

Risk Estimation and Portfolio Selection in Large Dimensions

Vincent Weng Choon Tan

A thesis submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Engineering Science
at the
University of Oxford
December 2023

© Copyright by Vincent Tan, 2023.
All rights reserved.

Supervisors: This thesis was supervised by Stefan Zohren.

Number of words: The main body of this thesis contains approximately 60,000 words.

Statement of authorship: The work presented in this thesis is the culmination of extensive research conducted by me during my tenure as a DPhil student at the University of Oxford. No portion of this thesis has been previously submitted for a degree at the University of Oxford or any other educational institution. Chapter 2 is the result of a collaboration with my supervisor, who contributed to the writing and publication efforts. Chapter 3 stems from a collaboration with Christian Bongiorno, who assisted in the implementation of the computer code and supporting algorithms, as well as in research discussions. Chapters 4 and 5 are joint efforts with Nick Firoozye, who contributed to the idea generation, methodological development, and writing. I am the primary and first author of all thesis chapters and take full responsibility for their creation, including conceptualization, methodology, analysis, investigation, implementation, and writing. The main chapters of this thesis have also greatly benefited from numerous discussions held at various seminars. Two of these chapters have been accepted for publication, while the remaining two are accessible as preprints. The corresponding articles for each chapter are listed below:

1. Tan, V., & Zohren, S. (2023). Estimation of large financial covariances: A cross-validation approach. *Journal of Portfolio Management. Forthcoming.*
2. Bongiorno, C., Tan, V., & Zohren, S. (2022). Large covariance estimation by hierarchical clustering and shrinkage. Available at SSRN 4233173.
3. Firoozye, N., Tan, V., & Zohren, S. (2023). Canonical portfolios: Optimal asset and signal combination. *Journal of Banking & Finance, 154*, 106952.
4. Firoozye, N., Tan, V., & Zohren, S. (2023). Dynamic portfolio selection under transaction costs and signal decay. Available at SSRN 4425407.

Acknowledgment of funding: I acknowledge the financial support provided by the Oxford-Man Institute of Quantitative Finance for this thesis.

Vincent Weng Choon Tan, 陳泳椿

Abstract

This integrated thesis consists of four distinct essays centered on the topics of risk estimation and portfolio selection. This thesis is organized into four parts. Part I serves as the introductory section, offering the motivation and background for the main chapters. Parts II and III constitute the core chapters of the thesis. Part IV provides concluding remarks and suggests directions for future research.

Part II addresses a challenging statistical issue often encountered in portfolio selection: the estimation of covariance matrices of security returns. This problem becomes especially acute when dealing with a large number of securities relative to the available observations, and when security returns exhibit temporal variations and cross-sectional structures. Part III then shifts the focus to the development of innovative dynamic portfolio selection techniques that are both theoretically grounded and suitable for large-scale financial applications.

In Part II, Chapter 2 introduces a novel non-parametric structure-agnostic shrinkage method. The method leverages the principle of cross-validation to mitigate the adverse effects of overfitting in large financial covariances that are measured using an exponentially-weighted average scheme. Chapter 3 expands on this concept by leveraging the hierarchical structure evident in financial returns to achieve additional efficiency improvements.

Transitioning to Part III, Chapter 4 delivers a closed-form investment solution to a dynamic portfolio selection problem that can ingests multiple return-predictive signals and is scalable to large dimensions. Moreover, it also studies the joint correlation structure of the solution through a dimension reduction technique known as Canonical Correlation Analysis. Chapter 5 extends the ideas from the previous chapter to a multi-period setting that considers transaction costs and multiple signals with different persistent properties. This is achieved by formulating the investment problem as a dynamic linear rational expectations model, for which a closed-form solution is provided. The chapter also studies the long-run implications of the optimal portfolio policy.

Acknowledgements

I wish to express my profound gratitude to the many individuals whose unwavering support has been crucial in my journey to attain this degree. Specifically, I extend my heartfelt thanks to:

My family—my parents, Donald and Rachael, as well as the many members of the Tan and Twong families—for your unconditional love and encouragement. I am deeply grateful to my parents for their relentless dedication and hard work, as well as the struggles they have endured over the years to provide my siblings and me with greater opportunities in life;

Stefan Zohren, for giving me the opportunity to pursue a PhD and the academic freedom to independently shape this thesis according to my creative vision and background in economics, ultimately transforming it into a reality;

Nick Firoozye, for his mentorship, wisdom, feedback on my research, and friendship. I have benefited immensely from his deep knowledge of finance and economics and extensive industry experience as well as his expertise in modern machine learning techniques applied to the discipline. His measured and disciplined approach to financial modeling, combined with his boundless creativity, has consistently served as a source of ongoing inspiration for me. I eagerly anticipate future collaborations. Terima Kasih;

Mihai Cucuringu, for his invaluable feedback during my Transfer and Confirmation of Status, which have enriched this thesis;

Fabio Caccioli and Doyne Farmer, for inspiring me to transition into economics, writing my references in support of my university and job applications, as well as guiding me on how to do rigorous academic research in the early phase of my career;

Seminar participants at Oxford University and various firms as well as academic researchers and professionals from around the world, for their invaluable comments and constructive feedback on my research.

To my family

Contents

Abstract iii

Acknowledgements v

List of Tables xi

List of Figures xiii

I Introduction 1

1 Motivation and Background 3
1.1 Motivation for Part II . . . 4
1.2 Background for Part II . . . 6
1.2.1 Structured-based Estimation . . . 6
1.2.2 Linear Shrinkage Estimation . . . 7
1.2.3 Nonlinear Shrinkage Estimation . . . 8
1.3 Motivation for Part III . . . 9
1.4 Background for Part III . . . 11
1.4.1 Dynamic Portfolio Selection . . . 11
1.4.2 Multi-Period Portfolio Selection and Rational Expectations . . . 12
1.5 Outline of Main Chapters . . . 14

II Risk Estimation 17

2 Estimation of Large Financial Covariances: A Cross-Validation Approach 19
2.1 Introduction . . . 19
2.2 Setting the Stage . . . 21
2.2.1 The Weighted Sample Covariance Matrix . . . 21
2.2.2 Exponential Weighted Average Covariance Matrix . . . 22
2.2.3 Decomposing the Covariance Matrix . . . 22
2.3 Spectrum of Exponential Weighted Average Covariances . . . 23
2.4 Shrinkage Estimation . . . 24
2.4.1 Covariance for Portfolio Selection . . . 24
2.4.2 Out-of-Sample Variance Estimation by Cross-Validation . . . 25
2.4.3 Related Literature . . . 26

2.4.4 Comparison to DCC-NL . . . 27
2.5 Simulation Study . . . 28
2.5.1 Return Generating Process . . . 28
2.5.2 Candidate Estimators . . . 28
2.5.3 Performance Assessment . . . 29
2.5.4 Simulation Results . . . 29
2.5.5 Sensitivity Analysis to the Number of Folds . . . 30
2.6 Empirical Analysis . . . 30
2.6.1 Data and Portfolio Construction Rules . . . 30
2.6.2 Base Candidate Portfolios . . . 31
2.6.3 Evaluation Methodology . . . 31
2.6.4 Global Minimum Variance Portfolio . . . 32
2.6.5 Robustness Checks . . . 33
2.6.6 Markowitz Portfolio with Signal . . . 35
2.6.7 Robustness Checks . . . 37
2.7 Conclusion . . . 38

3 Large Covariance Estimation by Hierarchical Clustering and Shrinkage 41
3.1 Introduction . . . 41
3.2 Hierarchically Nested Factor Model . . . 43
3.2.1 Notation . . . 43
3.2.2 Hierarchical Clustering . . . 43
3.2.3 Cluster Analysis . . . 44
3.2.4 Hierarchically Ordered Correlation Matrix . . . 45
3.2.5 The Regularized Covariance Matrix . . . 45
3.2.6 Factor Models . . . 46
3.2.7 Hierarchical Trees and Factor Models . . . 48
3.3 Proposed Estimation Scheme . . . 48
3.4 Empirical Analysis . . . 49
3.4.1 Data and Portfolio Construction Rules . . . 49
3.4.2 Candidate Portfolios . . . 50
3.4.3 Evaluation Methodology . . . 50
3.4.4 Global Minimum Variance Portfolio . . . 51
3.4.5 Robustness Checks . . . 52
3.4.6 Markowitz Portfolio with Signal . . . 53
3.5 Conclusion . . . 54

III Portfolio Selection 55

4 Canonical Portfolios: Optimal Asset and Signal Combination 57
4.1 Introduction . . . 57
4.2 Setting the Stage . . . 59
4.2.1 Notation . . . 59
4.2.2 Dynamic Strategies . . . 60
4.2.3 Fully Invested Constraint . . . 64

4.2.4 Relation to Existing Literature . . . 64
4.3 Strategy Diversification . . . 66
4.3.1 Canonical Correlation Analysis . . . 66
4.3.2 Relation to PCA, PLS, and RRR . . . 68
4.3.3 Reformulation as Canonical Portfolios . . . 69
4.3.4 Canonical Portfolio Analysis . . . 70
4.3.5 Static and Dynamic Returns Decomposition . . . 72
4.3.6 Estimation . . . 73
4.4 Empirical Analysis . . . 74
4.4.1 Data and Portfolio Construction Rules . . . 74
4.4.2 Signal Construction . . . 74
4.4.3 Candidate Portfolios . . . 75
4.4.4 Evaluation Methodology . . . 76
4.4.5 Application of Canonical Portfolio Analysis . . . 76
4.4.6 Empirical Results . . . 77
4.4.7 Robustness Checks . . . 78
4.4.8 Extension: Two-Signal Case . . . 80
4.5 Conclusion . . . 80

5 Dynamic Portfolio Selection under Transaction Cost and Signal Decay 83
5.1 Introduction . . . 83
5.2 Modeling Framework and Approach . . . 86
5.2.1 Notation . . . 86
5.2.2 The Investor's Problem . . . 86
5.2.3 The Euler Equation . . . 89
5.2.4 The Analytical Solution . . . 90
5.2.5 The Prediction Problem . . . 92
5.3 Modeling Extensions . . . 93
5.3.1 Portfolio Growth . . . 93
5.4 Impulse Response Functions . . . 95
5.4.1 Temporary Price Impact . . . 95
5.4.2 Temporary and Permanent Price Impact . . . 96
5.5 Performance in Steady-State . . . 97
5.5.1 Turnover . . . 97
5.5.2 Sharpe Ratio . . . 98
5.6 Conclusion . . . 99

IV Conclusion 101

6 Concluding Remarks 103

A Mathematical Derivations 107
A.1 Derivations for Chapter 4 . . . 107
A.2 Derivations for Chapter 5 . . . 112
A.2.1 Proof of Proposition 6 . . . 112

A.2.2 Proof of Proposition 7 . . . 112
A.2.3 Proof of Proposition 8 . . . 114
A.2.4 Proof of Proposition 9 . . . 115
A.2.5 Proof of Proposition 10 . . . 116
A.2.6 Proof of Proposition 11 . . . 116
A.2.7 Proof of Proposition 12 . . . 117
A.2.8 Proof of Proposition 13 . . . 118

B Tables **121**
B.1 Tables for Chapter 2 . . . 121
B.2 Tables for Chapter 3 . . . 133
B.3 Tables for Chapter 4 . . . 142
B.4 Tables for Chapter 5 . . . 153

C Figures **155**
C.1 Figures for Chapter 1 . . . 155
C.2 Figures for Chapter 2 . . . 156
C.3 Figures for Chapter 3 . . . 158
C.4 Figures for Chapter 4 . . . 161
C.5 Figures for Chapter 5 . . . 164

References **167**

List of Tables

B.1.1 Performance Summary . . . 121
B.1.2 Weight Statistics . . . 122
B.1.3 Subsample Analysis . . . 123
B.1.4 Sensitivity Analysis . . . 124
B.1.5 No Short-Sale Constraint . . . 125
B.1.6 Devolatization . . . 126
B.1.7 Performance Summary . . . 127
B.1.8 Weight Statistics . . . 128
B.1.9 Subsample Analysis . . . 129
B.1.10 Sensitivity Analysis . . . 130
B.1.11 No Short-Sale Constraint . . . 131
B.1.12 Devolatization . . . 132
B.2.1 GICS color code . . . 133
B.2.2 Performance Summary . . . 134
B.2.3 Weight Statistics . . . 135
B.2.4 Subsample Analysis . . . 136
B.2.5 Sensitivity Analysis . . . 137
B.2.6 Performance Summary . . . 138
B.2.7 Weight Statistics . . . 139
B.2.8 Subsample Analysis . . . 140
B.2.9 Sensitivity Analysis . . . 141
B.3.1 Summary of PCA and CCA . . . 142
B.3.2 Performance Summary . . . 143
B.3.3 Weight Statistics . . . 144
B.3.4 Portfolio Returns Decomposition . . . 145
B.3.5 Subsample Analysis . . . 146
B.3.6 Forecast Horizon . . . 147
B.3.7 Shrinkage Intensity . . . 148
B.3.8 Momentum Lookback Window . . . 149
B.3.9 Approximate versus Actual Solution . . . 150
B.3.10 Performance Summary for Two Signals Case . . . 151
B.3.11 Weight Statistics for Two Signals Case . . . 152
B.4.1 Parameter Values for Numerical Experiments . . . 153

List of Figures

C.1.1 Mean-Variance Efficient Frontier . . . 155
C.1.2 Marčhenko-Pastur Law . . . 155
C.2.1 Marčhenko-Pastur Law for the Exponentially-Weighted Sample Covariance Matrix . . . 156
C.2.2 Schematic Overview of Five-Fold Cross-Validation . . . 156
C.2.3 Out-of-Sample Risk Improvement of Different Estimators . . . 157
C.3.1 Binary Tree of Five Elements . . . 158
C.3.2 Dendrogram of Stocks Returns . . . 158
C.3.3 Heatmaps of Stock Returns Correlation Matrix . . . 159
C.3.4 Heatmap of Regularized Stock Returns Correlation Matrix . . . 159
C.3.5 Heatmaps of Residual Stock Returns Correlation Matrix . . . 160
C.3.6 Heatmap of Regularized Residual Stock Returns Correlation Matrix . . . 160
C.4.1 Wachter Law . . . 161
C.4.2 Optimal Singular Value Adjustment Versus Linear Approximation . . . 161
C.4.3 In-Sample and Out-of-Sample Canonical Portfolio Returns . . . 162
C.4.4 Weights for the Top Canonical Portfolio . . . 163
C.5.1 Impulse Response to Shock under Temporary Price Impact . . . 164
C.5.2 Steady-State Performance Analysis . . . 165

Part I

Introduction

Chapter 1

Motivation and Background

> The journey of a thousand miles begins with a single step.
>
> Lao Tzu, *Tao Te Ching*, ca. 400 BC

Investing is a human endeavor. It is a human enterprise. It involves the allocation of valuable financial resources sourced from a variety of stakeholders, including pensioners, government entities, charitable organizations, individual investors, and more, with the shared expectation of generating a return or profit over time. Since financial resources are often scarce, they must be responsibly managed at designated levels of risk. Investing also involves making difficult economic decisions under uncertainty when the future returns of securities are unknown. While investing itself can be done through various mechanisms, algorithms, and tools, it ultimately relies on human decision-making and action.

