	-0.1144	0.0458
	(0.5240)	(0.6785)	(0.9206)	(1.2197)	(1.5108)	(2.0466)	(2.8834)	(3.7655)

between the average total ESG score for $\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$ and the target $\bar{s}$. Thirdly, the average total ESG scores for $\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$ are generally larger than the target one $\bar{s}$, with an exception at $N = 100$ due to the severe estimation risk when N/T is close to one. This suggests that portfolios constructed under the inequality constraint typically achieve higher total ESG scores than the target one $\bar{s}$ across multiple rolling windows.

Furthermore, we estimate the expected out-of-sample mean and variance for any implementable portfolio $\hat{\boldsymbol{w}}$ as the sample mean (denoted by $\hat{\mu}_{\hat{\boldsymbol{w}}}$) and variance (denoted by $\hat{\sigma}^2_{\hat{\boldsymbol{w}}}$) of the $(X - T)$ out-of-sample daily returns, respectively. We annualize them by multiplying by 52 upon taking one year as 52 weeks. The certainty equivalent return (CER) is computed by

$$CER_{\hat{\boldsymbol{w}}} = \hat{\mu}_{\hat{\boldsymbol{w}}} - \frac{\gamma}{2}\hat{\sigma}^2_{\hat{\boldsymbol{w}}},$$

where the risk aversion coefficient is set to be $\gamma = 3$. The CER represents the risk-free rate of return that investors are willing to accept instead of taking on the risk associated with the risky portfolio $\hat{\boldsymbol{w}}$. Similarly, the Sharpe ratio is computed by the ratio between the sample mean $\hat{\mu}_{\hat{\boldsymbol{w}}}$ and the standard deviation $\hat{\sigma}_{\hat{\boldsymbol{w}}}$. The results are given in Tables 3.10, 3.11 and 3.12.

Based on the results in Table 3.10, several observations can be made regarding the annualized average out-of-sample portfolio returns across different portfolio sizes N and target ESG scores $\bar{s}$. Firstly, when comparing portfolios with smaller asset pools (specifically, when N is under 75) across different target ESG scores, the portfolio $\hat{\boldsymbol{w}}^{LW}_{esg}$ tends to generate higher returns but also shows significantly greater standard deviations. In some

Table 3.11: Annualized certainty equivalent returns across various portfolios.

N	6	10	15	25	35	55	75	100
Panel A: $\bar{s} = 0.7$								
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.0987	0.0978	0.0202	-0.0590	-0.0586	-0.0430	0.0584	0.0649
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.1120	0.1504	0.0847	0.0840	0.0827	0.0834	0.1038	0.0780
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.1331	0.0960	0.0130	-0.0648	-0.0653	-0.0534	0.0607	0.0652
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.1515	0.1570	0.0955	0.0832	0.0798	0.0742	0.1107	0.0722
$\hat{\boldsymbol{w}}^{LW}_{esg}$	-0.1113	-0.4015	-0.8432	-1.9850	-3.2510	-6.0309	-12.5788	-21.1953
Panel B: $\bar{s} = 0.8$								
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.1402	0.1027	0.0176	-0.0626	-0.0624	-0.0476	0.0560	0.0635
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.1590	0.1566	0.0821	0.0792	0.0775	0.0758	0.0967	0.0700
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.1361	0.0984	0.0103	-0.0685	-0.0702	-0.0573	0.0576	0.0640
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.1560	0.1596	0.0914	0.0774	0.0722	0.0666	0.1034	0.0702
$\hat{\boldsymbol{w}}^{LW}_{esg}$	-0.0687	-0.3992	-0.8471	-1.9887	-3.2570	-6.0387	-12.5818	-21.2088
Panel C: $\bar{s} = 0.9$								
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.1603	0.1063	0.0139	-0.0667	-0.0667	-0.0524	0.0535	0.0621
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}E}$	0.1819	0.1613	0.0782	0.0736	0.0713	0.0673	0.0890	0.0614
$\hat{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.1454	0.1008	0.0072	-0.0732	-0.0747	-0.0621	0.0545	0.0628
$\tilde{\boldsymbol{w}}^{LW}_{c\text{-}I}$	0.1678	0.1621	0.0871	0.0705	0.0646	0.0580	0.0932	0.0622
$\hat{\boldsymbol{w}}^{LW}_{esg}$	-0.0502	-0.3984	-0.8524	-1.9933	-3.2639	-6.0473	-12.5854	-21.2228

Table 3.12: Annualized Sharpe ratio across various portfolios.

N	6	10	15	25	35	50	75	100
Panel A: $\bar{s} = 0.7$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.8157	0.7796	0.6026	0.3909	0.3524	0.3769	0.5919	0.6770
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.8322	1.0397	0.7268	0.7765	0.7998	0.8125	0.9694	0.7863
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.9014	0.8016	0.6452	0.3972	0.3486	0.3517	0.6036	0.6794
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.9556	0.9956	0.7572	0.7446	0.7527	0.7382	1.0210	0.7478
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.6187	0.4086	0.4618	0.2046	0.1143	0.1235	-0.0340	0.0167
Panel B: $\bar{s} = 0.8$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.9203	0.7970	0.5956	0.3795	0.3402	0.3626	0.5796	0.6672
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.9846	1.0608	0.7139	0.7500	0.7695	0.7626	0.9284	0.7295
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.9102	0.8090	0.6377	0.3842	0.3325	0.3389	0.5879	0.6710
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.9713	1.0044	0.7406	0.7128	0.7067	0.6873	0.9740	0.7325
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.6628	0.4199	0.4586	0.1999	0.1101	0.1193	-0.0368	0.0144
Panel C: $\bar{s} = 0.9$								
$\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$	0.9809	0.8106	0.5871	0.3675	0.3275	0.3480	0.5668	0.6573
$\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$	1.0760	1.0697	0.6929	0.7134	0.7260	0.7009	0.8736	0.6645
$\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$	0.9391	0.8166	0.6300	0.3695	0.3182	0.3243	0.5720	0.6620
$\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$	1.0096	1.0118	0.7231	0.6731	0.6575	0.6257	0.9013	0.6710
$\hat{\boldsymbol{w}}_{esg}^{LW}$	0.6902	0.4305	0.4551	0.1952	0.1059	0.1151	-0.0397	0.0122

instances, while the average returns of $\hat{\boldsymbol{w}}_{esg}^{LW}$ may double those of the other portfolios, its standard deviations can increase tenfold. It suggests that the higher average returns hardly compensate for the increased risk and volatility, which leads to substantial fluctuations and less stable performances with $\hat{\boldsymbol{w}}_{esg}^{LW}$. Secondly, considering the two implementable combined portfolios with the same estimators across different portfolio sizes and target ESG scores, the average returns of $\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$ (or $\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$) are typically larger than those of $\hat{\boldsymbol{w}}_{c\text{-}E}^{LW}$ (or $\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$), with only a few exceptions. This performance can be attributed to the flexibility offered by the inequality constraint, allowing $\hat{\boldsymbol{w}}_{c\text{-}I}^{LW}$ (or $\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$) to adapt effectively to varying market conditions. Note that this finding contrasts with the one from the simulation study, and the difference highlights the considerable influence of real-world market dynamics on portfolio performance.

Tables 3.11 and 3.12 present the annualized certainty equivalent returns and Sharpe ratios, respectively. Notably, the four implementable combined portfolios consistently outperform the plug-in ESG portfolio $\hat{\boldsymbol{w}}_{esg}^{LW}$ across different numbers of assets and target ESG scores on both metrics. Furthermore, a comparison between the implementable combined portfolios using sample estimators and those using sophisticated estimators shows that portfolios employing sophisticated estimators generally perform better than those using sample estimators. Additionally, the portfolio $\tilde{\boldsymbol{w}}_{c\text{-}I}^{LW}$ outperform $\tilde{\boldsymbol{w}}_{c\text{-}E}^{LW}$ in over half of the scenarios in terms of annualized certainty equivalent returns. Theoretically, in scenarios characterized by a multivariate normal distribution assumption and the absence of estimation risk associated with ζ (see Theorem 3.2) and optimal combination coefficients, $\hat{\boldsymbol{w}}_{c\text{-}I}^{*}$ will not underperform $\hat{\boldsymbol{w}}_{c\text{-}E}^{*}$ in terms of the expected out-of-sample utility, given the same target ESG score $\bar{s}$. However, our empirical analysis presents two challenges: first, real data often deviate from the assumption of a multivariate normal distribution; second, there is estimation risk associated with the combination coefficients and ζ. To be more specific, given the target ESG score $\bar{s}$ and the realized threshold $\hat{\bar{s}}_i^{**}$ (see Corollary 3.2) estimated from the i-th window, the condition $\hat{\bar{s}}_i^{**} \geq \bar{s}$ or $\hat{\bar{s}}_i^{**} < \bar{s}$ may not consistently hold across all windows $i, i = 1, ..., (X - T)$. It is the windows $\{j : \hat{\bar{s}}_j^{**} < \bar{s}, j = 1, ..., X - T\}$ that that lead to the differences in the implementable combined portfolio between the inequality and equality constraints. These may lead to the implementable combined portfolio subject to the inequality constraint underperforming the portfolio with the equality constraint.

3.6 Conclusion

In this chapter, we introduce an MV optimization framework that incorporates both the weight constraint and the total ESG constraint. The resulting optimal ESG portfolio satisfies a three-fund separation. To address estimation risk, we propose a combined three-fund portfolio with components originating from the plug-in ESG portfolio. Under the assumption of a multivariate normal distribution for asset returns, we explore two scenarios: deriving the optimal combination coefficients through expected out-of-sample MV utility optimization, which incorporates either an inequality or equality constraint on the expected total ESG scores. Through both simulation and empirical studies, we find that the combined portfolios employing sophisticated estimators generally perform better than those using sample estimators in estimating the optimal combination coefficients. The implementable combined portfolios consistently outperform the plug-in ESG portfolios in terms of certainty equivalent return and Sharpe ratio.

Although our primary focus in this chapter is on ESG investing within the MV framework, it is also worthwhile to consider extending the discussion to the MVS framework as in Chapter 2, with particular attention to tail risks.

Chapter 4

Winning Probability Weighted Combined Portfolio

4.1 Introduction

In Chapters 2 and 3, we analytically determines the optimal combination coefficients for specific combined portfolios by maximizing the expected out-of-sample performance under the normality assumption of asset returns. However, such analytical methods are typically constrained to specific pairs of constituent portfolios due to the inherent complexity in deriving explicit expressions for the performance measures. In order to offer greater flexibility, this chapter introduces a novel Winning Probability Weighting (WPW) framework for constructing combined portfolios.

Unlike the aforementioned studies, we assume that each constituent portfolio exhibits weakly stationary out-of-sample returns, a premise commonly accepted in the literature (e.g., Györfi et al., 2008) and empirically validated in our numerical analysis through statistical hypothesis tests across multiple datasets. The WPW framework is founded on the innovative concept of characterizing combination coefficients as winning probabilities, which are the likelihoods of any one constituent portfolio outperforming the others in the out-of-sample scenarios. This innovation lays the groundwork for employing various machine learning methods to determine the combination coefficients, further allowing the incorporation of financial market factors, along with traditional historical return data in the portfolio construction process. Additionally, the WPW framework offers the flexibility to merge any two or more constituent portfolios, which is hardly achievable using the analytical methods.

The WPW framework stands out from the existing literature by offering advantages and flexibility from the following perspectives. Firstly, by framing the prediction of the winning probability as a classification problem, it becomes viable to utilize various machine learning methods, such as logistic regression and random forest, to leverage financial market data structures and movements to determine the combination coefficients. The application of machine learning to financial problems has garnered increasing attention in recent literature, as seen in studies such as Krauss et al. (2017), Fischer and Krauss (2018), Huck (2019), Chen et al. (2022), and Flori and Regoli (2021), among others. Notably, Lassance and Martin-Utrera (2023) propose calibrating the shrinkage intensity for the covariance matrix as the probability that the out-of-sample return of one portfolio will surpass that of the arbitrage portfolio, and then estimating it using the logistic regression with prior sentiment data. Contrasting with their work, our study focuses on the construction of combined portfolios and proposes to determine combination coefficients as the winning probabilities. In our WPW framework, we first adjust the constituent portfolios to share the same long-term variance so that they are comparable in long-term risk profile, and then we apply classification models to utilize financial market information for the estimation of the winning probabilities.

Secondly, the WPW framework enhances portfolio construction by facilitating the inclusion of exogenous factors, such as Fama-French factors, with historical asset returns to determine the combination coefficients. The external factors, as valuable sources of market information, can be effectively incorporated as covariates in the classification procedure to improve the accuracy of the winning probability predictions. Substantial evidence from the literature underscores the critical impact of financial market factors on the valuation of assets and portfolios, as demonstrated in studies by Nti et al. (2020) and Neely et al. (2014), among others. As revealed in our subsequent numerical analysis, the inclusion of fundamental market factors greatly contributes to improving portfolio performance. It is worth mentioning that Kazak and Pohlmeier (2023) present the combination coefficients in combined portfolios as bagged probabilities using bootstrap methods to assess the likelihood that a given constituent portfolio will be overall best performing. A crucial distinction between our methodology and theirs is that our framework can seamlessly incorporate exogenous factors into the analysis, moving beyond the conventional reliance on historical returns.

Thirdly, we further enhance the combined portfolio by scaling it when targeting the expected out-of-sample mean-variance (MV) utility optimization. As a relevant contribution, Nardari and Schüssler (2023) determine the combination coefficients using an ensemble approach that maximizes the utility generated jointly by the constituent rules. In contrast, our framework introduces a scaling parameter to achieve utility maximization rather than directly determine the associated combination coefficients.

In the empirical study, we utilized logistic regression and random forests as predictive models to estimate the winning probability, employing technical and fundamental features as input factors. The technical features comprised historical out-of-sample returns of the constituent portfolios, while the fundamental features encompassed widely recognized financial market factors from the existing literature. To evaluate the out-of-sample performance of our WPW combined portfolio, we explore various combined portfolios in the empirical study across different datasets. The results consistently demonstrate that our WPW combined portfolios outperformed both the analytical method-based portfolios and the corresponding constituent portfolios in terms of the certainty equivalent return (CER). There were only a few exceptions where we observed slight underperformance. The superiority of the WPW combined portfolios can be attributed to their data-driven nature, as they directly extract information from historical realizations to determine the combination coefficients without imposing strong assumptions like the normal distribution assumption.

The remainder of the chapter is organized as follows. Section 4.2 introduces combined portfolios and winning probability. Section 4.3 presents the winning probability weighting framework for the construction of combined portfolios. Section 4.4 introduces the prediction of the winning probability, including predictive models and feature exploration. Section 4.5 presents the empirical results based on various real datasets.

4.2 Combined Portfolios and Winning Probabilities

Consider an investor allocating wealth across N risky assets along with one risk-free asset. At time t, let $\boldsymbol{R}_t = [R_{1,t}, ..., R_{N,t}]^\top$ be the vector of asset excess returns with an N-dimensional mean vector $\boldsymbol{\mu} \in \mathbb{R}^N$ and a covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ of full rank. Here,

$R_{i,t}$ represents the excess return of risky asset i over the risk-free asset, $i = 1, \ldots, N$. The rate of return of the risk-free asset is denoted by $R_{f,t}$. A portfolio strategy is represented by a vector of investment weights (or fractions) $\boldsymbol{w} = [w_1, ..., w_N]^\top$, where w_i is the proportion of wealth allocated to the risky asset i. Consequently, the proportion of wealth invested in the risk-free asset is $1 - \boldsymbol{w}^\top \mathbf{1}$. Here, $\mathbf{1}$ is an N-dimensional vector of ones.

Over time, a wide range of portfolio strategies have been proposed in the literature to guide investors in making investment decisions. In the following, we first introduce estimation risk and a widely recognized strategy category known as combined portfolios and then delineate the connection between winning probabilities and combination coefficients of combined portfolios.

4.2.1 Estimation Risk and Combined Portfolios

The foundational investment framework introduced by Markowitz (1952) revolutionizes the design of optimal portfolio strategies by revealing the best trade-off between return and risk. Central to this seminal work is the mean variance (MV) portfolio $\boldsymbol{w}_{mv} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\gamma$, which maximizes the MV objective function with a specific risk aversion coefficient γ. Building upon this cornerstone, various investment models have continuously emerged to offer advanced insights into risk-reward optimization. Nonetheless, the implementation of most portfolio strategies often encounters practical challenges, primarily due to the reliance on the unknown mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of asset returns. Consequently, the effectiveness of these strategies is linked to the precise estimation of these unknown parameters or, more broadly, asset return distributions, which are complicated by the inherent uncertainty and fluctuation of financial markets.

A straightforward approach to circumvent this practical challenge is to replace the unknown parameters with their sample counterparts, giving rise to what are commonly known as plug-in portfolio strategies. Specifically, we consider a rolling window setup. At each time t, we first create estimates of mean vector and covariance matrix based on historical excess returns $\{\boldsymbol{R}_{t-M}, ..., \boldsymbol{R}_{t-1}\}$ of risky assets over the last M periods:

$$\hat{\boldsymbol{\mu}}_t = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{R}_{t-i} \qquad \text{and} \qquad \widehat{\boldsymbol{\Sigma}}_t = \frac{1}{M}\sum_{t=1}^{M}(\boldsymbol{R}_{t-i} - \hat{\boldsymbol{\mu}}_t)(\boldsymbol{R}_{t-i} - \hat{\boldsymbol{\mu}}_t)^\top. \tag{4.1}$$

The plug-in portfolios are then formulated using these two estimates, for example, the plug-in MV portfolio $\hat{\boldsymbol{w}}_{mv,t} = \widehat{\boldsymbol{\Sigma}}_t^{-1}\hat{\boldsymbol{\mu}}_t/\gamma$. However, given that the estimates from a certain historical period are unlikely to perfectly match the true parameters, the plug-in portfolio strategies are subject to temporal fluctuations and can significantly diverge from their theoretical optimal counterparts. The discrepancy between the theoretical optimal and plug-in portfolios, resulting from the inherent estimation error associated with the unknown parameters, often leads to deteriorated out-of-sample portfolio performance (see Frankfurter et al., 1971; Michaud, 1989; Kan and Zhou, 2007; DeMiguel et al., 2009b, for example).

To mitigate estimation risk and enhance out-of-sample performance for plug-in portfolio strategies, we focus on combined portfolios, a widely recognized class of strategies that aggregate two or more predetermined constituent portfolios through weighted combinations. The majority of existing literature typically focuses on combining two constituent portfolios due to the complexity and risks involved in estimating combination coefficients, where the associated estimation risk with additional constituent portfolios may outweigh their potential benefits (e.g., Tu and Zhou, 2011). Therefore, in this chapter we focus on combined portfolios with two constituent portfolios for which the weight vector for period t is given by:

$$\boldsymbol{w}_{c,t} = \delta_{1,t}\hat{\boldsymbol{w}}_{1,t} + \delta_{2,t}\hat{\boldsymbol{w}}_{2,t}, \tag{4.2}$$

where $\delta_{i,t}$ is the combination coefficient associated with the plug-in constituent portfolio $\hat{\boldsymbol{w}}_{i,t}, i = 1, 2$. Constituent portfolios can be flexibly selected from a variety of strategies documented in both academic research and industry practices, which include but are not limited to the MV portfolio, Bayesian portfolio rules (e.g., Jorion, 1986; Bodnar et al., 2017), portfolios incorporating constraints or regularization (e.g., Jagannathan and Ma, 2003; DeMiguel et al., 2009a), or the 1/N rule that allocates wealth uniformly across all available assets (DeMiguel et al., 2009b).

The development of combined portfolios can be traced back to methodologies like Bayesian approaches (Jorion, 1986; Frost and Savarino, 1986), robust optimization (Garlappi et al., 2007), or shrinkage estimation (DeMiguel et al., 2013). The principal benefit of employing a combined portfolio lies in its ability to effectively integrate the distinct advantages of its constituent portfolios, given that the combination coefficients are strategically constructed. Hence, the performance and effectiveness of the combined portfolio are significantly affected by these combination coefficients, where a set of well-chosen combination coefficients enables the combined portfolio to potentially outperform its individual components. Therefore, it is crucial to determine desirable combination coefficients to ensure the combined portfolio achieves superior out-of-sample performance.

Combined portfolios are theoretically appealing and have demonstrated robust performance in empirical studies as documented in the literature. Traditional analytical methods for determining combination coefficients typically rely on normality assumptions and are limited to specific pairs of constituent portfolios due to the complexity of deriving explicit expressions for performance measures (e.g., Kan and Zhou, 2007; Tu and Zhou, 2011; Lassance et al., 2024). To offer greater flexibility, we offer a novel perspective that utilizes winning probabilities to construct combined portfolios, which is in principle applicable to any pair of constituent portfolios.

4.2.2 Link Winning Probabilities with Combination Coefficients

Winning probability is introduced to quantify the likelihood that a particular constituent portfolio will surpass the out-of-sample performance of the other constituent strategy involved in the combined portfolio (4.2). A larger winning probability suggests a greater chance of this constituent portfolio outperforming the other during the out-of-sample period. By linking winning probabilities with combination coefficients, constituent portfolios with larger winning probabilities are generally assigned more weights in the construction of combined portfolios.

4.2.2.1 Winning Probabilities

To illustrate, it is crucial to first assume that we have adjusted the two constituent portfolios involved in the combined portfolio such that they have comparable risk profiles, such as volatility. Denote these two constituent portfolios after the risk adjustment as $\tilde{\boldsymbol{w}}_i, i = 1, 2$. More details about how to adjust the risk profile of constituent portfolios will be presented in Section 4.3.1. When the risk profiles of the components in a combined portfolio are equalized, it is intuitive to turn our attention to their potential rewards. Sitting at the end of the trading period t, we assess the performance of adjusted constituent portfolios using their out-of-sample returns from period $t+1$:

$$r(\tilde{\boldsymbol{w}}_{i,t}) = \tilde{\boldsymbol{w}}_{i,t}^\top \boldsymbol{R}_{t+1} + R_{f,t+1}, \;\; i = 1, 2.$$

The WPW framework is motivated by considering the following binary selection rule. Given the adjusted constituent portfolios $\tilde{\boldsymbol{w}}_i, i = 1, 2$, with comparable risk profiles, the binary selection rule is captured by selection indicators defined via a straightforward bang-bang method as follows:

$$\boldsymbol{I}_{i,t} = \mathbb{I}_{\{i=\text{argmax}_{j\in\{1,2\}} r(\tilde{\boldsymbol{w}}_{j,t})\}}, \quad i = 1, 2, \tag{4.3}$$

where $\mathbb{I}_{\{A\}}$ is the indicator function for any event A. The binary selection indicator $\boldsymbol{I}_{i,t}$ takes a value of one if the adjusted constituent portfolio i exhibits a higher out-of-sample return than the other one, and zero otherwise. If the selection indicators are used as the combination coefficients for the combined portfolio, it emphasizes the intent to invest fully in the adjusted constituent portfolio that yields a higher out-of-sample return.

However, such selection rule (4.3) is practically infeasible, as it relies on the future asset returns, $\boldsymbol{R}_{t+1}$, which remains unknown at the time of deploying the combined portfolio for period t. Additionally, given that the constituent portfolios are related to historical return data, the selection indicators are also random variables as functions of the historical return data. To overcome the implementation challenge and provide a more reliable alternative grounded in the selection indicator, we propose winning probabilities as defined below:

$$p_{i,t} = \mathbb{E}[\boldsymbol{I}_{i,t}|\mathcal{F}_t] = \mathbb{P}[r(\tilde{\boldsymbol{w}}_{i,t}) > r(\tilde{\boldsymbol{w}}_{j,t})|\mathcal{F}_t], \;\; i, j = 1, 2, \;\; \text{and} \;\; i \neq j, \tag{4.4}$$

where $\mathcal{F}_t$ generally denotes the σ-field of all relevant events (e.g., generated by all considered asset prices and market factors) considered up to time t. The winning probability $p_{i,t}$, derived from the expectation of the selection indicator conditional on available information, quantifies the probability that the adjusted constituent portfolio i will outperform the other one in terms of out-of-sample returns. Additionally, the winning probabilities are automatically normalized such that their sum across all adjusted constituent portfolios at any given period t equals one, i.e., $p_{1,t} + p_{2,t} = 1$.

4.2.2.2 Winning Probabilities as Combination Coefficients

Given that a portfolio with a higher winning probability is more likely to outperform the other in terms of out-of-sample return, there are several compelling reasons to motivate us to adopt the winning probabilities as weighting parameters assigned to the adjusted constituent portfolios in the combined portfolio. Firstly, the winning probability is intricately linked to the selection indicator, serving as the optimal predictor for minimizing the mean squared error associated with it. This connection refines the decision-making process by transitioning from simplistic binary indicators to the corresponding probabilities that effectively capture the inherent variability and unpredictability of financial markets. Secondly, the winning probability facilitates dynamic adjustments for portfolio strategies. As new information becomes available, the winning probability can be updated to support real-time decision-making. The flexibility ensures that investment decisions remain aligned with the latest market developments, potentially enhancing portfolio performance in a volatile environment. Finally, as highlighted by Kazak and Pohlmeier (2023), bagging can effectively stabilize binary decision processes, and they replace the binary indicators with bootstrapped probabilities to smooth decision-making. In contrast, we introduce winning probabilities to supplant the binary selection indicators, thereby enhancing stability within the binary selection framework in a comparable manner.

The winning probability $p_{i,t}$ is inherently unknown at time t since it depends on the out-of-sample return for the subsequent period but is potentially estimable. We propose to formulate the estimation of winning probabilities as a classification problem and detail a training-testing procedure in Section 4.3.3. This formulation allows us to employ various machine learning techniques and, in the meantime, enables us not only to utilize technical data, such as historical returns of assets or portfolios, but also to incorporate fundamental market features to enhance the estimation of winning probabilities, which are the key advantages of connecting the winning probabilities with the combined portfolio. Therefore, denote any chosen predictive model $f(\cdot)$ along with the feature variables $\boldsymbol{x}_t$ measurable with respect to $\mathcal{F}_t$ at any given time t, the estimated winning probabilities can be written as:

$$[\hat{p}_{1,t}, \hat{p}_{2,t}]^\top = [f(\boldsymbol{x}_t), 1 - f(\boldsymbol{x}_t)]^\top. \tag{4.5}$$

We explore different predictive models and assess a wide range of both fundamental and technical feature variables as detailed in Section 4.4 for estimating winning probabilities.

Upon obtaining the estimated winning probabilities, we denote the combined portfolio that is weighted by the winning probabilities across the adjusted constituent portfolios as

$$\tilde{\boldsymbol{w}}_{c,t} = \hat{p}_{1,t}\tilde{\boldsymbol{w}}_{1,t} + \hat{p}_{2,t}\tilde{\boldsymbol{w}}_{2,t} \tag{4.6}$$

This formulation ensures that the combination coefficients are designed to reflect the estimated relative performance in terms of out-of-sample returns of components. In the meantime, using the adjusted constituent portfolios which have comparable risk profiles reflect the adjustment for risk characteristics.

As an alternative approach, one might consider treating the constituent portfolios as individual assets and view the determination of the combination coefficients as a portfolio selection problem for two assets. This method requires estimating the expected returns and covariance matrix of portfolio returns of the constituent portfolios, which must be updated periodically for a dynamically rebalancing portfolio. Such estimates are notoriously imprecise due to their sensitivity to market fluctuations. Furthermore, it is plausible to consider a regression problem aimed at predicting the mean and variances of the constituent portfolio returns to inform the construction of the combined portfolio. However, the literature indicates that stock/portfolio returns, when treated as continuous response variables, are highly sensitive to historical data and challenging to predict accurately. In general, classification models tend to outperform regression models in financial market prediction; see Leung et al. (2000). One can also refer to Henrique et al. (2019) for the application of different machine learning techniques for financial market prediction.

Notably, Lassance and Martin-Utrera (2023) calibrate the shrinkage intensity of a shrinkage covariance matrix as the probability derived from comparing out-of-sample returns using logistic regression. Our work explores a different path by focusing on the construction of combined portfolios. Also, we adopt the winning probabilities as combination coefficients conditional on that the constituent portfolios have adjusted to possess comparable risk profiles. A more detailed elaboration of our framework for the construction of the WPW combined portfolio is given in the next section.

4.3 Winning Probability Weighting (WPW) Framework for Combined Portfolios

Besides taking winning probabilities as combination coefficients as described before, the WPW framework contains two additional features: one is long-run variance adjustment, and the other is expected out-of-sample utility optimization. Firstly, we assume that the out-of-sample returns of each constituent portfolio satisfy a weakly stationary process, which allow us to align the long-term risk profiles of constituent portfolios via the long-run variance adjustment. We then favor the adjusted constituent portfolio exhibiting a higher out-of-sample return than the other one, which links to the winning probability as discussed in Section 4.2.2. Secondly, for further potential improvement, we scale the combined portfolio to optimize the expected out-of-sample utility. This adjustment enables investors to tailor the portfolio in line with their specific risk-reward preferences. Finally, the last part of this section presents a general training-testing procedure for the WPW framework.

4.3.1 Long-run Variance Adjustment for Constituent Portfolios

To avoid the complexities involved in dynamically and actually estimating the conditional variance of portfolio returns, we utilize long-run variances (LV) to assess the long-term risk profiles of constituent portfolios. We introduce a LV adjustment parameter for each constituent portfolio to ensure that their risk profiles remain consistent over the long term. That is, by multiplying these adjustment parameters with the corresponding constituent portfolios, we standardize the out-of-sample return series of all adjusted constituent portfolios to exhibit uniform long-run variances.

4.3.1.1 Long-run Variance Adjusted Parameters

We assume that out-of-sample returns from each constituent portfolio adhere to a weakly stationary process, as defined in Definition 4.1. The premise of stationarity is prevalent in economic and financial modeling (e.g., Györfi et al., 2008) and is empirically validated through statistical hypothesis tests across various datasets and portfolio strategies in our numerical analysis.

Definition 4.1. *(Hamilton, 2020) A time series $\{r_t\}_t$ is a (weakly) stationary process if*

$$\mathbb{E}[r_t] = \mu \text{ for all } t,$$
$$\mathbb{E}[(r_t - \mu)(r_{t-j} - \mu)] = \gamma_j \text{ for all } t \text{ and any integer } j,$$

where μ and γ_j are constants, and the later is called the j^{th} order autocovariance.

Under the stationary assumption, LV is a critical measure that characterizes the variance behavior of a stationary time series as the number of observations approaches infinity. It provides valuable insights into the temporal stability and variability of the series. For a stationary time series $\boldsymbol{r} = \{r_t\}_{t=1}^T$ with size T, the LV can be expressed as:

$$\text{LV}(\boldsymbol{r}) = \lim_{T\to\infty} Var[\sqrt{T}(\bar{r} - \mu)] = \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j, \tag{4.7}$$

where $\bar{r} = \sum_{t=1}^T r_t/T$ and the last equation is due to a symmetry property: $\gamma_{-j} = \gamma_j$.

Let us denote the time series of out-of-sample returns for the constituent portfolio i as $\boldsymbol{r}_{\hat{\boldsymbol{w}}_i} = \{r(\hat{\boldsymbol{w}}_{i,t})\}_{t=1}^T$ and the corresponding long-run variance as $\text{LV}(\boldsymbol{r}_{\hat{\boldsymbol{w}}_i})$, $i = 1, 2$. To standardize the long-term risk profiles across all constituent portfolios, we introduce a LV adjustment parameter ξ_i for each constituent portfolio i. Our objective is to equalize the long-run variances of the adjusted constituent portfolios that are obtained by scaling the constituent portfolio weights, i.e., $\xi_i\hat{\boldsymbol{w}}_i$. That is, $\text{LV}(\boldsymbol{r}_{\xi_1\hat{\boldsymbol{w}}_1}) = \text{LV}(\boldsymbol{r}_{\xi_2\hat{\boldsymbol{w}}_2})$. Hence, the long-run variance adjustment parameter can be determined as follows:

$$\xi_i = \sqrt{\frac{\text{LV}(\boldsymbol{r}_{\hat{\boldsymbol{w}}_{\tilde{i}}})}{\text{LV}(\boldsymbol{r}_{\hat{\boldsymbol{w}}_i})}}, \quad i = 1, 2, \tag{4.8}$$

where $\tilde{i} = \text{argmin}_{i\in\{1,2\}} \text{LV}(\boldsymbol{r}_{\hat{\boldsymbol{w}}_i})$ represents the constituent portfolio with a smaller LV and serves as the benchmark for variance adjustment. This equalization of LV implies that all adjusted constituent portfolios are subject to the same level of risk (measured by the LV of out-of-sample return series) over the long term. As a relevant note, DeMiguel et al. (2013) consider scaling parameters for constituent portfolios so as to mitigate the bias of their shrinkage target. In contrast, our LV adjustment parameter ξ_i aims to align the long-term risk profiles of constituent portfolios.