Nevertheless, there is a systematic investment process that guides institutional investors to make informed decisions in the face of these challenges. This process was first conceived by the late American economist Professor Harry M. Markowitz, a recipient of the Nobel Prize in Economic Sciences in 1990. Markowitz's groundbreaking work marked a turning point in the study of finance, dividing it into two distinct eras: before and after 1952. His seminal article *Portfolio Selection* published in the Journal of Finance in 1952 laid the foundation for what is now known as Modern Portfolio Theory (MPT), which then created an entire investment management industry. With trillions of dollars worth of institutional assets invested according to his framework over the past century, it is evident that his ideas have withstood the test of time and continue to shape investment practices to this day.

The enduring popularity of Markowitz's framework can be attributed to two key tenets: its ability to account for the impact of diversification on risks of correlated securities, and its consideration of the risk or return trade-off in investment decisions. By emphasizing the importance of diversification, his framework recognizes that spreading investments across different assets can help mitigate risks associated with individual holdings.[1] By allowing investors to balance between risk and return, his framework acknowledges the fact that higher returns often accompany with higher levels of risk. Both considerations naturally align with

[1] The benefits of diversification have been recognized for centuries with references found in Sun Tzu's *Art of War*, the biblical texts, William Shakespeare's *The Merchant of Venice*, as well as many others, and is often expressed through the timeless adage, "Do not put all your eggs in one basket." MPT attempts to provide a formal description and an adequate theory that encompasses this concept.

the mindset of *real-world* money managers, as they seek to prioritize their fiduciary responsibilities by constructing investment portfolios that are in line with their risk tolerance and long-term financial goals.

Given the lasting significance of Markowitz's framework, this integrated thesis is divided into two parts with the objective of advancing his research paradigm. The goals for each part are summarized as follows:

- Part II delves into the issue of risk estimation within the mean-variance analysis framework. The most basic measure of risk as studied by Markowitz (1952) is given by the variance of a portfolio, which in turn relies on the variances and covariances of individual returns. The estimation of these covariances presents a significant challenge, particularly when dealing with a large number of securities relative to the number of available observations.
- Part III focuses on constructing portfolio selection methods that are *dynamic* in nature. While traditional mean-variance analysis facilitates the determination of an optimal weighting scheme to invest in each security, it provides a *static* (or fixed) set of weights that are often computed with historical averages of relevant statistics over a given period. However, in the realm of active management, dynamic methods are often preferred as they are able to adapt to the current state of the economy and respond accordingly.

Having outlined the goals of the two parts of the thesis, the motivation and background for each part will now be presented.

1.1 Motivation for Part II

Traditionally, the MPT framework requires two inputs given the investor's information set: (1) a vector of expected returns, and (2) a covariance matrix of security returns. With these inputs, MPT enables us to trace out the set of portfolios that offer the highest expected return for a specific level of risk or the lowest risk for a given level of expected return. This set of optimal portfolios is referred to as the *efficient frontier*. Figure C.1.1 illustrates the efficient frontiers for portfolios consisting solely of risky securities, as well as for combinations of risky securities and a risk-free security.[2]

Unfortunately, both expected returns and covariance of returns are unknown to investors and so they have to be estimated. Since individual expected returns can often be subjective and influenced by a manager's biases, it would be more judicious for the purpose of this thesis to refrain from taking a position on the source of these expected returns. Therefore, our focus is primarily directed toward the estimation covariance of security returns, given its pivotal role as the central input in Markowitz's scheme.

Generally, the concept of covariance is an unintuitive one but it can be better understood in terms of its normalized counterpart, correlation. Correlation is a quantitative measure that assesses the degree to which two random variables or series exhibit a mutual association in

[2] A risk-free asset refers to an investment instrument that entails no risk or uncertainty. When investors select this instrument, they expect a guaranteed rate of return, known as the risk-free rate. Although the concept of a completely risk-free asset does not hold true in reality, as every investment carries a certain level of risk, the return on a three-month Treasury bill commonly serves as the benchmark for the risk-free rate, given that it is backed by the U.S. government, perceived to have a low risk of default.

their directional fluctuations. In the scenario where they move perfectly in sync, the correlation attains a value of $+1$. Conversely, if they move in exact opposition, the correlation takes on a value of -1. However, when the average behavior of one variable fails to inform the other, the correlation is zero, indicating a lack of correlation or independence between the variables.

The significance of the covariance in finance cannot be overstated. To illustrate this point, let us consider a pathological example: suppose a particular security exhibits a high potential for substantial returns, albeit with a small probability of failure. Would it be considered a relatively safe investment if a small portion of the portfolio is allocated to this security? Unfortunately, the answer is not necessarily so, especially if the entire portfolio consists of similar investments that are susceptible to simultaneous failure. Indeed, it is possible that a shared underlying technology, legal decision, or some other economic factor could lead to a synchronous downfall of all such investments, resulting in a correlation of one. Consequently, the portfolio would lack the benefits of diversification.

Having acknowledged the important role the covariance matrix plays in risk management, obtaining reliable estimates of it with data does present significant challenges. Since the true (or population) covariance matrix is unknown, the first approach is to use the sample covariance matrix as an estimate for the true covariance. However, this approach becomes problematic when the number of variables is as large as the number of observations. To illustrate this issue, let us consider an example: suppose we have a dataset with $T = 2{,}000$ days of observations, which corresponds to a span of 8 trading 'years.' While this may seem like a substantial amount of observations, it fails to account for the breadth of the investment universe. For instance, with $N = 1{,}000$ securities, this situation is akin to using only two observations to estimate the variance of a single random variable.

To put things into perspective, imagine a set of 1,000 securities. Accurately estimating a $1{,}000 \times 1{,}000$ covariance matrix would require a significantly larger number of observations, such as at least 10,000. Achieving this amount of data would necessitate approximately 40 years of daily data. However, relying on data spanning such a long period (even if they are available) would also leave us expose to issues around nonstationarities, which can compromise the reliability of our estimates. Essentially, when the number of securities is of similar order as the available historical return observations, which is common in real-world scenarios, the sample covariance matrix becomes prone to estimation errors.

This issue is more than just theoretical; it has real-world financial consequences. The research of Jobson and Korkie (1980) underscores the inadequate results when using the sample covariance matrix for portfolio selection. In particular, when this matrix is used for portfolio optimization, it can inadvertently magnify its errors, a phenomenon known as "error maximization" as coined by Michaud (1989). Since the optimizer assigns higher weights to securities with high expected returns, low relative variance, and low correlations, the portfolio disproportionately leans towards securities with these characteristics. Consequently, it drives most securities out of the portfolio, resulting in 'corner solutions' that exhibit poor out-of-sample performance. This issue becomes more pronounced as the number of assets in the portfolio increases; see Chopra and Ziemba (1993), Britten-Jones (1999), Best and Grauer (1991), Kan and Zhou (2007), among others. Furthermore, as demonstrated by Kan and Zhou (2007) and Zhou (2008), optimized portfolios often underestimate risk due to estimation errors in the sample covariance matrix used for portfolio construction.

1.2 Background for Part II

In light of the previous discussion, numerous covariance estimators have been proposed in the literature to overcome the poor out-of-sample properties of the sample covariance matrix. Given the plethora of covariance estimators that have been proposed in the literature, this section discusses a selection of representative estimators that provide valuable insights into this area of study within the context of finance. These estimators fall into two categories: (1) structured-based estimation and (2) shrinkage estimation. Chapter 2 delves into the second category, whereas Chapter 3 tackles both.

1.2.1 Structured-based Estimation

Imposing restrictions on the covariance matrix is a straightforward way to reduce the degrees of freedom of a covariance. This results in a 'structured' matrix. Often, their design relies on domain knowledge and thus, have to be custom-tailored to each application. The advantage of this estimator is that they are better conditioned than the sample covariance matrix.

The simplest 'estimator' in this category is the isotropic covariance matrix, which is a scaled multiple of the identity matrix. This estimator imposes a great deal of restrictions on the covariance matrix: It assumes zero covariances and identical variances for all variables. There is only one free parameter as opposed to $O(N^2)$ for the sample covariance matrix. This parsimonious structure reduces the estimation error that is present in the sample covariance matrix, but introduces significant bias.

To introduce greater modeling flexibility, it is necessary to relax the isotropic restriction. For example, one can assume all stock returns have the same variance and all pairs of stocks have the same covariance; this approach contains two free parameters as discussed by Jobson and Korkie (1980) and Frost and Savarino (1986). Alternatively, one can consider a diagonal covariance matrix with different variances (requiring N parameters), or a covariance with different variances and equal correlations (requiring $N+1$ parameters) as suggested by Elton and Gruber (1973).

Another structured estimator that is of interest in financial economics is the factor model. These models have a long history in finance dating back to prominent applications in asset pricing such as the Capital Asset Pricing Model (CAPM) of Sharpe (1964), and the Arbitrage Pricing Theory (APT) of Ross (1976). The way factor models are classified in financial economics crucially depends on the assumptions that one is willing to take on the factors, factor loadings, and on the idiosyncratic component. For example, Sharpe (1963) single index model assumes that the factors are *observed* by a broad market-based index and the idiosyncratic risks between stocks are uncorrelated. The Sharpe's single index model results in $2N+1$ risk parameters that require estimation. More generally, if there are K (observed) indices, then there is a total of $(K+1)(N+K/2)$ free parameters. When $K \ll N$, then this represents much lower degrees of freedom than the sample covariance matrix.

While factors models have their place in finance, they can be easily misspecified and hence, be severely biased. For instance, the observation of stocks within the same industry moving in tandem directly contradicts the assumption of Sharpe's single-factor model, which requires the covariance matrix of the idiosyncratic terms to be diagonal. While introducing multiple factors might mitigate this issue, there is no guarantee that each additional factor will be free from measurement errors or serve as a reliable proxy for the true underlying factors.

1.2.2 Linear Shrinkage Estimation

The idea of shrinkage estimation, as introduced by Professor Charles Stein in the 1950s, is a statistical method that improves the estimation of unknown parameters by shrinking the estimated values towards a prior or target value. This concept was first conceived by Stein in his seminal paper Stein (1956a), and James and Stein (1961) in the context of estimation of the mean vector.

Stein's idea received widespread acclaim because it challenged the traditional approach of using maximum likelihood estimation (MLE), which tends to produce unstable or inefficient estimates, especially when dealing with large-dimensional data or when the number of parameters to be estimated is large relative to the sample size. Stein's shrinkage estimation addresses this issue by combining the MLE estimates with a biased estimator that pulls the estimates closer to a prior or target value. That is, by shrinking the estimates towards a common value, Stein's method reduces the overall mean squared error (MSE) of the estimates, resulting in improved efficiency, and a more favorable bias-variance trade-off.

The remarkable aspect of Stein's idea is that it applies even when the prior or target value is completely arbitrary. This concept gave rise to what is commonly known as Stein's 'paradox.' The paradox states that for dimensions $N > 3$, there exist estimators that outperform the MLE in terms of MSE. In other words, these alternative estimators have the potential to achieve a smaller expected error compared to the MLE. This finding shocked many researchers around the world in the 1960s because it challenged the prevailing notion that improving upon the MLE would be unlikely when analyzing data that come from independent sources governed by unrelated means. Stein's paradox demonstrated that, in a broader context, it is indeed possible to improve estimation in a global sense by leveraging the interplay between parameters, even when they are independent and have no direct relationship with each other.

Moreover, the concept of shrinkage can be extended to apply not only to the means but also to the covariance matrices. In this context, Haff (1980) proposed a linear shrinkage method with an empirical Bayesian interpretation, which shrinks the sample covariance matrix toward an identity matrix. However, Haff's method lacked optimality results. Further progress was made by Ledoit and Wolf (2004b), who obtained optimality results in terms of MSE in finite samples. The authors developed a fast and numerically robust method for estimating the optimal shrinkage intensity. Additionally, this approach can be extended to target matrices beyond the isotropic covariance. In the field of finance, for example, tailored target matrices have been devised, such as the constant correlation model proposed by Ledoit and Wolf (2004a) and the Sharpe single-factor model introduced by Ledoit and Wolf (2003).

The linear shrinkage estimator naturally leads to the classic bias-variance trade-off, which is a fundamental concept in statistics and machine learning. If the shrinkage intensity is close to one, it defines a low-variance but high-bias estimator, whilst if shrinkage intensity is close to zero, it defines a low-bias but high-variance estimator. Thus, shrinkage is a form of *regularization*, where its intensity controls the strength of regularization. One of the advantages of linear shrinkage is its ability to operate with only $O(1)$ degree of freedom. However, this convenience also brings about limitations in terms of its flexibility to effectively shrink the contents of the matrix, as it relies on a single shrinkage intensity. Furthermore, it requires careful consideration and selection of an appropriate target matrix.

1.2.3 Nonlinear Shrinkage Estimation

A useful approach to gaining insight into a covariance matrix is through an operation known as spectral decomposition. This operation allows us to extract a set of eigenvalues and their corresponding eigenvectors, which are key low-dimensional objects that summarize the contents of the matrix. Since the 1950s, Charles Stein and others observed that sample eigenvalues exhibit a systematic bias following a distinct pattern. Specifically, the smallest eigenvalues tend to be too small and the largest eigenvalues tend to be too large.

To address this issue, the concept of spectrum estimation arises, wherein the sample eigenvalues are shrunked by some nonlinear transformation while retaining the sample eigenvectors from the original covariance matrix. The idea is to protect the user from the consequences of the noisiness of those eigenvectors as well as the biases embedded in the eigenvalues. This concept goes back at least 60 years and was first understood by Stein (1956b) as being a possibility. Stein (1975) introduced an early nonlinear shrinkage method to correct the bias of each eigenvalue by optimizing Stein's unbiased risk estimate (SURE). Since then there have been hundreds of papers about ideas of how to shrink the eigenvalues of covariance matrices. At this point, it is worth noting that the spectrum estimators that Stein developed were from a decision-theoretic perspective. Here, the focus was on sample sizes where $T > N$ with small N (e.g., $N = 10$ or 50). However, since the 1990s, the advent of the modern computer had allowed a different kind of asymptotics to emerge, namely $N, T \to \infty$ (and possibly $N > T$), with their ratio N/T (known as the *concentration*) converging to a finite, nonzero limit. As a result, random matrix theory had become increasingly important for the analysis of large-dimensional systems.

The full extent to which the machinery of random matrix theory can be harnessed for the purpose of covariance matrix estimation proved to be challenging for many researchers. This difficulty arises due to a fundamental difference between the type of inference typically pursued in traditional random matrix theory research problems and the one relevant to covariance estimation. In traditional studies, the aim is to infer the distribution of sample eigenvalues from a postulated distribution of population eigenvalues. This *direct* inference problem allows researchers to make predictions from a phenomenological standpoint. However, in the context of covariance matrix estimation, the focus shifts towards the *inverse* of this inference problem: inferring the population eigenvalues from a set of given sample eigenvalues. Here, the underlying concept is based on the notion that if one possess knowledge about the true distribution of the eigenvalues of a matrix and consider the double asymptotic limit with N/T converging to a finite value, one can determine the biases that are embedded in the sample eigenvalues. This is achieved by inverting the distribution, allowing us to establish a mapping between the sample eigenvalues and their corresponding population counterparts. This mapping effectively shrinks the sample eigenvalues towards more central values.

This key insight led El Karoui (2008) and Ledoit and Péché (2011) to make the early progress to invert the celebrated Marčenko and Pastur (1967) law with the latter contributing to the development of the second-generation large-dimensional asymptotic spectrum estimator.[3] Notably, the seminal work of Ledoit and Péché (2011) provided an optimal theoretical shrinkage formula within the rotational equivariance class with respect to the

[3] A visualization of this law is provided in Figure C.1.2.

Frobenius loss. This was achieved by analyzing the inner product between true and sample eigenvectors in the large-dimensional asymptotic limit.[4] Bouchaud and Potters (2020) refer to the solution from Ledoit and Péché (2011) as a "mathematical miracle" because their formula depended only on observable quantities, which only can be achieved in the large-dimensional limit by virtue of the law of large numbers.

However, the implementation of the theoretical shrinkage formula of Ledoit and Péché (2011) with real *discrete* data proves to be a delicate problem. The reason is that formula hinges on the Marčhenko-Pastur law, which depends on a complex-valued function, the Stieltjes (1894) transform, that relates both the *limiting* sample and population spectral distribution. Instead, we would like a function that describes the relationship between the sample eigenvalues and its population counterparts. To address this problem, Ledoit and Wolf (2015) discretize the Marčhenko-Pastur law to derive a deterministic multivariate function called Quantized Eigenvalues Sampling Transform (QuEST) in order to model this relationship. A numerical inversion of the QuEST function provides an estimate for the population eigenvalues, which allows the computation of the Stieltjes transform, and hence, the optimal shrinkage function. More recently, Ledoit and Wolf (2020) borrowed the technology of nonparametric kernels from Jing et al. (2010) to estimate the limiting density of the sample covariance matrix in order to compute Stieltjes transform from discrete data. This analytical approach has an advantage over the numerical approach as it offers a more streamlined and transparent implementation of the estimator.

Finally, an interesting and different approach that optimizes the *same* Frobenius loss function as that of the analytical nonlinear shrinkage estimator is through a *numerical* procedure which was introduced by Abadir et al. (2014) and further developed by Lam (2016). What sets this method apart is its ability to bypass the estimation of the 'oracle' shrinkage function, a prerequisite for analytical nonlinear shrinkage. This innovative method draws inspiration from the well-established principle of cross-validation widely used in statistics and machine learning to tackle overfitting. One of the key advantages of cross-validation is its conceptual simplicity compared to analytical nonlinear shrinkage, making it accessible to users without prior knowledge of random matrix theory. This approach is what Chapter 2 aims to advance.

1.3 Motivation for Part III

The traditional mean-variance analysis assumes that investors are myopic, focusing only on the current period and a constant investment opportunity set (Mossin, 1968). It neglects the potential impact of future economic conditions and investment opportunities, which are subject to change and uncertainties over time. To illustrate, consider a scenario in January 2023 where the goal of an investment manager is to construct a portfolio for the upcoming five years. By applying the standard mean-variance framework and calculating its statistical moments using historical data, the manager would determine the fixed efficient portfolio weights. This approach represents an unconditional strategy, where the only motive to rebalance the portfolio

[4] A different derivation of the optimal shrinkage formula was provided by Bun et al. (2017) using the concept of free probabilities introduced in Voiculescu (1985, 1991). Furthermore, the recent work of Burda and Jarosz (2022) extended these derivations to covariances with correlated samples.

each period is to maintain the assigned fixed weights.[5] However, this treatment completely overlooks crucial information about the state of the economy at the end of December 2022, or any other present information, for that matter.

In fact, even with precise predictions of future returns, volatilities, and correlations (excluding higher moments such as skewness and kurtosis) and an accurate estimation of the investor's risk aversion, there is no guarantee that this myopic portfolio will consistently remain optimal.[6] Indeed, it is not difficult to imagine a portfolio that can outperform the traditional Markowitz implementation. For example, consider two assets: cash and the S&P 500 index. One can plot a standard efficient frontier that connects these two assets. Now, consider an alternative frontier achievable through dynamic trading. Specifically, consider a fairly reliable forecasting model that recommends allocating 100% of our investments to the S&P 500 when it predicts a positive return, and switching to a cash position when it predicts a negative return for the S&P 500. This dynamic strategy has the potential to outperform the standard frontier between the two points. Additionally, since the portfolio predominantly holds cash, its volatility is a fraction of that of the S&P 500. As a result, lower volatility is attained while generating returns that exceed those of the S&P 500.

The simple example above shows that the predictability of returns and/or of their variability can create new investment opportunities for the investor to improve upon the existing standard efficient frontier by pursuing dynamic trading strategies that are conditional on the current state of the economy. However, it is crucial to acknowledge that this discussion does not cover several other important factors, including transaction costs, taxes, an individual's labor income, and the stage of their investment cycle. These elements can significantly influence the overall performance and results of a portfolio.

Transaction costs are often overlooked or oversimplified in a portfolio selection algorithm by incorporating constant proportional transaction costs. Research articles that investigate transaction costs on portfolio selection include Davis and Norman (1990), Duffie and Sun (1990), Balduzzi and Lynch (1999), Leland (2000), and Lynch and Tan (2010). It is crucial to account for these costs as they can substantially influence short-term predictions. For instance, if a short-term forecast recommends complete divestment from all U.S. stocks in the upcoming month, this involves significant transaction fees. Additionally, if selling stocks in a different country were necessary, the process would become even more complex (involving legal requirements, currency conversion, and potentially dealing with foreign financial institutions), resulting in higher transaction costs. Thus, the potential gains from the forecasted strategy may be significantly diminished by the incurred transaction costs. In contrast, longer-term forecasts are less affected by transaction costs, as they allow for the gradual amortization of these expenses over an extended holding period; for example, if one holds the portfolio for five years, the impact of transaction costs is significantly reduced.

Taxes also play a pivotal role in financial decisions particularly when it comes to long- and short-term capital gains. The research conducted by Constantinides (1983, 1984) sheds light

[5] This is essentially a contrarian trading strategy, which involves buying underperforming assets and selling outperforming ones in order to maintain the (fixed) positions.

[6] Static and dynamic portfolios only coincide under specific conditions related to utility, such as unit elasticity of intertemporal substitution (e.g., log utility), or security returns, where changes in investment opportunities and asset returns are uncorrelated. However, research findings suggest that it is empirically implausible for either of these conditions to hold true; see for example Campbell and Viceira (1999).

on the significance of accounting for tax liabilities in these scenarios. The idea presented in these papers is that investors have the potential to minimize their tax obligations by strategically deferring capital gains and capitalizing on realized losses. This strategic approach allows investors to offset their current income or other capital gains, ultimately optimizing their overall tax liability.

Finally, human capital and its risk have the potential to affect investment decisions and optimal savings; see research studies from Viceira (2001), Cocco et al. (2005), Benzoni et al. (2007). Labor income refers to the continuous stream of income payments received by an investor over time. From an economic standpoint, an individual's labor income can be regarded as a dividend or yield on their implicit ownership of human capital. Human capital, being a nontradable security, cannot be sold in the form of claims against future labor income for moral hazard or ethical reasons. Additionally, the uncertainty surrounding future labor earnings is often prevalent among most individuals, stemming from various factors such as the nature of their profession or the prevailing economic conditions. Consequently, human capital becomes a risky, rather than a secure, nontradable security. This risk factor can impact decision-making since the trajectory of income determines how a portfolio evolves over time.

1.4 Background for Part III

Building on our previous discussion, this section examines the existing research on dynamic trading strategies. Our primary focus centers on dynamic strategies based on: (1) single-period optimization and (2) multi-period optimization that considers quadratic transaction costs. Chapter 4 addresses the former, while Chapter 5 delves into the latter.

1.4.1 Dynamic Portfolio Selection

As we have discussed, dynamic trading strategies provide clear benefits to investors by capitalizing on time-varying investment opportunities.[7] However, deriving these optimal strategies can be challenging. Several numerical methods have emerged as potential solutions to this problem given the dearth of readily available closed-form solutions. Unfortunately, the computational demands and complexity of these methods often make them impractical for most investment professionals.

Nevertheless, it is possible to create dynamic trading strategies as straightforward to implement as the static Markowitz portfolio. There are two approaches to addressing the mean-variance portfolio selection problem in the presence of conditioning information. The first approach entails specifying the joint conditional distribution of asset returns and determining the optimal portfolio based on this information. The second approach, proposed by Hansen and Richard (1987), suggests expanding the asset space by incorporating dynamic strategies and solving the relatively simpler unconditional mean-variance problem. Subsequent studies, such as those by Ferson and Siegel (2001), Cochrane (2001), and Brandt and Santa-Clara (2006), have expanded on this foundational perspective.

Elaborating on the second method, this part of the thesis is heavily influenced by a series of significant works by Brandt (1999), Aït-Sahalia and Brandt (2001), Brandt and Santa-Clara (2006), and Brandt et al. (2009), where the authors focus on modeling the portfolio weights

[7] This form of portfolio strategy is also referred to as tactical asset allocation in asset management practice.

directly. According to Brandt (1999): "*the return modeling step is, without doubt, the Achilles' heel of the traditional econometric approach.*" This statement serves as the primary motivation to adopt the single-step approach over the traditional two-step approach. The two-step approach first models asset returns and then forms optimal portfolios, whereas the single-step approach directly derives a set of dynamic portfolio weights.

In Brandt and Santa-Clara (2006), they solve a market timing problem where the portfolio weights are parameterized linearly in the vector of return-predictive signals. The key insight of their work is to expand the asset space to include simple managed portfolios and compute the optimal static portfolio within this extended asset space. The underlying idea behind this expansion is that selecting a static set of managed portfolios is essentially equivalent to employing a dynamic strategy. Consequently, the optimal dynamic strategy can be represented as a fixed combination of these managed portfolios, which encompasses both 'conditional' and 'timing' portfolios.

In the companion paper by Brandt et al. (2009), the authors expand on optimal portfolio policies using parameterized weights but with a more general utility function and set further restrictions on parameters. In this study, the authors model the weight invested in each asset using a common set of coefficients and asset-specific variables. This approach is useful for investors faced with the task of selecting among a large number of assets. Conversely, their previous approach from Brandt and Santa-Clara (2006) is more suitable for applications where only a few assets possess fundamentally distinct characteristics, such as the allocation of capital across different asset classes. Chapter 4 of this thesis, attempts to amend the methodology of Brandt and Santa-Clara (2006) to large-dimensional cross-sectional applications.

1.4.2 Multi-Period Portfolio Selection and Rational Expectations

A natural progression from the single-period approach of Markowitz (1952) is to a *multi-period* setting. In the context of multi-period portfolio selection, investors make decisions over multiple time periods, taking into account future uncertainties and intertemporal considerations. These intertemporal factors inclde elements such as the time horizon, time value of money, transaction costs, and tax planning, among others.

The subject of multi-period portfolio problems has been extensively studied since Samuelson (1969) and Merton (1969, 1971), albeit in the absence of transaction costs. However, when *both* trading cost frictions and time-varying investment opportunities are taken into account, it becomes necessary to adopt an alternative approach to multi-period trading. Most research efforts in this area have predominantly focused on settings involving a few assets and proportional transaction costs; see Lynch and Tan (2010). In these settings, numerical techniques are often employed to determine the optimal portfolio policy for each period, which is often characterized by a no-trade region, wherein the benefits of rebalancing are outweighed by the associated costs.

The seminal article by Gârleanu and Pedersen (2013) introduces a more analytically manageable multi-period framework by considering quadratic transaction costs, instead of proportional transaction costs.[8] This approach allows them to derive a closed-form solution for the optimal portfolio policy, which can effectively accommodate a large number of assets.

[8]Proportional transaction costs are generally used to model bid-ask spreads and other brokerage fees. These costs are suitable for modeling small trades that have no significant impact on market prices.

Under this framework, the investor maximizes the present value of the mean-variance utility of wealth changes over multiple time periods. The investor also has the ability to borrow without constraints and to predict future returns. The quadratic nature of the utility function and transaction costs allows their model to be formulated as a linear quadratic control (LQR) problem, which the authors solved analytically through dynamic programming. Chapter 5 contributes to this discussion by presenting a significant insight relating to the class of problems studied by Gârleanu and Pedersen (2013); these problems can be categorized as linear *rational expectations* models with costly adjustments, as presented in works from Sargent (1978) and Hansen and Sargent (1980).