4.3.1.2 Estimation of Long-run Variance Adjusted Parameters

Given that the LV of the out-of-sample return series from each constituent portfolio is not directly observable, we have to estimate these variances to obtain feasible LV adjustment parameters. The LV is typically estimated using a weighted sum of sample autocovariances, where the weights are obtained by a kernel function and a bandwidth parameter to guarantee positive semi-definiteness of the matrix. Early contributions for robust estimation of the LV include Newey and West (1987), Andrews and Monahan (1992), Müller (2007), among others. We adopt the Newey-West estimator (Newey and West, 1987), a popular and widely accepted method for estimating the LV of stationary time series. The Newey-West estimator for any time series $\boldsymbol{r}$ is calculated by

$$\widehat{\mathrm{LV}}(\boldsymbol{r}) = \widehat{\gamma}_0 + 2\sum_{j=1}^{m_T} \widehat{\gamma}_j \mathcal{K}(j), \tag{4.9}$$

where $\mathcal{K}(\cdot)$ is a symmetric kernel function with $\mathcal{K}(0) = 1$, m_T is the bandwidth parameter, and $\widehat{\gamma}_i$ represents the lag i sample autocovariance given by

$$\widehat{\gamma}_i = \frac{1}{T}\sum_{t=i+1}^{T} (r_t - \bar{r})(r_{t-i} - \bar{r}), \;\; i = 0, 1, \ldots, m_T.$$

In our empirical studies, we utilize the popular Bartlett kernel $\mathcal{K}(j) = 1 - j/(m_T + 1)$ for $1 \leq j \leq m_T$ and zero otherwise. The bandwidth parameter m_T is dependent on the sample size, and we set it equal to the integer part of $4(T/100)^{2/9}$ as recommended by Newey and West (1987). The resulting Newey-West estimator of the LV is asymptotically consistent and positive semi-definite for finite samples. Consequently, we calculate the empirical LV adjustment parameters using (4.9) as

$$\hat{\xi}_i = \sqrt{\frac{\widehat{\mathrm{LV}}(\boldsymbol{r}_{\hat{\boldsymbol{w}}_{\bar{i}}})}{\widehat{\mathrm{LV}}(\boldsymbol{r}_{\hat{\boldsymbol{w}}_i})}}, \;\; i = 1, 2, \tag{4.10}$$

and obtain the adjusted constituent portfolios for (4.6) as $\tilde{\boldsymbol{w}}_{i,t} = \hat{\xi}_i \hat{\boldsymbol{w}}_{i,t}$. Specifically, when we use the LV adjustment to obtain two adjusted constituent portfolios, the corresponding combined portfolio is

$$\hat{\boldsymbol{w}}_{c,t} = \hat{p}_{1,t}\hat{\xi}_1\hat{\boldsymbol{w}}_{1,t} + \hat{p}_{2,t}\hat{\xi}_2\hat{\boldsymbol{w}}_{2,t}. \tag{4.11}$$

4.3.2 Expected Out-of-sample Utility Optimization

Given that there may be potential biases from the specific risk-reward preferences of investors with the combined portfolio as (4.11), we further consider a scaled combined portfolio based on the historical information up to time t as follows:

$$_s\hat{\boldsymbol{w}}_{c,t} = a_t\hat{\boldsymbol{w}}_{c,t}, \tag{4.12}$$

where a_t is an adaptive scaling parameter. To determine the parameter a_t, we consider the expected out-of-sample MV utility optimization, which is widely used in the literature when considering estimation risk (e.g., Kan and Zhou, 2007). A rational decision maker should prefer the strategy with a greater expected out-of-sample MV utility. Hence, the scaling parameter a_t is determined by maximizing the expected out-of-sample MV utility through the following optimization problem:

$$\max_{a_t} \mathbb{E}\left[U({}_s\hat{\boldsymbol{w}}_{c,t})\right] = \max_{a_t} \mathbb{E}\left[a_t\hat{\boldsymbol{w}}_{c,t}^\top\boldsymbol{\mu} - \frac{\gamma}{2}a_t^2\hat{\boldsymbol{w}}_{c,t}^\top\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{c,t}\right]. \tag{4.13}$$

The corresponding optimal scaling parameter can be easily derived as follows:

$$a_t^* = \frac{\mathbb{E}[\hat{\boldsymbol{w}}_{c,t}^\top\boldsymbol{\mu}]}{\gamma\mathbb{E}[\hat{\boldsymbol{w}}_{c,t}^\top\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{c,t}]}. \tag{4.14}$$

In order to estimate the optimal parameter a_t^*, we need to estimate the expected out-of-sample mean and variance of the combined portfolio $\hat{\boldsymbol{w}}_{c,t}$. However, in practice, the first two moments of the out-of-sample portfolio returns are unknown. We consider a rolling window setup to estimate the moments. We first obtain estimates of the mean vector and covariance matrix of asset returns as $\hat{\boldsymbol{\mu}}_t$ and $\widehat{\boldsymbol{\Sigma}}_t$ using historical asset returns with window length K_1. Then, we use historical combined portfolios from the previous K_2 periods to obtain the estimates of $\mathbb{E}[\hat{\boldsymbol{w}}_{c,t}^\top \boldsymbol{\mu}]$ and $\mathbb{E}[\hat{\boldsymbol{w}}_{c,t}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_{c,t}]$ respectively as follows:

$$\hat{\mu}_{\hat{\boldsymbol{w}}_{c,t}} = \frac{1}{K_2} \sum_{i=1}^{K_2} \hat{\boldsymbol{w}}_{c,t-i}^\top \hat{\boldsymbol{\mu}}_t \quad \text{and} \quad \hat{\sigma}^2_{\hat{\boldsymbol{w}}_{c,t}} = \frac{1}{K_2} \sum_{i=1}^{K_2} \hat{\boldsymbol{w}}_{c,t-i}^\top \hat{\boldsymbol{\Sigma}}_t \hat{\boldsymbol{w}}_{c,t-i}. \tag{4.15}$$

Therefore, the scaled combined portfolio after the expected out-of-sample utility optimization is

$$_s\hat{\boldsymbol{w}}_{c,t} = \hat{a}_t \hat{\boldsymbol{w}}_{c,t} = \hat{a}_t (\hat{p}_{1,t} \hat{\xi}_1 \hat{\boldsymbol{w}}_{1,t} + \hat{p}_{2,t} \hat{\xi}_2 \hat{\boldsymbol{w}}_{2,t}), \tag{4.16}$$

where the scaling parameter is estimated by $\hat{a}_t = \hat{\mu}_{\hat{\boldsymbol{w}}_{c,t}} / (\gamma \hat{\sigma}^2_{\hat{\boldsymbol{w}}_{c,t}})$. For the window lengths K_1 and K_2, we recommend using a longer window K_1 to estimate the overall mean vector and covariance matrix of asset returns. Conversely, a shorter window length K_2 is utilized to capture the most recent information up to time t.[1]

This proposal of adjusting the combined portfolio with the scalar parameter a_t as shown in (4.12) and resulting in the single-variable optimization problem (4.13) is similar in spirit to the idea proposed by Kan and Zhou (2007). While our approach involves shrinking the combined portfolio, their approach shrinks the tangent portfolio. Additionally, we avoid the stringent multivariate normal distribution assumption on asset returns required to obtain the analytical expected out-of-sample MV utility. Instead, we obtain the mean and variance of the out-of-sample returns of the combined portfolio $\hat{\boldsymbol{w}}_c$ based on the historical realized information.

4.3.3 General Training-Testing Procedure

We outline the general training-testing procedure, which is summarized in Algorithm 1, for deploying and evaluating the WPW combined portfolio. Given the historical dataset for asset returns, we divide the entire dataset into two subsets: the training set, with time labels $1, 2, \ldots, T_1$, and the testing set, with time labels $T_1+1, T_1+2, \ldots, T$. The training set is utilized to estimate the long-run variance adjusted parameter and create the predictive model for the winning probability. Particularly, the scaling parameter that optimizes the expected out-of-sample utility is adaptively updated along with the testing set. Finally, we

[1]In our later empirical study, K_1 encompasses the entire available period up to time t for estimating the overall mean vector and covariance matrix of the asset returns, whereas K_2 is set to 250. We also evaluated the results with $K_2 = 500$, which did not significantly alter the outcomes.

evaluate the out-of-sample performance of the resulting scaled WPW combined portfolios from the testing set.

Algorithm 1: General procedure for the winning probability weighted combined portfolio

---

Split the dataset into the training set and the testing set;
Create rolling windows for both the training set and the testing set with a fixed length M;
For *the training set*
- Step 1. Calculate long run variance adjustment parameters $\hat{\xi}_i$:
  - 1.1 generate a realized out-of-sample return series for each constituent portfolio;
  - 1.2 estimate the long-run variance adjustment parameter (4.10);
- Step 2. Obtain the desirable predictive model $f(\cdot)$ for estimating winning probabilities:
  - 2.1 use the out-of-sample return for adjusted portfolios to create the label (4.17);
  - 2.2 create a predictive model and train it to obtain the desirable predictive model;

For *the testing set*
- Step 3: Calculate the scaling parameter $\hat{a}_t$ for each window:
  - 3.1 estimate the mean and covariance matrix using the previous K_1 asset returns;
  - 3.2 estimate the expectations by the averages of the previous K_2 combined portfolios (4.15);
  - 3.3 obtain the scaled WPW combined portfolio (4.16).

Firstly, we use the whole training set to estimate the long-run variance adjustment parameter (4.10) by implementing a rolling window approach with a window length of M. Specifically, we utilize the data from the most recent M days up to time t to obtain the plug-in constituent portfolios $\hat{\boldsymbol{w}}_{i,t}, i = 1, 2$. Given the daily excess returns and riskfree returns, the corresponding realized out-of-sample portfolio returns are represented by the series $\{r(\hat{\boldsymbol{w}}_{i,t})\}_{t=M}^{T_1-1}, i = 1, 2$. We then calculate the estimated long-run variance of these time series using (4.9) and calculate the estimated long-run variance adjusted parameter $\hat{\xi}_i$ using (4.10).

Secondly, we train a predictive model for the winning probabilities using the returns of the adjusted constituent portfolios along with the training dataset. We create a training label for each rolling window using the adjusted constituent portfolios as follows:

$$Y_{i,t} = \mathbb{I}_{\{i = \text{argmax}_{j \in \{1,2\}} r(\hat{\xi}_{j,t} \hat{\boldsymbol{w}}_{j,t})\}}, \qquad t = M, ..., T_1 - 1; \;\; i = 1, 2. \tag{4.17}$$

It is also the selection indicator in (4.3) with the adjusted constituent portfolios $\tilde{\boldsymbol{w}}_{i,t} = \hat{\xi}_i \hat{\boldsymbol{w}}_{i,t}$. We then take feature variables $\boldsymbol{x}_t$ as the input and the label $Y_{i,t}, i = 1, 2$, as the output to establish a predictive model $f(\cdot)$. In this process, we use 80 percent of the training set as training samples to build candidate models, and the remaining 20 percent serves as validation samples for model selection. The training details can be different when we use different predictive models. More information regarding the predictive models and feature variables are presented in Section 4.4. The output of the resulting predictive model is the winning probability that we desire.

Over the testing set from T_1+1 to T, we fix the estimate of long-run adjusted parameter $\hat{\xi}$ and the predictive model $f(\cdot)$ obtained from the training set. The scaling parameter a_t that optimizes the expected out-of-sample utility is adaptively estimated along with the testing set as introduced in Section 4.3.2. With these procedures, we obtain the realized returns of the WPW combined portfolio (4.11) and the scaled WPW combined portfolio (4.16) over the trading periods $t = T_1 + M, ..., T - 1$. These realized returns would enable us to evaluate the out-of-sample performance of the combined portfolios.

4.4 Prediction of Winning Probabilities

In this section, we introduce two specific predictive models for the winning probability and then explore the feature variables used in these predictive models, which include technical and fundamental features. Additionally, we introduce diagnostics for detecting multicollinearity among these feature variables to facilitate the construction of feature variables.

4.4.1 Predictive Models for Winning Probabilities

Among all of the predictive models in the literature, we focus on two of the most widely explored predictive models in our empirical studies: the logistic regression and the random forest.

4.4.1.1 Logistic Regression Model

Logistic regression is a widely utilized statistical classification model renowned for its simplicity and interoperability. It is commonly employed for binary classification tasks where the goal is to predict a binary outcome based on a set of continuous or categorical input variables. The model uses the logistic (sigmoid) function to transform a linear combination of the input features into a real-valued output ranging between 0 and 1, where the output value represents the likelihood of the input data being classified into one of the two predefined categories. The class that exhibits a higher probability is then selected as the predicted category. Comprehensive discussions and application examples of logistic regression are well-documented in the literature, such as Kleinbaum et al. (2002).

Within our WPW framework, the logistic regression model can be applied to forecast the winning probabilities. The model inputs are taken as both fundamental and technical factors that could potentially influence out-of-sample portfolio performances, which are represented by an n-dimensional vector of feature variables $\boldsymbol{x} = [x_1, ..., x_n]$. For binary classification tasks, where the combined portfolio consists of exactly two constituent portfolios, logistic regression assesses the probability that one portfolio will outperform the other in terms of out-of-sample returns. That is,

$$p(\boldsymbol{x}) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x})}}, \tag{4.18}$$

where β_0 and $\boldsymbol{\beta} = [\beta_1, ..., \beta_n]$ are the logistic regression coefficients that need to be estimated.

To find the logistic regression coefficients involved in (4.18), the logistic regression model is generally trained using the maximum likelihood estimation method, where the log-likelihood is expressed as:

$$l(\boldsymbol{\beta}) = \sum_i y_i \log(p(\boldsymbol{x}_i)) + (1 - y_i) \log(1 - p(\boldsymbol{x}_i)).$$

The training process can be enhanced by integrating variable selection with parameter estimation. This integration is achieved by incorporating a penalty term into the log-likelihood function, which not only simplifies the model but also improves its predictive accuracy. In our empirical studies, we employ the elastic net regularization technique, which refines the log-likelihood function by adding a penalty term:

$$L = \lambda_1 ||\boldsymbol{\beta}||_1 + \lambda_2 ||\boldsymbol{\beta}||^2. \tag{4.19}$$

The elastic net combines the properties of both lasso regression (when $\lambda_1 = 1$ and $\lambda_2 = 0$)

and ridge regression (when $\lambda_1 = 0$ and $\lambda_2 = 1$), which encourages the shrinkage of the coefficients for less influential features towards zero. To determine the two regularization parameters λ_1 and λ_2, we set up a parameter grid and utilize a randomized search strategy, which systematically explores different parameter combinations to find the most effective set through a ten-fold cross-validation process.

4.4.1.2 Random Forests

The random forest algorithm was originally proposed by Ho (1995) and subsequently refined and popularized by Breiman (2001). As a robust ensemble learning method, it utilizes multiple regression or classification decision trees to enhance predictive accuracy. In the random forest, each tree is independently constructed using distinct bootstrap samples drawn with replacement from the entire training dataset. Another layer of randomness to mitigate the risk of overfitting when creating each tree is that: the algorithm selects a random subset of the features, rather than all available features, to determine the best split at each node. During the prediction phase, random forest aggregates the predictions from all the individual trees, typically by majority voting for classification tasks or averaging for regression. This aggregation helps to stabilize and improve predictions by mitigating the noise and variance inherent in individual trees. Therefore, the random forest generally produces more reliable outputs than a single decision tree, which makes it a preferred choice for addressing a more complex structure.

The construction of the random forest model involves many hyperparameters. To effectively implement the random forest algorithm, we carefully control four key hyperparameters that are particularly important: number of trees, maximum depth, minimum samples split, and maximum features. These parameters critically influence the size of the forest, the complexity of the decision trees, and the degree of randomness within trees. Generally, increasing the number of trees in the forest enhances model stability and accuracy but also raises computational costs. The maximum depth controls the complexity of the trees, with deeper trees being able to capture more complex data patterns but may lead to overfitting. The last two parameters, minimum samples and maximum features, specify the minimum number of samples and the maximum number of features considered to split a node in the tree. A smaller number of samples or a higher number of features will generally lead to a more accurate model, but it can also lead to overfitting. In our empirical studies, we employ a randomized search cross-validation to determine these hyperparameters optimally.

The random forest method has become an important benchmark model in machine learning for prediction; see Fischer and Krauss (2018) and Kraus et al. (2020), to name a few. Notably, in the context of portfolio construction, Krauss et al. (2017) concludes that the random forest outperforms other machine learning algorithms, including deep neural networks, gradient-boosted trees, and other ensemble methods, when used for prediction to build portfolios. Particularly, DeMiguel et al. (2023) also find that gradient boosting and random forest outperform neural networks for constructing portfolios of mutual funds.

4.4.2 Feature Exploration

Our framework distinguishes itself from existing combined portfolio strategies due to its adaptability in incorporating valuable financial market information to predict the combination coefficient. As a portfolio strategy comprises a collection of assets, any economic market factors that influence asset values are likely to impact the portfolio value as well. Extensive research has been conducted on predicting financial markets using either technical or fundamental analysis, as extensively reviewed by Nti et al. (2020). In this context, technical analysis involves studying past asset trends, while fundamental analysis utilizes cross-sectional financial market data. Previous research, such as the work by Neely et al. (2014), has demonstrated that combining information from both technical and fundamental variables significantly enhances forecasts of equity risk premium. Building on these findings, we explore the use of technical and fundamental features as input variables in our predictive models for the combination coefficient. By integrating such relevant financial market information, our framework aims to improve the accuracy of predicting the winning probability and, consequently, enhance the performance of the resulting WPW combined portfolio.

4.4.2.1 Technical Features

The winning probability defined in (4.4) is inherently tied to the out-of-sample returns of the constituent portfolios. As a result, we incorporate the historical realized out-of-sample returns of these portfolios as technical features. To determine which historical out-of-sample portfolio returns should be included in the predictive model for the combination coefficient, we adopt a variable importance approach, which measures the relative predictive capacity or explanatory power of each feature. Our empirical findings consistently suggest that, regardless of the dataset or complex portfolio chosen, the most recent five trading days' realized out-of-sample returns exhibit superior predictive capabilities and yield higher explanatory power. This observation aligns with the conclusions drawn by Krauss et al. (2017), although their focus was on forecasting individual assets to outperform the overall market.

Additionally, we explore the potential efficacy of including the historical returns of individual assets in the predictive model for the combination coefficient. However, this investigation involves a large number of feature variables due to the many individual assets in the constituent portfolios. Dealing with a high-dimensional regression and variable selection problem, we employ feature selection methods like the sure independence screening method proposed by (Fan and Lv, 2008). Empirical results reveal that incorporating the historical excess returns of individual assets does not lead to an improvement in the performance of the resulting combined portfolio.

Therefore, we choose to use the historical realized out-of-sample returns of the two constituent portfolios from the preceding five trading days as the technical features in our predictive model. Specifically, for any trading period t, the technical features included in the predictive model are represented by:

$$\boldsymbol{x}_t^1 = [r(\tilde{\boldsymbol{w}}_{1,t-5}), ..., r(\tilde{\boldsymbol{w}}_{1,t-1}), r(\tilde{\boldsymbol{w}}_{2,t-5}), ..., r(\tilde{\boldsymbol{w}}_{2,t-1})], \tag{4.20}$$

where $r(\tilde{\boldsymbol{w}}_{i,t})$ denote the out-of-sample return of the long-run variance adjusted constituent portfolio i over the trading period t, where $\tilde{\boldsymbol{w}}_{i,t} = \hat{\xi}_i \hat{\boldsymbol{w}}_{i,t}, i = 1, 2$.

4.4.2.2 Fundamental Features

The literature presents a plethora of fundamental features for asset pricing, as demonstrated in Feng et al. (2020) for example. Given the wide array of available fundamental features, investing in all of them would be overly ambitious. Therefore, we narrow our focus to the market factors listed in Table 4.1.

The Fama-French five factors are widely acknowledged in the literature for explaining asset returns. They include the market factor, the return differential between small and

Table 4.1: List of fundamental factors.

No.	Abbreviation	Market factor
1	ERM	Fama-French five factors: excess return on the market (Fama and French, 2015)
2	SMB	Fama-French five factors: small minus big
3	HML	Fama-French five factors: high minus low
4	RMW	Fama-French five factors: robust minus weak
5	CMA	Fama-French five factors: conservative minus aggressive
6	LTR	Long term reversal
7	STR	Short term reversal (Gatev et al., 2006)
8	Mom	Momentum (Carhart, 1997)
9	1nidx	1/n favorable index (Guo et al., 2019)

The first eight market factors are obtained from Kenneth French data library, while the 1/N favorable index is calculated based on the definition in Guo et al. (2019).

large-cap stocks (small minus big), the return differential between high and low book-to-market stocks (high minus low), the disparity between robust and weak operating profitability companies (robust minus weak), and the contrast between conservatively and aggressively invested companies (conservative minus aggressive). Additionally, we include the long-term reversal, short-term reversal, and momentum factors as proposed by Fischer and Krauss (2018). Lastly, the 1/N favorable index is also included as it was introduced in Guo et al. (2019) and has shown the ability to outperform specific sophisticated portfolios when applied to the 1/N portfolio.

4.4.2.3 Multicollinearity Diagnostics

Given the multitude of technical and fundamental features discussed above, we utilize multicollinearity diagnostics to select the appropriate set of feature variables for the predictive model. We initially normalize all feature variables to mitigate the impact of varying scales. This normalization is achieved by removing the mean and scaling to unit variance. Multicollinearity refers to the presence of strong linear relationships among the feature variables. As a linear relation can encompass numerous feature variables, simple correlation coefficients are inadequate for identifying such a relationship. Chatterjee and Hadi (2006) delve into the effects of multicollinearity on statistical inference and review several criteria for detecting multicollinearity. One comprehensive diagnostic of multicollinearity is to use the variance inflation factor (VIF):

$$\text{VIF}_{x_i} = \frac{1}{1 - R_i^2}, i = 1, ..., I,$$

where I represents the total number of feature variables, and R_i^2 is the square of the multiple correlation coefficient when the feature variable x_i is regressed against all other feature variables. The VIF value ranges from 1 to infinity. A large VIF_{x_i} indicates a strong linear relationship between x_i and other feature variables, as R_i^2 approaches 1. A VIF value exceeding ten is generally considered a sign of severe collinearity.

To mitigate the adverse effects of multicollinearity, we compute the VIFs for both the technical features $\boldsymbol{x}_t^1$ defined in (4.20) and the fundamental features listed in Table 4.1 for each dataset in our empirical study. The results reveal severe multicollinearity between the ERM factor and the most recent return of the 1/N rule when it is taken as one of the constituent portfolios. This observation aligns well with the findings of Guo et al. (2019) that the performance of the 1/N portfolio is tied to market conditions. To alleviate the collinearity effect, we exclude the ERM factor and recalculate the VIFs. The VIFs of the remaining variables are all under ten, indicating no severe collinearity among the remaining feature variables. Therefore, in the later empirical study, we exclude the ERM factor from the list of fundamental features in Table 4.1 and include the following vector of fundamental feature variables in the predictive model for the winning probability:

$$\boldsymbol{x}_t^2 = [\text{SMB, HML, RMW, CMA, LTR, STR, Mon, 1nidx}]. \tag{4.21}$$

4.5 Empirical Study

We demonstrate the performance of our WPW combined portfolios in a comprehensive empirical analysis. In the following analysis, we first introduce the setup, including the datasets and portfolio strategies. We then investigate the parameter values involved in the WPW framework, including the long-run variance adjustment coefficients, the combination coefficients, and the scaling parameter. Finally, we present the out-of-sample performance of various portfolio strategies on these real datasets. We also provide an alpha-beta analysis for our WPW combined portfolios and relevant portfolios from the analytical method.

4.5.1 Datasets and Portfolio Strategies

We use eight datasets sourced from the Kenneth R. French data library as summarized in Table 4.2 with an acronym, a succinct explanation, and the number of assets for each dataset. Each dataset comprises a varying number of portfolios made up of stocks from NYSE, AMEX, and NASDAQ. We utilize the average value-weighted returns in daily frequency for each dataset.[2] Daily excess returns are calculated by subtracting the simple daily returns of a one-month treasury bill from the corresponding daily returns. Our study uses historical data ranging from 1970-01-01 to 2023-07-01. The start date (1970-01-01) is chosen as it is approximately the earliest available date across all datasets without missing values.

Table 4.2: List of datasets.

Abbreviation	Dataset	N
SLT6	6 Portfolios Formed on Size and Long-Term Reversal	6
LT10	10 Portfolios Formed on Long-Term Reversal	10
A17	17 Industry Portfolios	17
BMOP25	25 Portfolios Formed on Book-to-Market and Operating Profitability	25
SLT25	25 Portfolios Formed on Size and Long-Term Reversal	25
I30	30 Industry Portfolios	30
I49	30 Industry Portfolios	49
SOP100	100 Portfolios Formed on Size and Operating Profitability	100

While numerous portfolio strategies in the existing literature can serve as constituent portfolios, our subsequent numerical studies specifically concentrate on five well-known ones. Panel A of Table 4.3 outlines these constituent portfolios with a notation for portfolio weights and a description in two separate columns of the table. These constituent portfolios include the equally weighted (1/N) portfolio $\boldsymbol{w}_e$, the sample global minimum variance (GMV) portfolio $\hat{\boldsymbol{w}}_g$, the sample MV portfolio $\hat{\boldsymbol{w}}_{mv}$, the sample constrained MV portfolio $\hat{\boldsymbol{w}}_{mr}$ and the three-fund portfolio $\hat{\boldsymbol{w}}_{kz}$. Here, the MV portfolio refers to the optimal portfolio that maximizes the MV objective function including a risk-free asset, while the constrained MV portfolio represents the optimal one derived without a risk-free asset. Further details

[2]Here, the average value-weighted returns refer to the portfolio returns, where the portfolios are constructed by the value-weighted method according to features or industries of the underlying stocks. Further details can be found on the Kenneth R. French data library website.

Table 4.3: List of portfolio strategies.

Notation	Description
Panel A: constituent portfolios	
A.1: naive constituent portfolios	
$\boldsymbol{w}_e$	the 1/N portfolio
$\hat{\boldsymbol{w}}_g$	the sample global minimum variance portfolio
A.2: sophisticated constituent portfolios	
$\hat{\boldsymbol{w}}_{mv}$	the sample MV portfolio
$\hat{\boldsymbol{w}}_{mr}$	the sample constrained MV portfolio
$\hat{\boldsymbol{w}}_{kz}$	the three-fund portfolio (Kan and Zhou, 2007)
Panel B: combined portfolios	
B.1: Taking the 1/N portfolio as one of the components	
$\hat{\boldsymbol{w}}^{\#}_{e\text{-}mv}$	combined with the sample MV portfolio
$\hat{\boldsymbol{w}}^{\#}_{e\text{-}kz}$	combined with the three-fund portfolio
B.2: Taking the sample GMV portfolio as one of the components	
$\hat{\boldsymbol{w}}^{\#}_{g\text{-}mv}$	combined with the sample MV portfolio
$\hat{\boldsymbol{w}}^{\#}_{g\text{-}mr}$	combined with the sample constrained MV portfolio

This table lists various portfolio strategies considered in our empirical studies. Panel A outlines the constituent portfolios, and Panel B shows various combined portfolios using constituent portfolios from Panel A. The right superscript "#" of combined portfolios indicates the approach used in determining the combination coefficient, where we take "#" for "a", "u", "LG" and "RF" respectively in the analysis.

about these portfolios are provided in Appendix C.1. Without loss of generality, we refer to the 1/N rule and the sample GMV portfolio as "naive portfolios" and designate the remaining constituent portfolios as "sophisticated portfolios", which aligns with the spirit of portfolio combinations discussed in Tu and Zhou (2011).

To study the performance of the combined portfolios, we analyze different combined portfolios with two constituent portfolios from Panel A, as outlined in Panel B of Table 4.3. The combined portfolios are categorized into two groups: one includes the 1/N rule as one of the components, and the other incorporates the sample GMV portfolio. For each category, we investigate two combined portfolios. These specific combined portfolios are chosen because they have analytical combination coefficients well-documented in the existing literature, allowing us to analyze the performance differences between our WPW framework and the analytical approach (see Tu and Zhou, 2011; Kan and Zhou, 2007; Lassance et al., 2023). We denote the resulting combined portfolios as shown in the first column of Panel B in Table 4.3, where the subscript in each notation is formed by the corresponding subscript from the notations of the constituent portfolios linked by a hyphen. Notably, the three-fund portfolio itself is considered a combination of the sample GMV and MV portfolios (Kan and Zhou, 2007). Consequently, when the sample GMV portfolio is one of the components, we do not combine it with the three-fund portfolio.

The right superscript "#" in the notations for the combined portfolios in Table 4.3 takes values "a", "u", "LG", and "RF". The superscripts "a" and "u" signify the use of analytical methods for the determination of the combination coefficient under the normality assumption of asset returns from the existing literature, where a superscript "a" means that the combination coefficients assigned to the two constituent portfolios sum to one (Tu and Zhou, 2011; Kan and Zhou, 2007) while "u" indicates that the combination coefficients are not restricted as such (Lassance et al., 2023). Further details about combined portfolios with the two superscripts are reported in Appendix C.2. The superscripts "LG" and "RF" indicate that logistic regression and random forest are, respectively, employed to determine the combination coefficients (i.e., winning probabilities) within our WPW framework. Additionally, since the expected out-of-sample utility optimization step introduced in Section 4.3.2 is not limited to our WPW framework, it can be applied to the constituent portfolios. When incorporating the expected out-of-sample utility optimization step with a scaling parameter for any portfolio strategy, we prepend a left subscript "s" to the corresponding portfolio notation. Specifically, for any portfolio i for period t, $\hat{\boldsymbol{w}}_{i,t}$, the scaled portfolio with the corresponding scaling parameter $\hat{a}_{i,t}$ is represented by:

$$ {}_s\hat{\boldsymbol{w}}_{i,t} = \hat{a}_{i,t}\hat{\boldsymbol{w}}_{i,t}, \tag{4.22} $$

where $\hat{a}_{i,t} = \hat{\mu}_{\hat{\boldsymbol{w}}_{i,t}}/(\gamma\hat{\sigma}^2_{\hat{\boldsymbol{w}}_{i,t}})$ is the estimated scaling parameter as introduced in Section 4.3.2.

Following the training-testing procedure outlined in Section 4.3.3, we divided the dataset into a training set (1970-01-01 to 2013-07-01) and a testing set (the subsequent ten years). We employ a rolling window approach with a window length of $M = 120$ for parameter estimation in constituent portfolios. The risk aversion coefficient is set to be $\gamma = 3$. The training set is employed to estimate the LV adjustment parameters and train the classification models for winning probabilities. To evaluate and compare the out-of-sample performance across different portfolio strategies, the same testing set and rolling window approach are consistently applied throughout our empirical study.

4.5.2 Parameter Values in the WPW Framework

The WPW framework for combined portfolios involves three critical parameters: LV adjustment parameters, winning probabilities, and the scaling parameter. We analyze each of these parameters one by one, and compare them against those derived from analytical portfolio strategies if applicable.

Firstly, regarding the LV adjustment parameters discussed in Section 4.3.1, we start with the assumption that the out-of-sample returns of each constituent portfolio satisfy a weakly stationary time series. To validate this assumption, we conduct both the Augmented Dickey-Fuller (ADF) unit root test (Fuller, 1995) and the KPSS stationarity test (Kwiatkowski et al., 1992) on the out-of-sample return series of each constituent portfolio. The test results provide strong evidence of the stationarity of these series. We then calculate the LV for each constituent portfolio using the Bartlett (Newey-West) estimator, as specified in (4.9). Notably, the 1/N portfolio and the sample GMV portfolio, exhibit smaller LVs compared to the other constituent portfolios.

The LV adjustment parameters are designed to equalize the long-term risk profiles across each pair of constituent portfolios within combined portfolios. For each combined portfolio listed in Panel B of Table 4.3, the naive constituent portfolio - either the 1/N rule or the sample GMV portfolio - exhibits a smaller long-run variance compared to the sophisticated one and is thus designated as the benchmark. Accordingly, the LV adjustment parameter for each sophisticated constituent portfolio is calibrated to align its long-run variance with that of the naive constituent portfolio in each combined portfolio. Table 4.4 reports the LV adjustment parameters for the sophisticated constituent portfolios within the four combined portfolios. For example, $\hat{\xi}_{e\text{-}mv}$ represents the estimated LV adjustment parameter for the sophisticated component $\hat{\boldsymbol{w}}_{mv}$ in the combined portfolio $\hat{\boldsymbol{w}}_{e\text{-}mv}$, where $\hat{\boldsymbol{w}}_{e}$ is taken as the benchmark. Once adjusted, these sophisticated portfolios align with the same risk profile as their naive benchmarks. As the parameter values are all less than one, it indicates that the sophisticated portfolios before the adjustment inherently carry greater risk than the naive ones. Consequently, it is crucial to adjust the risk profiles of constituent portfolios to ensure that the comparisons of out-of-sample returns across constituent portfolios are on a comparable risk level. It avoids the potential misperception that portfolios yield higher returns due to higher risk exposure.

Table 4.4: The estimated LV adjustment parameters.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
$\hat{\xi}_{e\text{-}mv}$	0.0578	0.0727	0.0388	0.0462	0.0359	0.0328	0.0255	0.0124
$\hat{\xi}_{e\text{-}kz}$	0.0775	0.1399	0.0739	0.1275	0.0736	0.0859	0.1037	0.1774
$\hat{\xi}_{g\text{-}mv}$	0.0420	0.0629	0.0298	0.0374	0.0234	0.0240	0.0189	0.0118
$\hat{\xi}_{g\text{-}mr}$	0.0495	0.0620	0.0283	0.0314	0.0211	0.0191	0.0116	0.0020

Secondly, as previously discussed, we utilize logistic regression and random forest as predictive models to determine the combination coefficients (via the winning probabilities) for the WPW combined portfolios. Since including both technical features $\boldsymbol{x}_t^1$ (4.20) and fundamental features $\boldsymbol{x}_t^2$ (4.21) enhances the predictions, we incorporate both features as inputs for each predictive model. Table 4.5 displays the combination coefficients for the naive constituent portfolios within the four combined portfolios, while the coefficients for the corresponding sophisticated constituent portfolios are summarized in Appendix C.3.1. For comparative purposes, we also present combination coefficients derived from analytical methods if applicable. Since these combination coefficients vary over time, the table presents average values along with the corresponding standard deviations in brackets. The results indicate that our weighting for the naive constituent portfolio averages around 0.5, though it is subject to temporal adjustments. Notably, the weight assigned to the naive constituent portfolio through our WPW framework generally differs from those obtained through analytical methods, and the standard deviation of our method is consistently lower than that of the analytical methods across most scenarios.