At this juncture, it is worthwhile to briefly digress to discuss the essence of rational expectations. Rational expectations is an economic concept characterized by the assumption that economic agents are rational[9] *and* form their expectations about the future in a way that is consistent with the underlying economic model.[10] It implies that agents do not make systematic forecasting errors and that their expectations are unbiased. This concept has a long tradition in economics and was first introduced in a pioneering paper by Muth (1961) in order to tackle a crucial issue in macroeconomic theory: how economic agents form expectations of future variables in the presence of uncertainty. These models have had a profound influence on both economic theory and policy for analyzing and predicting the behavior of individuals and markets. Some notable applications include their utilization in models of asset pricing (Campbell and Shiller, 1988), consumption (Hall, 1978), real business cycles (Kydland and Prescott, 1982), labor demand (Sargent, 1978), inflation (Fischer, 1977; Taylor, 1979), hyperinflation (Sargent and Wallace, 1973), financial crisis and bank runs (Diamond and Dybvig, 1983), speculative bubbles (Tirole, 1982), exchange rates (Dornbusch, 1976), market microstructure (Grossman and Stiglitz, 1980), and many others. They are also extensively used by many respected central banks around the world to build large-scale dynamic models for serious real-world policy-making.

To overcome the limitations of traditional econometric and statistical models, economists adopted the notion that economic agents possess rational expectations and anticipate potential policy changes. This foundational assumption allows them to simulate various outcomes, providing insights into whether the intended policy would indeed achieve the desired effects or whether rational reactions from individuals might counteract its expected impact. It also serves as a practical, versatile, and readily implementable concept for incorporating microfoundations into dynamic economic models.

Building on this framework, Chapter 5 explores the application of the rational expectations hypothesis for the purpose of portfolio selection from the perspective of an *individual* investor. In this context, the hypothesis suggests that the investor possess all relevant information about the market and utilize it efficiently to make optimal predictions about future variables. This premise allows for the development of dynamic investment models with well-defined constraints, including the longevity of return-predictive signals.

[9]The concept of rationality means that economic agents make choices in a manner that maximizes their expected utility. Agents that are rational update their subjective beliefs according to Bayes' rule, and they make optimal decisions based on these updated beliefs.

[10]The use of rational expectations can also be justified through an alternative perspective. Consider an economy where agents initially maintain rather simplistic expectations but refine them through learning from their errors over time. This learning process could gradually steer their expectations to be rational.

Although the rational expectations hypothesis imposes strong constraints on the investor's knowledge, it serves as a valuable and convenient benchmark against which researchers can evaluate alternative methods.

Solving models of rational expectations presents several difficulties for researchers as they are characterized by their dynamic, stochastic, and forward-looking characteristics. Nevertheless, there are two common approaches available for solving the optimal decision rule in infinite-horizon dynamic linear rational expectations models with costly adjustments:

- The first approach is 'recursive,' which employs *dynamic programming* procedures; see Bertsekas (1976) and Sargent (1988). This involves casting the problem as a Bellman equation and solving the matrix algebraic Riccati equation to determine the optimal value function of the problem, and hence, the optimum decision rule.
- The second approach is 'classical,' which employs techniques from *stochastic discrete-time calculus of variations*; see Sargent (1987) and Hansen and Sargent (1980). There are two steps to this approach: (1) factoring the characteristic polynomial associated with the Euler equation[11] to express the solution in a feedback-feedforward form, and (2) solving the prediction problem using the Wiener-Kolmogorov formula.

The two approaches are intimately connected as they both rely on similar mathematical structures to find their solutions. Hansen and Sargent (1981) discuss the relationship between the recursive or classical solution methods. One difference is that the second approach tackles the 'optimization' and 'prediction' components of the problem *separately*, while the first approach tackles them *jointly* when solving the algebraic Riccati equation. However, the classical approach can provide several benefits over the recursive approach in settings where the Euler equation can be expressed as a set of time-invariant linear difference equations. Under these settings, the classical approach can provide more streamlined and transparent derivations, along with analytical decision rules. In Chapter 5, the focus will be on the classical approach, and its implications will be explored further.

1.5 Outline of Main Chapters

The research presented in this thesis makes a number of key contributions to the study of risk estimation and portfolio selection. Each chapter brings a fresh perspective, adding new layers to our understanding in the field of quantitative investment management. These contributions are summarized as follows:

- Chapter 2 introduces a novel covariance estimator for portfolio selection that adapts to the persistent heteroskedastic environments of financial time series by employing exponentially weighted averages and nonlinearly shrinking the sample eigenvalues through cross-validation. This estimator is structure agnostic, transparent, and computationally feasible in large dimensions. By correcting the biases in the sample eigenvalues and aligning the estimator with more recent risk, it is demonstrated that the estimator performs well in large dimensions against existing state-of-the-art static and dynamic covariance shrinkage estimators through simulations and with an empirical application in active portfolio management.

[11] Euler equations are a set of optimality conditions used to describe the intertemporal choices of agents in economic decision-making.

- Chapter 3 introduces a large-dimensional covariance estimator that exploits the hierarchical structure in financial returns. Prevailing methods that regularize the covariance matrix according to hierarchical agglomeration inordinately suppress important cross-sectional information. By contrast, rotation-equivariant estimators ignore the orientational content shaped by the intrinsic group organization of financial returns. Both modeling paradigms are integrated in order to produce an improved estimator that performs on large-dimensional portfolio selection applications.
- Chapter 4 presents a novel framework for solving and analyzing the optimal asset and signal combination problem in a single stage. It builds upon the dynamic portfolio selection problem of Brandt and Santa-Clara (2006) in two parts. First, their original investment problem is reformulated into a tractable one that allows us to derive a closed-form expression for the optimal portfolio policy that is scalable to large cross-sectional financial applications. Second, problem of selecting a portfolio of correlated assets and signals is recasted into selecting a set of uncorrelated managed portfolios through the lens of Canonical Correlation Analysis of Hotelling (1936). The new investment environment of uncorrelated managed portfolios offers unique economic insights into the joint correlation structure of our optimal portfolio policy. The theoretical framework is also operationalized to bridge the gap between theory and practice, showcasing the improved performance of our proposed method over natural competing benchmarks.
- Chapter 5 presents an analytical solution to the dynamic portfolio selection problem, considering transaction costs and signals with different persistence properties. The proposed optimal portfolio policy involves a smooth adjustment towards a forward-looking portfolio composed of a weighted average of current and future exogenously given target portfolios. Additionally, the expected long-run performance of the optimal portfolio policy is derived in terms of the model's structural parameters, such as investor time, risk, and liquidity preferences, as well as signals' persistence. These findings offer valuable insights for active managers aiming to optimize their portfolios in a dynamic and uncertain market environment.

Part II

Risk Estimation

Chapter 2

Estimation of Large Financial Covariances: A Cross-Validation Approach

A forthcoming version of this chapter to appear in the *Journal of Portfolio Management*

> My ventures are not in one bottom trusted, nor to one place, nor is my whole estate upon the fortune of this present year. Therefore my merchandise makes me not sad.
>
> —William Shakespeare, *The Merchant of Venice*, Act 1, Scene 1

2.1 Introduction

The covariance matrix of the security returns is a central object in the portfolio selection problem of Markowitz (1952). In practical applications, the covariance matrix is not known *a priori* and we have to resort to estimators. When estimating a large covariance matrix, one confronts the "curse of dimensionality," where the size of the problem grows quadratically as the number of securities increases. This means that we need to have a long enough historical data set to produce reliable estimates of the covariances. For example, to reliably estimate a covariance matrix of 1,000 securities, at least 40 years of daily data are needed. Estimation of large numbers of risk parameters relative to the number of observations introduces the possibility of aggregating many estimation errors into a global significant error. When such a matrix is used in the context of portfolio optimization, Jobson and Korkie (1980) and Chopra and Ziemba (1993), among others, showed that such optimizers exhibit poor out-of-sample properties, and inevitably become what Michaud (1989) regard as "error maximization" procedures.

To improve the estimation of a covariance matrix, it is thus imperative to reduce the effective number of risk parameters that require estimation. In finance, factor models are often employed for this purpose with early developments stemming from the single factor model of Sharpe (1963) and its multifactor extension in the context of the Arbitrage Pricing Theory

(APT) introduced by Ross (1976), and later generalized by Chamberlain and Rothschild (1983). Practical implementations of the APT concept include the fundamental risk factor model of Fama and French (1992) and commercial risk models that have been built and maintained by companies such as BARRA, Northfield, Axioma, and so on. What these approaches have in common is their focus on modeling the underlying structure of covariance. As a result, the estimation error is typically reduced, but at the expense of increased model bias.

Another line of attack is the shrinkage principle introduced by Stein (1956a) and James and Stein (1961) in the context of the estimation of the means. Shrinkage can be understood as a form of *regularization*–to use machine learning parlance–to control the effects of sampling variation, which results in overfitting. This idea was extended to covariance matrices by Haff (1980) and Ledoit and Wolf (2004b) by taking a linear combination of a sample covariance matrix and a (scaled) identity matrix to produce a composite matrix that has a favorable model bias-variance trade-off. In addition to the identity matrix, other structured target matrices have been explored in the literature, such as the constant correlation model (Ledoit and Wolf, 2004a) or the single-factor model (Ledoit and Wolf, 2003). However, a limitation of the linear shrinkage estimator is that it is constrained by a common shrinkage intensity and hence lacks the capacity to suppress further variations in the data.

A more recent and powerful approach that extends upon the linear shrinkage technology attempts to reduce the estimation error without assuming any *a priori* structure on the covariance. This is achieved by shrinking the eigenvalues at distinct intensities. The inception of this idea goes back to Stein (1975, 1986) and there are two modern approaches on this front. The first is an analytical approach taken by Ledoit and Péché (2011) and, subsequently, Bun et al. (2017), where the authors utilized tools from the spectral theory of large random matrices in order to provide a theoretical formula for shrinking the eigenvalues. In parallel to this development, there is also a numerical scheme from Abadir et al. (2014) and Lam (2016), where the authors demonstrate that we can accomplish a similar shrinkage effect by means of sample splitting and cross-validation.

Most of the aforementioned developments, however, revolve around a uniformly weighted covariance matrix that weighs every observation equally. However, in finance, it is known since Mandelbrot (1963) that financial returns exhibit certain empirical regularities, such as clustering of volatility. This volatility clustering effect can overwhelm our observations, which leaves us with effectively fewer observations than what we originally had because the stock returns are dominated by days with large fluctuations. Thus, practitioners often favor the use of exponential weighted average schemes such as RiskMetrics developed by J.P. Morgan/Reuters (1996) and Zumbach (2007), as they are considered more pertinent in heteroskedastic environments due to their ability to restrict the effective sample size and incorporate more localized observations. However, these models are also susceptible to estimation errors in large dimensions, since the effective time scale used to measure the observations is smaller than the sample size.

In this paper, we introduce a simple yet effective covariance estimator that is tailored specifically to address the curse of dimensionality present in exponential weighted average covariances. In particular, we extend the cross-validation shrinkage procedure of Abadir et al. (2014) and Lam (2016), *mutatis mutandis*, towards exponential weighted average covariances to robustify them against large dimensions. The broad appeal of cross-validation is that it is a well-known method with which many researchers and practitioners are familiar and it is an

effective procedure to combat overfitting. The key idea behind our method is to correct the biases that are embedded in the eigenvalues due to in-sample overfitting. This is achieved by splitting the original observations into two independent sets; the "training" set is used for computing the sample eigenvectors, while the "test" set is used for estimating the eigenvalues corresponding to these eigenvectors. As a result, extreme eigenvalues are pulled towards more central values to compensate for those biases. Our method is computationally scalable to large dimensions and it demonstrates superior results in this regime through an empirical application of active portfolio management.

The paper is structured as follows. Section 2.2 introduces the exponential weighted average covariance matrix and its decomposition. Section 2.3 explores the distributional behavior of the sample eigenvalues associated with these covariances. Section 2.4 describes our proposed cross-validation-based nonlinear shrinkage covariance estimator. Section 2.5 contains the simulation study. Section 2.6 provides the empirical experiments. Section 2.7 concludes. Appendices B.1 and C.2 contain all the tables and figures.

2.2 Setting the Stage

2.2.1 The Weighted Sample Covariance Matrix

Let $X \coloneqq (x_1, \ldots, x_T)'$ represent a matrix $T \times N$ that contains T observations of a vector of security returns, denoted as x_t, each with a dimensionality of N. Here, the symbol $(')$ denotes the transpose operator of a vector or matrix, the symbol $\coloneqq$ is a definition sign, and the symbol $=$ is the equal sign. It is assumed that the vector of security returns has zero mean and an unobserved population covariance matrix denoted as Σ.

The classical estimator for Σ is the sample covariance matrix defined as

$$S \coloneqq \frac{1}{T} X'X. \tag{2.1}$$

The standard sample covariance is an estimator of the population covariance since it is unbiased (that is, its expectation coincides with the population covariance matrix). In general, a standard sample covariance can be generalized to include some arbitrary weight profile assigned along the time dimension. In particular, it is expressed in the following form

$$S_W \coloneqq \frac{1}{T} X'WX, \tag{2.2}$$

where W is a fixed symmetric and positive-definite $T \times T$ matrix. This weighting scheme generalizes the standard sample covariance, and provides us with a way to incorporate prior knowledge about the non-independence and non-identicality nature of the observations in the covariance matrix. Indeed, if $W = \mathbb{I}_T$, where $\mathbb{I}_T$ is an T-dimensional identity matrix, then we recover the expression for the standard sample covariance in (2.1).

Without loss of generality, we assume that $T^{-1}\mathsf{Tr}(W) = 1$ where $\mathsf{Tr}(\cdot)$ represents the trace of a square matrix or the sum of its diagonal entries. This condition ensures that the weighted covariances are also unbiased estimators, that is $\mathbb{E}[S_W] = \Sigma$. In other words, the temporal weighting of observations should not influence the average recovery of the ground truth.

2.2.2 Exponential Weighted Average Covariance Matrix

To address the problem of non-stationarity in the time-series setting, we look at a (sample) covariance matrix that is measured using an exponential weighted average scheme, henceforth, EWA-SC, which is given by

$$E \coloneqq \frac{1-\delta}{1-\delta^T} \sum_{t=1}^{T} \delta^{T-t} x_t x_t' \tag{2.3}$$

where $\delta \in (0,1)$ is the exponential decay rate. The timescale for which the observations are "effectively" measured with this scheme can be determined using a characteristic time-scale defined as $\tilde{T} \coloneqq 1/(1-\delta)$. For example, if $\delta = 0.999$, then the timescale would be 1,000 days (or roughly 4 years). In the limit where δ tends to one from below the EWA-SC converges towards the standard sample covariance; conversely, when δ moves away from one, it assigns greater weight to recent cross-product observations.

The EWA-SC in Equation (2.3) has a weight matrix profile whose diagonal entries are

$$W_{t,t} = T \frac{1-\delta}{1-\delta^T} \delta^{T-t}, \tag{2.4}$$

for $t = 1, 2, \ldots, T$, and zero in the off-diagonal entries. Further, let us define the following auxiliary observation matrix:

$$\tilde{X} \coloneqq W^{1/2} X. \tag{2.5}$$

Since the matrix W is diagonal, its matrix factor $W^{1/2}$ is simply the square root of its diagonal entries taken with a non-negative sign. On the basis of our redefinition of the observation matrix, we can see that the EWA-SC can be expressed in a similar form as the standard uniformly weighted sample covariance. The advantage of recasting the EWA-SC in this way is that many of the refinements for the standard sample covariance that have been developed over the years are at our disposal; this includes shrinkage.

2.2.3 Decomposing the Covariance Matrix

The analysis of a covariance matrix would be easier if the entries were uncorrelated. While this is not generally true for security returns, it is nonetheless useful if we can summarize the contents of a matrix in terms of key low-dimensional objects with mutually uncorrelated components. Fortunately, the EWA-SC is symmetric and so for a fixed decay rate δ, they admit the following spectral decomposition:

$$E \coloneqq \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i', \quad \text{and} \quad \Sigma \coloneqq \sum_{i=1}^{N} \lambda_i u_i u_i', \tag{2.6}$$

where $(\hat{\lambda}_1, \ldots, \hat{\lambda}_N; \hat{u}_1, \ldots, \hat{u}_N)$ denotes a system of sample eigenvalues and eigenvectors of E, and $(\lambda_1, \ldots, \lambda_N; u_1, \ldots, u_N)$ denotes a system of eigenvalues and eigenvectors of the population covariance Σ.[1] The eigenvalues are assumed to be sorted in ascending order.

[1] Both the sample eigenvalues and sample eigenvectors are implicitly parameterized by the decay rate δ since the underlying EWA-SC depends on it. To maintain brevity in notation, we suppress their dependence on the

The main advantage of the sample eigenvalues and sample eigenvectors is that they are consistent estimators; that is, they recover the population eigenvalues and eigenvectors when the number of observations T becomes large while the number of assets N is kept fixed. However, this desirable statistical property breaks down in financial applications where the number of assets is of the same order of magnitude compared to the number of observations. In particular, the sample eigenvalues and sample eigenvectors end up being biased estimates of their population counterparts.

2.3 Spectrum of Exponential Weighted Average Covariances

In this section, we explore the qualitative behavior of the estimation error found in the EWA-SC by studying the shape of the distribution of the eigenvalues when the dimension of the matrix N is non-negligible with respect to the sample size T and effective time-scale $\tilde{T}$. To this end, it is instructive to define the following quantities $c \coloneqq N/T$ and $\tilde{c} \coloneqq N/\tilde{T}$. We shall also treat the observations in the data matrix X as independent and identically distributed (i.i.d) standard Gaussian random variables with a population covariance Σ that is an identity matrix.

Under these assumptions, there is an asymptotic formula derived by Burda et al. (2011), which we now use to study the distributional shape of the sample eigenvalues of the EWA-SC. The left panel of Figure C.2.1 plots this distribution for different values of $\tilde{c}$ and with $c = 0.5$. Note that for $\tilde{c} = 0$, the distribution corresponds to that of Marčenko and Pastur (1967) law for the standard sample covariance matrix. Overall, we see that the distribution of the sample eigenvalues are dispersed, where the sample eigenvalues are biased away from one.

However, what is more concerning is the fact that the excess dispersion of the sample eigenvalues in EWA-SC exceeds that of standard SC and increases as $\tilde{c}$ increases. This observation arises from the fact that the weights for EWA-SC introduce autocorrelations into an otherwise uncorrelated series:

$$\mathbb{E}[\tilde{x}_{ti}^2] = T\frac{1-\delta}{1-\delta^T}\delta^{T-t}, \quad \text{for } i = 1, \ldots, N. \tag{2.7}$$

Here, the variance of the auxiliary random variables $\tilde{x}_{ti}$ now exhibits a time-dependent structure, which decays geometrically over time. The right panel of Figure C.2.1 illustrates how different values of c also affect the distribution of the sample eigenvalues while keeping $\tilde{c}$ fixed at 0.5. We observe that at higher values of c, the excess dispersion also increases.

This brief analysis suggests that there are two channels that exaggerate the severity of the bias in the sample eigenvalues of the EWA-SC: the problem's dimensionality relative to the number of observations and its dimensionality in relation to the chosen decay rate. The phenomena that we described here hold under more general settings and are not limited to the identity matrix that we consider for the population covariance. By correcting these biases through shrinkage of the sample eigenvalues towards more central locations, it is possible to obtain a covariance matrix with improved out-of-sample properties for portfolio selection.

decay rate.

2.4 Shrinkage Estimation

2.4.1 Covariance for Portfolio Selection

We consider a class of covariance estimators due to Stein (1975, 1986), which suggests that the sample eigenvalues should be corrected while the sample eigenvectors of the original matrix are retained. This can be formulated mathematically as follows:

$$\hat{\Sigma} = \sum_{i=1}^{N} \xi_i \hat{u}_i \hat{u}_i', \tag{2.8}$$

where $\xi := (\xi_i)_{i=1,\dots,N}$ is an N-dimensional vector that we have to obtain. This framework serves as a reasonable starting point to address the systematic biases in the sample covariance. Indeed, absent any *a priori* knowledge about the structure of the covariance matrix, the most natural guess that we have about the population eigenvectors are the sample eigenvectors that we observe. Hence, the sample eigenvectors are the only ones that we get to keep and we have to try to make the best out of this situation by coming up with a set of bias-corrected eigenvalues so that the covariance matrix as a whole is closer to the truth.

The question is how we should choose the vector of bias-corrected eigenvalues ξ. We answer this question by choosing the one that minimizes a loss function between estimators of the form (2.8) and the population covariance matrix. To determine the optimal estimator within the postulated form of (2.8), we consider a loss function between the population and the estimated covariance matrix based on the minimum variance loss function proposed by Engle et al. (2019, Definition 1), which is tailored specifically for the purpose of portfolio selection:

$$\mathcal{L}(\hat{\Sigma}, \Sigma) := \frac{\mathsf{Tr}(\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1})/N}{(\mathsf{Tr}(\hat{\Sigma}^{-1})/N)^2} - \frac{1}{\mathsf{Tr}(\Sigma^{-1})/N}. \tag{2.9}$$

The loss function can be interpreted as the excess out-of-sample portfolio variance that would materialize by using an estimator instead of the population covariance matrix and should be reduced as much as possible.

If we insert the postulated estimator (2.8) into Equation (2.9), then from the cyclic property of the trace operator and orthogonality of the eigenvectors, we get

$$\mathcal{L}(\hat{\Sigma}, \Sigma) = N \frac{\sum_{i=1}^{N} \xi_i^{-2} \hat{u}_i' \Sigma \hat{u}_i}{\left(\sum_{i=1}^{N} \xi_i^{-1}\right)^2} + \text{constant}. \tag{2.10}$$

By minimizing this loss function with respect to ξ_i, we have

$$\xi_i^{\circ} = \text{scalar} \times \hat{u}_i' \Sigma \hat{u}_i, \quad \text{for } i = 1, \dots, N, \tag{2.11}$$

where the scalar value is independent of the asset index i. The scalar can be chosen such that the trace of the population covariance matrix is preserved. This can be achieved by setting the scalar to the value 1; for example, see Ledoit and Wolf (2017, Proposition 4.1).

We see that ξ_i° is neither the sample eigenvalue nor the population eigenvalue. Rather, it is an "eigenvalue" that internalizes both sample and population information to differentiate itself from its sample and population counterparts. From an economic perspective, one can interpret

each ξ_i° as an out-of-sample variance of a portfolio whose weights are the ith sample eigenvector $\hat{u}_i$. This is an enhancement over the sample eigenvalues $\hat{\lambda}_i$, which is an in-sample variance computed from a portfolio that uses the same observations as the weights $\hat{u}_i$. Further, Ledoit and Wolf (2004b, Section 2.3) showed that ξ_i° is less dispersed cross-sectionally than the sample eigenvalue $\hat{\lambda}_i$. Hence, it constitutes a form of shrinkage itself that corrects the biases in the sample eigenvalues.

If we plug ξ_i° into Equation (2.8), we arrive at the following covariance estimator:

$$\hat{\Sigma}^\circ = \sum_{i=1}^{N} \xi_i^\circ \hat{u}_i \hat{u}_i'. \tag{2.12}$$

An advantage of ξ_i° is that it is positive for $i = 1, \ldots, N$ as long as Σ is positive definite. In this case, the covariance estimator (2.12) is positive definite and invertible when $N > T$. Unfortunately, $\hat{\Sigma}^\circ$ is not attainable in practice since it depends on the population covariance Σ, which is the very object that we wanted to estimate in the first place. While this seems paradoxical at first sight, it is possible to substantiate the estimator $\hat{\Sigma}^\circ$ with real data and this is what we aim to achieve with our proposed method in the next subsection.

2.4.2 Out-of-Sample Variance Estimation by Cross-Validation

Another perspective on the biased nature of the sample eigenvalues and sample eigenvectors is that they are objects that have been obtained through numerical optimization routines. Hence, they are likely to overfit a given data. To address this problem, cross-validation is a powerful scheme to mitigate overfitting and estimate the out-of-sample performance for independent and identically distributed (i.i.d.) observations. We propose to apply this principle to estimate the out-of-sample variance corresponding to each eigenvector portfolio.

The estimation algorithm is described as follows:

1. Randomize the auxiliary observation matrix $\tilde{X}$ from Equation (2.5), along the time axis.
2. Split the (randomized) auxiliary observations into K non-overlapping subsets or "folds" of equal or nearly equal size.
3. For each of these K folds, perform the following steps:
   - i Designate the current fold as the validation set.
   - ii Use the remaining folds collectively as a training set.
   - iii Estimate the sample eigenvectors from the training set, and then compute an N-dimensional vector of out-of-sample variances using the validation set and the corresponding sample eigenvectors.
4. Average the out-of-sample variance estimates from all K folds to obtain the estimated bias-corrected eigenvalue of the ith sample eigenvector portfolio, denoted as $\xi_i^\bullet$, for all i.

Figure C.2.2 provides a schematic representation of the construction of the training and validation sets in $K = 5$ folds.

Formally, we start by splitting the T (randomized) observations into K non-overlapping index sets (or folds) given by $\{\mathcal{I}_k | \mathcal{I}_k \subset \{1, \ldots, T\}\}_{k=1}^K$. Each set indexed by $\mathcal{I}_k$ is called a "test" fold, while the set of indices of the remaining observations forms a "training" fold. From this

construction, we can estimate the out-of-sample variance ξ_i°; we introduce our K-fold cross-validation estimator, which is defined as

$$\xi_i^\bullet \coloneqq \frac{1}{K}\sum_{k=1}^{K}\sum_{t\in\mathcal{I}_k}\frac{1}{|\mathcal{I}_k|}\left(\hat{u}_i[k]'\tilde{x}_t\right)^2, \quad \text{for } i=1,\ldots,N, \tag{2.13}$$

where $|\mathcal{I}_k|$ denotes the cardinality of the kth validation set such that each of them is approximately equal in size, that is, $K|\mathcal{I}_k| \approx T$. Here, $\hat{u}_i[k]$ is the ith sample eigenvector of a sample covariance matrix that is obtained from the training fold, and $\tilde{x}_t$ is a sample vector of the auxiliary observation matrix in Equation (2.5) from the test fold. This is a more general estimator of Abadir et al. (2014) and Lam (2016) since if δ tends to one, we recover the cross-validation estimator of the standard sample covariance as the auxiliary observation vectors become the original ones.

Finally, with $\{\xi_i^\bullet\}$ determined, we plug them into the estimator (2.8) to yield our operational covariance estimator, which we term EWA-CV:

$$\hat{\Sigma}^\bullet = \sum_{i=1}^{N}\xi_i^\bullet\hat{u}_i\hat{u}_i'. \tag{2.14}$$

An estimate for the inverse covariance can be obtained by taking the inverse of the covariance estimator.

Note that our covariance estimator (2.14) is not completely restricted to weight profiles that are exponentially declining over time. In principle, our auxiliary variable definition (2.5) can accommodate any weight matrix design W, provided it can be factorized into a matrix and a conjugate transpose. Thus, our estimator is general enough to incorporate other forms of weighting schemes to model the time dependence of security returns. We leave this investigation for future research.

Remark 1 (Positive semi-definiteness). Since all the shrunked eigenvalues $\{\xi_i^\bullet\}_{i=1}^N$ are non-negative by construction, the EWA-CV covariance estimator $\hat{\Sigma}^\bullet$ is positive semi-definite.

Remark 2 (Monotonicity of eigenvalues). The ordering of cross-validated eigenvalues is often violated in practice. This may be unsettling for some researchers since it can be argued that eigenvalues should satisfy a ranking where they are ordered from the smallest to the largest; for example, see Sheena and Takemura (1992). To address this concern, we adopt the approach from Stein (1975) and apply an isotonic regression of the set cross-validated eigenvalues on the sample eigenvalues of E in order to enforce a monotonic structure on the estimated out-of-sample variances.

2.4.3 Related Literature

Our proposed estimator corresponds to a specific choice of a split location, denoted as M, representing the size of the training fold. The approach presented in Abadir et al. (2014) involves averaging the estimated eigenvalues across various folds and split locations, where $M \in \mathcal{M}$ and $\mathcal{M} = [0.2T, 0.8T]$ represents the range of split locations. Their estimator can be described as

the following grand average:

$$\breve{\Sigma} := \sum_{M \in \mathcal{M}} \breve{\Sigma}_M, \tag{2.15}$$

where $\breve{\Sigma}_M \equiv \hat{\Sigma}^{\bullet}$ is the estimator (2.14) with $W = \mathbb{I}_T$ for a given split location M. This estimator is asymptotically optimal if the split location satisfies $M, T - M \to \infty$ with $M/T \to c \in (0, 1)$; see Abadir et al. (2014, Proposition 3).