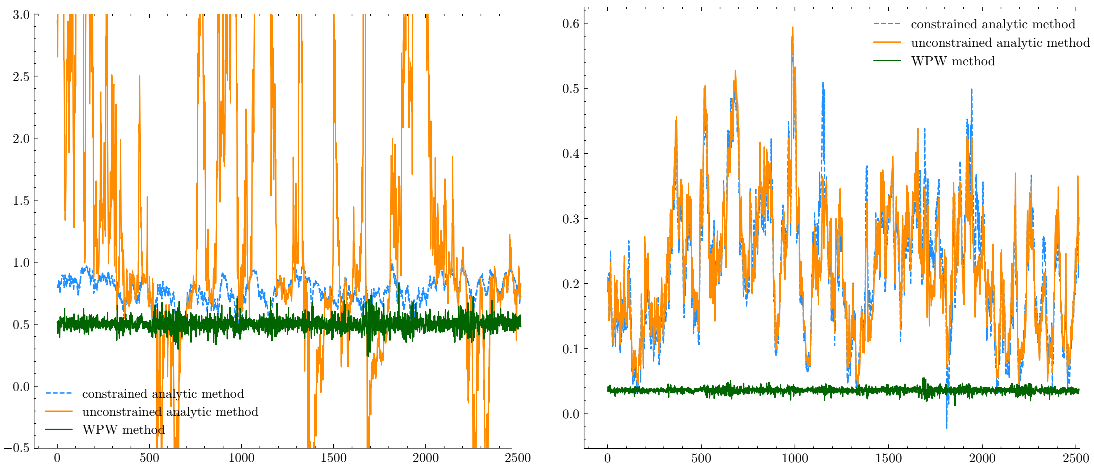
Table 4.5: The average combination coefficients for naive constituent portfolios.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}_e^{LG}$	0.4859	0.5033	0.4955	0.4932	0.4892	0.4979	0.4919	0.5101
	(0.0628)	(0.0521)	(0.0739)	(0.0471)	(0.0763)	(0.0813)	(0.0464)	(0.0509)
$\hat{\boldsymbol{w}}_e^{RF}$	0.5016	0.5048	0.5048	0.4964	0.5006	0.5126	0.5076	0.5116
	(0.0645)	(0.0514)	(0.0632)	(0.0622)	(0.067)	(0.0732)	(0.0742)	(0.0557)
Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_e^{a}$	0.6907	0.7725	0.8705	0.8675	0.8384	0.9044	0.9386	0.9778
	(0.1672)	(0.1038)	(0.0663)	(0.0721)	(0.076)	(0.0559)	(0.0367)	(0.0156)
$\hat{\boldsymbol{w}}_e^{u}$	1.2221	1.5671	1.6752	1.5276	1.3375	1.7688	1.9286	1.5131

	(1.4251)	(1.8392)	(2.026)	(1.6715)	(1.527)	(2.3385)	(2.4578)	(1.7443)
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}_e^{LG}$	0.4926	0.5086	0.5085	0.4932	0.498	0.5071	0.5063	0.5171
	(0.058)	(0.0332)	(0.0327)	(0.0497)	(0.055)	(0.014)	(0.0346)	(0.0617)
$\hat{\boldsymbol{w}}_e^{RF}$	0.5139	0.5137	0.5317	0.502	0.5154	0.526	0.5195	0.5182
	(0.0627)	(0.0452)	(0.0502)	(0.0537)	(0.0551)	(0.0522)	(0.0569)	(0.0724)
Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_e^{a}$	0.2973	0.3177	0.3941	0.3358	0.2939	0.4063	0.4398	0.4428
	(0.2766)	(0.217)	(0.3074)	(0.255)	(0.17)	(0.3162)	(0.3455)	(0.4442)
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}_g^{LG}$	0.5006	0.4971	0.4986	0.5045	0.5076	0.4928	0.5023	0.5141
	(0.0844)	(0.046)	(0.0637)	(0.0588)	(0.0717)	(0.0645)	(0.0484)	(0.0453)
$\hat{\boldsymbol{w}}_g^{RF}$	0.5175	0.5024	0.5114	0.5088	0.512	0.5022	0.5109	0.5163
	(0.0617)	(0.0447)	(0.0586)	(0.0664)	(0.0622)	(0.0709)	(0.0711)	(0.0361)
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_g^{a}$	2.5541	2.7445	2.8369	3.7849	4.0498	3.2352	3.3581	1.1247
	(3.411)	(3.1071)	(3.3606)	(4.3837)	(4.1289)	(3.9534)	(4.251)	(2.1993)
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}_g^{LG}$	0.5092	0.5024	0.5049	0.5099	0.5182	0.4964	0.507	0.5182
	(0.0777)	(0.0496)	(0.0774)	(0.0681)	(0.0717)	(0.0711)	(0.0694)	(0.0484)
$\hat{\boldsymbol{w}}_g^{RF}$	0.5201	0.5067	0.5137	0.5133	0.5136	0.5032	0.5094	0.5175
	(0.0608)	(0.0467)	(0.0707)	(0.0583)	(0.0633)	(0.0644)	(0.0718)	(0.0501)
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_e^{a}$	0.6872	0.7834	0.8825	0.8966	0.8757	0.9278	0.9638	0.9964
	(0.1631)	(0.0921)	(0.0586)	(0.0561)	(0.0586)	(0.038)	(0.021)	(0.0025)

To illustrate the temporal variations in combination coefficients, Figure 4.1 shows the combination coefficients for the two constituent portfolios within the combined portfolio $\hat{\boldsymbol{w}}_{e\text{-}mv}$ across the testing period of the dataset LT10. We compare three different approaches: the WPW framework utilizing logistic regression (i.e.,$\hat{\boldsymbol{w}}_{e\text{-}mv}^{LG}$), the analytical method with constraints (i.e., $\hat{\boldsymbol{w}}_{e\text{-}mv}^{a}$), and the analytical method without constraints (i.e., $\hat{\boldsymbol{w}}_{e\text{-}mv}^{u}$). Figure 4.1a details the combination coefficients for the 1/N rule. Our WPW method maintains a relatively stable combination coefficient (winning probability) around 0.5, in contrast to the more variable coefficients observed with the analytical methods. Figure 4.1b presents the combination coefficients for the sample MV portfolio. In our WPW framework, the allocation to the sample MV portfolio is small due to the application of the variance-adjusted

Figure 4.1: Combination coefficients for $\hat{\boldsymbol{w}}_{e\text{-}mv}$ from different methods.



(a) the 1/N portfolio **(b)** the sample mean variance portfolio

This set of figures compares the combination coefficients for the two constituent portfolios in $\hat{\boldsymbol{w}}_{e\text{-}mv}$ from: 1) the constrained analytical method ($\hat{\boldsymbol{w}}^{a}_{e\text{-}mv}$); 2) the unconstrained analytical method ($\hat{\boldsymbol{w}}^{u}_{e\text{-}mv}$); 3) our WPW framework ($\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$).

parameter $\xi_{e\text{-}mv}$. This adjustment strategically modulates exposure to the sophisticated portfolio, taking into account the long-term risk profile. Conversely, the analytical methods generally allocate higher exposure and exhibit greater volatility for both the 1/N rule and the sample MV portfolio.

Finally, we explore the optimization of expected out-of-sample utility by examining scaling parameters for various portfolio strategies, as detailed in Equation (4.22). This analysis includes combined portfolios from the WPW framework alongside naive constituent portfolios (i.e., the 1/N rule, or the sample GMV portfolio). Table 4.6 displays the average scaling parameters, with standard deviations noted in parentheses. The results indicate that all scaling parameters exceed one. For naive constituent portfolios, which inherently have lower variance, the scaling parameters are inversely proportional to variance and tend to be larger than one. Meanwhile, for the WPW combined portfolios, we have implemented the LV adjustment to align the risk levels of sophisticated constituent portfolios with those of naive constituent portfolios. These adjustments typically result in lower variance for the combined portfolios, consequently leading to scaling parameters being larger than one.

Table 4.6: The average scaling parameter a from various datasets.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	3.0680	2.5039	2.6007	2.9314	3.4668	2.5446	2.2429	2.7010
	(0.0780)	(0.0810)	(0.2883)	(0.0348)	(0.0893)	(0.2101)	(0.1451)	(0.0274)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	3.0331	2.4869	2.5766	2.8902	3.1945	2.5178	2.0444	2.6579
	(0.0799)	(0.0828)	(0.2872)	(0.0321)	(0.1103)	(0.2147)	(0.1531)	(0.0237)
Panel B: Combining the 1/N rule with the three-fund portfolio								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	2.9695	2.5526	2.9268	2.4863	3.3985	2.8712	2.3777	2.6432
	(0.0917)	(0.1157)	(0.3765)	(0.0182)	(0.1447)	(0.3071)	(0.2437)	(0.0378)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	2.9386	2.5362	2.8491	2.4668	3.366	2.8003	2.3213	2.3587
	(0.0911)	(0.1156)	(0.3673)	(0.0188)	(0.147)	(0.306)	(0.2393)	(0.0492)
Panel C: Combining the sample GMV rule with the sample MV portfolio								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	3.9096	2.7407	3.3683	3.1323	4.9666	3.1156	2.6069	2.3315
	(0.0486)	(0.0636)	(0.2973)	(0.0701)	(0.0583)	(0.17)	(0.0523)	(0.0988)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	3.8945	2.7312	3.3192	2.8924	4.947	2.5588	2.3715	2.3190
	(0.0471)	(0.066)	(0.2971)	(0.0438)	(0.0524)	(0.1935)	(0.0698)	(0.0964)
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	4.5262	2.8897	3.3477	3.438	5.4066	3.1627	2.6642	2.3837
	(0.0751)	(0.0888)	(0.3191)	(0.0789)	(0.0895)	(0.1863)	(0.0517)	(0.1128)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	4.4794	2.8664	2.8880	3.4074	5.3809	3.1083	2.6194	2.1444
	(0.073)	(0.0931)	(0.2992)	(0.0723)	(0.0781)	(0.1898)	(0.0535)	(0.0741)
Panel E: Constituent portfolios								
${}_s\hat{\boldsymbol{w}}_{e}$	3.1314	2.7763	2.3982	2.5377	3.3734	2.5197	2.4455	2.8224
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
${}_s\hat{\boldsymbol{w}}_{g}$	3.2948	2.5324	2.7194	2.3327	4.0501	2.5036	2.0252	2.0114
	(0.1147)	(0.0167)	(0.0253)	(0.0463)	(0.1412)	(0.0144)	(0.0533)	(0.1063)

4.5.3 Portfolio Performance Assessment

We evaluate the performance of the various portfolios previously introduced by first analyzing the out-of-sample portfolio returns. Subsequently, we examine the certainty equivalent returns and then consider the effect of transaction costs.

4.5.3.1 Out-of-sample Portfolio Returns

As previously mentioned, the out-of-sample portfolio return for a given strategy $\hat{\boldsymbol{w}}_{i,t}, i = 1, 2$, formed at the end of period t and implemented in the subsequent time period $t+1$ is represented by $r(\hat{\boldsymbol{w}}_{i,t})$. From the testing dataset, we obtain a time series of these out-of-sample returns, denoted as $\{r(\hat{\boldsymbol{w}}_{i,t})\}_{t=1}^{T}$, where T represents the length of the testing dataset. The mean and variance of the out-of-sample portfolio returns are calculated as follows:

$$\hat{\mu}_{\hat{\boldsymbol{w}}_i} = \frac{1}{T}\sum_{t=1}^{T} r(\hat{\boldsymbol{w}}_{i,t}),$$

$$\hat{\sigma}^2_{\hat{\boldsymbol{w}}_i} = \frac{1}{T}\sum_{t=1}^{T} [r(\hat{\boldsymbol{w}}_{i,t}) - \hat{\mu}_{\hat{\boldsymbol{w}}_i}]^2.$$

Table 4.7 reports the average out-of-sample portfolio returns for various datasets and portfolios, with the corresponding standard deviations included in parentheses. This table highlights several key findings. Firstly, as seen in Panel E, the 1/N portfolio and the sample GMV portfolio both exhibit increased returns and variance when scaling parameters are included. It is attributable to the scaling parameters being greater than one. Secondly, a comparison between the returns of combined portfolios from the WPW framework and those derived from the analytical method shows that our WPW combined portfolios consistently demonstrate lower volatility. This suggests that the WPW combined portfolios experience smaller fluctuations compared to those managed by the analytical method. Interestingly, however, the WPW framework yields higher average returns for datasets including LT10, A17, and SOP100.

Table 4.7: Portfolio return on various datasets (in percentage).

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	0.1508	0.0979	0.0728	0.1599	0.1747	0.0999	0.0885	0.0829
	(2.155)	(1.8577)	(1.7586)	(2.2917)	(2.5814)	(1.7267)	(1.5303)	(2.243)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	0.1505	0.0921	0.0668	0.1543	0.1551	0.0959	0.0802	0.0824
	(2.1534)	(1.8401)	(1.7857)	(2.3017)	(2.5863)	(1.7147)	(1.4126)	(2.2250)
Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}mv}$	0.3089	0.0743	-0.0306	0.1972	0.1814	0.0878	0.1118	0.054
	(5.0112)	(3.8355)	(3.1114)	(4.0291)	(4.5624)	(3.2639)	(3.164)	(3.7044)
$\hat{\boldsymbol{w}}^{u}_{e\text{-}mv}$	0.3154	0.0737	-0.0205	0.2006	0.1898	0.1039	0.1319	0.065

	(5.4471)	(4.4758)	(3.9257)	(4.6124)	(5.1578)	(4.236)	(4.1964)	(4.2972)
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	0.1152	0.0802	0.0687	0.1199	0.1343	0.0782	0.0775	0.0692
	(2.0702)	(1.8608)	(1.9626)	(1.8968)	(2.44)	(1.87)	(1.583)	(2.1201)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	0.1172	0.0826	0.0734	0.1133	0.133	0.0791	0.0799	0.0738
	(2.0688)	(1.865)	(1.9719)	(1.8622)	(2.4415)	(1.7744)	(1.5684)	(2.1024)
Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}kz}$	0.2555	0.0743	-0.0585	0.2242	0.2314	0.0591	0.1127	0.0242
	(5.1364)	(4.0051)	(3.2044)	(4.2801)	(4.9535)	(3.4935)	(3.6108)	(3.8774)
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	0.1362	0.0801	0.1013	0.1922	0.2518	0.0863	0.0837	0.0516
	(2.263)	(1.8211)	(1.7915)	(2.0969)	(2.8537)	(1.6443)	(1.4994)	(2.0829)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	0.1323	0.0788	0.0952	0.1862	0.2293	0.0838	0.0855	0.0452
	(2.2958)	(1.8109)	(1.7931)	(2.0606)	(2.8157)	(1.6446)	(1.4856)	(2.0606)
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mv}$	0.3047	0.1106	-0.0774	0.3653	0.3321	0.0962	0.1747	0.0258
	(5.8577)	(5.0126)	(4.3663)	(5.6128)	(6.1861)	(4.7455)	(4.8275)	(4.8979)
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	0.1688	0.0807	0.1059	0.2027	0.2583	0.0911	0.0879	0.0534
	(2.6534)	(1.9079)	(1.7967)	(2.2611)	(3.082)	(1.6632)	(1.5013)	(2.0893)
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	0.1268	0.083	0.1054	0.198	0.2398	0.0835	0.0856	0.0452
	(2.2382)	(1.8918)	(1.7977)	(2.2189)	(3.0363)	(1.6616)	(1.4857)	(2.0601)
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mr}$	0.3031	0.0449	-0.0036	0.1899	0.1488	0.0697	0.0864	0.0396
	(4.8318)	(3.5735)	(3.1919)	(3.8902)	(4.3454)	(3.2338)	(3.1357)	(3.6796)
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_{e}$	0.0506	0.0508	0.0521	0.0475	0.0509	0.0489	0.0499	0.0485
	(1.2455)	(1.1386)	(1.1822)	(1.2544)	(1.3)	(1.1662)	(1.1585)	(1.2937)
$\hat{\boldsymbol{w}}_{g}$	0.0399	0.0395	0.0587	0.0722	0.0639	0.037	0.0444	0.0289
	(1.0166)	(1.0167)	(0.8654)	(0.9666)	(0.9709)	(0.8558)	(0.9277)	(1.3479)
$\hat{\boldsymbol{w}}_{mv}$	0.6948	0.3181	-0.1182	1.1939	1.0051	0.5538	0.9998	0.6276
	(10.4623)	(11.9538)	(15.4332)	(20.8936)	(20.3676)	(22.1976)	(31.4195)	(86.0407)
$\hat{\boldsymbol{w}}_{mr}$	0.707	0.2622	-0.055	1.1934	0.9453	0.6733	1.3811	3.5877
	(10.1626)	(12.1814)	(17.5215)	(25.3762)	(24.72)	(29.0558)	(52.387)	(521.3039)
$\hat{\boldsymbol{w}}_{kz}$	0.3047	0.1106	-0.0774	0.3653	0.3321	0.0962	0.1747	0.0258
	(5.8577)	(5.0126)	(4.3663)	(5.6128)	(6.1861)	(4.7455)	(4.8275)	(4.8979)
Pabel E.2: Scaled naive constituent portfolios								
${}_s\hat{\boldsymbol{w}}_{e}$	0.1509	0.1347	0.1201	0.115	0.1632	0.1178	0.1169	0.1303
	(3.9003)	(3.1614)	(2.8353)	(3.1834)	(4.3857)	(2.9385)	(2.8332)	(3.6515)
${}_s\hat{\boldsymbol{w}}_{g}$	0.1249	0.0943	0.1535	0.1638	0.2494	0.0872	0.087	0.0562
	(3.3334)	(2.5797)	(2.3625)	(2.2561)	(3.9065)	(2.1436)	(1.8712)	(2.6732)

To examine how portfolio returns are exposed to systematic risk sources, we assess the exposure of various combined portfolios to common systematic risk factors using the Fama-French three-factor model (FF3). This model includes a market factor (Market: excess return of the market portfolio), a size factor (SMB: small minus big capitalization stocks), and a value factor (HML: high minus low book-to-market stocks). The objective is to determine how returns from different combined portfolios respond to these systematic factors. The analysis is conducted using the following regression model for the portfolio's excess return r_t:

$$r_t = \alpha + \beta_1 \cdot \text{Market}_t + \beta_2 \cdot \text{SMB}_t + \beta_3 \text{HML}_t.$$

We study the out-of-sample returns of combined portfolios $\hat{\boldsymbol{w}}_{e\text{-}mv}$ from different methods involved in Panel A of Table 4.7. The results of the dataset LT10 are reported in Table 4.8. Panel A shows the regression coefficients (α, β_1, β_2, and β_3), the coefficient of determination (R^2), the number of observations used in the regression, and the root mean squared error (RMSE). Panel B summarizes the combination coefficients and portfolio performance metrics.

Table 4.8: Exposure to systematic sources of risk of different methods.

	$\hat{\boldsymbol{w}}^a_{e\text{-}mv}$	$\hat{\boldsymbol{w}}^u_{e\text{-}mv}$	${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$
Panel A: Regression Results				
(Intercept)	0.0003	0.0003	0.0004	0.0003
	(0.001)	(0.001)	(0.000)	(0.000)
Market	0.8763***	0.7058***	1.2902***	1.2888***
	(0.066)	(0.078)	(0.020)	(0.020)
SMB	-0.1207	0.3078*	0.0328	0.0380
	(0.121)	(0.144)	(0.037)	(0.037)
HML	−0.3985***	−0.4038***	0.1224***	0.1204***
	(0.085)	(0.101)	(0.026)	(0.026)
R^2	0.077	0.046	0.629	0.631
Num. obs.	2517	2517	2517	2517
RMSE	0.0369	0.0437	0.0113	0.0111
Panel B: Combination Coefficients and Performance				
avg. coeff. $\hat{\boldsymbol{w}}_e$	0.7725	1.5671	1.2601	1.2552
	(0.1038)	(1.8392)	(0.1387)	(0.1339)
avg. coeff. $\hat{\boldsymbol{w}}_{mv}$	0.2275	0.2274	0.0904	0.0896
	(0.1038)	(0.0984)	(0.0099)	(0.0099)
avg. port. return	0.0743	0.0737	0.0979	0.0929
	(3.8355)	(4.4758)	(1.8577)	(1.8332)

$^{***}p < 0.001$. The average portfolio return (avg. port. return) is presented in percentage.

From Panel A of Table 4.8, it is evident that for portfolio returns generated by our WPW framework, R^2 exceeds 0.5, indicating a substantial portion of the portfolio's returns can be explained by the FF3 factors. This is attributed to the inclusion of fundamental factors in determining winning probabilities. In contrast, R^2 for portfolio returns from the analytical method is below 0.1, suggesting minimal explanatory power from the same factors. Furthermore, the RMSE of the combined portfolio obtained from the WPW framework is smaller than that from analytical methods, indicating more precise predictions. Panel B of Table 4.8 reveals that the combined portfolio from our WPW framework allocates more to the 1/N rule than to the sample MV portfolio on average. However, as pointed out earlier from Table 4.7, the lower volatility of portfolio returns does not necessarily come with the lower average return when comparing the WPW combined portfolios with the analytical ones.

4.5.3.2 Certainty Equivalent Return

For every involved portfolio $\hat{\boldsymbol{w}}_i$, we calculate the Certainty Equivalent Return (CER) according to the following formula:

$$\text{CER}_{\hat{\boldsymbol{w}}_i} = \mu_{\hat{\boldsymbol{w}}_i} - \frac{\gamma}{2}\sigma^2_{\hat{\boldsymbol{w}}_i},$$

where $\mu_{\hat{\boldsymbol{w}}_i}$ and $\sigma^2_{\hat{\boldsymbol{w}}_i}$ are the mean return and variance of the out-of-sample portfolio returns, and γ denotes the risk aversion coefficient. We also present the out-of-sample Sharpe ratio in Appendix C.3.3.

Table 4.9 presents the CER in basis points (bps) for various datasets and portfolios, highlighting the comparative performance across different strategies. The results demonstrate a notable advantage of the WPW combined portfolios over others. Firstly, for Panels A through D, when comparing the four combined portfolios constructed in different ways, those structured within the WPW framework consistently exhibit superior performance relative to those developed via analytical methods. Secondly, a comparison between the first four panels (i.e., A through D) with Panel E shows that the WPW combined portfolios outperform their constituent counterparts for a majority of cases, and the exceptions happen only with dataset SOP100 and a few other cases with datasets A17 and LT10.

Table 4.9: Certainty equivalent return on various datasets.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	8.1118	4.6161	2.6423	8.1121	7.4720	5.5228	5.3389	0.7421
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	8.0905	4.1318	1.8988	7.4824	5.4797	5.1796	5.0225	0.8177
Panel A.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}mv}$	-6.7759	-14.6328	-17.5856	-4.6324	-13.0845	-7.1985	-3.8312	-15.1847
$\hat{\boldsymbol{w}}^{u}_{e\text{-}mv}$	-12.9696	-22.6833	-25.1643	-11.8484	-20.9291	-16.5226	-13.2284	-21.1974
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	5.0936	2.8241	1.088	6.5913	4.4999	2.576	3.992	0.1395
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.3002	3.0409	1.512	6.1322	4.354	3.1912	4.2963	0.7485
Panel B.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}kz}$	-14.0236	-16.6264	-21.251	-5.0561	-13.6619	-12.3954	-8.2873	-20.1338
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	5.9381	3.0368	5.3907	12.6263	12.9614	4.5761	4.9936	-1.3452
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.3240	2.9559	4.6963	12.2518	11.0369	4.3207	5.2354	-1.8529
Panel C.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mv}$	-20.9957	-26.6327	-36.3343	-10.7279	-24.1954	-24.1601	-17.4903	-33.3993
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	6.3738	2.6063	5.7481	12.6061	11.5854	4.9609	5.4068	-1.2032
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.1690	2.9357	5.694	12.4141	10.1553	4.2119	5.2457	-1.8450
Panel D.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mr}$	-4.7062	-14.6634	-15.6463	-3.7154	-13.4442	-8.712	-6.1106	-16.3479
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_{e}$	2.7317	3.1338	3.1186	2.3868	2.5514	2.8468	2.9763	2.3350
$\hat{\boldsymbol{w}}_{g}$	2.436	2.3951	4.7502	5.8198	4.9772	2.6008	3.1473	0.1609
$\hat{\boldsymbol{w}}_{mv}$	-94.711	-182.5342	-369.097	-535.4249	-521.7458	-683.7145	-1380.7968	-11041.7506
$\hat{\boldsymbol{w}}_{mr}$	-84.2156	-196.3545	-466.0029	-846.5937	-822.0931	-1199.0283	-3978.4773	-407277.9177
$\hat{\boldsymbol{w}}_{kz}$	-20.9957	-26.6327	-36.3343	-10.7279	-24.1954	-24.1601	-17.4903	-33.3993
Pabel E.2: Scaled constituent portfolios								
${}_s\hat{\boldsymbol{w}}_{e}$	-7.7326	-1.5206	-0.0462	-3.699	-12.5325	-1.1773	-0.3504	-6.9695
${}_s\hat{\boldsymbol{w}}_{g}$	-4.1806	-0.5501	6.9828	8.749	2.0497	1.8279	3.4466	-5.1025

4.5.3.3 Certainty Equivalent Return Net of Trading Costs

To account for transaction costs, we introduce a proportional transaction cost parameter κ. The net excess return of the out-of-sample portfolio, after accounting for transaction costs, is defined as follows (Kircher and Rösch, 2021):

$$r^{\kappa}(\hat{\boldsymbol{w}}_{c,t}) = (1 + \hat{\boldsymbol{w}}_{c,t}^{\top}\boldsymbol{R}_{t+1} + R_{f,t+1})\left(1 - \kappa\sum_{i=1}^{N}|\hat{w}_{t+1,i} - \hat{w}_{t^+,i}|\right) - 1 - R_{f,t+1},$$

where $\hat{w}_{t^+,i}$ denotes the weight of risky asset i at time $t+1$ immediately before rebalancing, calculated as

$$\hat{w}_{t^+,i} = \hat{w}_{t,i}\frac{1 + R_{f,t+1} + R_{t+1,i}}{1 + \hat{\boldsymbol{w}}_{c,t}^{\top}\boldsymbol{R}_{t+1} + R_{f,t+1}},$$

and $w_{t+1,i}$ is the weight at time $t+1$ after rebalancing.

Consistent with insights from Kan et al. (2022), current transaction costs can be much lower than previous studies suggested. For large institutional investors or when trading high volumes, the proportional transaction costs can be as low as 1 to 5 bps, see, for example, the published analysis on trading costs from Bloomberg or Reuters. Hence, we set 1 bps and recalculate the CER as detailed in Table 4.10, with additional CER results for scenarios involving 2 or 10 bps provided in Appendix C.3.2. Meanwhile, the turnover of various portfolio rules is also reported in Appendix C.3.2. The results from Table 4.10 indicate that the introduction of transaction costs tends to diminish the performance of most portfolio strategies. Nevertheless, when $\kappa = 1$ bps, the scaled WPW combined portfolios generally maintain superior performance across most scenarios.

Table 4.10: Certainty equivalent return with transaction cost $\kappa = 1$ bps.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}_{e\text{-}mv}^{LG}$	6.5062	2.7347	1.3515	5.5627	3.4707	3.713	3.5036	-7.5215
${}_s\hat{\boldsymbol{w}}_{e\text{-}mv}^{RF}$	6.4762	2.2764	0.6684	4.7348	1.9598	3.434	3.0707	-7.4387
Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_{e\text{-}mv}^{a}$	-11.7766	-19.1841	-20.3147	-10.0171	-22.3472	-10.8688	-8.0095	-31.6954
$\hat{\boldsymbol{w}}_{e\text{-}mv}^{u}$	-17.9715	-27.2523	-27.9719	-17.314	-30.2917	-20.3139	-17.5318	-37.8174
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}_{e\text{-}kz}^{LG}$	3.9429	1.5352	0.3092	4.964	2.5357	1.6157	2.7656	-6.7278
${}_s\hat{\boldsymbol{w}}_{e\text{-}mv}^{RF}$	4.1481	1.6964	0.693	4.4153	2.4695	2.0956	3.0014	-6.0884
Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_{e\text{-}kz}^{a}$	-18.7909	-20.8002	-23.9176	-10.5428	-22.6697	-16.2392	-13.2168	-40.5961

Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
${}_{s}\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	3.7221	1.0914	3.7808	9.7278	8.0938	2.5193	2.6726	-12.3477
${}_{s}\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	3.3925	1.0444	3.1866	9.3365	6.4373	2.3163	2.7955	-12.6382
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mv}$	-26.517	-31.7922	-39.9897	-18.0411	-35.7188	-29.4512	-24.7101	-61.6506
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
${}_{s}\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	3.6817	0.4283	4.0225	9.1763	6.178	2.7836	2.8129	-12.2624
${}_{s}\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	2.9087	0.8144	4.0184	9.2196	4.795	2.1116	2.6859	-12.9251
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mr}$	-9.7745	-19.271	-18.4628	-8.9641	-22.3835	-12.3112	-10.3666	-34.5197
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_{e}$	2.3498	2.7378	2.7269	1.9972	2.175	2.4559	2.5852	1.9569
$\hat{\boldsymbol{w}}_{g}$	1.8764	1.7629	4.1436	4.9242	3.9157	1.8119	2.0028	-6.2152
$\hat{\boldsymbol{w}}_{mv}$	-107.0712	-199.3307	-386.187	-573.8948	-580.5723	-719.6729	-1440.5418	-9053.5828
$\hat{\boldsymbol{w}}_{mr}$	-96.7800	-214.5916	-486.6779	-895.5905	-901.4563	-1246.7491	-3957.2526	-76154.4388
$\hat{\boldsymbol{w}}_{kz}$	-26.5170	-31.7922	-39.9897	-18.0411	-35.7188	-29.4512	-24.7101	-61.6506
Pabel E.2: Scaled constituent portfolios								
${}_{s}\hat{\boldsymbol{w}}_{e}$	-8.2263	-2.022	-0.509	-4.1687	-13.0297	-1.6451	-0.8134	-7.4308
${}_{s}\hat{\boldsymbol{w}}_{g}$	-5.2526	-1.6285	5.9306	7.1264	-1.2499	0.3721	1.4847	-17.5842

4.6 Conclusion

In this chapter, we introduced a flexible statistical framework for constructing combined portfolio rules to offer greater flexibility and better performance. Our proposed approach utilizes winning probability weighting, where the combination coefficients are characterized as winning probabilities and determined using machine learning techniques. This methodology allows for the incorporation of both technical and fundamental feature variables in determining the combination coefficients. Our extensive empirical studies have demonstrated the superior performance of the WPW combined portfolio over most scenarios.

Although our discussion primarily focuses on combining two constituent portfolios, our WPW framework is readily adaptable for combining more than two constituent portfolios by employing multiple classification models to determine the winning probabilities. When there are more than two constituent portfolios, the combination coefficient assigned to each constituent could be the probability for it to beat the rest of the constituents in the out-of-sample return. As a future research direction on the topic, it is interesting to explore more machine-learning techniques for determining the combination coefficient within the WPW framework. For instance, deep neural networks, long short-term memory networks, or ensemble models that combine several algorithms could be investigated. Moreover, there is room for exploring more cross-sectional financial market feature variables to enhance the construction of the combination coefficient and further improve the out-of-sample performance of the combined portfolio.

Chapter 5

Conclusion and Future Work

In this thesis, we delve into three distinct perspectives within the field of portfolio selection, focusing on risk estimation and optimal combined portfolio strategies. We begin with an introduction and some preliminary discussions in Chapter 1. Subsequently, we present three original works in Chapters 2-4. In this last chapter, we summarize the main results of the thesis work in Section 5.1 and present open questions along with challenges for future research in Section 5.2.

5.1 Summary of the Thesis Work

Chapter 2 discusses the Tail Mean-Variance (TMV) portfolio selection with estimation risk, where the TMV risk measure is specifically focused on addressing extreme losses. We introduce the Mean-Variance-Standard-deviation (MVS) optimization criterion, which includes the TMV, MV, and Mean-Standard-Deviation as special cases. To mitigate estimation risk, we propose a combined three-fund portfolio consisting of the 1/N rule and the other two components corresponding to the global minimum variance portfolio and the zero investment portfolio. We derive optimal combination coefficients based on the expected out-of-sample performance that aligns with the MVS optimization criterion under the multivariate normal distribution assumption of asset returns. From the simulation and empirical studies, we find that the combined three-fund portfolio outperforms the combined two-fund portfolios, the plug-in MVS portfolio, and the 1/N rule in most scenarios in terms of the expected out-of-sample MVS utility. Additionally, it remains competitively close even in scenarios where it does not achieve the best performance. We conclude that for investors interested in TMV portfolio selection, it is advisable to opt for the combined three-fund portfolio rather than relying on the traditional plug-in rule.

Chapter 3 focuses on ESG investing with the consideration of estimation risk. ESG investing has emerged as a global trend, enabling investors and financial institutions to capture sustainable benefits. Firstly, we introduce an MV optimization framework that incorporates both the weight constraint (ensuring the sum of portfolio weights equals one) and the total ESG constraint (aligning the total ESG score of the portfolio with a target ESG score). The resulting optimal ESG portfolio satisfies a three-fund separation. To address estimation risk, we then propose a combined three-fund portfolio with components originating from the plug-in ESG portfolio. Under the multivariate normal distribution assumption of asset returns, we derive the optimal combination coefficients using the expected out-of-sample MV utility optimization, incorporating either an inequality or equality constraint on the expected total ESG scores. Through both simulation and empirical studies, we conclude the implementable combined portfolios outperform the traditional plug-in ESG portfolio in terms of certainty equivalent return and Sharpe ratio.