In contrast, Lam (2016) considers a similar setup but studies the large-dimensional setting where $N \to \infty$ as $T \to \infty$ such that $N/T \to c > 0$. Instead of averaging over a range of splits, Lam (2016, Theorem 5) proposed a split $M = T - aT^{1/2}$, with some constant $a > 0$, that is asymptotically efficient. However, this asymptotic property may be challenged in finite samples or as M/T approaches a constant smaller than 1, and there is no practical guide for setting the constant a. Lam (2016, Section 4.2) proposes to search M over seven candidate split locations $\mathcal{M} = [2T^{1/2}, 0.2T, 0.4T, 0.6T, 0.8T, T - 2.5T^{1/2}, T - 1.5T^{1/2}]$ in order to minimize a criterion similar to that of Bickel and Levina (2008).

In our paper, we avoid the non-trivial task of choosing the split locations entirely to reduce the number of required computations to build the estimator.

2.4.4 Comparison to DCC-NL

Another estimator that is often compared to a covariance matrix measured through an exponential weighted scheme is the Dynamic Conditional Correlation (DCC) model proposed by Engle (2002) to model the time-varying dynamics of multivariate financial returns due to conditional heteroskedasticity. As is well known, the DCC model also suffers from the "curse of dimensionality." There have been efforts from Engle et al. (2019) to successfully enable the estimation of the DCC to be both (i) computationally feasible in the large dimensions through the use of composite likelihood functions from Pakel et al. (2021), and (ii) with reduced estimation error through the nonlinear shrinkage (NL) of Ledoit and Wolf (2015). This estimator is referred to as the DCC-NL.

Both EWA-CV and DCC-NL are intimately related by the concept of shrinkage but they also differ in a few other ways.[2] First, the latter explicitly applies devolatization to the returns. Therefore, in order to allow for a more fair comparison between both estimators, EWA-CV should also be applied to a panel of returns series that have been devolatized through a univariate GARCH model of Bollerslev (1986) and then reconstructed to form a covariance matrix. Second, the dynamic nature of DCC requires that the covariance be determined recursively, while the EWA-SC requires only matrix multiplications. Third, the DCC-NL is governed by two parameters, while EWA-CV has just one. Our view is that maintaining a single parameter can offer a more parsimonious representation of security returns, which provides insights into the sensitivity of the estimator's performance to only changes in the decay rate. Moreover, since the decay rate is typically specified by the user in practice, it possesses less implementation overhead over the DCC, whose dynamic parameters are typically estimated via maximum likelihood. Admittedly, EWA-CV is a numerical scheme compared to NL, an analytical method, but it only employs a few spectral decompositions and

[2] Lam (2016) proved the data splitting scheme of Abadir et al. (2014) is a nonparametric approach of achieving a similar nonlinear shrinkage effect in Ledoit and Wolf (2012).

vectorized operations to estimate an N-dimensional vector, which are fast to execute. Finally, EWA-CV is conceptually simpler since it relies on the principle of cross-validation, which many practitioners are familiar with, whereas the DCC-NL requires recourse to the theory of random matrices. Notwithstanding the similarities and differences between both approaches, we formally compare their performances in Section 2.6.

2.5 Simulation Study

2.5.1 Return Generating Process

We now study the effectiveness of our proposed estimator against different competing estimators through a challenging simulation environment, where the risk dynamics are driven by the RiskMetrics model of J.P. Morgan/Reuters (1996).[3] This allows us to assess if our proposed estimator is accurate enough to be useful in an ideal environment where it is supposed to thrive. In particular, we consider a return-generating process (RGP) that is drawn from a multivariate standard normal distribution given by

$$x_t = \Sigma_t^{1/2} z_t \quad \text{with} \quad z_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbb{I}_N), \tag{2.16}$$

where Σ_t is a population covariance matrix that evolves according to

$$\Sigma_t \coloneqq \delta_\Sigma \Sigma_{t-1} + (1 - \delta_\Sigma) x_{t-1} x_{t-1}'. \tag{2.17}$$

The parameter δ_Σ is the "intrinsic" decay rate of the population covariance matrix, which we set to 0.996 in this simulation study. We use this RGP to generate a data set of $N = 500$ security returns with $T = 1{,}250$ daily return observations, which corresponds to roughly 5 years of daily data.

2.5.2 Candidate Estimators

From the generated data set of security returns, we can estimate a covariance matrix. Given the plethora of covariance estimators in the literature, we shall limit our analysis to the following candidate estimators for $\widehat{\Sigma}$, which are either uniformly weighted or exponential weighted:

- LS: the linear shrinkage estimator of Ledoit and Wolf (2004b), with the shrinkage target being (a scaled multiple of) the identity matrix.
- NL: the quadratic inverse shrinkage estimator of Ledoit and Wolf (2022).
- CV: the cross-validation shrinkage estimator with $W = \mathbb{I}_T$ and $K = 10$ folds.
- EWA-SC: the estimator of Equation (2.3).
- EWA-CV: based on our proposed cross-validation shrinkage estimator from Equation (2.14) with $K = 10$ folds.

For EWA-based estimators, we consider various specifications of decay rates δ ranging from 0.999 to 0.99. This would allow us to see how they behave under model misspecifications.

[3] The RiskMetrics 1996 methodology is a covariance matrix that evolves recursively according to an exponential weighted moving average. It has a similar form to our EWA-SC (2.3) up to timing in the measurement.

2.5.3 Performance Assessment

To compare the performance of various covariance estimators with population covariance, we need a loss function and an evaluation metric. The loss between an estimator $\hat{\Sigma}$ and the population covariance matrix Σ is based on Equation (2.9).

A simple metric to quantify an improvement of an estimator over a benchmark model is the so-called PRIAL (percentage of relative improvement in average loss). If we let the standard sample covariance be the benchmark model, then the PRIAL is defined as

$$\text{PRIAL}(\hat{\Sigma}, S) \coloneqq \left(1 - \frac{\mathbb{E}[\mathcal{L}(\hat{\Sigma}, \Sigma_T)]}{\mathbb{E}[\mathcal{L}(S, \Sigma_T)]}\right) \times 100\%, \tag{2.18}$$

where Σ_T is a population covariance obtained at the end of the sample period, and $\mathcal{L}$ is a loss function. The PRIAL is 100% when our estimator coincides with the population covariance Σ_T at time T, and 0% with the uniformly-weighted sample covariance S. A negative value for PRIAL is possible, which indicates that the estimator did not improve over the reference matrix S. The notation $\mathbb{E}[\cdot]$ denotes an expected value of the losses, which is taken to be the average of the 100 simulation trials.

2.5.4 Simulation Results

We now present the results of the Monte Carlo simulations. Figure C.2.3 shows the PRIAL for different covariance estimators against the different decay rate specifications. Starting with the uniformly weighted-based estimators, we see that LS, NL, and CV perform better than the standard sample covariance S as seen by their positive PRIAL values. This improvement owes to the benefits of shrinkage. However, their performances remain flat for all decay rate specifications, since, by their uniformly weighted construction, they are not parameterized by them. This lack of flexibility of the uniformly-weighted estimators fundamentally limits their ability to align themselves toward more recent risk, which gives rise to higher estimation errors.

In contrast, EWA-based estimators beat uniformly weighted estimators when their decay rate δ is in the vicinity of the intrinsic decay rate at $\delta_\Sigma = 0.996$ and EWA-CV performs the best with PRIAL values greater than 90%. This is reassuring for us because the EWA-based estimators (2.3) have the same form as the RGP in (2.17) (up to timing in measurement). Moreover, this also speaks to the importance of capturing time-series variation of risk since the EWA-based estimators differ from their uniformly weighted counterparts by how the observations are measured across time.

A further inspection of the performance of EWA-SC shows that there is very little margin for error in specifying δ. In particular, EWA-SC starts to rapidly underperform all other estimators as δ decreases and even takes on negative values for δ less than 0.993. Fortunately, our proposed EWA-CV does not share that fragile behavior. The shrinkage technology that we have introduced in Section 2.4.2 has allowed it to be robust against a large ensemble of noisy observations causing it to yield stellar outperformance even with a misspecified decay rate. This is a desirable feature because, in practice, the actual decay rate of an ensemble of non-stationary return series is not known, and we want to be protected from placing bets on risky sample eigenvectors by mistake.

Finally, we ran this experiment on a single server with an (256GB RAM 48 cores) Intel

Xeon Gold 6252N 2.3GHz CPU. To estimate a 500-dimensional covariance of this set-up, EWA-CV takes about 0.5 seconds wall-clock time and, by comparison, NL takes about 0.1 seconds. Although our method is slower than NL, it is still fast to execute and incurs negligible runtime overhead.

2.5.5 Sensitivity Analysis to the Number of Folds

We also conduct a sensitivity analysis of EWA-CV to the different number of folds K. Figure C.2.3 shows that, in general, the performance of EWA-CV is not highly sensitive to the number of folds. Although its performance can be suboptimal for $K < 10$, it is consistently optimal for $K \geq 10$. The intuition is that for $K < 10$, there is not enough averaging across the folds to further suppress the noise from the projections of the test observation onto the sample eigenvectors. However, for $K \geq 10$, we find that the performances remain similar. In light of these findings, we will use $K = 10$ folds for all of our experiments, which is sufficient for us to reap the most benefits from averaging and to strike an appropriate balance between computational tractability and precision.

2.6 Empirical Analysis

2.6.1 Data and Portfolio Construction Rules

We obtain daily stock returns data from the Center for Research in Security Prices (CRSP), starting at 01/01/1981 and ending at 12/31/2019. Our analysis is limited to stocks traded on the NYSE, AMEX, and NASDAQ stock exchanges. The size of the investment universe that we consider is $N \in \{100, 500, 1{,}000\}$.

We define one "month" and "year" as 21 and 252 consecutive trading days, respectively. The portfolios are rebalanced on a monthly basis using a rolling window scheme where only past information is used to avoid a look-ahead bias. During the month, the number of shares is kept fixed in order to avoid incurring unnecessary transaction costs. The covariance matrices are estimated using an in-sample period of size $T = 1{,}250$, which is approximately 5 years. The out-of-sample period starts from 09/04/1986 and ends at 12/31/2019. This provides us with a total of 400 months (or 8,400 trading days) of consecutive non-overlapping observations of length 21 days for which the portfolios are rebalanced.

To construct a well-defined investment universe on which we can estimate the covariances on, we use a similar procedure described in Engle et al. (2019) for reproducibility. For each rebalancing month $h = 0, 1, \ldots, 399$ (using zero-based indexing), we first select stocks that have complete data over the past 1,250 days and over subsequent 21 days. The backward and forward restrictions ensure that we have data to estimate our models and to evaluate out-of-sample. Then, we search for pairs of highly correlated stocks (that is, those with a sample correlation exceeding 0.95) and remove the stock with the lower volume in each pair. From this remaining set of stocks, we select the largest N stocks as measured by their market capitalization on the rebalancing month h to include in our investment universe.

2.6.2 Base Candidate Portfolios

To test the performance of different covariance matrix estimators in the context of portfolio selection, we consider two investment problems: (1) the global minimum variance (GMV) portfolio and (2) the Markowitz portfolio with a given signal. Both problems are described in Section 2.6.4 and 2.6.6, respectively.

In either setting, we employ the following candidate estimators as input:

- SC: the sample covariance matrix (2.1).
- LS: the linear shrinkage estimator of Ledoit and Wolf (2004a), with the shrinkage target being (a scaled multiple of) the identity matrix.
- NL: the quadratic inverse shrinkage estimator of Ledoit and Wolf (2022).
- EWA-SC: the exponentially-weighted average sample covariance (2.3).
- EWA-CV: our proposed method based on estimator (2.14) with $K = 10$.
- AFM1-NL: an approximate single factor model of De Nard et al. (2021) with NL applied to the residuals.
- AFM1-EWA-CV: our second proposed method based on an approximate single factor model with EWA-CV applied to the residuals.

Note that for the EWA-based estimators, we choose a fixed decay rate that is relatively high at $\delta = 0.997$, which makes them slightly closer to the uniformly weighted setting. This corresponds to a half-life of approximately one year, a parameter choice that practitioners commonly employ. Although this choice may seem *ad-hoc*, our robustness checks indicate that values of δ above 0.995 yield results that do not significantly differ from one another.

Remark 3 (Factor models). While many active portfolio managers tend to hold relatively concentrated portfolios of less than 100 stocks, there are quantitative portfolio managers that invest in a broad investment universe spanning thousands of single stock equities in order to benefit from efficiency gains in large dimensions. In such a case, they tend to exploit some form of factor structure in order to make the investment problem more manageable. To accommodate for this practice, we can also extend our proposed estimator towards data that have been generated by a factor model in a similar spirit to De Nard et al. (2021) but with EWA–CV applied to the residual terms. In particular, we will utilize the market factor, which is easily accessible to researchers, and term our second proposed covariance estimator in this paper AFM1-EWA-CV.

2.6.3 Evaluation Methodology

To evaluate the performance of the different portfolios, we report four main out-of-sample performance measures: the average out-of-sample returns (AV), the standard deviation of portfolio returns (SD), the information ratio (IR), and the Sharpe ratio (SR). For ease of interpretability, all performance measures are annualized with 252 trading 'days'. The information ratio is computed with respect to the actual returns (as opposed to returns in excess of a given benchmark) since we believe it is more relevant in our context where the portfolios are formed solely on the basis of risky assets. On the other hand, the Sharpe ratio is the information ratio with the risk-free rate as the benchmark. In addition, we also include the maximum drawdown (MDD), which measures the largest cumulative loss from peak to

trough over the entire out-of-sample period.

We also report the information ratio and Sharpe ratio net transaction costs denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. In particular, we assume a constant transaction cost of 5 bps for all stocks, which is a reasonable approximation today. Even though the investment problems that we consider do not explicitly account for transaction costs at the portfolio pre-formation stage, it is nevertheless worthwhile to investigate if our methods survive transaction costs in an unconstrained setup.

Additionally, we report the following portfolio weight statistics averaged over the 400 trading months: turnover (TO), gross exposure (GE; computed as the sum of the absolute value of long and short positions), proportional leverage (PL; computed as the fraction of negative weights), minimum weight (MIN), maximum weight (MAX), the standard deviation of weights (SD), and mean absolute deviation from an equally-weighted portfolio (MAD). Note that these statistics are not our primary focus since our proposed method is not optimized to account for these measures. Nevertheless, they are provided to give a better overview of the different methods.