Note that in both Chapters 2 and 3, the combined portfolios are constrained to a specific structure due to the complexities involved in deriving the analytical expected out-of-sample performance measures, and the calculation of combination coefficients relies on the assumption of a multivariate normal distribution of asset returns. These limitations raise questions about how we can develop a more flexible framework that can determine combination coefficients for any combined portfolios, irrespective of the structure of the constituent portfolios. To address it, Chapter 4 introduces the Winning Probability Weighted (WPW) combined portfolios, which presents a flexible statistical framework for constructing combined portfolios that offer greater flexibility and enhanced out-of-sample performance. The WPW framework treats combination coefficients as winning probabilities, which are determined using machine learning techniques. This method allows for the integration of both technical and fundamental feature variables in determining the combination coefficients. We examine various combined portfolios across different datasets in the empirical study. The results consistently show that our WPW combined portfolios provide superior out-of-sample performance in terms of certainty equivalent return, and consistently outperform both the combined portfolios based on analytical methods and the individual constituent portfolios, with only a few exceptions of slight underperformance. This chapter provides a valuable framework for investors looking to tailor combined portfolios based on their preferences in practice.

5.2 Future Research Avenues

Building on the research work presented in this thesis, there are many directions worth further exploration for future research. In this section, we first outline potential future research avenues for Chapters 2-4 that align closely with our interests, and then discuss broader research avenues.

5.2.1 Potential Research Avenues for Each Chapter

Chapter 2 (Tail Mean-Variance Portfolio Selection with Estimation Risk):

- The Tail Mean-Variance (TMV) risk measure focuses specifically on the tail portion of the loss distribution, which involves a weighted combination of the expectation and deviation of losses exceeding a certain threshold. In actuarial science, heavy tails are prevalent and significantly impact the assessment of tail risks. However, the adopted normal distribution assumption does not adequately capture these heavy tails. Given this limitation, it may be beneficial to explore combined portfolios that account for the effect of heavy tails in the TMV portfolio selection with estimation risk. A potential approach is to allow for more realistic distributional assumptions. Specifically, Owadally and Landsman (2013) explore TMV portfolio selection under the multivariate elliptical distribution, and Kan and Lassance (2024) examine the impact of fat tails in return distributions under parameter uncertainty. Building on their work, we could further enhance the TMV portfolio to comprehensively account for both heavy tails and estimation risk, making it better suited for real-world scenarios.

- We introduce the Mean-Variance-Standard-deviation (MVS) framework which encompasses the TMV, MV, and Mean-Standard-deviation as special cases. The MVS framework can also address the ambiguity of the mean vector, as discussed by Garlappi et al. (2007). Specifically, a max-min problem with the MV objective, subject to a weight constraint that the sum of portfolio weights equals one and an ambiguity constraint related to the uncertainty about the expected return, can be reformulated into a maximization problem with the MVS objective function. In light of this connection, it may be intriguing to explore whether the idea of combined portfolios could be effectively applied to the MV problem under ambiguity. Additionally, it may be worthwhile to look at the feasibility of incorporating ambiguity into an expected out-of-sample MV utility function. These explorations could potentially enhance the robustness and applicability of combined portfolios in handling real-world uncertainties.

- As a further extension, it would be interesting to explore the minimization of various deviation measures. The deviation measures quantify the nonconstancy of a random variable, and can be obtained from the risk quadrangle framework as proposed by Rockafellar and Uryasev (2013), such as the quantile-based and expectile-based quadrangles, among others. Unlike standard deviation, these deviation measures are asymmetric, making them suitable for capturing the tail behavior of the loss (or return) distribution in portfolio optimization.

Chapter 3 (ESG Investing with Estimation Risk):

- When addressing estimation risk, we utilize a constraint based on the expected total ESG score (i.e., $\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}] = \bar{s}$) rather than a pointwise total ESG score (i.e., $\hat{\boldsymbol{w}}_c^\top \boldsymbol{S} = \bar{s}$). It is due to the design of the combined portfolio, which adheres to a weight constraint ensuring that the sum of the portfolio weights equals one, making it impossible to satisfy the pointwise ESG constraint directly. A potential solution to meet the pointwise ESG constraint is to adjust the composition of the constituent portfolios within the combined portfolio. For instance, if we remove the weight constraint, we could redefine the combined portfolio as follows:

$$\hat{\boldsymbol{w}}_c = \bar{s}\frac{\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}}{\boldsymbol{S}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}} + \delta\widehat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\mu}} - \frac{\boldsymbol{S}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\boldsymbol{S}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}}\boldsymbol{S}\right)$$

  where δ is the combination coefficient and it satisfies the relationship $\hat{\boldsymbol{w}}_c^\top \boldsymbol{S} = \bar{s}$. If we treat the ESG score $\boldsymbol{S}$ as constant, we end up with a combined portfolio similar to that considered in Kan et al. (2022), where the ESG score vector is replaced by the vector of ones. Additionally, it may provide valuable insights to further explore the ambiguity of the ESG score $\boldsymbol{S}$. It would introduce additional dimensions of parameter uncertainty, adding complexity and depth to the analysis.

- The reason we prefer adding the weight constraint - that the sum of portfolio weights equals one - is to exclude the consideration of the risk-free asset in the portfolio strategy. Including a risk-free asset presents challenges, particularly because its ESG score is unknown. It may be overly cautious to simply assume that the ESG score of the risk-free asset is zero and that it does not contribute to the total ESG score of the portfolio. Given this limitation, it may be valuable to specifically investigate the ESG score or the "greenness" of the risk-free asset. To our knowledge, such research has not yet been conducted in the existing literature so far. Building upon the developed ESG score for the risk-free asset, it is intriguing to extend our analysis in this chapter to include the risk-free asset in portfolio construction.

Chapter 4 (Winning Probability Weighted Combined Portfolio):

- The WPW framework is a promising candidate for enhancement in a multiperiod setting. Given that portfolio gains or losses are immediately achievable after each investment period, this inherent nature allows us to utilize advanced machine learning methods, such as reinforcement learning which possesses real-time adaptability and predictive prowess. Notably, DeMiguel et al. (2015) address estimation risk in the multiperiod setting using traditional analytical methods. It would be interesting to explore how the WPW framework could be further extended in a multiperiod context to leverage both technical and fundamental features without relying on the strict normal distribution assumption.

- When defining the winning probability, we currently consider it as the scenario where the out-of-sample return of the adjusted constituent portfolio is greater than that of the other. However, it may be interesting to explore different criteria for defining the winning probability. For example, we could define the winning probability using an out-of-sample performance function $g(\cdot)$ as follows:

$$p_{i,t} = \mathbb{P}[g(\tilde{\boldsymbol{w}}_{i,t}) > g(\tilde{\boldsymbol{w}}_{j,t})|\mathcal{F}_t], \quad i, j = 1, 2, \text{ and } i \neq j,$$

where the function $g(\cdot)$ could directly be the out-of-sample MV utility or Sharpe ratio, or the ratio of changes in utility or Sharpe ratio. A potential challenge in this approach lies in accurately estimating these metrics, which would require historical data from several previous days. The results might be highly volatile when the estimation window changes. Hence, this extension necessitates careful consideration of the methods used to estimate these values to ensure they do not compromise the reliability of the classification.

5.2.2 Broader Research Avenues

Centered around the main topics of this thesis, the estimation risk and optimal combined portfolio strategies, there are more interesting directions to explore.

- Instead of combining pre-specified constituent portfolios and determining the optimal combination coefficients, it is also interesting to explore the potentials of simply averaging various portfolios. This averaging method resembles a fund-of-funds structure, with each strategy functioning as a distinct fund. Basic statistical principles suggest that averaging uncorrelated strategies, treated as random variables, effectively reduces variance. For example, averaging can be strategically applied to existing portfolios that possess distinct information and are inherently unlikely to be perfectly correlated. Additionally, the averaging method can accommodate portfolios that lack analytical tractability, such as those subject to no-short-sale constraints.
- Under the assumption of non-normality in asset returns and the presence of higher-order co-moments, it is worthwhile to consider the dynamic risk aversion parameters for both ESG and traditional investment frameworks. For example, when working with a certainty equivalent derived from a fourth-order Taylor expansion of the constant absolute risk aversion (CARA) utility function, we can capture not only the expected return but also the variance, skewness, and kurtosis of the portfolio's returns. Consequently, the portfolio optimization problem can be formulated as a

time-varying maximization program of the certainty equivalent. Even though such an optimization problem lacks an analytical solution and must be solved numerically when a non-normal distribution is assumed, it could still potentially be computed using various types of quadrature rules.

- It is worthwhile to explore how leverage can be incorporated into the portfolio optimization process, particularly under parameter uncertainty. A leveraged portfolio fundamentally differs from a traditional one without leverage due to the unique risks associated with leverage. Investors seeking to manage portfolio leverage often incorporate a leverage constraint into the MV optimization to achieve a desired leverage level based on the volatility of the securities. Integrating leverage aversion into portfolio optimization typically results in portfolios with lower leverage compared to those generated by conventional MV optimization. It benefits leverage-averse investors because it aligns the portfolios more closely with their preferences regarding leverage.

References

Abdi, H., Williams, L.J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* **2**(4), 433–459.

Acerbi, C., Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance* **26**(7), 1487–1503.

Ait-Sahalia, Y., Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics* **201**(2), 384–399.

Alexander, G.J., Baptista, A.M. (2002). Economic implications of using a mean-var model for portfolio selection: A comparison with mean-variance analysis. *Journal of Economic Dynamics and Control* **26**(7-8), 1159–1193.

Alexander, G.J., Baptista, A.M. (2004). A comparison of var and cvar constraints on portfolio selection with the mean-variance model. *Management Science* **50**(9), 1261–1273.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(3), 370–374.

Anderson, T.W. (1994). *The statistical analysis of time series*. John Wiley & Sons.

Andrews, D.W., Monahan, J.C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica: Journal of the Econometric Society* pp. 953–966.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9**(3), 203–228.

Avramov, D., Cheng, S., Lioui, A., Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics* **145**(2), 642–664.

Ba, J.L., Kiros, J.R., Hinton, G.E. (2016). Layer normalization. *arXiv preprint arXiv: 1607.06450*.

Bauder, D., Bodnar, T., Parolya, N., Schmid, W. (2021). Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty. *Quantitative Finance* **21**(2), 221–242.

Bauer, C. (2000). Value at risk using hyperbolic distributions. *Journal of Economics and Business* **52**(5), 455–467.

Belsley, D.A., Kuh, E., Welsch, R.E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Benartzi, S., Thaler, R.H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review* **91**(1), 79–98.

Berg, F., Koelbel, J.F., Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance* **26**(6), 1315–1344.

Berger, J.O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Best, M.J., Grauer, R.R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The Review of Financial Studies* **4**(2), 315–342.

Bingham, N.H., Kiesel, R. (2002). Semi-parametric modelling in finance: theoretical foundations. *Quantitative Finance* **2**(4), 241–250.

Bingham, N.H., Kiesel, R., Schmidt, R. (2003). A semi-parametric approach to risk management. *Quantitative Finance* **3**(6), 426–441.

Bodnar, T., Mazur, S., Okhrin, Y. (2017). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research* **256**(1), 292–307.

Bodnar, T., Okhrin, Y. (2011). On the product of inverse wishart and normal distributions with applications to discriminant analysis and portfolio theory. *Scandinavian Journal of Statistics* **38**(2), 311–331.

Bodnar, T., Okhrin, Y., Parolya, N. (2022). Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics* **41**(1), 140–156.

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

Bodnar, T., Schmid, W., Zabolotskyy, T. (2012). Minimum var and minimum cvar optimal portfolios: estimators, confidence regions, and tests. *Statistics & Risk Modeling* **29**(4), 281–314.

Bodnar, T., Schmid, W., Zabolotskyy, T. (2013). Asymptotic behavior of the estimated weights and of the estimated performance measures of the minimum var and the minimum cvar optimal portfolios for dependent data. *Metrika* **76**, 1105–1134.

Boyd, S., Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

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

Boyle, P., Feng, S., Melkuev, D., Yang, S., Zhang, J. (2018). Short positions in the first principal component portfolio. *North American Actuarial Journal* **22**(2), 223–251.

Boyle, P., Tian, W. (2008). The design of equity-indexed annuities. *Insurance: Mathematics and Economics* **43**(3), 303–315.

Branger, N., Lučivjanská, K., Weissensteiner, A. (2019). Optimal granularity for portfolio choice. *Journal of Empirical Finance* **50**, 125–146.

Breiman, L. (2001). Random forests. *Machine Learning* **45**, 5–32.

Briec, W., Kerstens, K., Van de Woestyne, I. (2013). Portfolio selection with skewness: A comparison of methods and a generalized one fund result. *European Journal of Operational Research* **230**(2), 412–421.

Brown, S.J. (1976). *Optimal portfolio choice under uncertainty: a Bayesian approach.*. The University of Chicago.

Carhart, M.M. (1997). On persistence in mutual fund performance. *The Journal of Finance* **52**(1), 57–82.

Chakrabarti, D. (2021). Parameter-free robust optimization for the maximum-sharpe portfolio problem. *European Journal of Operational Research* **293**(1), 388–399.

Chatterjee, S., Hadi, A.S. (2006). *Regression analysis by example.* John Wiley & Sons.

Chen, W., Zhang, H., Jia, L. (2022). A novel two-stage method for well-diversified portfolio construction based on stock return prediction using machine learning. *North American Journal of Economics and Finance* **63**, 101818.

Cheung, Y.W., Chinn, M.D. (1996). Deterministic, stochastic, and segmented trends in aggregate output: A cross-country analysis. *Oxford Economic Papers, New Series* **48**(1), 134–162.

Chiragiev, A., Landsman, Z. (2007). Multivariate pareto portfolios: TCE-based capital allocation and divided differences. *Scandinavian Actuarial Journal* **2007**(4), 261–280.

Conrad, J., Kaul, G. (1988). Time-variation in expected returns. *Journal of Business* pp. 409–425.

Cui, H., Tan, K.S., Yang, F., Zhou, C. (2022). Asymptotic analysis of portfolio diversification. *Insurance: Mathematics and Economics* **106**, 302–325.

Das, S., Markowitz, H., Scheid, J., Statman, M. (2010). Portfolio optimization with mental accounts. *Journal of Financial and Quantitative Analysis* **45**(2), 311–334.

De Caigny, A., Coussement, K., De Bock, K.W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research* **269**(2), 760–772.

De Nard, G., Ledoit, O., Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics* **19**(2), 236–257.

De Spiegeleer, J., Höcht, S., Jakubowski, D., Reyners, S., Schoutens, W. (2023). ESG: A new dimension in portfolio allocation. *Journal of Sustainable Finance & Investment* **13**(2), 827–867.

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

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

DeMiguel, V., Gil-Bazo, J., Nogales, F.J., Santos, A.A. (2023). Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics* **150**(3), 103737.

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

DeMiguel, V., Martín-Utrera, A., Nogales, F.J. (2015). Parameter uncertainty in multiperiod portfolio optimization with transaction costs. *Journal of Financial and Quantitative Analysis* **50**(6), 1443–1471.

Dichtl, H., Drobetz, W., Wendt, V.S. (2021). How to build a factor portfolio: Does the allocation strategy matter?. *European Financial Management* **27**(1), 20–58.

Dickey, J.M. (1967). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *The Annals of Mathematical Statistics* **38**(2), 511–518.

Dickson, D.C., Hardy, M.R., Waters, H.R. (2019). *Actuarial mathematics for life contingent risks*. Cambridge University Press.

Doan, B., Papageorgiou, N., Reeves, J.J., Sherris, M. (2018). Portfolio management with targeted constant market volatility. *Insurance: Mathematics and Economics* **83**, 134–147.

Dowd, K., Blake, D. (2006). After var: the theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance* **73**(2), 193–229.

Duchin, R., Levy, H. (2009). Markowitz versus the talmudic portfolio diversification strategies. *The Journal of Portfolio Management* **35**(2), 71–74.

Dumitrescu, E., Hue, S., Hurlin, C., Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research* **297**(3), 1178–1192.

Eini, E.J., Khaloozadeh, H. (2021*a*). The tail mean–variance optimal portfolio selection under generalized skew-elliptical distribution. *Insurance: Mathematics and Economics* **98**, 44–50.

Eini, E.J., Khaloozadeh, H. (2021*b*). The tail mean-variance model and extended efficient frontier. *Advances in Mathematical Finance and Applications* **6**(1), 179–193.

Eini, E.J., Khaloozadeh, H. (2022*a*). Dynamic portfolio selection: Tail mean-variance model and stochastic lq approach.

Eini, E.J., Khaloozadeh, H. (2022*b*). Tail variance for generalized skew-elliptical distributions. *Communications in Statistics-Theory and Methods* **51**(2), 519–536.

Enke, D., Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications* **29**(4), 927–940.

Fabozzi, F.J., Gupta, F., Markowitz, H.M. (2002). The legacy of modern portfolio theory. *The Journal of Investing* **11**(3), 7–22.

Fama, E.F. (1965). The behavior of stock-market prices. *The Journal of Business* **38**(1), 34–105.

Fama, E.F., French, K.R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116**(1), 1–22.

Fan, J., Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(5), 849–911.

Fan, J., Song, R. (2010). Sure independence screening in generalized linear models with np-dimensionality. *The Annals of Statistics* **38**(6), 3567–3604.

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

Fang, K.T., Kotz, S., Ng, K.W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall/CRC.

Feng, G., Giglio, S., Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance* **75**(3), 1327–1370.

Fischer, T., Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* **270**(2), 654–669.

Flori, A., Regoli, D. (2021). Revealing pairs-trading opportunities with long short-term memory networks. *European Journal of Operational Research* **295**(2), 772–791.

Frahm, G., Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics* **159**(2), 289–302.

Frankfurter, G.M., Phillips, H.E., Seagle, J.P. (1971). Portfolio selection: the effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis* **6**(5), 1251–1262.

Freund, Y., Schapire, R., Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence* **14**(771-780), 1612.

Frost, P.A., Savarino, J.E. (1986). An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis* **21**(3), 293–305.

Fuller, W.A. (1995). *Introduction to statistical time series*. Vol. 230. John Wiley & Sons.

Furman, E., Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin: The Journal of the IAA* **36**(2), 433–462.

Garlappi, L., Uppal, R., Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies* **20**(1), 41–81.

Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies* **19**(3), 797–827.

Goldstein, I., Spatt, C.S., Ye, M. (2021). Big data in finance. *The Review of Financial Studies* **34**(7), 3213–3225.

Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep learning*. MIT press.

Green, R.C., Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified?. *The Journal of Finance* **47**(5), 1785–1809.

Gu, S., Kelly, B., Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies* **33**(5), 2223–2273.

Guo, D., Boyle, P.P., Weng, C., Wirjanto, T.S. (2019). When does the 1/n rule work?. *Available at SSRN 3111531*.

Gupta, A.K., Nagar, D.K. (2018). *Matrix variate distributions*. Chapman and Hall/CRC.

Gurney, K. (1997). *An introduction to neural networks*. CRC press.

Györfi, L., Udina, F., Walk, H. (2008). Nonparametric nearest neighbor based empirical portfolio selection strategies. *Statistics and Decisions* **26**, 145–157.

Haff, L. (1979). An identity for the wishart distribution with applications. *Journal of Multivariate Analysis* **9**(4), 531–544.

Halbritter, G., Dorfleitner, G. (2015). The wages of social responsibility—where are they? a critical review of esg investing. *Review of Financial Economics* **26**, 25–35.

Hamilton, J.D. (2020). *Time series analysis*. Princeton university press.

Hao, J., He, F., Ma, F., Zhang, S., Zhang, X. (2023). Machine learning vs deep learning in stock market investment: an international evidence. *Annals of Operations Research* pp. 1–23.

Hardy, M.R. (2006). An introduction to risk measures for actuarial applications. *SOA Syllabus Study Note* **19**.

Hashiguchi, H., Takayama, N., Takemura, A. (2018). Distribution of the ratio of two wishart matrices and cumulative probability evaluation by the holonomic gradient method. *Journal of Multivariate Analysis* **165**, 270–278.

He, X.D., Jin, H., Zhou, X.Y. (2015). Dynamic portfolio choice when risk is measured by weighted VaR. *Mathematics of Operations Research* **40**(3), 773–796.

Hellmich, M., Kassberger, S. (2011). Efficient and robust portfolio optimization in the multivariate generalized hyperbolic framework. *Quantitative Finance* **11**(10), 1503–1516.

Henrique, B.M., Sobreiro, V.A., Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications* **124**, 226–251.

Ho, T.K. (1995). Random decision forests. *in* 'Proceedings of 3rd international conference on document analysis and recognition'. Vol. 1. IEEE. pp. 278–282.

Hofmann, T., Schölkopf, B., Smola, A.J. (2008). Kernel methods in machine learning. *The Annals of Statistics* **36**(3), 1171–1220.

Huang, H., Milevsky, M.A., Wang, J. (2008). Portfolio choice and life insurance: The crra case. *Journal of Risk and Insurance* **75**(4), 847–872.

Huck, N. (2019). Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research* **278**(1), 330–342.

Hwang, I., Xu, S., In, F. (2018). Naive versus optimal diversification: Tail risk and performance. *European Journal of Operational Research* **265**(1), 372–388.

Ignatieva, K., Landsman, Z. (2015). Estimating the tails of loss severity via conditional risk measures for the family of symmetric generalised hyperbolic distributions. *Insurance: Mathematics and Economics* **65**, 172–186.

Ignatieva, K., Landsman, Z. (2019). Conditional tail risk measures for the skewed generalised hyperbolic family. *Insurance: Mathematics and Economics* **86**, 98–114.

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

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An introduction to statistical learning*. Vol. 112. Springer.

Javier, W., Gupta, A. (1985). On generalized matric variate beta distributions. *Statistics* **16**(4), 549–558.

Jiang, C., Du, J., An, Y. (2019). Combining the minimum-variance and equally-weighted portfolios: Can portfolio performance be improved?. *Economic Modelling* **80**, 260–274.

Jiang, C.F., Peng, H.Y., Yang, Y.K. (2016). Tail variance of portfolio under generalized laplace distribution. *Applied Mathematics and Computation* **282**, 187–203.

Jobson, J. (1979). Improved estimation for markowitz portfolios using james-stein type estimators. *in* 'Proceedings of the American Statistical Association, Business and Economics Statistics Section'. Vol. 71. pp. 279–284.

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

Jorion, P. (1991). Bayesian and capm estimators of the means: Implications for portfolio selection. *Journal of Banking & Finance* **15**(3), 717–727.

Kalymon, B.A. (1971). Estimation risk in the portfolio selection model. *Journal of Financial and Quantitative Analysis* **6**(1), 559–582.

Kan, R., Lassance, N. (2024). Optimal portfolio choice with fat tails and parameter uncertainty. *Available at SSRN 4652814*.

Kan, R., Wang, X. (2019). On the economic value of alphas. *Available at SSRN 1785161*.

Kan, R., Wang, X. (2023). Optimal portfolio choice with unknown benchmark efficiency. *Management Science* pp. 1–22.

Kan, R., Wang, X., Zhou, G. (2016). *On the value of portfolio optimization in the presence of estimation risk: the case with and without risk-free asset*. University of Toronto-Rotman School of Management.

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

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

Kazak, E., Pohlmeier, W. (2023). Bagged pretested portfolio selection. *Journal of Business & Economic Statistics* **41**(4), 1116–1131.

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

Kirby, C., Ostdiek, B. (2012). It is all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis* **47**(2), 437–467.

Kircher, F., Rösch, D. (2021). A shrinkage approach for sharpe ratio optimal portfolios with estimation risks. *Journal of Banking & Finance* **133**, 106281.

Klein, R.W., Bawa, V.S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics* **3**(3), 215–231.

Klein, R.W., Bawa, V.S. (1977). The effect of limited information and estimation risk on optimal portfolio diversification. *Journal of Financial Economics* **5**(1), 89–111.

Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M. (2002). *Logistic regression.* Springer.

Kraus, M., Feuerriegel, S., Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research* **281**(3), 628–641.

Krauss, C., Do, X.A., Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* **259**(2), 689–702.

Kwiatkowski, D., Phillips, P.C., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of Econometrics* **54**(1-3), 159–178.

Landier, A., Lovo, S. (2020). ESG investing: How to optimize impact?. *HEC Paris Research Paper No. FIN-2020-1363* .

Landsman, Z. (2008*a*). Minimization of the root of a quadratic functional under a system of affine equality constraints with application to portfolio management. *Journal of Computational and Applied Mathematics* **220**(1-2), 739–748.

Landsman, Z. (2008*b*). Minimization of the root of a quadratic functional under an affine equality constraint. *Journal of Computational and Applied Mathematics* **216**(2), 319–327.

Landsman, Z. (2010). On the tail mean–variance optimal portfolio selection. *Insurance: Mathematics and Economics* **46**(3), 547–553.

Landsman, Z., Makov, U. (2016). Minimization of a function of a quadratic functional with application to optimal portfolio selection. *Journal of Optimization Theory and Applications* **170**, 308–322.

Landsman, Z., Makov, U., Shushi, T. (2016). Multivariate tail conditional expectation for elliptical distributions. *Insurance: Mathematics and Economics* **70**, 216–223.

Landsman, Z., Makov, U., Shushi, T. (2020). Portfolio optimization by a bivariate functional of the mean and variance. *Journal of Optimization Theory and Applications* **185**, 622–651.

Landsman, Z., Pat, N., Dhaene, J. (2013). Tail variance premiums for log-elliptical distributions. *Insurance: Mathematics and Economics* **52**(3), 441–447.

Landsman, Z.M., Valdez, E.A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal* **7**(4), 55–71.

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

Lassance, N. (2022). Reconciling mean-variance portfolio theory with non-gaussian returns. *European Journal of Operational Research* **297**(2), 729–740.

Lassance, N., Martin-Utrera, A. (2023). Shrinking against sentiment: Exploiting latent asset demand in portfolio selection. *Available at SSRN 3551224*.

Lassance, N., Martín-Utrera, A., Simaan, M. (2024). The risk of expected utility under parameter uncertainty. *Management Science* pp. 1–20.

Lassance, N., Vanderveken, R., Vrins, F. (2023). On the combination of naive and mean-variance portfolio strategies. *Journal of Business & Economic Statistics,* pp. 1–15.

Lassance, N., Vrins, F. (2023). Portfolio selection: A target-distribution approach. *European Journal of Operational Research* **310**(1), 302–314.

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

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

Lee, G.Y. (2023). Multivariate insurance portfolio risk retention using the method of multipliers. *North American Actuarial Journal* pp. 1–19.

Leung, M.T., Daouk, H., Chen, A.S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* **16**(2), 173–190.

Li, H., Huang, Q., Wu, B. (2021). Improving the naive diversification: An enhanced indexation approach. *Finance Research Letters* **39**, 101661.

Liu, Y., Zhou, G., Zhu, Y. (2020). Maximizing the sharpe ratio: A genetic programming approach. *Available at SSRN 3726609*.

Luo, D., Shan, X., Yan, J., Yan, Q. (2023). Sustainable investment under ESG volatility and ambiguity. *Economic Modelling* **128**, 106471.

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

Mahayni, A., Schneider, J.C. (2012). Variable annuities and the option to seek risk: Why should you diversify?. *Journal of Banking & Finance* **36**(9), 2417–2428.

Mainik, G., Embrechts, P. (2013). Diversification in heavy-tailed portfolios: properties and pitfalls. *Annals of Actuarial Science* **7**(1), 26–45.

Makariou, D., Barrieu, P., Chen, Y. (2021). A random forest based approach for predicting spreads in the primary catastrophe bond market. *Insurance: Mathematics and Economics* **101**, 140–162.

Mandelbrot, B.B. (1997). The variation of certain speculative prices. *in* 'Fractals and scaling in finance'. Springer. pp. 371–418.

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

Meenakshi, A., Rajian, C. (1999). On a product of positive semidefinite matrices. *Linear algebra and its applications* **295**(1-3), 3–6.

Mei, X., Zhu, H., Chen, C. (2023). Mean-variance portfolio selection with estimation risk and transaction costs. *Applied Economics* **55**(13), 1436–1453.

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

Michaud, R.O. (1989). The markowitz optimization enigma: Is 'optimized' optimal?. *Financial Analysts Journal* **45**, 31–42.

Midi, H., Sarkar, S.K., Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics* **13**(3), 253–267.

Milevsky, M.A., Kyrychenko, V. (2008). Portfolio choice with puts: Evidence from variable annuities. *Financial Analysts Journal* **64**(3), 80–95.

Morgan, J. (1996). Reuters. RiskMetrics-Technical Document.

Muirhead, R.J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons.

Müller, U.K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics* **141**(2), 1331–1352.

Mustafa, A.A., Lin, C.Y., Kakinaka, M. (2022). Detecting market pattern changes: A machine learning approach. *Finance Research Letters* **47**, 102621.

Nardari, F., Schüssler, R.A. (2023). Ensembles of portfolio rules. *Available at SSRN 4217088*.

Neely, C.J., Rapach, D.E., Tu, J., Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science* **60**(7), 1772–1791.

Newey, W.K., West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**(3), 703–708.

Novy-Marx, R., Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies* **29**(1), 104–147.

Nti, I.K., Adekoya, A.F., Weyori, B.A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review* **53**(4), 3007–3057.

Oehmke, M., Opp, M.M. (2020). A theory of socially responsible investment. *Swedish House of Finance Research Paper* (20-2).

Okhrin, Y., Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of econometrics* **134**(1), 235–256.

Olkin, I., Rubin, H. (1964). Multivariate beta distributions and independence properties of the wishart distribution. *The Annals of Mathematical Statistics* pp. 261–269.

Olson, D., Mossman, C. (2003). Neural network forecasts of canadian stock returns using accounting ratios. *International Journal of Forecasting* **19**(3), 453–465.

Orsi, C. (2017). New insights into non-central beta distributions. *arXiv preprint arXiv:1706.08557*.

Owadally, I., Landsman, Z. (2013). A characterization of optimal portfolios under the tail mean–variance criterion. *Insurance: Mathematics and Economics* **52**(2), 213–221.

Panjer, H.H. (2002). Measurement of risk, solvency requirements and allocation of capital within financial conglomerates. *Institute of Insurance and Pension Research, University of Waterloo, Research Report*.

Panjer, H.H., Boyle, P.P., Cox, S.H., Dufresne, D., Gerber, H.U., Mueller, H.H., Pedersen, H.W., Pliska, S.R., Sherris, M., Shiu, E.S., Tan, K.S. (1998). *Financial Economics: With Applications to Investments, Insurance, and Pensions*. Actuarial Foundation Schaumburg, Ill.

Pástor, L., Stambaugh, R.F., Taylor, L.A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics* **142**(2), 550–571.

Pedersen, L.H., Fitzgibbons, S., Pomorski, L. (2021). Responsible investing: The esg-efficient frontier. *Journal of Financial Economics* **142**(2), 572–597.

Rockafellar, R.T., Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* **18**(1-2), 33–53.

Saldana, D.F., Feng, Y. (2018). Sis: An r package for sure independence screening in ultrahigh-dimensional statistical models. *Journal of Statistical Software* **83**, 1–25.

Sharpe, W.F. (1994). The sharpe ratio. *Journal of Portfolio Management* **21**(1), 49–58.

Smith, L.N. (2017). Cyclical learning rates for training neural networks. *2017 IEEE winter conference on applications of computer vision*.

Smith, L.N. (2018). A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.

Smith, S.L., Kindermans, P.J., Ying, C., Le, Q.V. (2017). Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *in* 'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics'. Vol. 3. University of California Press. pp. 197–207.

Tian, R., Cox, S.H., Lin, Y., Zuluaga, L.F. (2010). Portfolio risk management with cvar-like constraints. *North American Actuarial Journal* **14**(1), 86–106.

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

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics* **38**(2), 413–426.

Wei, P. (2018). Risk management with weighted VaR. *Mathematical Finance* **28**(4), 1020–1060.

White, H. (2014). *Asymptotic theory for econometricians.* Academic press.

Windcliff, H., Boyle, P.P. (2004). The 1/n pension investment puzzle. *North American Actuarial Journal* **8**(3), 32–45.

Wu, W.B., Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(3), 391–410.

Xu, M., Mao, T. (2013). Optimal capital allocation based on the tail mean–variance model. *Insurance: Mathematics and Economics* **53**(3), 533–543.

Yuan, M., Zhou, G. (2022). Why naive diversification is not so naive, and how to beat it?. *Journal of Financial and Quantitative Analysis* pp. 1–32.

Zine, R. (2012). On the matrix-variate beta distribution. *Communications in Statistics-Theory and Methods* **41**(9), 1569–1582.

APPENDICES

Appendix A

Appendix for Chapter 2

A.1 Proofs

A.1.1 Proof of Proposition 2.1

Lemma A.1. *For $\gamma_1 > 0$ and $\gamma_2 > 0$, equation (2.4) has a unique positive solution η_{mvs} located in the range $(2\gamma_1, \infty)$.*

Proof of Lemma A.1. We refer to Lemma 3 in Eini and Khaloozadeh (2021*b*) for the proof. □

Proof of Proposition 2.1. Define the Lagrangian

$$\mathcal{L}(\boldsymbol{w}, \lambda_1) = \boldsymbol{w}^\top \boldsymbol{\mu} - \gamma_1 \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} - \gamma_2 \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} + \lambda_1(\boldsymbol{w}^\top \mathbf{1}_N - 1),$$

where $\lambda_1 \in \mathbb{R}$ is the Lagrangian multiplier. For any $\lambda_1 \in \mathbb{R}$, $\mathcal{L}(\boldsymbol{w}, \lambda_1)$ is a concave function of $\boldsymbol{w}$ since both the variance and standard deviation are convex functionals. The proof of the concavity is similar to that for the convexity of the approximated expected out-of-sample MVS objective function in Problem 2.12 as shown in Appendix A.1.4 for a similar proof. Therefore, we omit the details.

Consequently, for the optimality of $\boldsymbol{w}^*$, it is sufficient to verify that $\boldsymbol{w}^*$ given in (2.3) satisfies the first-order optimality condition:

$$\begin{cases} \dfrac{\partial \mathcal{L}(\boldsymbol{w}, \lambda_1)}{\partial \boldsymbol{w}} = \boldsymbol{\mu} - \left(2\gamma_1 + \dfrac{\gamma_2}{\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}}\right) \boldsymbol{\Sigma} \boldsymbol{w} + \lambda_1 \mathbf{1}_N = \mathbf{0}, \\ \dfrac{\partial \mathcal{L}(\boldsymbol{w}, \lambda_1)}{\partial \lambda_1} = \boldsymbol{w}^\top \mathbf{1}_N - 1 = 0. \end{cases} \tag{A.1}$$

The second condition in (A.1) is obviously satisfied by $\boldsymbol{w}^* = \boldsymbol{w}_{gmv} + \boldsymbol{w}_z/\eta_{mvs}$ since $\boldsymbol{w}_{gmv}^\top \mathbf{1}_N = 1$ and $\boldsymbol{w}_z^\top \mathbf{1}_N = 0$. Let

$$\nabla(\boldsymbol{w}, \lambda_1) = \boldsymbol{\mu} - \left(2\gamma_1 + \frac{\gamma_2}{\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}}\right) \boldsymbol{\Sigma} \boldsymbol{w} + \lambda_1 \mathbf{1}_N.$$

It remains to verify $\nabla(\boldsymbol{w}^*, \lambda_1) = \mathbf{0}$ for some constant $\lambda_1 \in \mathbb{R}$. To proceed, we define

$$h = \frac{\eta_{mvs}\sigma_{gmv}^2 - \mu_{gmv}}{\eta_{mvs}}$$

so that we can rewrite $\boldsymbol{w}^*$ and obtain $\boldsymbol{\Sigma}\boldsymbol{w}^*$ as follows:

$$\boldsymbol{w}^* = \frac{1}{\eta_{mvs}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + h \boldsymbol{\Sigma}^{-1} \mathbf{1}_N \quad \text{and} \quad \boldsymbol{\Sigma} \boldsymbol{w}^* = \frac{1}{\eta_{mvs}} \boldsymbol{\mu} + h \mathbf{1}_N.$$

Consequently, we take $\lambda_1 = h\left(2\gamma_1 + \frac{\gamma_2}{\sqrt{\boldsymbol{w}^{*\top} \boldsymbol{\Sigma} \boldsymbol{w}^*}}\right)$ to get

$$\nabla(\boldsymbol{w}^*, \lambda_1) = \boldsymbol{\mu} - \left(2\gamma_1 + \frac{\gamma_2}{\sqrt{\boldsymbol{w}^{*\top} \boldsymbol{\Sigma} \boldsymbol{w}^*}}\right) \left(\frac{1}{\eta_{mvs}} \boldsymbol{\mu} + h \mathbf{1}_N\right) + \lambda_1 \mathbf{1}_N = \nu \boldsymbol{\mu},$$

where the constant ν writes

$$\nu = 1 - \left(2\gamma_1 + \frac{\gamma_2}{\sqrt{\sigma_{gmv}^2 + \frac{\psi^2}{\eta_{mvs}^2}}}\right) \frac{1}{\eta_{mvs}}. \tag{A.2}$$

In the above, we applied the fact $\boldsymbol{w}^{*T} \boldsymbol{\Sigma} \boldsymbol{w}^* = \sigma_{gmv}^2 + \psi^2/\eta_{mvs}^2$ as a result from (2.3).

To complete the proof, it suffices to show $\nu = 0$ for η_{mvs} given in each of the three cases defined in the proposition according to the magnitudes of γ_1 and γ_2.

(a) When $\gamma_1 > 0$ and $\gamma_2 = 0$, we are given $\eta_{mvs} = 2\gamma_1$ and therefore, $\nu = 0$ follows trivially.

(b) When $\gamma_1 = 0$ and $\gamma_2 > 0$, we are given $\eta_{mvs} = \sqrt{(\gamma_2^2 - \psi^2)/\sigma_{gmv}^2}$, which implies

$$\eta_{mvs}^2 \sigma_{gmv}^2 + \psi^2 = \gamma_2^2.$$

Therefore, in this case, we have

$$\nu = 1 - \frac{\gamma_2}{\sqrt{\sigma_{gmv}^2 + \frac{\psi^2}{\eta_{mvs}^2}}} \frac{1}{\eta_{mvs}} = 1 - \frac{\gamma_2}{\sqrt{\eta_{mvs}^2 \sigma_{gmv}^2 + \psi^2}} = 0.$$

(c) When $\gamma_1 > 0$ and $\gamma_2 > 0$, the existence of the unique solution η_{mvs} over $(2\gamma_1, \infty)$ to equation (2.4) is guaranteed by Lemma A.1. It remains to show ν in (A.2) is zero. Indeed, (2.4) implies

$$(\eta_{mvs} - 2\gamma_1)\sqrt{\eta_{mvs}^2 \sigma_{gmv}^2 + \psi^2} - \gamma_2 \eta_{mvs} = 0.$$

Therefore,

$$\begin{aligned} \nu &= 1 - \left(2\gamma_1 + \frac{\gamma_2}{\sqrt{\sigma_{gmv}^2 + \frac{\psi^2}{\eta_{mvs}^2}}}\right) \frac{1}{\eta_{mvs}} \\ &= \frac{1}{\eta_{mvs}\sqrt{\eta_{mvs}^2 \sigma_{gmv}^2 + \psi^2}} \left[(\eta_{mvs} - 2\gamma_1)\sqrt{\eta_{mvs}^2 \sigma_{gmv}^2 + \psi^2} - \gamma_2 \eta_{mvs}\right] \\ &= 0. \end{aligned}$$

□

A.1.2 Proof of Proposition 2.2

Proof of Proposition 2.2. The out-of-sample mean of the combined three-fund portfolio $\hat{\boldsymbol{w}}_c$ can be calculated as follows:

$$\begin{aligned}\mu_{\hat{\boldsymbol{w}}_c} &= \hat{\boldsymbol{w}}_c^\top \boldsymbol{\mu} = [\delta\hat{\boldsymbol{w}}_g + \beta\hat{\boldsymbol{w}}_m + (1-\delta\hat{\boldsymbol{w}}_g^\top \mathbf{1}_N - \beta\hat{\boldsymbol{w}}_m^\top \mathbf{1}_N)\boldsymbol{w}_e]^\top \boldsymbol{\mu} \\ &= \delta(\mu_{\hat{\boldsymbol{w}}_g} - \hat{s}_g\mu_N) + \beta(\mu_{\hat{\boldsymbol{w}}_m} - \hat{s}_m\mu_N) + \mu_N \\ &= \delta H_{11} + \beta H_{21} + \mu_N,\end{aligned} \tag{A.3}$$

as desired. The out-of-sample variance of the combined three-fund portfolio $\hat{\boldsymbol{w}}_c$ can be computed as follows:

$$\begin{aligned}\sigma^2_{\hat{\boldsymbol{w}}_c} &= [\delta\hat{\boldsymbol{w}}_g + \beta\hat{\boldsymbol{w}}_m + (1-\delta\hat{\boldsymbol{w}}_g^\top \mathbf{1}_N - \beta\hat{\boldsymbol{w}}_m^\top \mathbf{1}_N)\boldsymbol{w}_e]^\top \boldsymbol{\Sigma}[\delta\hat{\boldsymbol{w}}_g + \beta\hat{\boldsymbol{w}}_m + (1-\delta\hat{\boldsymbol{w}}_g^\top \mathbf{1}_N - \beta\hat{\boldsymbol{w}}_m^\top \mathbf{1}_N)\boldsymbol{w}_e] \\ &= [\delta(\hat{\boldsymbol{w}}_g - \hat{s}_g\boldsymbol{w}_e) + \beta(\hat{\boldsymbol{w}}_m - \hat{s}_m\boldsymbol{w}_e) + \boldsymbol{w}_e]^\top \boldsymbol{\Sigma}[\delta(\hat{\boldsymbol{w}}_g - \hat{s}_g\boldsymbol{w}_e) + \beta(\hat{\boldsymbol{w}}_m - \hat{s}_m\boldsymbol{w}_e) + \boldsymbol{w}_e] \\ &= \sigma_N^2 + \delta^2(\hat{\boldsymbol{w}}_g - \hat{s}_g\boldsymbol{w}_e)^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{w}}_g - \hat{s}_g\boldsymbol{w}_e) + \beta^2(\hat{\boldsymbol{w}}_m - \hat{s}_m\boldsymbol{w}_e)^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{w}}_m - \hat{s}_m\boldsymbol{w}_e) \\ &\quad + 2\delta\beta(\hat{\boldsymbol{w}}_g - \hat{s}_g\boldsymbol{w}_e)^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{w}}_m - \hat{s}_m\boldsymbol{w}_e) + 2\delta(\hat{\boldsymbol{w}}_g - \hat{s}_g\boldsymbol{w}_e)^\top\boldsymbol{\Sigma}\boldsymbol{w}_e + 2\beta(\hat{\boldsymbol{w}}_m - \hat{s}_m\boldsymbol{w}_e)^\top\boldsymbol{\Sigma}\boldsymbol{w}_e \\ &= \sigma_N^2 + \delta^2(\sigma^2_{\hat{\boldsymbol{w}}_g} + \hat{s}_g^2\sigma_N^2 - 2\hat{s}_g\sigma^2_{\hat{\boldsymbol{w}}_{g,e}}) + \beta^2(\sigma^2_{\hat{\boldsymbol{w}}_m} + \hat{s}_m^2\sigma_N^2 - 2\hat{s}_m\sigma^2_{\hat{\boldsymbol{w}}_{m,e}}) \\ &\quad + 2\delta\beta(\sigma^2_{\hat{\boldsymbol{w}}_{g,m}} + \hat{s}_g\hat{s}_m\sigma_N^2 - \hat{s}_g\sigma^2_{\hat{\boldsymbol{w}}_{m,e}} - \hat{s}_m\sigma^2_{\hat{\boldsymbol{w}}_{g,e}}) + 2\delta(\sigma^2_{\hat{\boldsymbol{w}}_{g,e}} - \hat{s}_g\sigma_N^2) + 2\beta(\sigma^2_{\hat{\boldsymbol{w}}_{m,e}} - \hat{s}_m\sigma_N^2) \\ &= \sigma_N^2 + \delta^2 A_1 + \beta^2 A_3 + 2\delta\beta A_2 + \delta H_{12} + \beta H_{22},\end{aligned} \tag{A.4}$$

as desired. □

A.1.3 Proof of Proposition 2.3

Recall that when asset returns follow the multivariate normal distribution assumption, the sample mean and sample covariance matrix are independent of each other and satisfy

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{R}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T), \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}})^\top \sim W_N(T-1, \boldsymbol{\Sigma})/T,$$

where $W_N(T-1, \boldsymbol{\Sigma})$ denotes a Wishart distribution with $T-1$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$. The following two lemmas present some properties of the sample mean and covariance matrix.

Lemma A.2. *The random variable* $\hat{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} \sim \chi^2_N(T\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})/T$*, where* $\chi^2_N(x)$ *denotes the noncentral chi-squared distribution with degree of freedom* N *and non-centrality parameter* x*, and therefore,* $\mathbb{E}[\hat{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}] = N/T + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$*.*

Proof. The proof essentially follows from Theorem 1.4.1 in Muirhead (1982). Because the random variable $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$, we can transform the components of $\hat{\boldsymbol{\mu}}$ to a set of independent normal variables. Specifically, we consider the Cholesky decomposition $\boldsymbol{\Sigma} = CC^\top$, where C is a nonsingular $N \times N$ matrix satisfying $(CC^\top)^{-1} = (C^{-1})^\top C^{-1}$. Then, we denote $U = \sqrt{T} C^{-1} \hat{\boldsymbol{\mu}}$ to have $U \sim N(\sqrt{T} C^{-1} \boldsymbol{\mu}, \boldsymbol{I}_N)$, where $\boldsymbol{I}_N$ is the $N \times N$-dimensional identity matrix with ones on the main diagonal and zeros elsewhere. Consequently, by the relation between independent normal random variables and chi-squired distribution, we obtain

$$T\hat{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}} = U^\top U \sim \chi^2_N(\zeta),$$

where the noncentral parameter $\zeta = T(C^{-1}\boldsymbol{\mu})^\top (C^{-1}\boldsymbol{\mu}) = T\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$. Finally, it is well known that the expectation of the noncentral chi-square distributed random variable is given by

$$\mathbb{E}[T\hat{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}] = N + \zeta = N + T\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}.$$

□

Lemma A.3. *Let* $\boldsymbol{W} = \boldsymbol{\Sigma}^{-\frac{1}{2}} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-\frac{1}{2}}$*. Then, we have* $\boldsymbol{W} \sim W_N(T-1, \boldsymbol{I}_N)/T$*, where* $\boldsymbol{I}_N$ *is the* $N \times N$ *identity matrix with ones on the main diagonal and zeros elsewhere. When* $T > N + 4$*, we denote* $c_1 = T/(T - N - 2)$ *and* $c_2 = T/(T - N - 1)(T - N - 4)$ *to have the following formulae:*

(a). $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}] = c_1 \boldsymbol{\Sigma}^{-1}$ *and* $\mathbb{E}[\boldsymbol{W}^{-1}] = c_1 \boldsymbol{I}_N$*,*

(b). $\mathbb{E}[\boldsymbol{W}^{-2}] = c_1 c_2 (T - 2) \boldsymbol{I}_N$*,*

(c). $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_N \mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}] = c_1 c_2 (\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_N) \boldsymbol{\Sigma}^{-1} + c_2 T \boldsymbol{\Sigma}^{-1} \mathbf{1}_N \mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1}$*,*

(d). $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_N \mathbf{1}_N^\top \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1}] = c_1 c_2 [(T - N - 3) \boldsymbol{\Sigma}^{-1} \mathbf{1}_N \mathbf{1}_N^\top + \mathbf{1}_N \mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1} + N \boldsymbol{\Sigma}^{-1}]$*.*

Proof. By Theorem 3.2.5 in Muirhead (1982), given the Wishart distributed matrix $\widehat{\boldsymbol{\Sigma}} \sim W_N(T-1, \boldsymbol{\Sigma})/T$, we have $\boldsymbol{W} = \boldsymbol{\Sigma}^{-\frac{1}{2}} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-\frac{1}{2}} \sim W_N(T-1, \boldsymbol{I}_N)/T$.

(a). The inverse of the Wishart distributed matrix $\widehat{\boldsymbol{\Sigma}}^{-1}$ follows the inverted Wishart distribution. The expectation of the inverted Wishart distributed matrix $\widehat{\boldsymbol{\Sigma}}^{-1}$ can be obtained by using Theorem 3.2.12 in Muirhead (1982).

(b). By Theorem 3.2 in Haff (1979), we have

$$\mathbb{E}[\boldsymbol{W}^{-2}] = c_1 c_2 N \boldsymbol{I}_N + T c_2 \boldsymbol{I}_N = c_2 \boldsymbol{I}_N (c_1 N + T) = c_1 c_2 (T-2) \boldsymbol{I}_N.$$

(c). By Theorem 3.3.18 in Gupta and Nagar (2018), we have

$$\begin{aligned}
&\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}] \\
=&c_1c_2[\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N^\top)\boldsymbol{\Sigma}^{-1}] + c_1c_2[(T-N-3)+1]\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\Sigma}^{-1} \\
=&c_1c_2(\mathbf{1}_N^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}_N)\boldsymbol{\Sigma}^{-1} + c_2T\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\Sigma}^{-1},
\end{aligned}$$

where the notation 'tr' denotes the trace of a matrix.

(d). By Theorem 3.3.18 in Gupta and Nagar (2018), we have

$$\begin{aligned}
&\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}] = \mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}\widehat{\boldsymbol{\Sigma}}^{-1}] \\
=&(T-N-3)c_1c_2\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} + c_1c_2[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}^\top\boldsymbol{\Sigma}^{-1} + \mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}^{-1}] \\
=&c_1c_2[(T-N-3)\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N^\top + \mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\Sigma}^{-1} + N\boldsymbol{\Sigma}^{-1}],
\end{aligned}$$

where $\boldsymbol{S} = \mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\Sigma}$, $\boldsymbol{S}^\top = \boldsymbol{\Sigma}\mathbf{1}_N\mathbf{1}_N^\top$ and $\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) = \mathrm{tr}(\mathbf{1}_N\mathbf{1}_N^\top) = N$.

□

Proof of Proposition 2.3. First of all, by part (a) of Lemma A.3 and the independence between $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, we can easily obtain:

$$\begin{aligned}
\mathbb{E}[H_{11}] &= \mathbb{E}[\mu_{\hat{\boldsymbol{w}}_g} - \hat{s}_g\mu_N] = \mathbb{E}[\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}] - \mu_N\mathbb{E}[\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N] = c_1(\theta_2^2 - \mu_N\theta_3^2), \\
\mathbb{E}[H_{21}] &= \mathbb{E}[\mu_{\hat{\boldsymbol{w}}_m} - \hat{s}_m\mu_N] = \mathbb{E}[\hat{\boldsymbol{\mu}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}] - \mu_N\mathbb{E}[\hat{\boldsymbol{\mu}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N] = c_1(\theta_1^2 - \mu_N\theta_2^2).
\end{aligned}$$

Furthermore, for computing the remaining quantities in the proposition, we apply the formulae in Lemmas A.2 and A.3 to obtain:

$$\begin{aligned}
\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_g}] &= \mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N] = \mathbb{E}[\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}^{-2}\boldsymbol{\Sigma}^{-1/2}\mathbf{1}_N] = c_1c_2(T-2)\theta_3^2,\\
\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_m}] &= \mathbb{E}[\hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}] = \mathbb{E}[\mathbb{E}[\hat{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}^{-2}\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}}]] = c_1c_2(T-2)\left(\frac{N}{T}+\theta_1^2\right),\\
\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{g,m}}] &= \mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}] = \mathbb{E}[\mathbb{E}[\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}^{-2}\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}}]] = c_1c_2(T-2)\theta_2^2,\\
\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{g,e}}] &= \mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}_N/N] = \frac{1}{N}\mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}_N] = c_1,\\
\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{m,e}}] &= \mathbb{E}[\hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}_N/N] = \frac{1}{N}\mathbb{E}[\hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}_N] = c_1\mu_N.
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{E}[\hat{s}_g] &= \mathbb{E}[\hat{\boldsymbol{w}}_g^\top \mathbf{1}_N] = \mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N] = c_1\theta_3^2,\\
\mathbb{E}[\hat{s}_m] &= \mathbb{E}[\hat{\boldsymbol{w}}_m^\top \mathbf{1}_N] = \mathbb{E}[\hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N] = c_1\theta_2^2,\\
\mathbb{E}[\hat{s}_g^2] &= \mathbb{E}[(\hat{\boldsymbol{w}}_g^\top \mathbf{1}_N)^2] = \mathbf{1}_N^\top\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}]\mathbf{1}_N = c_2(c_1+T)\theta_3^4,\\
\mathbb{E}[\hat{s}_m^2] &= \mathbb{E}[(\hat{\boldsymbol{w}}_m^\top \mathbf{1}_N)^2] = \mathbb{E}[\mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N|\widehat{\boldsymbol{\Sigma}}]]\\
&= \mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}\boldsymbol{\mu}^\top+\boldsymbol{\Sigma}/T)\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N] = c_2T\theta_2^4 + c_1c_2\theta_3^2\left(\theta_1^2+\frac{T-2}{T}\right),\\
\mathbb{E}[\hat{s}_g\hat{s}_m] &= \mathbb{E}[\hat{\boldsymbol{w}}_g^\top \mathbf{1}_N\hat{\boldsymbol{w}}_m^\top \mathbf{1}_N] = \mathbb{E}[\mathbb{E}[\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}}]] = c_2(c_1+T)\theta_2^2\theta_3^2,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\hat{s}_g\sigma^2_{\hat{\boldsymbol{w}}_{g,e}}] &= \frac{1}{N}\mathbf{1}_N^\top\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}]\mathbf{1}_N = c_1c_2(T-N-1)\theta_3^2,\\
\mathbb{E}[\hat{s}_g\sigma^2_{\hat{\boldsymbol{w}}_{m,e}}] &= \frac{1}{N}\mathbb{E}[\mathbb{E}[\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}|\hat{\boldsymbol{\mu}}]] = c_1c_2(T-N-3)\mu_N\theta_3^2 + 2c_1c_2\theta_2^2,\\
\mathbb{E}[\hat{s}_m\sigma^2_{\hat{\boldsymbol{w}}_{g,e}}] &= \frac{1}{N}\mathbb{E}[\mathbb{E}[\hat{\boldsymbol{\mu}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}_N|\hat{\boldsymbol{\mu}}]] = c_2T\theta_2^2 + c_1c_2\theta_3^2\mu_N,\\
\mathbb{E}[\hat{s}_m\sigma^2_{\hat{\boldsymbol{w}}_{m,e}}] &= \frac{1}{N}\mathbb{E}[\mathbb{E}[\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}_N|\widehat{\boldsymbol{\Sigma}}]] = \frac{1}{N}\mathbb{E}[\mathbf{1}_N^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}\boldsymbol{\mu}^\top+\boldsymbol{\Sigma}/T)\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{1}_N]\\
&= c_2T\theta_2^2\mu_N + c_1c_2\theta_1^2 + \frac{c_1c_2(T-2)}{T}.
\end{aligned}$$

Consequently, substituting the above expressions into the formulae given in Proposition 2.2 immediately yields the desired formulae for $\mathbb{E}[A_1]$, $\mathbb{E}[A_2]$, $\mathbb{E}[A_3]$, $\mathbb{E}[H_{12}]$, and $\mathbb{E}[H_{22}]$. □

A.1.4 Proof of Lemma 2.1

Proof of Lemma 2.1. We put $\boldsymbol{x} = [\delta, \beta]^\top$ and express the combined portfolio (2.7) as an affine function of $\boldsymbol{x}$:

$$\hat{\boldsymbol{w}}_c = \boldsymbol{Q}^\top \boldsymbol{x} + \boldsymbol{q},$$

where $\boldsymbol{Q} = [\hat{\boldsymbol{w}}_g - \hat{s}_g \boldsymbol{w}_e, \hat{\boldsymbol{w}}_m - \hat{s}_m \boldsymbol{w}_e]^\top$ and $\boldsymbol{q} = \boldsymbol{w}_e$, with $\hat{s}_g = \hat{\boldsymbol{w}}_g^\top \mathbf{1}_N$ and $\hat{s}_m = \hat{\boldsymbol{w}}_m^\top \mathbf{1}_N$ as previously defined. In view of (2.13), we can write the objective function $\tilde{G}(\delta, \beta)$ as a compound function:

$$\tilde{G}(\delta, \beta) = g(h(\boldsymbol{x})),$$

where

$$g(\boldsymbol{w}) = \mathbb{E}[\boldsymbol{w}^\top \boldsymbol{\mu}] - \gamma_1 \mathbb{E}[\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}] - \gamma_2 \sqrt{\mathbb{E}[\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}]}.$$

Given the property that the composition of a convex function and an affine function results in a convex function, it is sufficient for us to show that $g(\boldsymbol{w})$ is convex in $\boldsymbol{w}$ which could be a random weights vector.

Note that the first item ($\mathbb{E}[\boldsymbol{w}^\top \boldsymbol{\mu}]$ in the expression of $g(\boldsymbol{w})$ is a linear mapping of $\boldsymbol{w}$, and the second term $\mathbb{E}[\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}]$ is convex in $\boldsymbol{w}$ since $\boldsymbol{\Sigma}$ is positive semi-definite. Therefore, it remains to show that the third term $\sqrt{\mathbb{E}[\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}]}$ is also convex in $\boldsymbol{w}$. This must be true since the standard deviation is a convex functional. Alternatively, we verify its convexity directly by definition. We check, for any $t \in [0, 1]$ and two portfolios $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, whether it holds that

$$t\sqrt{\mathbb{E}[\boldsymbol{w}_1^\top \boldsymbol{\Sigma} \boldsymbol{w}_1]} + (1-t)\sqrt{\mathbb{E}[\boldsymbol{w}_2^\top \boldsymbol{\Sigma} \boldsymbol{w}_2]} \geq \sqrt{\mathbb{E}[(t\boldsymbol{w}_1 + (1-t)\boldsymbol{w}_2)^\top \boldsymbol{\Sigma}(t\boldsymbol{w}_1 + (1-t)\boldsymbol{w}_2)]},$$

which is equivalent to (by taking squares on both sides and canceling some items)

$$2t(1-t)\sqrt{\mathbb{E}[\boldsymbol{w}_1^\top \boldsymbol{\Sigma} \boldsymbol{w}_1]}\sqrt{\mathbb{E}[\boldsymbol{w}_2^\top \boldsymbol{\Sigma} \boldsymbol{w}_2]} \geq 2t(1-t)\mathbb{E}[\boldsymbol{w}_1^\top \boldsymbol{\Sigma} \boldsymbol{w}_2].$$

For $t = 0$ or 1, the above holds with an equality. Otherwise, we only need to ensure:

$$\sqrt{\mathbb{E}[\boldsymbol{w}_1^\top \boldsymbol{\Sigma} \boldsymbol{w}_1]}\sqrt{\mathbb{E}[\boldsymbol{w}_2^\top \boldsymbol{\Sigma} \boldsymbol{w}_2]} \geq \mathbb{E}[\boldsymbol{w}_1^\top \boldsymbol{\Sigma} \boldsymbol{w}_2], \tag{A.5}$$

which is obvious by the Cauchy-Schwartz inequality. Therefore, the function $g(\hat{\boldsymbol{w}})$ is a concave function of $\hat{\boldsymbol{w}}$ and the proof is complete.

□

A.1.5 Proof of Theorem 2.1

Proof of Theorem 2.1. The first-order condition with respect to (δ, β) gives

$$\begin{cases} \dfrac{\partial \tilde{G}(\delta,\beta)}{\partial \delta} = \dfrac{\partial \mathbb{E}[\mu_{\hat{\boldsymbol{w}}_c}]}{\partial \delta} - \left(\gamma_1 + \dfrac{\gamma_2}{2\sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]}}\right) \dfrac{\partial \mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]}{\partial \delta} = 0, \\ \dfrac{\partial \tilde{G}(\delta,\beta)}{\partial \beta} = \dfrac{\partial \mathbb{E}[\mu_{\hat{\boldsymbol{w}}_c}]}{\partial \beta} - \left(\gamma_1 + \dfrac{\gamma_2}{2\sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]}}\right) \dfrac{\partial \mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]}{\partial \beta} = 0. \end{cases} \tag{A.6}$$

Letting

$$\eta = \left(\gamma_1 + \frac{\gamma_2}{2\sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]}}\right), \tag{A.7}$$

and noticing the expressions of $\mathbb{E}[\mu_{\hat{\boldsymbol{w}}_c}]$ and $\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_c}]$ in (2.8), we can write the system of linear equations (A.6) into:

$$\begin{cases} \mathbb{E}[H_{11}] - \eta(2\delta\mathbb{E}[A_1] + \mathbb{E}[H_{12}] + 2\beta\mathbb{E}[A_2]) = 0, \\ \mathbb{E}[H_{21}] - \eta(2\beta\mathbb{E}[A_3] + \mathbb{E}[H_{22}] + 2\delta\mathbb{E}[A_2]) = 0. \end{cases}$$

The optimal combination coefficients are then given by

$$\delta_c^* = \frac{\mathbb{E}[H_1]\mathbb{E}[A_3] - \mathbb{E}[H_2]\mathbb{E}[A_2]}{2\eta(\mathbb{E}[A_1]\mathbb{E}[A_3] - \mathbb{E}[A_2]^2)}, \quad \beta_c^* = \frac{\mathbb{E}[H_1] - 2\mathbb{E}[A_1]\eta\delta_c^*}{2\eta\mathbb{E}[A_2]},$$

where $\mathbb{E}[H_1] = \mathbb{E}[H_{11}] - \eta\mathbb{E}[H_{12}]$ and $\mathbb{E}[H_2] = \mathbb{E}[H_{21}] - \eta\mathbb{E}[H_{22}]$, and positive constant $\eta \in (\gamma_1, \infty)$ satisfies (A.7), or equivalently (2.15). □

A.1.6 Proof of Theorem 2.2

Proof of Theorem 2.2.

1. The combined portfolio that consists of $\hat{\boldsymbol{w}}_g$ and 1/N rule is given by

$$\hat{\boldsymbol{w}}_{cg} = \delta\hat{\boldsymbol{w}}_g + (1 - \delta\hat{\boldsymbol{w}}_g^\top \mathbf{1}_N)\boldsymbol{w}_e.$$

The approximated expected out-of-sample MVS objective is

$$\tilde{G}(\delta, 0) = \mathbb{E}[H_{11}]\delta + \mu_N - \gamma_1(\sigma_N^2 + \delta^2\mathbb{E}[A_1] + \delta\mathbb{E}[H_{12}]) - \gamma_2\sqrt{\sigma_N^2 + \delta^2\mathbb{E}[A_1] + \delta\mathbb{E}[H_{12}]}.$$

Setting the first-order derivative with respect to the combination coefficient δ to zero yields

$$\frac{\partial\tilde{G}(\delta, 0)}{\partial\delta} = \frac{\partial\mathbb{E}[\mu_{\hat{\boldsymbol{w}}_{cg}}]}{\partial\delta} - \left(\gamma_1 + \frac{\gamma_2}{2\sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cg}}]}}\right)\frac{\partial\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cg}}]}{\partial\delta} = 0.$$

Letting $\eta_{cg} = \left(\gamma_1 + \frac{\gamma_2}{2\sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cg}}]}}\right)$ and noticing the expressions of $\mathbb{E}[\mu_{\hat{\boldsymbol{w}}_{cg}}]$ and $\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cg}}]$ as indicated in (2.8), we can write the above equation into:

$$\mathbb{E}[H_{11}] - \eta_{cg}(2\delta\mathbb{E}[A_1] + \mathbb{E}[H_{12}]) = 0.$$

The optimal combination coefficient is then given by

$$\delta^*_{cg} = \frac{\mathbb{E}[H_{11}] - \eta_{cg}\mathbb{E}[H_{12}]}{2\eta_{cg}\mathbb{E}[A_1]},$$

where $\eta_{cg} \in (\gamma_1, \infty)$ solves $2(\eta_{cg} - \gamma_1)\sqrt{\sigma_N^2 + (\delta^*_{cg})^2\mathbb{E}[A_1] + \delta^*_{cg}\mathbb{E}[H_{12}]} = \gamma_2$.

2. The combined portfolio with the ingredient portfolio $\hat{\boldsymbol{w}}_m$ and 1/N rule is given by

$$\hat{\boldsymbol{w}}_{cm} = \delta\hat{\boldsymbol{w}}_m + (1 - \delta\hat{\boldsymbol{w}}_m^\top \mathbf{1}_N)\boldsymbol{w}_e.$$

The approximated expected out-of-sample MVS objective is

$$\tilde{G}(0, \beta) = \mathbb{E}[H_{21}]\beta + \mu_N - \gamma_1(\sigma_N^2 + \beta^2\mathbb{E}[A_3] + \beta\mathbb{E}[H_{22}]) - \gamma_2\sqrt{\sigma_N^2 + \beta^2\mathbb{E}[A_3] + \beta\mathbb{E}[H_{22}]}.$$

The first-order condition is

$$\frac{\partial \tilde{G}(0,\beta)}{\partial \beta} = \frac{\partial \mathbb{E}[\mu_{\hat{\boldsymbol{w}}_{cm}}]}{\partial \beta} - \left(\gamma_1 + \frac{\gamma_2}{2\sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cm}}]}}\right)\frac{\partial \mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cm}}]}{\partial \beta} = 0.$$

Letting $\eta_{cm} = \left(\gamma_1 + \frac{\gamma_2}{2\sqrt{\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cm}}]}}\right)$, and noticing the expressions of $\mathbb{E}[\mu_{\hat{\boldsymbol{w}}_{cm}}]$ and $\mathbb{E}[\sigma^2_{\hat{\boldsymbol{w}}_{cm}}]$ as indicated in (2.8), we can write the above equation into:

$$\mathbb{E}[H_{21}] - \eta_{cm}(2\beta\mathbb{E}[A_3] + \mathbb{E}[H_{22}]) = 0.$$

The optimal combination coefficient is then given by

$$\beta^*_{cm} = \frac{\mathbb{E}[H_{21}] - \eta_{cm}\mathbb{E}[H_{22}]}{2\eta_{cm}\mathbb{E}[A_3]},$$

where $\eta_{cm} \in (\gamma_1, \infty)$ solves $2(\eta_{cm} - \gamma_1)\sqrt{\sigma^2_N + (\beta^*_{cm})^2\mathbb{E}[A_3] + \beta^*_{cm}\mathbb{E}[H_{22}]} = \gamma_2$.

□

A.1.7 Proof of Lemma 2.2

Proof of Lemma 2.2.

1. The proof can be found in the Appendix of Kan and Zhou (2007). We repeat it here for completeness. By Theorem 3.2.13 of Muirhead (1982), given $T\widehat{\boldsymbol{\Sigma}} \sim W_N(T-1, \boldsymbol{\Sigma})$ and $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ and they are independent, we have

$$\frac{\hat{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\hat{\boldsymbol{\mu}}^\top (T\widehat{\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{\mu}}} \sim \chi^2_{T-N},$$

where χ^2_{T-N} is the chi-square distribution with $T-N$ degree of freedom. Since $T\hat{\boldsymbol{\mu}}^\top\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} \sim \chi^2_N(T\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$, we have

$$\hat{\theta}^2_1 = \hat{\boldsymbol{\mu}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} \sim \left(\frac{N}{T-N}\right)F_{N,T-N}(T\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}),$$

where $F_{a,b}(c)$ represent a noncentral F distribution with a and b degrees of freedom and a noncentrality parameter c.