2.6.4 Global Minimum Variance Portfolio

The global minimum variance (GMV) portfolio is expressed as

$$\begin{aligned} &\min_{w} w'\Sigma w \\ \text{such that} \quad &w'\mathbf{1} = 1, \end{aligned} \tag{2.19}$$

where $w = (w_1, \ldots, w_N)'$ is a vector of portfolio weights and $\mathbf{1}$ is an N-dimensional conformable vector of ones. The GMV portfolio allocation framework is useful for isolating the estimation error stemming from the expected returns and allows us to focus solely on evaluating the quality of a covariance matrix estimator. Empirical studies of Haugen and Baker (1991) and Jagannathan and Ma (2003), among others, have found that the GMV portfolios enjoy favorable out-of-sample performance both in terms of risk and Sharpe ratio when compared with other benchmark portfolios.

In this setting, the solution for optimal weights has an explicit form given by

$$w = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \tag{2.20}$$

To yield a feasible portfolio $\hat{w}$ that can be computed from the returns data, we replace Σ with a covariance matrix estimator. We also include the equal-weighted portfolio promoted by DeMiguel et al. (2009) as another benchmark, denoted as $1/N$, since it is argued to be difficult to beat.

The primary metric for which we evaluate the performance of the GMV portfolio will be the out-of-sample standard deviation. We also report the statistical significance of the differences between the SD values of EWA-SC and EWA-CV derived by using the two-sided p-value of the prewhitened HAC_{PW} test described by Ledoit and Wolf (2011, Section 3.1). Since the purpose of this paper is to demonstrate that EWA-CV is an improvement over EWA-SC, we confine our attention to the comparison of EWA-SC with EWA-CV to avoid the issue of multiple testing.

The out-of-sample performances are reported in Table B.1.1 and the results from the SD column are summarized as follows:

- EWA-CV outperforms EWA-SC for all N and is statistically significant at the 1% level for $N = 500, 1{,}000$.
- Overall, the EWA-CV portfolio delivers the best results, owing in part to its ability to adapt to the non-stationary nature of the financial returns. Its performance can be seen to improve with increasing universe size N.
- The EWA-SC portfolio is competitive for $N = 100$ and even outperforms portfolios that are regularized with shrinkage. This is probably due to the fact that at relatively small dimensions, a covariance estimator may benefit more from picking up more recent time series variations. However, as N increases, the EWA-SC portfolio starts to underperform. This also corroborates the empirical findings of Engle et al. (2019, Table 7) in their application of the RiskMetrics 2006 methodology of Zumbach (2007), which is a more elaborate version to the EWA-SC but also inherits the same cross-sectional estimation errors.
- Even though AFM1-EWA-CV has slightly better performance than the NL estimators, it does not lead to any material improvement above EWA-CV.
- The reduction in the standard deviation of EWA-CV over NL in $N = 100, 500, 1{,}000$ amounts to 5%, 4%, and 2%, respectively. A similar conclusion holds for EWA-CV and AFM1-NL since both AFM1-NL and NL share similar performance, which is consistent with the findings of De Nard et al. (2021, Table A.1).

Additionally, we find that EWA-CV is the best performer in terms of information ratio and Sharpe ratio for $N = 1{,}000$ although it is competitive for $N = 100, 500$. Once we have taken into account the transaction costs at the post-formation stage of the portfolio, we see a decline in the information and Sharpe ratios across all portfolios but the relative performance remain similar. In terms of the maximum drawdown incurred by the portfolio, we see that both EWA-CV and AFM1-EWA-CV suffer the least.

Table B.1.2 describes the distribution of the portfolio weights of the estimated portfolios. We see that EWA-CV has the least dispersed weight while both SC and EWA-SC have the most dispersed weight. Furthermore, EWA-CV and AFM1-EWA-CV have the lowest turnover, which is interesting given that we make no effort to control the trajectory of the weights. Moreover, EWA-CV also has the lowest amount of gross exposure and proportion of leverage.

2.6.5 Robustness Checks

In this section, we inspect whether the outperformance of our proposed estimator is robust to different revisions in the current empirical setup. In particular, we will be interested in parameters that dictate (1) sub-sample period, (2) exponential decay rate, (3) no short-sale constraint, and (4) devolatization.

2.6.5.1 Subsample Analysis

Here, we check if there are any peculiar subsample effects that may drive the performances of our proposed scheme. We divide the out-of-sample period into four equally-sized subsamples of 100 months (that is, roughly 8 trading years) each: (1) 09/04/1986 to 12/21/1994, (2) 12/22/1994 to 04/25/2003, (3) 04/28/2003 to 08/25/2011, and (4) 08/26/2011 to 12/31/2019. Then we perform the same procedure in each subsample. The results are provided in Table B.1.3.

We see that the ranking of performance across different portfolios is relatively consistent over time. Generally, EWA-CV is the best performer although there are brief episodes where it is not. For example, for $N = 100$, EWA-SC outperforms in Sample 1, while for $N = 1{,}000$, NL (and AFM1-NL) outperforms in Sample 4. However, the performance of EWA-CV is close to the leading in these scenarios.

2.6.5.2 Sensitivity Analysis

In our base experiments, we parameterize the EWA-based estimators with a decay rate of 0.997. We now repeat the same backtest exercise as in the previous section, but focus only on EWA-based estimators for different decay rates $\delta \in \{0.999, 0.997, 0.995, 0.992, 0.990, 0.985\}$ and $N \in \{100, 1{,}000\}$. This allows us to examine if the performance of our estimator is sensitive to changes in the decay rate specification or if there are other choices that may lead to better results.

Table B.1.4 shows that the outperformance of EWA-CV over EWA-SC is robust across different decay rates for $N = 1{,}000$ while for $N = 100$, it only marginally outperforms EWA-CV in Sample 1 for $\delta \geq 0.99$. Moreover, EWA-CV tends to produce a lower standard deviation with decay rates greater than 0.995. The results of the unreported simulation also indicate that decay rates lower than 0.985 consistently give a higher standard deviation. This observation holds for the full out-of-sample period as well as across different subsample periods. These empirical findings are certainly at odds with the recommended value of 0.94 suggested by J.P. Morgan/Reuters (1996) to model exponential weighted moving average covariances.

2.6.5.3 No Short-Sale Constraint

In our original investment problem (2.19), we allow for the possibility of short-selling. However, shortening in the equity markets is difficult and a margin of 50% must be posted. Since some portfolio managers or institutional investors may encounter a no-short-sale constraint in practice (for example, university endowment funds), we now impose a nonnegative constraint on the portfolio weights in the optimization objective.

Table B.1.5 presents the results. As can be seen, SC and EWA-SC are now competitive in terms of standard deviation, even though they differ only slightly from other estimators. By comparing these results to Table B.1.1, it is evident that prohibiting short sales benefits SC and EWA-SC but harms other shrinkage estimators, including EWA-CV. Based on Jagannathan and Ma (2003), these results agree with them that enforcing a no-short-sales constraint has an implicit shrinkage effect on sample covariances when estimating the global minimum-variance portfolio.

2.6.5.4 Devolatization

The base case experiments focus exclusively on the performance improvement that exponential weighting adds over a uniformly weighted shrinkage estimator. We now consider applying EWA-based estimators to a panel of returns series that have been devolatized through a univariate GARCH model. For this effort, we use Kevin Sheppard's Python `arch` package to implement the GARCH(1,1) model.

We are now in a position to compare against the DCC-NL model of Engle et al. (2019) as well as the AFM1-DCC-NL model of De Nard et al. (2021). To this end, we use the Matlab implementation of DCC-NL found on Michael Wolf's website but replaced the correlation

targeting matrix with the QIS estimator.[4] Note that if at any rebalancing date, the estimation of the GARCH model (via the Matlab routine) produces a convergence issue, we add to the returns a noise term that is normally distributed with an expected value of 0 and a level of standard deviation, which increases from 10^{-8} to 10^{-4} in step size of 20 according to a geometric progression until convergence is achieved.

Finally, given that we exclude one of the two assets whose correlations are in excess of 0.95, our investment universe includes only stocks with moderate correlations. In such an environment, it may be the case that having an accurate correlation matrix estimate is of second-order importance relative to estimating the variances of each asset accurately. Thus, as another benchmark, we consider a GMV portfolio, which assumes a diagonal correlation matrix with the variances estimated with a GARCH(1,1) model, denoted as VOL.

Table B.1.6 summarizes the performances of the five portfolios: VOL, DCC-NL, EWA-CV, AFM1-DCC-NL, and AFM1-EWA-CV. Compared to Table B.1.1, the standard deviation of EWA-CV estimators has reduced with devolatization for $N = 500, 1{,}000$. For example, for $N = 1{,}000$, the standard deviation for EWA-CV and AFM1-EWA-CV decreased by 9.7% and 16.4%, respectively. Moreover, AFM1-EWA-CV appears to benefit the most from devolatization and is now the best performer for $N = 500, 1{,}000$.

We also see that the performance of EWA-CV estimators is better than VOL for all N by a wide margin. Hence, the performance improvements we observe in our estimator are not just due to better estimates of individual stocks' volatility; they also stem from better estimates of correlations.

Interestingly, the performances of EWA-CV and DCC-NL estimators are quite similar for all investment sizes N. For instance, for $N = 1{,}000$, the standard deviation reduction of EWA-CV over DCC-NL is 3%, whereas for AFM1-EWA-CV over AFM1-DCC-NL it is 1%. This observation is not unexpected given that both estimators employ a similar dynamic capture and shrinkage technology. The EWA-CV estimators also tend to be accompanied by a marginal improvement in average returns compared to their DCC-NL analogs, and also slightly better gross and net information and Sharpe ratios.

2.6.6 Markowitz Portfolio with Signal

The Markowitz problem with a fully-invested constraint is expressed as

$$\text{such that} \quad \begin{array}{l} \min_{w} w'\Sigma w \\ w'm \geq \mathcal{G}, \ \text{and} \\ w'\mathbf{1} = 1, \end{array} \tag{2.21}$$

where m is a predictive signal and $\mathcal{G}$ is an investor's minimum target expected return. We consider the momentum signal as there is substantial empirical evidence that documents this anomaly in the returns of individual stocks (Jegadeesh, 1990; Jegadeesh and Titman, 1993). A momentum signal is calculated by taking the geometric average of the past year's stock returns, excluding the most recent month's returns. Taking the current momentum values for all N stocks in the universe produces a cross-sectional predictive signal m. This application is similar to Engle et al. (2019) for reproducibility.

[4]The code can be downloaded at http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html.

The well-known solution to this optimization problem is given by

$$\begin{aligned} w &= c_1\Sigma^{-1}\mathbf{1} + c_2\Sigma^{-1}m, \\ \text{where } c_1 &:= \frac{C - \mathcal{G}B}{AC - B^2}, \quad \text{and} \quad c_2 := \frac{\mathcal{G}A - B}{AC - B^2}, \\ \text{with } A &:= \mathbf{1}'\Sigma^{-1}\mathbf{1}, \quad B := \mathbf{1}'\Sigma^{-1}m, \quad \text{and} \quad C := m'\Sigma^{-1}m, \end{aligned} \tag{2.22}$$

where $\mathbf{1}$ is a vector of ones. A feasible portfolio $\hat{w}$ can be obtained by replacing the unknown covariance Σ with a covariance matrix estimator. We also include an equally-weighted portfolio of the top quintile stocks sorted according to their momentum value, which is labeled as EW-TQ. Further, we set the target expected return $\mathcal{G}$ to be the average of the momentum values of the stocks from in this portfolio.

The primary metric for which we evaluate the performance of Markowitz portfolios will be the out-of-sample information ratio. We also report the statistical significance of the differences between the IR values of EWA-SC and EWA-CV derived by using the two-sided p-value of the prewhitened HAC_{PW} test described by Ledoit and Wolf (2008, Section 3.1). Since the purpose of this paper is to demonstrate that EWA-CV is an improvement over EWA-SC, we confine our attention to the comparison of EWA-SC with EWA-CV to avoid a multiple testing problem.

The out-of-sample performances are reported in Table B.1.7 and the results from the IR column are summarized as follows:

- EWA-CV outperforms EWA-SC for all N and is statistically and economically significant at the 0.01 level for $N = 500, 1{,}000$.
- Overall, the EWA-CV portfolio delivers the best results, owing in part to its ability to adapt to the non-stationary nature of the financial returns. Its performance can be seen to improve with increasing universe size N.
- Even though AFM1-EWA-CV has slightly better performance than the NL estimators, it does not lead to any material improvement above EWA-CV.
- The increase in information ratio of EWA-CV over NL in $N = 100, 500, 1{,}000$ amounts to 1%, 6%, and 9%, respectively. A similar conclusion holds for the comparison between EWA-CV and AFM1-NL since both AFM1-NL and NL share similar performance, which is consistent with the findings of De Nard et al. (2021, Table A.3).

In addition, we find that EWA-CV produces the highest out-of-sample information ratio and Sharpe ratio after incorporating transaction costs at the post-formation stage of the portfolio for $N = 500, 1{,}000$. Further, Engle and Colacito (2006) suggests assessing the performance of a Markowitz portfolio in terms of the out-of-sample standard deviation. Based on this metric, we find that EWA-CV has the lowest value across different N. In terms of the maximum drawdown incurred by the portfolio, we see that both EWA-CV and AFM1-EWA-CV suffer the least for $N = 500, 1{,}000$. For instance, if $N = 1{,}000$, the maximum drawdown for EW-TQ is 58.56%, while for EWA-CV, it is 23.19%, which is a 60% reduction.

Table B.1.8 describes the distribution of the portfolio weights of the estimated portfolios. We see that EWA-CV has the least dispersed weight while both SC and EWA-SC have the most dispersed weight. Furthermore, EWA-CV and AFM1-EWA-CV have the lowest turnover, which is interesting given that we make no effort to control the trajectory of the weights. Moreover, EWA-CV also has the lowest amount of gross exposure and the lowest proportion of leverage.

2.6.7 Robustness Checks

In this section, we inspect whether the outperformance of our proposed estimator is robust to different revisions in the current empirical setup. In particular, we will be interested in parameters that dictate (1) sub-sample period, (2) exponential decay rate, and (3) devolatization.

2.6.7.1 Subsample Analysis

Here, we check if there are any peculiar subsample effects that may drive the performances of our proposed scheme. We divide the out-of-sample period into four equally-sized subsamples of 100 months (that is, roughly 8 trading years) each: (1) 09/04/1986 to 12/21/1994, (2) 12/22/1994 to 04/25/2003, (3) 04/28/2003 to 08/25/2011, and (4) 08/26/2011 to 12/31/2019. Then we perform the same procedure in each subsample. The results are provided in Table B.1.9.