2. By Theorem 3.3.14 in Gupta and Nagar (2018), given $T\widehat{\boldsymbol{\Sigma}} \sim W_N(T-1, \boldsymbol{\Sigma})$, we have

$$\frac{\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_N}{\mathbf{1}_N^\top (T\widehat{\boldsymbol{\Sigma}})^{-1} \mathbf{1}_N} \sim \chi^2_{T-N}.$$

   Hence, the sample estimator $\hat{\theta}_3^2$ follows a inverse chi-squared distribution as

$$\frac{\hat{\theta}_3^2}{T} = \mathbf{1}_N^\top (T\widehat{\boldsymbol{\Sigma}})^{-1} \mathbf{1}_N \sim (\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_N) \text{Inv-}\chi^2_{T-N},$$

   where Inv-χ^2_{T-N} is the inverse chi-square distribution with $T-N$ degrees of freedom.

3. By Theorem 3.3.12 in Gupta and Nagar (2018), given $T\widehat{\boldsymbol{\Sigma}} \sim W_N(T-1, \boldsymbol{\Sigma})$, we have

$$\frac{\mathbf{1}_N^\top (T\widehat{\boldsymbol{\Sigma}}) \mathbf{1}_N}{\mathbf{1}_N^\top \boldsymbol{\Sigma} \mathbf{1}_N} \sim \chi^2_{T-1}.$$

   Hence, the sample estimator $\hat{\sigma}_N^2$ follows a chi-squared distribution as

$$TN^2 \hat{\sigma}_N^2 = \mathbf{1}_N^\top (T\widehat{\boldsymbol{\Sigma}}) \mathbf{1}_N \sim (\mathbf{1}_N^\top \boldsymbol{\Sigma} \mathbf{1}_N) \chi^2_{T-1},$$

   where χ^2_{T-1} is the chi-square distribution with $T-1$ degree of freedom.

4. By Corollary 1 in Bodnar and Okhrin (2011), given $T\widehat{\boldsymbol{\Sigma}} \sim W_N(T-1, \boldsymbol{\Sigma})$ and $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$, we have

$$\hat{\theta}_2^2/T = \hat{\boldsymbol{\mu}}^\top (T\widehat{\boldsymbol{\Sigma}})^{-1} \mathbf{1}_N \stackrel{d}{=} \frac{1}{u_1} \left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_N + \sqrt{\left(\frac{1}{T} + \frac{N-1}{T(T-N+1)} u_3 \right) (\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_N) u_2} \right),$$

   where $u_1 \sim \chi^2_{T-N}$, $u_2 \sim N(0,1)$ and

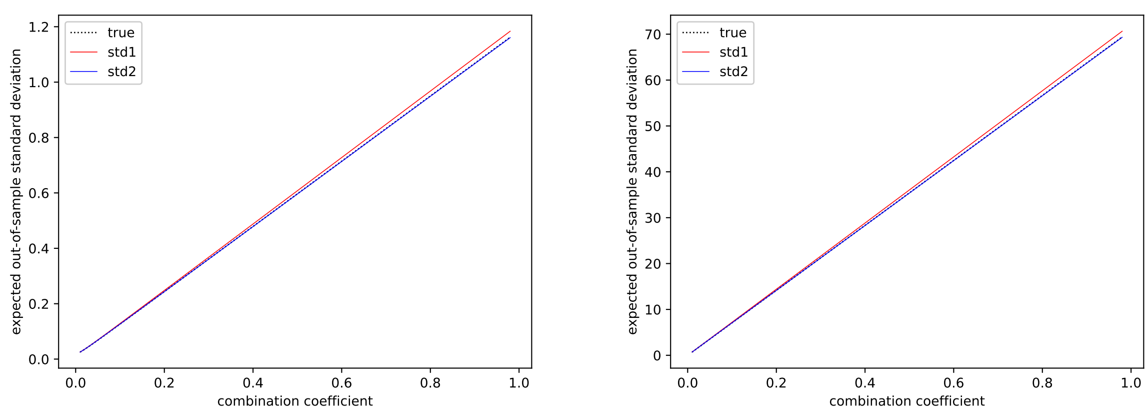
$$u_3 \sim F_{(N-1)/2,(T-N+1)/2}(T\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - T(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_N)^2 / \mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_N).$$

□

A.2 Comparison between zero-order and first-order Taylor series expansions

Figure A.1 compares the first-order and second-order Taylor series expansions with the corresponding true value of the expected out-of-sample standard deviation for the combined market portfolio in Figure A.1a and the combined global minimum variance portfolio in Figure A.1b.

Figure A.1: The approximations of expected out-of-sample standard deviation.



(a) combined market portfolio **(b)** combined global minimum variance portfolio

This figure depicts the approximations of the expected out-of-sample standard deviation for the combined market portfolio and the combined global minimum variance portfolio. The horizontal axis represents the combination coefficient that appeared in the corresponding combined portfolio. The true expected out-of-sample standard deviation (labeled as 'true') is computed numerically based on 250,000 sets of $T \times N$ asset returns. The first-order Taylor series expansion (labeled as 'std1') is obtained based on the analytic form of the expected out-of-sample variance, while the second-order Taylor series expansion (labeled as 'std2') is computed by both the analytic form of the expected out-of-sample variance and the numerical estimator of the variance of the out-of-sample variance.

We fix the number of risky assets at $N = 25$ and employ the linear factor model for asset returns with the same parameter settings as detailed in Section 2.5.1. Similar to Lassance (2021), the true expected out-of-sample standard deviation is depicted numerically based on 250,000 sets of asset return vectors. Specifically, we simulate 250,000 sets of $T \times N$

asset returns from the five-factor model specified in (2.23), where the estimation window is fixed at $T = 120$. For each set of $T \times N$ simulated asset returns, we first obtain the implementable combined portfolio (either $\hat{\boldsymbol{w}}_{cm}$ or $\hat{\boldsymbol{w}}_{cg}$) and evaluate the out-of-sample standard deviation by definition. Then we take an average of all out-of-sample standard deviations as the estimate of the true expected out-of-sample standard deviation. The first-order Taylor series expansion is calculated from the analytic form of the expected out-of-sample variance (presented in Section 2.3.2), while the second-order Taylor series expansion is computed by both the analytic form of the expected out-of-sample variance and the numerical approximation of the variance of the out-of-sample variance. We observe that the second-order Taylor series expansion is closer to the true value but the difference between the first-order and second-order expansions is small.

Appendix B

Appendix for Chapter 3

B.1 Proofs

B.1.1 Proof of Theorem 3.1

Proof of Theorem 3.1. Define the Lagrangian

$$\mathcal{L}(\boldsymbol{w}, \lambda_1, \lambda_2) = \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} + \lambda_1(\boldsymbol{w}^\top \boldsymbol{S} - \bar{s}) + \lambda_2(\boldsymbol{w}^\top \mathbf{1} - 1)$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ are the Lagrange multipliers. Using the first-order optimality condition, we have

$$\begin{cases} \dfrac{\partial \mathcal{L}(\boldsymbol{w}, \lambda_1, \lambda_2)}{\partial \boldsymbol{w}} = \boldsymbol{\mu} - \gamma \boldsymbol{\Sigma} \boldsymbol{w} + \lambda_1 \boldsymbol{S} + \lambda_2 \mathbf{1} = \mathbf{0}, \\ \dfrac{\partial \mathcal{L}(\boldsymbol{w}, \lambda_1, \lambda_2)}{\partial \lambda_1} = \boldsymbol{w}^\top \boldsymbol{S} - \bar{s} = 0, \\ \dfrac{\partial \mathcal{L}(\boldsymbol{w}, \lambda_1, \lambda_2)}{\partial \lambda_2} = \boldsymbol{w}^\top \mathbf{1} - 1 = 0. \end{cases}$$

Therefore, we have the expression for the optimal portfolio weight as

$$\boldsymbol{w}^*_{esg} = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} + \lambda_1 \boldsymbol{S} + \lambda_2 \mathbf{1}).$$

To find the two Lagrange multipliers λ_1 and λ_2, we have

$$\begin{cases} \dfrac{1}{\gamma}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} + \lambda_1 \boldsymbol{S} + \lambda_2 \mathbf{1}) = \bar{s}, \\ \dfrac{1}{\gamma}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} + \lambda_1 \boldsymbol{S} + \lambda_2 \mathbf{1}) = 1. \end{cases}$$

When $\mathbf{1}^\top \boldsymbol{\Omega} \boldsymbol{S} \neq 0$ with $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}$, it gives that

$$\begin{aligned} \lambda_1 &= \frac{\gamma \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \gamma \bar{s}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1} + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}} \\ &= \frac{\gamma \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{S} - \bar{s}\mathbf{1}) + \boldsymbol{\mu}^\top \boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}}, \\ \lambda_2 &= \frac{\gamma \bar{s}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1} - \gamma \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}} \\ &= \frac{\gamma \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}(\bar{s}\mathbf{1} - \boldsymbol{S}) - \boldsymbol{\mu}^\top \boldsymbol{\Omega}\boldsymbol{S}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}}. \end{aligned}$$

Denote $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}$ with $\boldsymbol{Z} = \boldsymbol{S}\mathbf{1}^\top - \mathbf{1}\boldsymbol{S}^\top$. The matrix $\boldsymbol{Z}$ is skew-symmetric, and it satisfies that $\boldsymbol{Z}^\top = -\boldsymbol{Z}$. In a skew-symmetric matrix, the diagonal elements are zero, resulting in a trace of zero for $\boldsymbol{Z}$. Additionally, $\boldsymbol{Z}^2$ is a symmetric and negative semi-definite matrix. We can also establish the following relationships:

$$\mathbf{1}^\top \boldsymbol{\Omega}\mathbf{1} = 0, \qquad \text{and} \qquad \boldsymbol{S}^\top \boldsymbol{\Omega}\boldsymbol{S} = 0.$$

We substitute the values of the two Lagrange multipliers, λ_1 and λ_2, into the optimal portfolio weight, which gives that

$$\begin{aligned} \boldsymbol{w}_{esg}^* &= \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\mu} + \left(\frac{\gamma \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{S} - \bar{s}\mathbf{1}) + \boldsymbol{\mu}^\top \boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}}\right)\boldsymbol{S} + \left(\frac{\gamma \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}(\bar{s}\mathbf{1} - \boldsymbol{S}) - \boldsymbol{\mu}^\top \boldsymbol{\Omega}\boldsymbol{S}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}}\right)\mathbf{1}\right] \\ &= \bar{s}\frac{-\boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}} + \frac{\boldsymbol{\Omega}\boldsymbol{S}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}} + \frac{1}{\gamma}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}}\right) \\ &= \bar{s}\boldsymbol{w}_e + \boldsymbol{w}_a + \frac{1}{\gamma}\boldsymbol{w}_o, \end{aligned}$$

where we have

$$\boldsymbol{w}_a = \frac{\boldsymbol{\Omega S}}{\mathbf{1}^\top \boldsymbol{\Omega S}}, \qquad \boldsymbol{w}_e = -\frac{\boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega S}},$$
$$\boldsymbol{w}_o = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega \mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega \mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega S}}.$$

For the constituent portfolio $\boldsymbol{w}_o$, we have that

$$\begin{aligned}
&\sigma^2_{\boldsymbol{w}_o} = \boldsymbol{w}_o^\top \boldsymbol{\Sigma} \boldsymbol{w}_o \\
&= \left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega \mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega \mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega S}}\right]^\top \boldsymbol{\Sigma} \left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega \mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega \mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega S}}\right] \\
&= \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega \mu})\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega \mu})\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega S}} = \mu_{\boldsymbol{w}_o} = \boldsymbol{w}_o^\top \boldsymbol{\mu}.
\end{aligned}$$

□

B.1.2 Proof of Corollary 3.1

Proof of Corollary 3.1. The mean-variance utility from holding the optimal portfolio $\boldsymbol{w}^*_{esg}$ is

$$\begin{aligned}
U^*_{esg} &= (\boldsymbol{w}^*_{esg})^\top \boldsymbol{\mu} - \frac{\gamma}{2}(\boldsymbol{w}^*_{esg})^\top \boldsymbol{\Sigma} \boldsymbol{w}^*_{esg} \\
&= \bar{s}\boldsymbol{w}_e^\top \boldsymbol{\mu} + \boldsymbol{w}_a^\top \boldsymbol{\mu} + \frac{1}{\gamma}\boldsymbol{w}_o^\top \boldsymbol{\mu} - \frac{\gamma}{2}\left(\bar{s}\boldsymbol{w}_e + \boldsymbol{w}_a + \frac{1}{\gamma}\boldsymbol{w}_o\right)^\top \boldsymbol{\Sigma}\left(\bar{s}\boldsymbol{w}_e + \boldsymbol{w}_a + \frac{1}{\gamma}\boldsymbol{w}_o\right) \\
&= \bar{s}\mu_{\boldsymbol{w}_e} + \mu_{\boldsymbol{w}_a} + \frac{1}{\gamma}\mu_{\boldsymbol{w}_o} - \frac{\gamma}{2}\left(\bar{s}^2\sigma^2_{\boldsymbol{w}_e} + 2\bar{s}\sigma^2_{\boldsymbol{w}_{a,e}} + \frac{2\bar{s}}{\gamma}\sigma^2_{\boldsymbol{w}_{e,o}} + \sigma^2_{\boldsymbol{w}_a} + \frac{2}{\gamma}\sigma^2_{\boldsymbol{w}_{a,o}} + \frac{1}{\gamma^2}\sigma^2_{\boldsymbol{w}_o}\right) \\
&= \bar{s}\mu_{\boldsymbol{w}_e} + \mu_{\boldsymbol{w}_a} + \frac{1}{\gamma}\mu_{\boldsymbol{w}_o} - \frac{\gamma}{2}\left(\bar{s}^2\sigma^2_{\boldsymbol{w}_e} + 2\bar{s}\sigma^2_{\boldsymbol{w}_{a,e}} + \sigma^2_{\boldsymbol{w}_a} + \frac{1}{\gamma^2}\mu_{\boldsymbol{w}_o}\right) \\
&= \bar{s}\mu_{\boldsymbol{w}_e} + \mu_{\boldsymbol{w}_a} - \frac{\gamma}{2}\left(\bar{s}^2\sigma^2_{\boldsymbol{w}_e} + 2\bar{s}\sigma^2_{\boldsymbol{w}_{a,e}} + \sigma^2_{\boldsymbol{w}_a}\right) + \frac{\mu_{\boldsymbol{w}_o}}{2\gamma},
\end{aligned}$$

where we use $\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{B} = -\boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{A}$ for any vectors $\boldsymbol{A}$ and $\boldsymbol{B}$, as well as $\mathbf{1}^\top\boldsymbol{\Omega}\mathbf{1} = 0$ and $\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{S} = 0$ to obtain that

$$\begin{aligned}
\sigma^2_{\boldsymbol{w}_{e,o}} &= \boldsymbol{w}_o^\top\boldsymbol{\Sigma}\boldsymbol{w}_e = -\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S}}\right]^\top \boldsymbol{\Sigma}\left(\frac{\boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S}}\right) \\
&= -\left[\frac{\boldsymbol{\mu}^\top\boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S}} - \frac{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{\mu})(\boldsymbol{S}^\top\boldsymbol{\Omega}\mathbf{1})}{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S})^2} + \frac{(\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{\mu})(\mathbf{1}^\top\boldsymbol{\Omega}\mathbf{1})}{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S})^2}\right] = 0, \\
\sigma^2_{\boldsymbol{w}_{a,o}} &= \boldsymbol{w}_o^\top\boldsymbol{\Sigma}\boldsymbol{w}_a = \left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S}}\right]^\top \boldsymbol{\Sigma}\left(\frac{\boldsymbol{\Omega}\boldsymbol{S}}{\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S}}\right) \\
&= \left[\frac{\boldsymbol{\mu}^\top\boldsymbol{\Omega}\boldsymbol{S}}{\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S}} - \frac{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{\mu})(\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{S})}{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S})^2} + \frac{(\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{\mu})(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S})}{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S})^2}\right] = 0.
\end{aligned}$$

□

B.1.3 Proof of Proposition 3.1

Proof of Proposition 3.1.

1. Let the utility U^*_{esg} be a function of $\bar{s}$. That is,

$$f(\bar{s}) = \bar{s}\mu_{\boldsymbol{w}_e} + \mu_{\boldsymbol{w}_a} - \frac{\gamma}{2}\left(\bar{s}^2\sigma^2_{\boldsymbol{w}_e} + 2\bar{s}\sigma^2_{\boldsymbol{w}_{a,e}} + \sigma^2_{\boldsymbol{w}_a}\right) + \frac{\mu_{\boldsymbol{w}_o}}{2\gamma}.$$

   This function is quadratic in terms of $\bar{s}$, with a negative coefficient for $\bar{s}^2$. Take the first-order derivative with respect to $\bar{s}$ and we have

$$\mu_{\boldsymbol{w}_e} - \gamma\sigma^2_{\boldsymbol{w}_e}\bar{s} - \gamma\sigma^2_{\boldsymbol{w}_{a,e}} = 0.$$

   Solving for $\bar{s}$, we find that

$$\bar{s}^* = \frac{\mu_{\boldsymbol{w}_e} - \gamma\sigma^2_{\boldsymbol{w}_{a,e}}}{\gamma\sigma^2_{\boldsymbol{w}_e}}.$$

   Hence, when $\bar{s} = \bar{s}^*$, the utility U^*_{esg} reaches its maximum value.

2. The target ESG score $\bar{s}^*$ can be expressed as

$$\begin{aligned}\bar{s}^* &= \frac{\boldsymbol{\mu}^\top \boldsymbol{\Omega} \mathbf{1}\mathbf{1}^\top \boldsymbol{\Omega} \boldsymbol{S} + \gamma \mathbf{1} \boldsymbol{\Omega}(\boldsymbol{S}\mathbf{1}^\top - \mathbf{1}\boldsymbol{S}^\top)\boldsymbol{\Sigma}^{-1}\boldsymbol{S}}{\gamma \mathbf{1}^\top \boldsymbol{\Omega}(\boldsymbol{S}\mathbf{1}^\top - \mathbf{1}\boldsymbol{S}^\top)\boldsymbol{\Sigma}^{-1}\mathbf{1}} \\ &= \frac{\boldsymbol{\mu}^\top \boldsymbol{\Omega}\mathbf{1}}{\gamma \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} + \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}},\end{aligned}$$

where $\boldsymbol{\Omega\Sigma\Omega} = \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega}(\boldsymbol{S}\mathbf{1}^\top - \mathbf{1}\boldsymbol{S}^\top)\boldsymbol{\Sigma}^{-1}$ and $\mathbf{1}^\top\boldsymbol{\Omega}\mathbf{1} = 0$. We plug this value into the optimal portfolio $\boldsymbol{w}^*_{esg}$. It gives that

$$\begin{aligned}\boldsymbol{w}^{**}_{esg} =& - \bar{s}^* \frac{\boldsymbol{\Omega}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}} + \frac{\boldsymbol{\Omega}\boldsymbol{S}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}} + \frac{1}{\gamma}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S}}\right] \\ =& \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{\boldsymbol{\mu}^\top\boldsymbol{\Omega}\mathbf{1}\boldsymbol{\Omega}\mathbf{1} + (\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})\left[(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top\boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{1}\right]}{\gamma(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S})} \\ &+ \frac{(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})\boldsymbol{\Omega}\boldsymbol{S} - (\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{S})\boldsymbol{\Omega}\mathbf{1}}{(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})(\mathbf{1}^\top\boldsymbol{\Omega}\boldsymbol{S})} \\ =& \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\mu} - \frac{\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}\mathbf{1}\right].\end{aligned}$$

This optimal portfolio $\boldsymbol{w}^{**}_{esg}$ obtained under the target ESG score $\bar{s}^{**}$ coincides precisely with the solution derived from the following traditional mean-variance optimization problem:

$$\begin{aligned}\max_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^\top\mathbf{1} = 1.\end{aligned}$$

□

B.1.4 Proof of Proposition 3.2

Proof of Proposition 3.2.

1. Define the Lagrangian

$$\mathcal{L}_1(\boldsymbol{w}, \lambda_1, \lambda_2) = \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w} + \lambda_1(\boldsymbol{w}^\top\boldsymbol{S} - 1) + \lambda_2\boldsymbol{w}^\top\mathbf{1}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ are the Lagrange multipliers. Using the first-order optimality condition, we have

$$\begin{cases} \dfrac{\partial \mathcal{L}_1(\boldsymbol{w}, \lambda_1, \lambda_2)}{\partial \boldsymbol{w}} = \boldsymbol{\Sigma w} + \lambda_1 \boldsymbol{S} + \lambda_2 \mathbf{1} = \mathbf{0}, \\ \dfrac{\partial \mathcal{L}_1(\boldsymbol{w}, \lambda_1, \lambda_2)}{\partial \lambda_1} = \boldsymbol{w}^\top \boldsymbol{S} - 1 = 0. \\ \dfrac{\partial \mathcal{L}_1(\boldsymbol{w}, \lambda_1, \lambda_2)}{\partial \lambda_2} = \boldsymbol{w}^\top \mathbf{1} = 0. \end{cases}$$

Therefore, the optimal portfolio weight has the expression as

$$\boldsymbol{w}^* = -\boldsymbol{\Sigma}^{-1}(\lambda_1 \boldsymbol{S} + \lambda_2 \mathbf{1}). \tag{B.1}$$

To find the two Lagrange multipliers, we have

$$\begin{cases} \lambda_1 \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S} + \lambda_2 \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} = -1 \\ \lambda_1 \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S} + \lambda_2 \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} = 0 \end{cases}$$

It gives that

$$\lambda_1 = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S})^2 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S}}$$
$$\lambda_2 = \frac{-\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S}}{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S})^2 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S}}$$

We substitute the values of the two Lagrange multipliers, λ_1 and λ_2, into the optimal portfolio weight, which gives that

$$\boldsymbol{w}^* = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S} \boldsymbol{\Sigma}^{-1} \mathbf{1} - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} \boldsymbol{\Sigma}^{-1} \boldsymbol{S}}{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S})^2 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1} \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{S}} = \boldsymbol{w}_e$$

2. Define the Lagrangian

$$\mathcal{L}_2(\boldsymbol{w}, \lambda_1, \lambda_2) = \boldsymbol{w}^\top \boldsymbol{\Sigma w} + \lambda_1 \boldsymbol{w}^\top \boldsymbol{S} + \lambda_2 (\boldsymbol{w}^\top \mathbf{1} - 1)$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ are the Lagrange multipliers. Similarly, using the first-order optimality condition, the optimal portfolio weight has the expression as (B.1). The two

Lagrange multipliers satisfy

$$\begin{cases} \lambda_1 \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} + \lambda_2 \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1} = 0 \\ \lambda_1 \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} + \lambda_2 \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1} = -1 \end{cases}$$

It gives that

$$\lambda_1 = \frac{-\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}}{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S})^2 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}}$$
$$\lambda_2 = \frac{\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}}{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S})^2 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}}$$

Therefore, the optimal portfolio is

$$\boldsymbol{w}^* = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\mathbf{1}}{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S})^2 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}} = \boldsymbol{w}_a$$

3. For the zero-weight-ESG portfolio $\boldsymbol{w}_o$, it is straightforward to calculate that

$$\mathbf{1}^\top \boldsymbol{w}_o = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}} = 0,$$
$$\boldsymbol{S}^\top \boldsymbol{w}_o = \boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - (\boldsymbol{S}^\top \boldsymbol{\Omega}\boldsymbol{\mu})\boldsymbol{S}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Omega}\boldsymbol{S}} = 0.$$

□

B.1.5 Proof of Proposition 3.3

Lemma B.1. *For the covariance matrix $\boldsymbol{\Sigma}$, we have the following relationships for the Hadamard product (element-wise product):*

1. $\mathbf{1}^\top (\boldsymbol{\Sigma}^{-1} \circ \boldsymbol{\Sigma})\mathbf{1} = N$
2. $\mathbf{1}^\top \left(\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} \circ \boldsymbol{\Sigma}\right)\mathbf{1} = 1$

Proof. Denote $\det(\boldsymbol{\Sigma})$ and $\boldsymbol{C}$ as the determinant of $\boldsymbol{\Sigma}$ and the cofactor matrix formed by all of the cofactors of $\boldsymbol{\Sigma}$, respectively. The inverse of $\boldsymbol{\Sigma}$ satisfies that

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\det(\boldsymbol{\Sigma})}\boldsymbol{C}^{\top}.$$

Meanwhile, $\det(\boldsymbol{\Sigma})$ can be expanded in terms of its cofactors along any row or column:

$$\det(\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \boldsymbol{\Sigma}_{ij}\boldsymbol{C}_{ij} \quad \text{along the } j^{\text{th}} \text{ column, or} \quad \sum_{j=1}^{N} \boldsymbol{\Sigma}_{ij}\boldsymbol{C}_{ij} \quad \text{along the } i^{\text{th}} \text{ row.}$$

1. The $(i,j)^{\text{th}}$ element in $\boldsymbol{\Sigma}^{-1} \circ \boldsymbol{\Sigma}$ satisfies that $(\boldsymbol{\Sigma}^{-1} \circ \boldsymbol{\Sigma})_{ij} = \frac{1}{\det(\boldsymbol{\Sigma})}\boldsymbol{C}_{ji}\boldsymbol{\Sigma}_{ij}$. Therefore, for the symmetric matrix $\boldsymbol{\Sigma}$ with $\boldsymbol{C}^{\top} = \boldsymbol{C}$, we have

$$\mathbf{1}^{\top}(\boldsymbol{\Sigma}^{-1} \circ \boldsymbol{\Sigma})\mathbf{1} = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{1}{\det(\boldsymbol{\Sigma})}\boldsymbol{C}_{ji}\boldsymbol{\Sigma}_{ij} = N.$$

2. Denote $u_i = \sum_{k=1}^{N} C_{ik}$ as the sum of the i^{th} row of $\boldsymbol{C}$. Hence, the $(i,j)^{\text{th}}$ element in $\boldsymbol{C}^{\top}\mathbf{1}\mathbf{1}^{\top}\boldsymbol{C}^{\top}$ is obtained by $(\boldsymbol{C}^{\top}\mathbf{1}\mathbf{1}^{\top}\boldsymbol{C}^{\top})_{ij} = u_i u_j$. Meanwhile, the sum of all elements of $\boldsymbol{C}$ satisfies that $\mathbf{1}^{\top}\boldsymbol{C}\mathbf{1} = \sum_{i=1}^{N} u_i$. Hence, we have

$$\mathbf{1}^{\top}\left(\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^{\top}\boldsymbol{\Sigma}^{-1}}{\mathbf{1}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}} \circ \boldsymbol{\Sigma}\right)\mathbf{1} = \frac{1}{\det(\boldsymbol{\Sigma})}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{u_i u_j}{\sum_{i=1}^{N} u_i}\boldsymbol{\Sigma}_{ij} = 1$$

□

Proof of Proposition 3.3.

1. We first calculate the explicit forms for the expectations $\mathbb{E}[\boldsymbol{\mu}^{\top}\hat{\boldsymbol{w}}_c]$ and $\mathbb{E}[\boldsymbol{S}^{\top}\hat{\boldsymbol{w}}_c]$, which are extended from Theorem 1 in Okhrin and Schmid (2006). For ease of reference, we provide detailed calculations below.

   Let $\hat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\Sigma}}^{-1} - \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\mathbf{1}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}/\mathbf{1}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$. The combined portfolio can be written as

$$\hat{\boldsymbol{w}}_c = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}} + \delta\hat{\boldsymbol{\Upsilon}}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right).$$

Conditional on $\hat{\boldsymbol{\mu}}$, we denote the i^{th} component of the combined portfolio as

$$\hat{w}_i = \hat{w}_{c,i}|\hat{\boldsymbol{\mu}} = \frac{\boldsymbol{e}_i^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}} + \delta \boldsymbol{e}_i^\top \hat{\boldsymbol{\Upsilon}} \left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right),$$

where $\boldsymbol{e}_i$ is an N-dimensional vector whose i^{th} element is 1 and other elements are 0. Let $\boldsymbol{M} = \left[\boldsymbol{e}_i, \hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}, \mathbf{1}\right]$ be an $N \times 3$ matrix and $\boldsymbol{M}_1 = \left[\boldsymbol{e}_i, \hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right]$ be the first two columns. Denote that

$$\hat{\boldsymbol{G}} = \boldsymbol{M}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{M} = \begin{bmatrix} \hat{\boldsymbol{G}}_{11} & \hat{\boldsymbol{G}}_{12} \\ \hat{\boldsymbol{G}}_{21} & \hat{\boldsymbol{G}}_{22} \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{H}} = (\hat{\boldsymbol{G}})^{-1} = \begin{bmatrix} \hat{\boldsymbol{H}}_{11} & \hat{\boldsymbol{H}}_{12} \\ \hat{\boldsymbol{H}}_{21} & \hat{\boldsymbol{H}}_{22} \end{bmatrix}, \tag{B.2}$$

where $\hat{\boldsymbol{G}}_{11}$ is a 2×2 matrix.

We aim to obtain $\hat{\boldsymbol{w}}_i$ using the elements from $\hat{\boldsymbol{H}}$. To achieve this, we utilize the following relationships:

$$\begin{aligned} \hat{\boldsymbol{G}}_{22} &= \mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1} = (\hat{\boldsymbol{H}}_{22} - \hat{\boldsymbol{H}}_{21}\hat{\boldsymbol{H}}_{11}^{-1}\hat{\boldsymbol{H}}_{12})^{-1}, \\ \hat{\boldsymbol{G}}_{12} &= -\hat{\boldsymbol{H}}_{11}^{-1}\hat{\boldsymbol{H}}_{12}(\hat{\boldsymbol{H}}_{22} - \hat{\boldsymbol{H}}_{21}\hat{\boldsymbol{H}}_{11}^{-1}\hat{\boldsymbol{H}}_{12})^{-1}, \\ \hat{\boldsymbol{H}}_{11}^{-1} &= \hat{\boldsymbol{G}}_{11} - \hat{\boldsymbol{G}}_{12}\hat{\boldsymbol{G}}_{22}^{-1}\hat{\boldsymbol{G}}_{21} = \boldsymbol{M}_1^\top \hat{\boldsymbol{\Upsilon}} \boldsymbol{M}_1. \end{aligned}$$

Firstly, the first element of $-\hat{\boldsymbol{H}}_{11}^{-1}\hat{\boldsymbol{H}}_{12}$ satisfies that

$$(-\hat{\boldsymbol{H}}_{11}^{-1}\hat{\boldsymbol{H}}_{12})_1 = \left(\frac{\hat{\boldsymbol{G}}_{12}}{\hat{\boldsymbol{G}}_{22}}\right)_1 = \frac{\boldsymbol{e}_i^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}. \tag{B.3}$$

Secondly, the first element of $\hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta}$ where $\boldsymbol{\delta} = [0, \delta]^\top$ satisfies that

$$(\hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta})_1 = \delta \boldsymbol{e}_i^\top \hat{\boldsymbol{\Upsilon}} \left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right). \tag{B.4}$$

Therefore, $\hat{w}_i$ can be constructed by the first element of $\boldsymbol{z} = -\hat{\boldsymbol{H}}_{11}^{-1}(\hat{\boldsymbol{H}}_{12} - \boldsymbol{\delta})$.

Denote $\boldsymbol{G}$ and $\boldsymbol{H}$ by substituting $\widehat{\boldsymbol{\Sigma}}$ in $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{H}}$ with $\boldsymbol{\Sigma}$. Since $T\widehat{\boldsymbol{\Sigma}} \sim \mathcal{W}_N(T-1, \boldsymbol{\Sigma})$ and conditional on $\hat{\boldsymbol{\mu}}$, according to Corollary 3.2.6, Theorems 3.2.10 and 3.2.11 of

Muirhead (1982), we have

$$\begin{aligned} T\hat{\boldsymbol{H}} &\sim \mathcal{W}_3(T-N+2, \boldsymbol{H}), \\ T\hat{\boldsymbol{H}}_{11} &\sim \mathcal{W}_2(T-N+2, \boldsymbol{H}_{11}), \\ T\hat{\boldsymbol{H}}_{12}|T\hat{\boldsymbol{H}}_{11} &\sim N_2(T\hat{\boldsymbol{H}}_{11}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}, (\boldsymbol{H}_{22}-\boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12})T\hat{\boldsymbol{H}}_{11}). \end{aligned} \tag{B.5}$$

Therefore, using the tower rule of the expectation, we obtain that the expectation of $\boldsymbol{z}$ as

$$\begin{aligned} \mathbb{E}[\boldsymbol{z}] &= \mathbb{E}[\mathbb{E}[-\hat{\boldsymbol{H}}_{11}^{-1}\hat{\boldsymbol{H}}_{12} + \hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta})|\hat{\boldsymbol{H}}_{11}]] = \mathbb{E}[-\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12} + \hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta}] \\ &= -\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12} + k_1\boldsymbol{H}_{11}^{-1}\boldsymbol{\delta}, \end{aligned}$$

where $k_1 = T/(T-N-1)$. Correspondingly, the first element of the expectation of $\boldsymbol{z}$ is the expectation of $\hat{w}_i$. That is,

$$\mathbb{E}[\hat{w}_i] = \frac{\boldsymbol{e}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} + k_1\delta\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right). \tag{B.6}$$

Since we have $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ and $\hat{\boldsymbol{\mu}}$ is independent with $\widehat{\boldsymbol{\Sigma}}$, the unconditional expectation of the portfolio weight is

$$\mathbb{E}[\hat{\boldsymbol{w}}_c] = \mathbb{E}[\mathbb{E}[\hat{\boldsymbol{w}}_c|\hat{\boldsymbol{\mu}}]] = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} + k_1\delta\boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right).$$

Hence, we have the expected out-of-sample portfolio mean and ESG score as

$$\begin{aligned} \mathbb{E}[\boldsymbol{\mu}^\top\hat{\boldsymbol{w}}_c] &= \frac{\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} + k_1\delta\boldsymbol{\mu}^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right) = \mu_g + k_1\delta\theta_1^2 + k_1\beta\theta_2^2, \\ \mathbb{E}[\boldsymbol{S}^\top\hat{\boldsymbol{w}}_c] &= \frac{\boldsymbol{S}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} + k_1\delta\boldsymbol{S}^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right) = \mu_s + k_1\delta\theta_2^2 + k_1\beta\theta_3^2, \end{aligned}$$

where $\mu_g = \boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}/\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}$, $\mu_s = \boldsymbol{S}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}/\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}$, $\theta_1^2 = \boldsymbol{\mu}^\top\boldsymbol{\Upsilon}\boldsymbol{\mu}, \theta_2^2 = \boldsymbol{\mu}^\top\boldsymbol{\Upsilon}\boldsymbol{S}$, and $\theta_3^2 = \boldsymbol{S}^\top\boldsymbol{\Upsilon}\boldsymbol{S}$.

2. We then calculate the expectation of the out-of-sample variance $\mathbb{E}[\hat{\boldsymbol{w}}_c^\top\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_c]$. Let σ_{ij}

represent the $(i,j)^{th}$ element of the covariance matrix $\boldsymbol{\Sigma}$. We have

$$\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_c] = \sum_{i=1}^{N} \sigma_{ii} \mathbb{E}\left[\mathbb{E}\left[\hat{w}_{c,i}^2 | \hat{\boldsymbol{\mu}}\right]\right] + 2\sum_{i>j} \sigma_{ij} \mathbb{E}\left[\mathbb{E}\left[\hat{w}_{c,i}\hat{w}_{c,j} | \hat{\boldsymbol{\mu}}\right]\right]. \tag{B.7}$$

Following previous notations, we denote $\mathbb{E}\left[\hat{w}_i^2\right] = \mathbb{E}\left[\hat{w}_{c,i}^2 | \hat{\boldsymbol{\mu}}\right]$. Since $\hat{w}_i$ is the first element of $\boldsymbol{z} = -\hat{\boldsymbol{H}}_{11}^{-1}(\hat{\boldsymbol{H}}_{12} - \boldsymbol{\delta})$, based on (B.5), we have

$$\begin{aligned}
\boldsymbol{z} | T\boldsymbol{H}_{11} &\sim N_2(-\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12} + \hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta}, (\boldsymbol{H}_{22} - \boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12})(T\hat{\boldsymbol{H}}_{11})^{-1}),\\
\hat{w}_i | T\hat{\boldsymbol{H}}_{11} &\sim N_1(\mu_{\boldsymbol{z},1}, \sigma_{\boldsymbol{z},1}^2),
\end{aligned}$$

where $\mu_{\boldsymbol{z},1}$ and $\sigma_{\boldsymbol{z},1}^2$ are the first and $(1,1)^{th}$ elements of the mean vector and covariance matrix of $\boldsymbol{z}|T\boldsymbol{H}_{11}$, respectively. Therefore, the expectation $\mathbb{E}\left[\hat{w}_i^2\right]$ satisfies that

$$\mathbb{E}[\hat{w}_i^2] = \mathbb{E}[\mathbb{E}[\hat{w}_i^2 | T\hat{\boldsymbol{H}}_{11}]] = \mathbb{E}[\mu_{\boldsymbol{z},1}^2] + \mathbb{E}[\sigma_{\boldsymbol{z},1}^2] \tag{B.8}$$

The first expectation $\mathbb{E}[\mu_{\boldsymbol{z},1}^2]$ is the $(1,1)^{th}$ element of the following expectation:

$$\begin{aligned}
&\mathbb{E}[(-\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12} + \hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta})(-\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12} + \hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta})^\top]\\
=&\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}\boldsymbol{H}_{12}^\top\boldsymbol{H}_{11}^{-1} - \mathbb{E}[\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}\boldsymbol{\delta}^\top\hat{\boldsymbol{H}}_{11}^{-1}] - \mathbb{E}[\hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta}\boldsymbol{H}_{12}^\top\boldsymbol{H}_{11}^{-1}] + \mathbb{E}[\hat{\boldsymbol{H}}_{11}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^\top\hat{\boldsymbol{H}}_{11}^{-1}]\\
=&\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}\boldsymbol{H}_{12}^\top\boldsymbol{H}_{11}^{-1} - k_1\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}\boldsymbol{\delta}^\top\boldsymbol{H}_{11}^{-1} - k_1\boldsymbol{H}_{11}^{-1}\boldsymbol{\delta}\boldsymbol{H}_{12}^\top\boldsymbol{H}_{11}^{-1} + Tk_2\boldsymbol{H}_{11}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{H}_{11}^{-1}\\
&\quad + k_1k_2\mathrm{tr}(\boldsymbol{H}_{11}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^\top)\boldsymbol{H}_{11}^{-1},
\end{aligned}$$

where $k_2 = T/(T-N)(T-N-3)$ and we use Theorem 3.3.16 of Gupta and Nagar (2018). Using (B.3), (B.4) and $\mathrm{tr}(\boldsymbol{H}_{11}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^\top) = \boldsymbol{\delta}^\top\boldsymbol{H}_{11}^{-1}\boldsymbol{\delta}$, we have

$$\begin{aligned}
\mathbb{E}[\mu_{\boldsymbol{z},1}^2] =& 2k_1\delta\left(\frac{\boldsymbol{e}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}\right)\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right] + Tk_2\delta^2\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2\\
&+\left(\frac{\boldsymbol{e}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}\right)^2 + k_1k_2\delta^2\hat{\psi}^2\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\boldsymbol{e}_i,
\end{aligned} \tag{B.9}$$

where $\hat{\psi}^2 = \left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right).$

The second expectation $\mathbb{E}[\sigma^2_{\boldsymbol{z},1}]$ is the $(1,1)^{th}$ element of the following expectation:

$$\mathbb{E}[(\boldsymbol{H}_{22} - \boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12})(T\hat{\boldsymbol{H}}_{11})^{-1}] = \frac{k_1}{T}(\boldsymbol{H}_{22} - \boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12})\boldsymbol{H}_{11}^{-1}.$$

Hence, we have

$$\mathbb{E}[\sigma^2_{\boldsymbol{z},1}] = \frac{k_1}{T}\sigma_g^2 \boldsymbol{e}_i^\top \boldsymbol{\Upsilon} \boldsymbol{e}_i, \tag{B.10}$$

where $\sigma_g^2 = 1/\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}$.

Since we have $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ and $\hat{\boldsymbol{\mu}}$ is independent of $\widehat{\boldsymbol{\Sigma}}$, the unconditional expectation of $\mathbb{E}[\hat{w}_i^2]$ (B.8) using (B.9) and (B.10) is

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[\hat{w}_i^2\right]\right] = &\left(\frac{\boldsymbol{e}_i^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}} + k_1\delta\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]\right)^2 \\ &+ \frac{(T-N+1)}{T}k_1^2k_2\delta^2\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2 \\ &+ \frac{\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\boldsymbol{e}_i}{T}\left[k_1\sigma_g^2 + k_1k_2(T-2)\delta^2 + k_1k_2T\delta^2\psi^2\right],\end{aligned} \tag{B.11}$$

where we use the following two expectations:

$$\begin{aligned}\mathbb{E}\left[\hat{\psi}^2\right] &= \frac{N-1}{T} + \psi^2, \\ \mathbb{E}\left[\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2\right] &= \frac{\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\boldsymbol{e}_i}{T} + \left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2.\end{aligned} \tag{B.12}$$

Here, the first expectation is obtained from the distribution $\hat{\psi}^2 \sim \chi^2_k(\psi^2)$ with the non-centrality parameter $\psi^2 = \left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right) = \theta_1^2 + \beta^2\theta_3^2/\delta^2 + 2\beta\theta_2^2/\delta$ and the degree of freedom $k = \mathrm{tr}(\boldsymbol{\Upsilon}\boldsymbol{\Sigma}/T) = (N-1)/T$ (Theorem 1.4.2, Muirhead, 1982), and $\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right) \sim N\left(\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right), \boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\boldsymbol{\Sigma}\boldsymbol{\Upsilon}\boldsymbol{e}_i^\top/T\right)$ (Theorem 1.2.6, Muirhead, 1982) with $\boldsymbol{\Upsilon}\boldsymbol{\Sigma}\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}$ allows us to obtain the second expectation.

Similarly, we denote $\mathbb{E}[\hat{w}_i, \hat{w}_j] = \mathbb{E}\left[\hat{w}_{c,i}\hat{w}_{c,j}|\hat{\boldsymbol{\mu}}\right]$. Let $\boldsymbol{M}^* = \left[\boldsymbol{e}_i, \boldsymbol{e}_j, \hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}, \mathbf{1}\right]$ be an $N \times 4$ matrix. We decomposite $\boldsymbol{M}^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{M}^*)^\top$ and its inverse in a similar way to (B.2), where $\hat{\boldsymbol{G}}_{11}^*$ is the 3×3 matrix. The notations introduced in (1) are correspondingly adjusted with a superscript *. According to Corollary 3.2.6 of Muirhead

(1982), we have

$$T\hat{\boldsymbol{H}}^* \sim \mathcal{W}_4(T-N+3, \boldsymbol{H}),$$
$$T\hat{\boldsymbol{H}}^*_{11} \sim \mathcal{W}_3(T-N+3, \boldsymbol{H}^*_{11}).$$

The portfolio weights $\hat{w}_i$ and $\hat{w}_j$ are the first two elements in $\boldsymbol{z}^* = -(\hat{\boldsymbol{H}}^*_{11})^{-1}(\hat{\boldsymbol{H}}^*_{12} - \boldsymbol{\delta}^*)$, where $\boldsymbol{\delta}^* = [0,0,\delta]^\top$. The expectation $\mathbb{E}\left[\hat{w}_i\hat{w}_j\right]$ satisfies that

$$\mathbb{E}[\hat{w}_i\hat{w}_j] = \mathbb{E}[\mathbb{E}[\hat{w}_i\hat{w}_j | T\hat{\boldsymbol{H}}^*_{11}]] = \mathbb{E}[\mu_{\boldsymbol{z}^*,1}\mu_{\boldsymbol{z}^*,2}] + \mathbb{E}[\sigma_{\boldsymbol{z},12}], \tag{B.13}$$

where $\mu_{\boldsymbol{z}^*,1}$ and $\mu_{\boldsymbol{z}^*,2}$ are the first two elements of the mean vector $\mathbb{E}[\boldsymbol{z}^*|T\boldsymbol{H}^*_{11}]$ and $\sigma_{\boldsymbol{z},12}$ is the $(1,2)^{\text{th}}$ element of the covariance matrix $\text{Cov}(\boldsymbol{z}^*|T\boldsymbol{H}^*_{11})$.

Similarly to (B.9) and (B.10), we have

$$\begin{aligned}\mathbb{E}[\mu_{\boldsymbol{z}^*,1}\mu_{\boldsymbol{z}^*,2}] &= k_1\delta\left(\frac{\boldsymbol{e}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}\right)\left[\boldsymbol{e}_j^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right] + k_1\delta\left(\frac{\boldsymbol{e}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}\right)\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right] \\ &+ \frac{(\boldsymbol{e}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1})(\boldsymbol{e}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1})}{(\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1})^2} + Tk_2\delta^2\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\left[\boldsymbol{e}_j^\top\boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right] + k_1k_2\delta^2\hat{\psi}^2\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\boldsymbol{e}_j, \\ \mathbb{E}[\sigma_{\boldsymbol{z},12}] &= \frac{k_1}{T}\sigma_g^2\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\boldsymbol{e}_j.\end{aligned}$$

Since we have $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ and $\hat{\boldsymbol{\mu}}$ is independent of $\widehat{\boldsymbol{\Sigma}}$, the unconditional expectation of $\mathbb{E}[\hat{w}_i\hat{w}_j]$ is therefore obtained by

$$\begin{aligned}&\mathbb{E}\left[\mathbb{E}\left[\hat{w}_i\hat{w}_j\right]\right] \\ &= \left[\frac{\boldsymbol{e}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} + k_1\delta\left(\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right)\right]\left[\frac{\boldsymbol{e}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{1}} + k_1\delta\left(\boldsymbol{e}_j^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right)\right] \\ &\quad + \frac{(T-N+1)}{T}k_1^2k_2\delta^2\left[\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\left[\boldsymbol{e}_j^\top\boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right] \\ &\quad + \frac{\boldsymbol{e}_i^\top\boldsymbol{\Upsilon}\boldsymbol{e}_j}{T}\left[k_1\sigma_g^2 + k_1k_2(T-2)\delta^2 + k_1k_2T\delta^2\psi^2\right]\end{aligned} \tag{B.14}$$

where we use the following expectation:

$$\mathbb{E}\left[\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}}+\frac{\beta}{\delta}\boldsymbol{S}\right)\left(\hat{\boldsymbol{\mu}}+\frac{\beta}{\delta}\boldsymbol{S}\right)^\top \boldsymbol{\Upsilon}\boldsymbol{e}_j\right]=\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)^\top \boldsymbol{\Upsilon}\boldsymbol{e}_j+\frac{1}{T}\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_j$$

Given that $\boldsymbol{I}_N = [\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_N]$ and according to (B.11) and (B.14), we obtain the expected out-of-sample portfolio variance (B.7) by

$$\begin{aligned}
&\mathbb{E}[\hat{\boldsymbol{w}}_c^\top \boldsymbol{\Sigma}\hat{\boldsymbol{w}}_c]\\
=&\sum_{i=1}^N \sigma_{ii}\left[\frac{\boldsymbol{e}_i^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}+k_1\delta\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2+\sum_{i=1}^N \sigma_{ii}\frac{(T-N+1)k_2}{T}\left[k_1\delta\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2\\
&+2\sum_{i>j}\sigma_{ij}\left[\frac{\boldsymbol{e}_i^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}+k_1\delta\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\left[\frac{\boldsymbol{e}_j^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}+k_1\delta\boldsymbol{e}_j^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\\
&+2\sum_{i>j}\sigma_{ij}\frac{(T-N+1)k_2}{T}\left[k_1\delta\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\left[k_1\delta\boldsymbol{e}_j^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\\
&+\frac{k_1\sigma_g^2+k_1k_2(T-2)\delta^2+k_1k_2T\delta^2\psi^2}{T}\left[\sum_{i=1}^N \sigma_{ii}\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i+2\sum_{i>j}\sigma_{ij}\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_j\right]\\
=&\left[\frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}+k_1\delta\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)^\top \boldsymbol{\Upsilon}\right]\boldsymbol{\Sigma}\left[\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}+k_1\delta\boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\\
&+\frac{(T-N+1)k_2}{T}\left[k_1\delta\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)^\top \boldsymbol{\Upsilon}\right]\boldsymbol{\Sigma}\left[k_1\delta\boldsymbol{\Upsilon}\left(\boldsymbol{\mu}+\frac{\beta}{\delta}\boldsymbol{S}\right)\right]\\
&+\frac{k_1\sigma_g^2+k_1k_2(T-2)\delta^2+k_1k_2T\delta^2\psi^2}{T}\mathbf{1}^\top(\boldsymbol{\Upsilon}\circ\boldsymbol{\Sigma})\mathbf{1}\\
=&\sigma_g^2+\frac{T+(T-N+1)k_2}{T}k_1^2\delta^2\psi^2+\frac{N-1}{T}[k_1\sigma_g^2+k_1k_2(T-2)\delta^2+k_1k_2T\delta^2\psi^2]\\
=&k_1k_3(\sigma_g^2+Tk_2\delta^2\psi^2+k_2(N-1)\delta^2)
\end{aligned} \tag{B.15}$$

where $k_3 = (T-2)/T$ and $\boldsymbol{\Upsilon}\circ\boldsymbol{\Sigma}$ represents the Hadamard product (element-wise product) of two matrices $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Sigma}$. We use the following two relationships:

$$\mathbf{1}^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right) = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right) - \frac{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1})\left[\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} = 0$$

$$\mathbf{1}^\top(\boldsymbol{\Upsilon} \circ \boldsymbol{\Sigma})\mathbf{1} = \mathbf{1}^\top\left[\left(\boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}\right) \circ \boldsymbol{\Sigma}\right]\mathbf{1} = N - 1$$

Here, the last equation is achieved using Lemma B.1.

□

B.1.6 Proof of Theorem 3.2

Proof of Theorem 3.2. Let $\boldsymbol{c} = [\delta, \beta]^\top$, $\boldsymbol{h} = [k_1\theta_1^2, k_1\theta_2^2]^\top$, $O = \mu_g - \frac{\gamma k_1 k_3 \sigma_g^2}{2}$, and the matrix $\boldsymbol{A} = \begin{bmatrix} A_1 & A_2 \\ A_2 & A_3 \end{bmatrix}$ with $A_1 = \gamma k_1 k_2 k_3 (T\theta_1^2 + N - 1)$, $A_2 = \gamma T k_1 k_2 k_3 \theta_2^2$, $A_3 = \gamma T k_1 k_2 k_3 \theta_3^2$. Also, let $\boldsymbol{p} = [k_1\theta_2^2, k_1\theta_3^2]^\top$ and $q = \mu_s - \bar{s}$. Hence, the objective function in the optimization problem (3.13) for the combined portfolio $\hat{\boldsymbol{w}}_c$ can be written by

$$\begin{aligned} G(\boldsymbol{c}) &= \mu_g - \frac{\gamma k_1 k_3 \sigma_g^2}{2} + k_1\theta_1^2\delta + k_1\theta_2^2\beta - \frac{\gamma k_1 k_2 k_3}{2}[(T\theta_1^2 + N - 1)\delta^2 + T\theta_3^2\beta^2 + 2T\theta_2^2\delta\beta] \\ &= -\frac{1}{2}\boldsymbol{c}^\top \boldsymbol{A}\boldsymbol{c} + \boldsymbol{h}^\top\boldsymbol{c} + O, \end{aligned}$$

and the constraint is

$$\boldsymbol{p}^\top\boldsymbol{c} + q \geq 0.$$

The optimization problem (without the constant O since it does not affect the optimal solution) can be expressed as

$$\begin{aligned} \max_{\boldsymbol{c}} \quad & -\frac{1}{2}\boldsymbol{c}^\top\boldsymbol{A}\boldsymbol{c} + \boldsymbol{h}^\top\boldsymbol{c}, \\ s.t. \quad & \boldsymbol{p}^\top\boldsymbol{c} + q \geq 0. \end{aligned} \tag{B.16}$$

A square matrix is positive definite if it is symmetric and all its eigenvalues are positive. Since $\boldsymbol{A}$ is a 2×2 symmetric matrix, a necessary and sufficient condition for $\boldsymbol{A}$ to be positive definite is that $A_1 > 0$ and the determinant of $\boldsymbol{A}$ is positive. Given that $N \geq 2$ and $\theta_1^2 = \boldsymbol{\mu}^\top \boldsymbol{\Upsilon} \boldsymbol{\mu}$ is non-negative, we have $A_1 > 0$. Meanwhile, the determinant of $\boldsymbol{A}$ satisfies that $\det(\boldsymbol{A}) = \gamma^2 k_1^2 k_2^2 k_3^2 [T^2(\theta_1^2\theta_3^2 - \theta_2^4) + T(N-1)\theta_3^2] > 0$. Therefore, we conclude that the matrix $\boldsymbol{A}$ is positive definite. The problem is a convex optimization problem.

The Lagrangian function with a Lagrange multiplier ζ is

$$\mathcal{L}(\boldsymbol{c}, \lambda) = -\frac{1}{2}\boldsymbol{c}^\top \boldsymbol{A}\boldsymbol{c} + \boldsymbol{h}^\top \boldsymbol{c} + \zeta(\boldsymbol{p}^\top \boldsymbol{c} + q). \tag{B.17}$$

If the constraint is not satisfied, the last term is negative, effectively penalizing the objective function (since our goal is to maximize it). The Karush-Kuhn-Tucker conditions for this problem are

$$\begin{aligned} -\boldsymbol{A}\boldsymbol{c} + \boldsymbol{h} + \zeta\boldsymbol{p} &= 0, \\ \boldsymbol{p}^\top \boldsymbol{c} + q &\geq 0, \\ \zeta &\geq 0, \\ \zeta(\boldsymbol{p}^\top \boldsymbol{c} + q) &= 0. \end{aligned}$$

For convex problems, any point that satisfies the KKT conditions is a global optimum. Since $\boldsymbol{A}$ is positive definite, it is invertible. The first condition indicates that

$$\boldsymbol{c} = \boldsymbol{A}^{-1}(\boldsymbol{h} + \zeta\boldsymbol{p}). \tag{B.18}$$

If the Lagrange multiplier $\zeta \neq 0$, the last condition indicates that $\boldsymbol{p}^\top \boldsymbol{c} + q = 0$. Substituting (B.18) into this equation gives that

$$\boldsymbol{p}^\top \boldsymbol{A}^{-1}(\boldsymbol{h} + \zeta\boldsymbol{p}) + q = 0,$$

which gives that

$$\zeta = -\frac{\boldsymbol{p}^\top \boldsymbol{A}^{-1}\boldsymbol{h} + q}{\boldsymbol{p}^\top \boldsymbol{A}^{-1}\boldsymbol{p}} = -\frac{k_1\theta_2^2 + \gamma k_2 k_3 T(\mu_s - \bar{s})}{k_1\theta_3^2}.$$

Because of the dual feasibility condition $\zeta \geq 0$, we consider two cases:

- If $\zeta > 0$, the inequality constraint effectively functions as an equality constraint.

That is, the optimal solution is obtained at the boundary. The solution is

$$\boldsymbol{c} = \boldsymbol{A}^{-1}\left(\boldsymbol{h} + \zeta\boldsymbol{p}\right) = \boldsymbol{A}^{-1}\boldsymbol{h} + \zeta\boldsymbol{A}^{-1}\boldsymbol{p}.$$

The corresponding combination coefficients are obtained by

$$\begin{aligned}\delta_I^* &= \frac{\theta_1^2\theta_3^2 - \theta_2^4}{\gamma k_2 k_3\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]},\\ \beta_I^* &= \frac{\theta_2^2(\theta_1^2\theta_3^2 - \theta_2^4)}{\gamma k_2 k_3 \theta_3^2\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]} - \frac{\mu_s - \bar{s}}{k_1\theta_3^2}.\end{aligned}$$

- If $\zeta \leq 0$, it indicates that the inequality constraint is not active, which means that the solution is the same as the one obtained by maximizing the same objective function without any constraint. Hence, we let the Lagrange multiplier be zero. The solution is

$$\boldsymbol{c}^* = \boldsymbol{A}^{-1}\boldsymbol{h}.$$

The corresponding combination coefficients are obtained by

$$\begin{aligned}\delta_I^* &= \frac{\theta_1^2\theta_3^2 - \theta_2^4}{\gamma k_2 k_3\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]},\\ \beta_I^* &= \frac{(N-1)\theta_2^2}{\gamma k_2 k_3 T\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]}.\end{aligned}$$

□

B.1.7 Proof of Theorem 3.3

Proof. The optimal combination coefficients are the same as those when $\lambda > 0$ in Appendix B.1.6. That is,

$$\begin{aligned}\delta_I^* &= \frac{\theta_1^2\theta_3^2 - \theta_2^4}{\gamma k_2 k_3\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]},\\ \beta_I^* &= \frac{\theta_2^2(\theta_1^2\theta_3^2 - \theta_2^4)}{\gamma k_2 k_3 \theta_3^2\left[T\theta_1^2\theta_3^2 - T\theta_2^4 + \theta_3^2(N-1)\right]} - \frac{\mu_s - \bar{s}}{k_1\theta_3^2}.\end{aligned}$$

□

B.1.8 Proof of Proposition 3.4

Proof. The variance of the out-of-sample portfolio ESG score can be written as

$$Var[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}] = \boldsymbol{S}^\top Var[\hat{\boldsymbol{w}}_c]\boldsymbol{S}$$

Hence, we need to calculate $Var[\hat{\boldsymbol{w}}_c]$. We follow the notations in Appendix B.1.5. Let $\hat{w}_i = \hat{w}_{c,i}|\hat{\boldsymbol{\mu}}$ be the i^{th} conditional portfolio weight, using (B.6) and (B.8), we obtain that

$$\begin{aligned}
&Var[\hat{w}_i] = \mathbb{E}[\hat{w}_i^2] - (\mathbb{E}[\hat{w}_i])^2 \\
=&\left(\frac{\boldsymbol{e}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{1}}\right)^2 + 2k_1\delta\left(\frac{\boldsymbol{e}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{1}}\right)\left[\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right] + Tk_2\delta^2\left[\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2 \\
&+ \frac{k_1}{T}\sigma_g^2 \boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i + k_1k_2\delta^2\hat{\psi}^2\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i - \left[\frac{\boldsymbol{e}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{1}}{\boldsymbol{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{1}} + k_1\delta\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2 \\
=&\frac{k_1}{T}\sigma_g^2 \boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i + k_1k_2\delta^2\hat{\psi}^2\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i + (Tk_2 - k_1^2)\delta^2\left[\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\hat{\boldsymbol{\mu}} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2
\end{aligned}$$

Since we have $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ and $\hat{\boldsymbol{\mu}}$ is independent of $\widehat{\boldsymbol{\Sigma}}$, utilizing (B.12), the unconditional variance is obtained by

$$\begin{aligned}
&Var[\hat{w}_{c,i}] = \mathbb{E}[Var[\hat{w}_i]] + Var[\mathbb{E}[\hat{w}_i]] \\
=&k_1k_2\delta^2\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i\left(\frac{N-1}{T} + \psi^2\right) + (Tk_2 - k_1^2)\delta^2\left[\left(\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right)^2 + \frac{\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i}{T}\right] \\
&+ \frac{k_1}{T}\sigma_g^2 \boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i + \frac{k_1^2\delta^2}{T}\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i \qquad \text{(B.19)} \\
=&k_1k_2\delta^2\psi^2\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i + \left[k_1\sigma_g^2 + k_1^2\delta^2 + k_1k_2\delta^2(N-1) + (Tk_2 - k_1^2)\delta^2\right]\frac{\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\boldsymbol{e}_i}{T} \\
&+ (Tk_2 - k_1^2)\delta^2\left[\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]^2
\end{aligned}$$

Similarly, for the i^{th} and j^{th} conditional portfolio weights $\hat{w}_i = \hat{w}_{c,i}|\hat{\boldsymbol{\mu}}$ and $\hat{w}_j = \hat{w}_{c,j}|\hat{\boldsymbol{\mu}}$, using (B.6) and (B.13), the conditional covariance is

$$
\begin{aligned}
&Cov[\hat{w}_i, \hat{w}_j] = \mathbb{E}[\hat{w}_i \hat{w}_j] - \mathbb{E}[\hat{w}_i]\mathbb{E}[\hat{w}_j] \\
=&\frac{k_1}{T}\sigma_g^2 \boldsymbol{e}_i^\top \boldsymbol{\Upsilon} \boldsymbol{e}_j + k_1 k_2 \delta^2 \hat{\psi}^2 (\boldsymbol{e}_i^\top \boldsymbol{\Upsilon} \boldsymbol{e}_j) + (Tk_2 - k_1^2)\delta^2 \left[\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]\left[\boldsymbol{e}_j^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]
\end{aligned}
$$

Since we have $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$ and $\hat{\boldsymbol{\mu}}$ is independent of $\widehat{\boldsymbol{\Sigma}}$, the unconditional covariance is

$$
\begin{aligned}
&Cov[\hat{w}_{c,i}, \hat{w}_{c,j}] = \mathbb{E}[Cov[\hat{w}_i, \hat{w}_j]] + Cov[\mathbb{E}[\hat{w}_i], \mathbb{E}[\hat{w}_i]] \\
=&k_1 k_2 \delta^2 \psi^2 (\boldsymbol{e}_i^\top \boldsymbol{\Upsilon} \boldsymbol{e}_j) + [k_1 \sigma_g^2 + k_1 k_2 \delta^2 (N-1) + (Tk_2 - k_1^2)\delta^2 + k_1^2 \delta^2] \frac{\boldsymbol{e}_i^\top \boldsymbol{\Upsilon} \boldsymbol{e}_j}{T} \\
&+ (Tk_2 - k_1^2)\delta^2 \left[\boldsymbol{e}_i^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]\left[\boldsymbol{e}_j^\top \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\right]
\end{aligned} \tag{B.20}
$$

Therefore, based on (B.19) and (B.20), the covariance matrix $Var[\hat{\boldsymbol{w}}_c]$ is obtained by

$$
Var[\hat{\boldsymbol{w}}_c] = \left[\frac{k_1 \sigma_g^2}{T} + \delta^2 k_1 k_2 (k_3 + \psi^2)\right] \boldsymbol{\Upsilon} + (Tk_2 - k_1^2)\delta^2 \boldsymbol{\Upsilon}\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)\left(\boldsymbol{\mu} + \frac{\beta}{\delta}\boldsymbol{S}\right)^\top \boldsymbol{\Upsilon}
$$

The variance of the out-of-sample portfolio ESG score is then given by

$$
Var[\hat{\boldsymbol{w}}_c^\top \boldsymbol{S}] = \boldsymbol{S}^\top Var[\hat{\boldsymbol{w}}_c]\boldsymbol{S} = \left[\frac{k_1 \sigma_g^2}{T} + \delta^2 k_1 k_2 (k_3 + \psi^2)\right] \theta_3^2 + (Tk_2 - k_1^2)\left(\delta\theta_2^2 + \beta\theta_3^2\right)^2
$$

□

B.1.9 Calculation of the Sophisticated Estimators

We follow the notations in Appendix B.1.5. Let $\dot{\boldsymbol{M}} = [\hat{\boldsymbol{\mu}}, \boldsymbol{S}, \boldsymbol{1}]$ be an $N \times 3$ matrix and $\dot{\boldsymbol{M}}_1 = [\hat{\boldsymbol{\mu}}, \boldsymbol{S}]$ be the first two columns. We decompose $\dot{\boldsymbol{M}}\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{M}}^\top$ and its inverse similarly to (B.2), where $\dot{\boldsymbol{G}}_{11}$ is the 2×2 matrix. The notations introduced in (1) of the proof of Proposition 3.3 in Appendix B.1.5 are correspondingly adjusted with a point above the symbol. We obtain that

$$\hat{\dot{\boldsymbol{H}}}_{11}^{-1} = \hat{\dot{\boldsymbol{G}}}_{11} - \hat{\dot{\boldsymbol{G}}}_{12}\hat{\dot{\boldsymbol{G}}}_{22}^{-1}\hat{\dot{\boldsymbol{G}}}_{21} = \dot{\boldsymbol{M}}_1^\top \hat{\boldsymbol{\Upsilon}} \dot{\boldsymbol{M}}_1 = \begin{bmatrix} \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} & \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S} \\ \hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S} & \boldsymbol{S}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S} \end{bmatrix}.$$

Conditional on $\hat{\boldsymbol{\mu}}$, since $T\widehat{\boldsymbol{\Sigma}} \sim \mathcal{W}_N(T-1, \boldsymbol{\Sigma})$, according to Corollary 3.2.6 and Theorem 3.2.11 of Muirhead (1982), we have

$$\begin{aligned} T\hat{\dot{\boldsymbol{H}}} &\sim \mathcal{W}_3(T-N+2, \dot{\boldsymbol{H}}), \\ T\hat{\dot{\boldsymbol{H}}}_{11} &\sim \mathcal{W}_2(T-N+2, \dot{\boldsymbol{H}}_{11}). \end{aligned}$$

Therefore, the expectation of $\hat{\dot{\boldsymbol{H}}}_{11}^{-1}$ is $\mathbb{E}[\hat{\dot{\boldsymbol{H}}}_{11}^{-1}] = k_1 \dot{\boldsymbol{H}}_{11}^{-1}$. Accordingly, we have

$$\begin{aligned} &\mathbb{E}[\hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}] = \mathbb{E}[\mathbb{E}[\hat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}|\hat{\boldsymbol{\mu}}] = k_1 \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} \\ &\mathbb{E}[\boldsymbol{S}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}] = k_1 \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S} \end{aligned} \tag{B.21}$$

The unbiased estimators of θ_2^2 and θ_3^2 are correspondingly obtained by

$$\tilde{\theta}_2^2 = \frac{T-N-1}{T}\hat{\theta}_2^2 \quad \text{and} \quad \tilde{\theta}_3^2 = \frac{T-N-1}{T}\hat{\theta}_3^2.$$

B.1.10 Proof of Proposition 3.5

Proof of Proposition 3.5.

- Denote the implementable version of the optimal combination coefficients δ_E^* and β_E^* using the sample estimators $\hat{\theta}_1^2$, $\hat{\theta}_2^3$, $\hat{\theta}_3^2$ as

$$\begin{aligned} \hat{\delta}_E &= \frac{\hat{\theta}_1^2\hat{\theta}_3^2 - \hat{\theta}_2^4}{\gamma k_2 k_3 \left[T\hat{\theta}_1^2\hat{\theta}_3^2 - T\hat{\theta}_2^4 + \hat{\theta}_3^2(N-1)\right]}, \\ \hat{\beta}_E &= \frac{\bar{s} - \hat{\mu}_s - k_1\hat{\theta}_2^2\hat{\delta}_E}{k_1\hat{\theta}_3^2}. \end{aligned}$$

  Given that $\hat{\theta}_1^2 = \hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Upsilon}}\hat{\boldsymbol{\mu}}, \hat{\theta}_2^2 = \hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{\Upsilon}}\boldsymbol{S}, \hat{\theta}_3^2 = \boldsymbol{S}^\top \hat{\boldsymbol{\Upsilon}}\boldsymbol{S}$ and $\hat{\mu}_s = \boldsymbol{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}/\boldsymbol{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{1}$, the total ESG score of the implementable combined portfolio $\hat{\boldsymbol{w}}_{c\text{-}E}$ satisfies that

$$\begin{aligned}\hat{\boldsymbol{w}}_{c\text{-}E}^\top \boldsymbol{S} &= \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}} + \hat{\delta}_E \boldsymbol{S}^\top \hat{\boldsymbol{\Upsilon}}\hat{\boldsymbol{\mu}} + \frac{\bar{s} - \hat{\mu}_s - k_1\hat{\theta}_2^2\hat{\delta}_E}{k_1\hat{\theta}_3^2}\boldsymbol{S}^\top \hat{\boldsymbol{\Upsilon}}\boldsymbol{S} \\ &= \frac{\bar{s} + (k_1 - 1)\hat{\mu}_s}{k_1}\end{aligned}$$

Therefore, as the coefficient of the target ESG score $\bar{s}$ in $\hat{\boldsymbol{w}}_{c\text{-}E}^\top \boldsymbol{S}$ is constant (i.e., $1/k_1$), the variance of the total ESG score $\hat{\boldsymbol{w}}_{c\text{-}E}^\top \boldsymbol{S}$ will not be affected by the change of $\bar{s}$.

- Similarly, denote the implementable version of the optimal combination coefficients δ_E^* and β_E^* using the sophisticated estimators $\tilde{\theta}_1^2$, $\tilde{\theta}_2^3$, $\tilde{\theta}_3^2$ as

$$\begin{aligned}\tilde{\delta}_E &= \frac{\tilde{\theta}_1^2\tilde{\theta}_3^2 - \tilde{\theta}_2^4}{\gamma k_2 k_3 \left[T\tilde{\theta}_1^2\tilde{\theta}_3^2 - T\tilde{\theta}_2^4 + \tilde{\theta}_3^2(N-1)\right]}, \\ \tilde{\beta}_E &= \frac{\bar{s} - \hat{\mu}_s - k_1\tilde{\theta}_2^2\tilde{\delta}_E}{k_1\tilde{\theta}_3^2}.\end{aligned}$$

Since $\tilde{\theta}_2^2 = (T-N-1)\hat{\theta}_2^2/T$, $\tilde{\theta}_3^2 = (T-N-1)\hat{\theta}_3^2/T$ and $k_1 = T/(T-N-1)$, the total ESG score of the implementable combined portfolio $\tilde{\boldsymbol{w}}_{c\text{-}E}$ satisfies that

$$\begin{aligned}\tilde{\boldsymbol{w}}_{c\text{-}E}^\top \boldsymbol{S} &= \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}} + \tilde{\delta}_E \boldsymbol{S}^\top \hat{\boldsymbol{\Upsilon}}\hat{\boldsymbol{\mu}} + \frac{\bar{s} - \hat{\mu}_s - k_1\tilde{\theta}_2^2\tilde{\delta}_E}{k_1\tilde{\theta}_3^2}\boldsymbol{S}^\top \hat{\boldsymbol{\Upsilon}}\boldsymbol{S} \\ &= \bar{s}\end{aligned}$$

□

Appendix C

Appendix for Chapter 4

C.1 Constituent portfolios

This section summarizes all the constituent portfolios considered in the combined portfolios in this work. Given the historical excess returns of the N risky assets from T periods, denoted as $\{\boldsymbol{r}_1, ...\boldsymbol{r}_T\}$, the plug-in strategies estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, by

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \qquad \text{and} \qquad \widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})^\top. \tag{C.1}$$

C.1.1 The 1/N portfolio

The 1/N rule is a portfolio strategy that allocates total wealth equally among N risky assets. It does not involve optimization or require historical information for asset returns. The portfolio strategy is

$$\boldsymbol{w}_e = \frac{\mathbf{1}_N}{N}.$$

DeMiguel et al. (2009b) compare the 1/N portfolio with various optimal rules derived in the case where a risk-free asset is available and find the superior performance of the 1/N rule in many empirical data sets. The portfolio mean and variance are

$$\hat{\mu}_e = \boldsymbol{w}_e^\top \hat{\boldsymbol{\mu}} \quad \text{and} \quad \hat{\sigma}_e^2 = \boldsymbol{w}_e^\top \widehat{\boldsymbol{\Sigma}} \boldsymbol{w}_e.$$

Additionally, we denote $\hat{\theta}_e^2 = \hat{\mu}_e^2/\hat{\sigma}_e^2$.

C.1.2 The sample GMV portfolio

Since the estimation error in the sample mean is generally large, some studies focus on the global minimum variance (GMV) portfolio (see Jobson, 1979; Jagannathan and Ma, 2003; Kempf and Memmel, 2006; Bodnar et al., 2018, for example). The GMV portfolio minimizes the variance and is subject to the constraint $\boldsymbol{w}^\top \mathbf{1}_N = 1$. The sample GMV portfolio weight of the optimal counterpart is

$$\hat{\boldsymbol{w}}_g = \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}{\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}. \tag{C.2}$$

The expected return of the global minimum variance portfolio is $\mu_g = \mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\mathbf{1}_N^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_N$ and the corresponding estimator of the sample minimum variance portfolio is:

$$\hat{\mu}_g = \frac{\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_N^\top \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}.$$

C.1.3 The sample constrained MV portfolio

When the investor chooses the portfolio weight to maximize the mean-variance utility function with the constraint $\boldsymbol{w}^\top \mathbf{1}_N = 1$, the optimal portfolio rule can be expressed as a combination of the GMV portfolio and an efficient zero-investment portfolio. The sample version of the optimal portfolio is

$$\hat{\boldsymbol{w}}_{mr} = \hat{\boldsymbol{w}}_g + \frac{1}{\gamma}\hat{\boldsymbol{w}}_z,$$

where γ is the risk aversion coefficient, $\hat{\boldsymbol{w}}_g$ is the sample GMV portfolio defined in (C.2), and $\hat{\boldsymbol{w}}_z = \widehat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \hat{\mu}_g \mathbf{1}_N)$ is the sample zero-investment portfolio.

C.1.4 The sample MV portfolio

The sample MV portfolio with a risk-free asset is

$$\hat{\boldsymbol{w}}_{mv} = \frac{1}{\gamma}\tilde{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}, \tag{C.3}$$

where the unbiased estimator of the covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\tilde{\boldsymbol{\Sigma}} = \frac{T\widehat{\boldsymbol{\Sigma}}}{T-N-2}.$$

C.1.5 The three-fund rule

The three-fund rule is proposed by Kan and Zhou (2007) as

$$\hat{\boldsymbol{w}}_{kz} = \frac{1}{\gamma}(c\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} + d\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N),$$

where c and d are constants determined to maximize the expected out-of-sample mean-variance utility. The optimal combination coefficients of c and d are respectively given by

$$c^* = C_1\frac{\psi_g^2}{\psi_g^2 + \frac{N}{T}} \quad \text{and} \quad d^* = C_1\mu_g\frac{\frac{N}{T}}{\psi_g^2 + \frac{N}{T}},$$

where $C_1 = [(T-N-1)(T-N-4)]/[T(T-2)]$ and $\psi_g^2 = (\boldsymbol{\mu} - \mu_g\mathbf{1}_N)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_g\mathbf{1}_N)$ is the squared slope of the asymptote to the ex ante minimum-variance frontier. The optimal values of c and d depend on the values of μ_g and ψ^2, which require the true values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We follow Kan and Zhou (2007) and replace the unknown μ_g and ψ^2 by the corresponding estimators $\hat{\mu}_g$ and

$$\hat{\psi}_{a,g}^2 = \frac{(T-N-1)\hat{\psi}_g^2 - (N-1)}{T} + \frac{2(\hat{\psi}_g^2)^{\frac{N-1}{2}}(1+\hat{\psi}_g^2)^{-\frac{T-2}{2}}}{TB_{\hat{\psi}_g^2/(1+\hat{\psi}_g^2)}((N-1)/2, (T-N+1)/2)}$$

to obtain the estimated optimal combination coefficients in the calculation, where $\hat{\psi}_g^2 = (\hat{\boldsymbol{\mu}} - \hat{\mu}_g\mathbf{1}_N)^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \hat{\mu}_g\mathbf{1}_N)$ and $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\,\mathrm{d}y$ is the incomplete beta function.

C.2 Analytical combined portfolios

This section summarizes the analytical combined portfolios from the existing literature considered in this work. We summarize them from the following two perspectives: using the 1/N rule or the sample GMV portfolios as one of the constituent portfolios.

C.2.1 Taking the 1/N portfolio as a component

C.2.1.1 Combined with the sample MV portfolio

From Tu and Zhou (2011), on the combination of the 1/N rule with the sample MV portfolio, with the convex constraint on the combination coefficients, the estimated optimal one is

$$\hat{\boldsymbol{w}}^{a}_{e\text{-}mv} = (1 - \hat{\delta}^{a}_{e\text{-}mv})\boldsymbol{w}_e + \hat{\delta}^{a}_{e\text{-}mv}\hat{\boldsymbol{w}}_{mv},$$

where $\hat{\delta}^{a}_{e\text{-}mv} = \hat{\pi}_1/(\hat{\pi}_1 + \hat{\pi}_2)$ with $\hat{\pi}_1$ and $\hat{\pi}_2$ respectively given by

$$\hat{\pi}_1 = \hat{\sigma}^2_e - \frac{2}{\gamma}\hat{\mu}_e + \frac{1}{\gamma^2}\hat{\theta}^2_a,$$
$$\hat{\pi}_2 = \frac{1}{\gamma^2}(C_2 - 1)\hat{\theta}^2_a + \frac{C_2}{\gamma^2}\frac{N}{T},$$

with $C_2 = (T-2)(T-N-2)/((T-N-1)(T-N-4))$ and

$$\hat{\theta}^2_a = \frac{(T-N-2)\hat{\theta}^2 - N}{T} + \frac{2(\hat{\theta}^2)^{\frac{N}{2}}(1+\hat{\theta}^2)^{-\frac{T-2}{2}}}{TB_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2, (T-N)/2)},$$

Here, $\hat{\theta}^2 = \hat{\boldsymbol{\mu}}^{\top}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$.

C.2.1.2 Combined with the sample MV portfolio - unconstrained

From Lassance et al. (2023), on the combination of the 1/N rule with the sample MV portfolio, without the convexity constraint on the combination coefficient, the estimated optimal one is

$$\hat{\boldsymbol{w}}^{u}_{e\text{-}mv} = \hat{\delta}^{u1}_{e\text{-}mv}\boldsymbol{w}_e + \hat{\delta}^{u2}_{e\text{-}mv}\hat{\boldsymbol{w}}_{mv},$$

where the combination coefficients are respectively given by

$$\hat{\delta}^{u1}_{e\text{-}mv} = \frac{1 + \frac{\hat{\mu}_e}{\gamma\hat{\sigma}_e^2} T\hat{l}^2}{1 + T\hat{l}^2}(1 - \hat{\delta}^{u2}_{e\text{-}mv}) \quad \text{and} \quad \hat{\delta}^{u2}_{e\text{-}mv} = \frac{\hat{\psi}^2_{a,e}}{\hat{\psi}^2_{a,e} + \hat{d}}.$$

In the above,

$$\begin{aligned}
\hat{l} &= \hat{\sigma}_e\left(\gamma - \frac{\hat{\mu}_e}{\hat{\sigma}_e^2}\right),\\
\hat{d} &= \frac{C_2 N}{T} + (C_2 - 1)\hat{\theta}_a^2,\\
\hat{\psi}^2_{a,e} &= \frac{(T-N-2)\hat{\psi}_e^2 - (N-1)(1+\hat{\theta}_e^2)}{T} + \frac{2(1+\hat{\theta}_e^2)\Delta^{\frac{N-1}{2}}(1+\Delta)^{-\frac{T-3}{2}}}{TB_{\Delta/(1+\Delta)}((N-1)/2,(T-N)/2)},
\end{aligned}$$

where $\hat{\psi}_e^2 = \hat{\theta}^2 - \hat{\theta}_e^2$ and $\Delta = \hat{\psi}_e^2/(1+\hat{\theta}_e^2)$.

C.2.1.3 Combined with the three-fund portfolio

From Tu and Zhou (2011), on the combination of the 1/N rule with the three-fund portfolio, the estimated optimal one is

$$\hat{\boldsymbol{w}}^a_{e\text{-}kz} = (1 - \hat{\delta}^a_{e\text{-}kz})\boldsymbol{w}_e + \hat{\delta}^a_{e\text{-}kz}\hat{\boldsymbol{w}}_{kz},$$

where $\hat{\delta}^a_{e\text{-}kz} = (\hat{\pi}_1 - \hat{\pi}_{13})/(\hat{\pi}_1 - 2\hat{\pi}_{13} + \hat{\pi}_3)$ with $\hat{\pi}_{13}$ and $\hat{\pi}_3$ respectively given by

$$\begin{aligned}
\hat{\pi}_{13} &= \frac{1}{\gamma^2}\hat{\theta}_a^2 - \frac{1}{\gamma}\hat{\mu}_e + \frac{1}{\gamma C_2}\left[\hat{\eta}\left(\hat{\mu}_e - \frac{1}{\gamma}\hat{\theta}_a^2\right) + (1-\hat{\eta})\hat{\mu}_g\left(\boldsymbol{w}_e^\top \mathbf{1}_N - \frac{1}{\gamma}\hat{\boldsymbol{\mu}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\right)\right]\\
\hat{\pi}_3 &= \frac{1}{\gamma^2}\hat{\theta}_a^2 - \frac{1}{\gamma^2 C_2}\left(\hat{\theta}_a^2 - \frac{N}{T}\hat{\eta}\right)
\end{aligned}$$

with $\hat{\eta} = \hat{\psi}^2_{a,g}/(\hat{\psi}^2_{a,g} + N/T)$.

C.2.2 Taking the sample GMV as a component

C.2.2.1 Combined with the sample MV portfolio

From Kan and Zhou (2007), on the combination of the sample GMV with the sample MV portfolio, the estimated optimal one is also the three-fund rule as we discussed before, i.e.,

$$\hat{\boldsymbol{w}}^a_{g\text{-}mv} = \hat{\boldsymbol{w}}_{kz}.$$

C.2.2.2 Combined with the sample constrained MV portfolio

From Kan et al. (2022) and Lassance et al. (2024), on the combination of the sample GMV with the sample constrained MV portfolio, the estimated optimal one is

$$\hat{\boldsymbol{w}}^a_{g\text{-}mr} = (1 - \hat{\delta}^a_{g\text{-}mr})\hat{\boldsymbol{w}}_g + \hat{\delta}^a_{g\text{-}mr}\hat{\boldsymbol{w}}_{mr},$$

where the combination coefficient $\hat{\delta}^a_{g\text{-}mr} = C_3\hat{\psi}^2_{a,g}/(\hat{\psi}^2_{a,g} + \frac{N-1}{T})$ with $C_3 = (T-N)(T-N-3)/(T(T-2))$.

C.3 Additional Empirical Results

C.3.1 Parameter Values in the Combined Portfolios

Table C.1: The average combination coefficients for sophisticated constituent portfolios.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}^{LG}_{mv}$	0.0297	0.0361	0.0196	0.0234	0.0183	0.0165	0.0129	0.0061
	(0.0036)	(0.0038)	(0.0029)	(0.0022)	(0.0027)	(0.0027)	(0.0012)	(0.0006)
$\hat{\boldsymbol{w}}^{RF}_{mv}$	0.0288	0.036	0.0192	0.0232	0.0179	0.016	0.0125	0.0061
	(0.0037)	(0.0037)	(0.0025)	(0.0029)	(0.0024)	(0.0024)	(0.0019)	(0.0007)
Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{mv}$	0.3093	0.2275	0.1295	0.1325	0.1616	0.0956	0.0614	0.0222
	(0.1672)	(0.1038)	(0.0663)	(0.0721)	(0.076)	(0.0559)	(0.0367)	(0.0156)
$\hat{\boldsymbol{w}}^{u}_{mv}$	0.3041	0.2274	0.1304	0.1334	0.1628	0.0972	0.0622	0.0223

	(0.1679)	(0.0984)	(0.0643)	(0.0708)	(0.0752)	(0.0549)	(0.0363)	(0.0157)
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}_{kan}^{LG}$	0.0393	0.0687	0.0363	0.0646	0.0369	0.0424	0.0512	0.0857
	(0.0045)	(0.0046)	(0.0024)	(0.0063)	(0.004)	(0.0012)	(0.0036)	(0.0109)
$\hat{\boldsymbol{w}}_{kan}^{RF}$	0.0377	0.068	0.0346	0.0635	0.0357	0.0407	0.0498	0.0855
	(0.0049)	(0.0063)	(0.0037)	(0.0068)	(0.0041)	(0.0045)	(0.0059)	(0.0128)
Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_{kan}^{a}$	0.7027	0.6823	0.6059	0.6642	0.7061	0.5937	0.5602	0.5572
	(0.2766)	(0.217)	(0.3074)	(0.255)	(0.17)	(0.3162)	(0.3455)	(0.4442)
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}_{mv}^{LG}$	0.021	0.0316	0.0149	0.0185	0.0115	0.0122	0.0094	0.0057
	(0.0035)	(0.0029)	(0.0019)	(0.0022)	(0.0017)	(0.0016)	(0.0009)	(0.0005)
$\hat{\boldsymbol{w}}_{mv}^{RF}$	0.0203	0.0313	0.0145	0.0184	0.0114	0.012	0.0092	0.0057
	(0.0026)	(0.0028)	(0.0017)	(0.0025)	(0.0015)	(0.0017)	(0.0013)	(0.0004)
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_{mv}^{a}$	0.2966	0.2194	0.1304	0.1264	0.1522	0.0936	0.0601	0.0210
	(0.164)	(0.0956)	(0.0658)	(0.0692)	(0.0725)	(0.0494)	(0.035)	(0.0149)
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
$\hat{\boldsymbol{w}}_{mr}^{LG}$	0.0243	0.0308	0.014	0.0154	0.0102	0.0096	0.0057	0.001
	(0.0038)	(0.0031)	(0.0022)	(0.0021)	(0.0015)	(0.0014)	(0.0008)	(0.0001)
$\hat{\boldsymbol{w}}_{mr}^{RF}$	0.0237	0.0306	0.0137	0.0153	0.0103	0.0095	0.0057	0.001
	(0.003)	(0.0029)	(0.002)	(0.0018)	(0.0013)	(0.0012)	(0.0008)	(0.0001)
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}_{mr}^{a}$	0.3128	0.2166	0.1175	0.1034	0.1243	0.0722	0.0362	0.0036
	(0.1631)	(0.0921)	(0.0586)	(0.0561)	(0.0586)	(0.038)	(0.021)	(0.0025)

C.3.2 Turnover and Certainty Equivalent Return Net of Trading Costs

Table C.2: Turnover on various datasets.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}_{e\text{-}mv}^{LG}$	1.1872	1.461	0.8563	2.1079	3.5875	1.377	1.4213	7.8727
${}_s\hat{\boldsymbol{w}}_{e\text{-}mv}^{RF}$	1.1959	1.4318	0.7964	2.308	3.1069	1.3143	1.542	7.8682

Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}mv}$	4.5152	4.063	2.2328	4.8314	8.7219	3.1255	3.6802	15.8042
$\hat{\boldsymbol{w}}^{u}_{e\text{-}mv}$	4.508	4.0741	2.289	4.8996	8.8215	3.2231	3.8026	15.9215
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1:The winning probability weighted framework								
${}_{s}\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	0.7381	0.8684	0.3452	1.2023	1.5627	0.5229	0.8112	6.4794
${}_{s}\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	0.7395	0.9245	0.3807	1.2951	1.4824	0.6627	0.8803	6.4328
Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}kz}$	4.2588	3.7066	2.1608	4.9066	8.4522	3.2642	4.4596	19.7271
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
${}_{s}\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	1.7572	1.5386	1.1072	2.4268	4.4247	1.6156	1.9071	10.5956
${}_{s}\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	1.4753	1.5027	1.0665	2.4432	4.1421	1.5688	2.0295	10.3788
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mv}$	4.9787	4.6662	3.0737	6.6639	10.9004	4.6285	6.6671	27.4495
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
${}_{s}\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	2.1477	1.7653	1.2804	2.9501	4.9519	1.7427	2.1805	10.6551
${}_{s}\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	1.8105	1.7067	1.2348	2.7126	4.8824	1.6652	2.1501	10.6717
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mr}$	4.5902	4.1261	2.3229	4.6932	8.4273	3.0821	3.7693	17.4654
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_{e}$	0.0035	0.0036	0.006	0.0056	0.0038	0.0067	0.0069	0.0051
$\hat{\boldsymbol{w}}_{g}$	0.1681	0.2541	0.2215	0.4927	0.6816	0.4051	0.7636	5.9923
$\hat{\boldsymbol{w}}_{mv}$	11.9954	16.4792	18.0441	48.8191	68.2112	44.2371	136.2154	6751.2106
$\hat{\boldsymbol{w}}_{mr}$	12.2131	17.9532	21.7982	67.5194	119.0629	90.5544	542.794	18264.2523
$\hat{\boldsymbol{w}}_{kz}$	4.9787	4.6662	3.0737	6.6639	10.9004	4.6285	6.6671	27.4495
Pabel E.2: Scaled constituent portfolios								
${}_{s}\hat{\boldsymbol{w}}_{e}$	0.0581	0.0395	0.031	0.0369	0.0731	0.0356	0.0334	0.0487
${}_{s}\hat{\boldsymbol{w}}_{g}$	0.6032	0.6624	0.6145	1.1653	2.8807	1.027	1.5551	12.0796

Table C.3: Certainty equivalent return with transaction cost $\kappa = 2bps$.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
${}_{s}\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	5.3149	1.2679	0.4927	3.4439	-0.1385	2.3296	2.0754	-15.449
${}_{s}\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.1969	0.5417	0.0964	2.2221	-1.3052	2.1016	1.661	-15.8962
Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}mv}$	-16.3158	-23.3134	-22.5799	-14.9438	-31.2384	-14.0594	-11.7684	-48.1042
$\hat{\boldsymbol{w}}^{u}_{e\text{-}mv}$	-22.5014	-31.3887	-30.2948	-22.3132	-39.2783	-23.6079	-21.4036	-54.3347

Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1:The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	3.2021	0.6625	-0.0373	3.7551	0.9632	1.0905	1.9507	-13.2486
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	3.6052	0.5216	0.8794	3.4756	0.7062	1.3406	1.8195	-13.0608
Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}kz}$	-23.0755	-24.5704	-26.1092	-15.5439	-31.3055	-19.5669	-17.7136	-60.9617
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	1.9546	-0.4561	2.6118	7.2847	3.6457	0.8902	0.7542	-23.0048
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	1.9511	-0.4241	2.1322	7.1221	2.3144	0.5057	0.8222	-21.3944
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mv}$	-31.5304	-36.5422	-43.1122	-24.8268	-46.8625	-34.1735	-31.4609	-89.9019
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	1.4508	-1.3468	2.737	6.2071	1.1969	1.0298	0.6198	-22.9779
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	0.9529	-0.9903	2.2652	6.797	-0.2424	0.5955	0.5083	-22.7495
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mr}$	-14.3853	-23.4553	-20.8212	-13.7468	-30.9408	-15.4567	-14.2128	-52.5758
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_{e}$	2.3463	2.7342	2.721	1.9916	2.1711	2.4492	2.5783	1.9518
$\hat{\boldsymbol{w}}_{g}$	1.7079	1.5082	3.922	4.4318	3.2344	1.4062	1.2384	-12.2126
$\hat{\boldsymbol{w}}_{mv}$	-118.829	-215.5969	-402.4018	-611.5851	-639.273	-754.7011	-1500.2136	-7756.1298
$\hat{\boldsymbol{w}}_{mr}$	-108.7917	-232.2735	-506.4778	-943.9888	-981.1904	-1293.7267	-3945.7133	-671158.9086
$\hat{\boldsymbol{w}}_{kz}$	-31.5304	-36.5422	-43.1122	-24.8268	-46.8625	-34.1735	-31.4609	-89.9019
Pabel E.2: Scaled constituent portfolios								
${}_s\hat{\boldsymbol{w}}_{e}$	-8.2844	-2.0616	-0.5401	-4.2057	-13.1026	-1.6808	-0.8469	-7.4796
${}_s\hat{\boldsymbol{w}}_{g}$	-5.8567	-2.2923	5.3173	5.9656	-4.0946	-0.6572	-0.0728	-29.6731

Table C.4: Certainty equivalent return with transaction cost $\kappa = 10bps$.

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	-4.2258	-10.475	-6.3806	-13.5198	-29.0639	-8.7456	-9.3566	-79.2049
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	-4.3407	-10.6635	-6.4588	-15.9227	-26.6799	-8.1195	-10.2994	-79.1016
Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}mv}$	-52.9417	-56.4994	-40.7623	-54.6149	-103.1193	-39.7297	-42.0361	-187.1263
$\hat{\boldsymbol{w}}^{u}_{e\text{-}mv}$	-59.057	-64.6244	-48.9366	-62.5625	-111.9245	-50.1055	-52.576	-194.2769
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	-2.732	-6.3255	-2.8101	-5.9275	-11.6372	-3.1132	-4.5738	-66.1954
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	-2.5713	-6.9812	-2.1406	-6.1511	-11.2614	-3.3348	-5.3204	-59.8637

Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}kz}$	-57.6464	-54.8684	-43.6975	-55.8109	-101.1344	-46.326	-53.9148	-232.2251
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	-12.2117	-12.8488	-6.241	-12.2841	-32.0425	-12.132	-14.6053	-108.7273
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	-10.0833	-12.4334	-6.354	-12.9591	-32.1355	-10.5212	-15.5447	-107.1887
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mv}$	-71.9435	-74.7007	-68.1725	-79.4654	-136.9147	-72.1423	-85.8353	-326.5249
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	-15.7976	-15.5639	-7.5559	-17.5847	-38.7681	-13.0136	-16.9435	-109.1292
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	-15.7912	-14.4752	-6.5672	-15.4665	-39.7418	-13.1148	-18.0605	-100.0212
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mr}$	-51.5803	-57.0742	-39.7462	-52.2527	-100.0643	-40.732	-45.1628	-204.1112
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_{e}$	2.3186	2.7056	2.6734	1.9472	2.1405	2.3959	2.5230	1.9112
$\hat{\boldsymbol{w}}_{g}$	0.3590	-0.5301	2.1487	0.4911	-2.2204	-1.8400	-4.8802	-60.3230
$\hat{\boldsymbol{w}}_{mv}$	-213.8900	-346.8129	-533.2948	-921.9033	-1129.0008	-1042.1254	-2024.3138	-22284.0754
$\hat{\boldsymbol{w}}_{mr}$	-205.9582	-375.0456	-666.7662	-1351.0717	-1661.9102	-1690.4056	-4292.25	-38774723.7567
$\hat{\boldsymbol{w}}_{kz}$	-71.9435	-74.7007	-68.1725	-79.4654	-136.9147	-72.1423	-85.8353	-326.5249
Pabel E.2: Scaled constituent portfolios								
${}_s\hat{\boldsymbol{w}}_{e}$	-8.7497	-2.3785	-0.7886	-4.5021	-13.6864	-1.9663	-1.1146	-7.8699
${}_s\hat{\boldsymbol{w}}_{g}$	-10.6951	-7.6081	0.4071	-3.3317	-26.9264	-8.8977	-12.5454	-126.9214

C.3.3 Out-of-sample Sharpe Ratio

For every involved portfolio $\hat{\boldsymbol{w}}_i$, we calculate the out-of-sample Sharpe Ratio according to the following formula:

$$\mathrm{SR}_{\hat{\boldsymbol{w}}_i} = \frac{\mu_{\hat{\boldsymbol{w}}_i}}{\sigma_{\hat{\boldsymbol{w}}_i}},$$

where $\mu_{\hat{\boldsymbol{w}}_i}$ and $\sigma^2_{\hat{\boldsymbol{w}}_i}$ are the mean return and variance of the out-of-sample portfolio returns. Although scaling the portfolio does not change the Sharpe ratio, it aligns with the specific risk aversion desired by the investor. Alternatively, we may rescale the portfolio such that it has a specific level of variance desired by the investor.

Table C.5: Out-of-sample Sharpe Ratio on various datasets (in percentage).

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100

Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	6.9967	5.2714	4.1402	6.9773	6.7668	5.7885	5.7842	3.6954
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	6.9872	5.0056	3.7419	6.7034	5.9982	5.5927	5.6744	3.7050
Panel A.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}mv}$	6.1646	1.9381	-0.985	4.894	3.9757	2.6904	3.535	1.4575
$\hat{\boldsymbol{w}}^{u}_{e\text{-}mv}$	5.7896	1.6456	-0.5216	4.3499	3.679	2.4535	3.1422	1.5129
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	5.5658	4.3089	3.4983	6.3202	5.5042	4.1825	4.8963	3.2461
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.6652	4.428	3.7246	6.0863	5.4456	4.4601	5.0919	3.5096
Panel B.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{e\text{-}kz}$	4.9744	1.8563	-1.8253	5.2388	4.6722	1.6922	3.1211	0.6235
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	6.0184	4.3992	5.6562	9.1668	8.8223	5.2469	5.5794	2.4833
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.7627	4.3486	5.3087	9.0366	8.1433	5.0941	5.7525	2.1916
Panel C.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mv}$	5.2023	2.2057	-1.772	6.5078	5.3679	2.027	3.6181	0.5278
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: the winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	6.3614	4.2279	5.8943	8.9669	8.3821	5.4776	5.8534	2.5581
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.6668	4.3895	5.8639	8.9231	7.8991	5.0272	5.7594	2.1947
Panel D.2: the analytical method in the existing literature								
$\hat{\boldsymbol{w}}^{a}_{g\text{-}mr}$	6.2737	1.2568	-0.114	4.8803	3.4242	2.1568	2.7548	1.0767
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_{e}$	4.0615	4.4602	4.4113	3.7843	3.9126	4.1904	4.3069	3.7455
$\hat{\boldsymbol{w}}_{g}$	3.9211	3.8808	6.787	7.471	6.5828	4.3228	4.7841	2.1413
$\hat{\boldsymbol{w}}_{mv}$	6.6408	2.6607	-0.7658	5.7143	4.9348	2.495	3.182	0.7294
$\hat{\boldsymbol{w}}_{mr}$	6.9571	2.1528	-0.3138	4.7027	3.8239	2.3172	2.6364	0.6882
$\hat{\boldsymbol{w}}_{kz}$	5.2023	2.2057	-1.772	6.5078	5.3679	2.027	3.6181	0.5278
Pabel E.2: Scaled constituent portfolios								
${}_s\hat{\boldsymbol{w}}_{e}$	3.8679	4.2611	4.2366	3.6132	3.7209	4.0072	4.1261	3.5686
${}_s\hat{\boldsymbol{w}}_{g}$	3.7459	3.6563	6.4994	7.2621	6.3845	4.0681	4.6487	2.1010

Table C.6: Out-of-sample Sharpe Ratio with transaction cost $\kappa = 1$ bps (in percentage).

$M = 120, \gamma = 3$	SLT6	LT10	A17	BMOP25	SLT25	I30	I49	SOP100
Panel A: Combining the 1/N rule with the sample MV portfolio								
Panel A.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}mv}$	6.2517	4.2597	3.4068	5.8659	5.2205	4.7408	4.5849	0.0412
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	6.2376	3.9985	3.0535	5.5108	4.6416	4.5751	4.2926	0.0231

Panel A.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^a_{e\text{-}mv}$	5.1805	0.783	-1.8385	3.5864	2.0125	1.5936	2.249	-2.6128
$\hat{\boldsymbol{w}}^u_{e\text{-}mv}$	4.8859	0.6554	-1.2169	3.1953	1.9245	1.5853	2.1497	-2.0357
Panel B: Combining the 1/N rule with the three-fund portfolio								
Panel B.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{e\text{-}kz}$	5.0104	3.6176	3.1021	5.4625	4.7017	3.6696	4.1221	0.0441
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	5.1087	3.7083	3.3097	5.1646	4.6758	3.843	4.2666	0.2874
Panel B.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^a_{e\text{-}kz}$	4.0616	0.8444	-2.6342	3.9832	2.9126	0.6202	1.7945	-4.2351
Panel C: Combining the sample GMV rule with the sample MV portfolio								
Panel C.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mv}$	5.0407	3.3341	4.79	7.7824	7.1183	3.9992	4.0326	-2.7507
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	4.9225	3.2965	4.4673	7.6199	6.5135	3.8771	4.1109	-2.9883
Panel C.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^a_{g\text{-}mv}$	4.2777	1.2103	-2.582	5.2331	3.5709	0.9444	2.1637	-4.8236
Panel D: Combining the sample GMV rule with the sample constrained MV portfolio								
Panel D.1: The winning probability weighted framework								
${}_s\hat{\boldsymbol{w}}^{LG}_{g\text{-}mr}$	5.3901	3.1607	5.0699	7.4385	6.6447	4.2135	4.1752	-2.6438
${}_s\hat{\boldsymbol{w}}^{RF}_{e\text{-}mv}$	4.6583	3.272	4.9321	7.4821	6.1414	3.7652	4.0374	-3.1298
Panel D.2: The analytical method in the existing literature								
$\hat{\boldsymbol{w}}^a_{g\text{-}mr}$	5.2345	-0.0028	-0.9731	3.5586	1.4408	1.0729	1.4365	-3.4508
Panel E.1: Constituent portfolios								
$\hat{\boldsymbol{w}}_e$	3.7548	4.1123	4.0799	3.4738	3.6231	3.8551	3.9692	3.4533
$\hat{\boldsymbol{w}}_g$	3.3706	3.2589	6.0855	6.5459	5.4894	3.4007	3.55	-2.5855
$\hat{\boldsymbol{w}}_{mv}$	5.4771	1.3108	-1.8451	3.7913	2.0037	0.8793	0.8413	-7.2078
$\hat{\boldsymbol{w}}_{mr}$	5.7322	0.7035	-1.47	2.5481	0.4739	0.4867	-0.2114	-11.6756
$\hat{\boldsymbol{w}}_{kz}$	4.2777	1.2103	-2.582	5.2331	3.5709	0.9444	2.1637	-4.8236
Pabel E.2: Scaled constituent portfolios								
${}_s\hat{\boldsymbol{w}}_e$	3.7428	4.103	4.074	3.4667	3.6095	3.8487	3.9634	3.4439
${}_s\hat{\boldsymbol{w}}_g$	3.4258	3.2399	6.0541	6.5427	5.5354	3.3899	3.6011	-2.5664