We see that the performance ranking across different portfolios is relatively consistent over time for $N = 500, 1{,}000$ with EWA-CV estimators being the best performers. For $N = 100$, however, there are brief episodes where it is not; EWA-SC outperforms in Sample 1, EW-TQ in Sample 2, and NL (and AFM1-NL) in Sample 4. However, the performance of EWA-CV is close to the leading ones in these scenarios. In terms of standard deviation, EWA-CV consistently has the lowest value across all subsamples and universes.

2.6.7.2 Sensitivity Analysis

In our base experiments, we parameterize the EWA-based estimators with a decay rate of 0.997. We now repeat the same backtest exercise as in the previous section, but focus only on EWA-based estimators for different decay rates $\delta \in \{0.999, 0.997, 0.995, 0.992, 0.990, 0.985\}$ and $N \in \{100, 1{,}000\}$. This allows us to examine if the performance of our estimator is sensitive to changes in the decay rate specification or if there are other choices that may lead to better results.

According to Table B.1.10, EWA-CV outperforms EWA-SC across different decay rates for $N = 1{,}000$, whereas for $N = 100$, EWA-SC outperforms EWA-CV only marginally in Sample 1. Moreover, for $N = 1{,}000$, EWA-CV tends to deliver a higher information ratio with decay rates greater than 0.995. Unreported simulation results also indicate that decay rates lower than 0.985 consistently give a lower information ratio. This observation holds for the full sample period and for different subsamples. For $N = 100$, there appears to be some performance benefit from tilting the EWA-CV towards more recent observations in some periods. For example, in Sample 2 and Sample 4, $\delta = 0.985$ gives the highest information ratio, which is a 23% and 15% improvement over using $\delta = 0.997$.

2.6.7.3 Devolatization

The base case experiments focus exclusively on the performance improvement exponential weighting adds over a uniform weighted shrinkage estimator. We now consider applying EWA-based estimators to a panel of returns series that have been devolatized through a univariate GARCH model. This will allow us to facilitate comparison with the DCC-NL model of Engle et al. (2019). As another benchmark, we consider a Markowitz portfolio, which assumes a diagonal correlation matrix with the variances estimated with a GARCH(1,1) model, denoted as VOL.

Table B.1.12 summarizes the performances of the five portfolios: VOL, DCC-NL, EWA-CV, AFM1- DCC-NL, and AFM1-EWA-CV. Compared to Table B.1.7, it can be seen that the information ratio of EWA-CV has increased with devolatization for $N = 100, 500$ but reduced for $N = 1{,}000$. AFM1-EWA-CV appears to benefit the most from devolatization and is now the best performer for $N = 500, 1{,}000$.

We also see that the information ratios of the EWA-CV estimators are higher than VOL for all N by a wide margin. Hence, the performance improvements we observe in our estimator are not just due to better estimates of individual stocks' volatility; they also stem from better estimates of correlations.

Interestingly, the performances of EWA-CV and DCC-NL estimators are quite similar for all investment sizes N. For example, the improvement in the information ratio of EWA-CV over DCC-NL for $N = 1{,}000$ is 4%, while for AFM1-EWA-CV over AFM1-DCC-NL it is 1%. EWA-CV estimators also have slightly better net information and Sharpe ratios.

2.7 Conclusion

This paper introduces a novel covariance estimator that performs in large dimensions and adapts to the persistent heteroskedastic environments of financial stock returns. Such scenarios are often found in practical applications of portfolio selection for which our estimator is highly relevant. Our approach involves weighting the sample covariance matrix towards more recent data using exponential weighted averages (EWA), combined with a novel use of the existing cross-validation (CV) shrinkage technology. The former allows us to capture the persistent heteroskedastic dynamics of financial time series with more flexible sample eigenvectors, while the latter attenuates the adverse bias effects that are embedded in the sample eigenvalues. We call our estimator EWA-CV. Beyond these statistical benefits, our proposed estimator possesses several appealing features such as the simplicity of implementation and transparency for independent verification.

We conducted simulation experiments to gain insight into the channels driving the outperformance of our estimator over the exponentially weighted average covariance and other uniformly weighted estimators within the sample. Our empirical experiments on real financial data further reinforce its superiority, exhibiting improved realized measures in larger investment universe sizes when compared to the exponential weighted sample covariance and uniformly weighted shrinkage estimators. Additionally, we demonstrated that our proposed shrinkage estimator(s) rivals the performance of the state-of-the-art dynamic conditional correlation nonlinearnear shrinkage (DCC-NL) estimator introduced by Engle et al. (2019) and its factor model extension by De Nard et al. (2021) while incurring less implementation and computational overhead as well as simpler routines. Therefore, our proposal provides practitioners in risk and asset management with an attractive, reduced-form alternative to existing large-dimensional multivariate dynamic volatility models for financial returns.

In general, the adverse noise effects from a large-dimensional regime can be further reduced by incorporating additional prior knowledge about the orientation of the eigenvectors. For example, this can include overlaying the covariance matrix with economically driven models that possess a small number of risk parameters or with an appropriately structured block matrix reflecting the sector relationships of the stock returns. A cross-pollination of a postulated covariance matrix with nonlinear shrinkage technology provides an interesting avenue for further

research.

Chapter 3

Large Covariance Estimation by Hierarchical Clustering and Shrinkage

> The greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgment with which it is anywhere directed, or applied, seem to have been the effects of the division of labour.
>
> —Adam Smith, *The Wealth of Nations* (1776)

3.1 Introduction

The challenge of estimating the covariance matrix of financial returns for the purpose of Markowitz (1952) portfolio selection is well known amongst academics and practitioners. The standard approach is to use the sample covariance matrix. However, when the dimensionality of the problem is large relative to the number of observations, estimation errors of the sample covariance matrix can create issues for portfolio optimizers (Jobson and Korkie, 1980) as they tend to place extreme bets on low-risk coefficients. This observation led Michaud (1989) to refer to mean-variance optimizers as "error maximization."

To address the issues in the sample covariance matrix, one must reduce the number of effective risk parameters needed to be estimated. This can be achieved by imposing some structure on the estimation problem. In finance, factor models are often employed for this purpose. The Sharpe (1963) single-index model serves as an early example, which assumes that the comovements of returns are driven only by the market-wide risk factors. However, as King (1966) demonstrated in his pioneering study, there are important covariances among stock returns beyond those due to market-wide risk. Indeed, the study showed that industry-wide risk factors – that is, the comovements of groups of stocks within their industries – also explain the returns on individual stocks. The study also indicates an extra factor beyond industry may

be required to completely explain common stock returns.

Rather than strictly assigning stocks to the respective industries to build industry factors, groups or clusters of stocks can also be built solely from their propensity to co-move together. The seminal work of Mantegna (1999) demonstrated that a meaningful economic taxonomy that links the stocks together can be obtained by embedding the stocks into a minimum spanning tree through hierarchical clustering. Tumminello et al. (2007) showed that the taxonomy that arises from this construction can be associated to a factor model. This structure is encoded in a correlation matrix that has a smaller degree of freedom, but retains the hierarchical backbone of the original correlation matrix, which can be leveraged for portfolio selection; for example, see Tola et al. (2008) for such an application. Ahn et al. (2009) relied on such clustering methods to group individual stocks into a set of portfolios that better represent an investor's opportunity set (also referred to as basis stocks) and find that it improves the conditioning property of the covariance matrix.

A different approach to the estimation problem is to not take any stance on the underlying structure of the covariance. These estimators are called structure-free and they typically belong to the rotational equivariance framework of Stein (1975, 1986), which postulates that the sample eigenvectors should be preserved but sample eigenvalues are allowed to be modified. An example that falls under this class of estimators is the linear shrinkage of Ledoit and Wolf (2004b), which shrinks the sample eigenvalues toward their grand mean. A more elaborate extension of this framework that allows for greater flexibility in the shrinkage would be the analytical-based approach from Ledoit and Wolf (2012, 2015, 2020), and Bun et al. (2017) or numerical-based approach of Abadir et al. (2014), Lam (2016), and Tan and Zohren (2024). In particular, the former approach leverages the machinery of large-dimensional asymptotics to consistently estimate the optimal shrinkage function, while the latter appeals to brute-force spectral decompositions for this purpose.

The aim of this paper is to provide an estimated covariance matrix that takes advantage of the hierarchical structure of financial returns that is also robust against large dimensions. We achieve this aim by combining methods from two different modeling paradigms. The first method is from Tola et al. (2008) to build a regularized covariance matrix that retains the hierarchical structure of the original covariance matrix. The second method is the nonlinear shrinkage of Tan and Zohren (2024), which is used to suppress variations in the data that are unexplained by the hierarchical structure. Even though both methods exist in the literature, we present a hybrid method that leverages both innovations in order to produce an estimator that possesses greater efficiency gains over its stand-alone counterparts.

The remainder of the paper is organized as follows. Section 3.2 provides a brief description of hierarchical agglomerative clustering and the regularized covariance matrix generated from such a procedure. Section 3.3 describes our proposed estimation scheme. Section 3.4 provides the backtesting results using real-life stock return data. Section 3.5 concludes. Appendices B.2 and C.3 contain all the figures and tables.

3.2 Hierarchically Nested Factor Model

3.2.1 Notation

In this section, we introduce the necessary notation to accompany our analysis. Let the subscript i index the stock returns such that $i \in \{1, \ldots, N\}$, where N denotes the dimension of the investment universe. The subscript t indexes the dates such that $t \in \{1, \ldots, T\}$, where T denotes the number of observations. Let $x_{t,i}$ be the return for a risky stock i at date t, stacked into a vector $x_t \coloneqq (x_{t,1}, \ldots, x_{t,N})'$. Here, the symbol $(')$ denotes the transpose operator of a vector or matrix, the symbol $\coloneqq$ is a definition sign, and the symbol $=$ is the equal sign.

Let $\Sigma \coloneqq \mathsf{Cov}(x_t)$ and $\Omega \coloneqq \mathsf{Corr}(x_t)$ be the population covariance and correlation matrices of the returns, respectively. Generic estimators of them are denoted as $\hat{\Sigma}$ and $\hat{\Omega}$. Let S and R denote the sample covariance matrix and sample correlation matrix, respectively. Due to the symmetric positive semidefinite property of the covariance and correlation matrices, they admit a spectral decomposition. For example, $\Sigma = V \Lambda V'$, where $V \coloneqq [v_1, \ldots, v_N]$ is an orthogonal matrix whose columns contain the eigenvectors and Λ is a diagonal matrix containing the eigenvalues. The eigenvalues are assumed to be sorted in ascending order. The matrix factor $\Sigma^{1/2}$ can be obtained by retaining the eigenvectors of Σ, but reassembling them with the square roots of the eigenvalues given by $(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})'$.

3.2.2 Hierarchical Clustering

Cluster analysis is a dimension reduction tool used to organize data into groups within which they have a high correlation and between them, they have a low correlation without any prior knowledge about their predefined groups. The notion of similarity of the data can be expressed quantitatively through a distance measure, which can take on different forms depending on the task at hand. Since we are interested to detect patterns of comovements in the stock universe, we select a distance measure that is conditioned on the correlation between any two stock returns.

We consider the following distance measure $d(i, j)$ between stocks i and j

$$d(i, j) = 1 - r_{i,j}, \tag{3.1}$$

where $r_{i,j}$ denotes the correlation coefficient between two stocks. Stocks with perfectly correlated returns have a distance of 0 from each other, while perfectly anti-correlated stocks have a maximum distance of 2. Since this distance measure fulfills the axioms of a metric distance, it can be input into a clustering algorithm.

Once we calculated the distances, we need to make make a hypothesis about the underlying structure of returns and select a method for linking these stocks together. Hierarchical agglomerative clustering is a simple algorithm employed by Mantegna (1999) to systematically examine the correlation matrix of returns and group stocks according to their mutual similarity. The hierarchical structure is an attractive hypothesis from the point of view of modeling stock returns as it allows the possibility of detecting relationships in stocks that are not simply due to stringent sector (or industry) effects, but also broader than sector effects.

The agglomerative algorithm starts by assigning each stock its own cluster. Then, given a way of measuring the distance between a pair of clusters groups of stocks, we iteratively merge the closest pair of clusters and update our cluster distance for $N-1$ steps until all the stocks belong to a single cluster.

In this paper, we consider the average linkage criterion, which is defined as the average distance between each stock in one cluster to every point in the other cluster:

$$d(p,q) \coloneqq \frac{\sum_{i\in C_p}\sum_{j\in C_q} d(i,j)}{|C_q||C_p|}, \tag{3.2}$$

where C_p and C_q denote the sets of elements belonging to the clusters p and q respectively, and $|C_p|$ and $|C_q|$ are their cardinality. There are other linkage methods for computing the cluster distances, which can alter the clustering characteristics; examples include single linkage, complete linkage, the centroid method, and Ward's method. We chose the average linkage criterion due to its ability to produce more structured and balanced clusters as it uses information from all pairs of distances

Remark 4 (Correlation measures). Two commonly used measures to indicate the strength and direction of the relationship between two variables are the Pearson and Spearman correlation coefficients. Pearson correlation measures linear relationships between continuous variables, while Spearman correlation measures monotonic relationships between variables. We shall use the Spearman correlation in the empirical exercises since it is often considered to be more robust in situations where the assumptions of linearity and normality are not met, and when outliers are present. The coefficient matrix R is calculated by first mapping each variable into its rank, and then finding the Pearson correlation coefficient between the two sets of ranked values.

3.2.3 Cluster Analysis

The sequence of clustering assignments can be visualized through a *dendrogram*. The dendrogram is a binary tree data structure with the following nomenclature. The "leaf nodes" correspond to the data points, which are also indexed by $i \in \{1,\dots,N\}$. The "internal nodes" represent the clusters of its children leaves; we denote the first internal node in which leaf nodes i and j are merged together by $(g_k)_{k=1\dots N-1}$, where the index k tracks the step of the merging operation. The "root node" is the group containing the whole data set. Given N variables, there are $N-1$ nodes in the dendrogram given a succession of $N-1$ pairwise merge operations. We can organize the sequence of $N-1$ clustering assignments in a set known as the *genealogy*, which is denoted by G. The genealogy is an ordered set of internal nodes connecting the leaf i (or an internal node g_k) to the root node g_1.

A hypothetical dendrogram, with labeled internal nodes and root, for $N=5$ is shown in the left panel of Figure C.3.1. In this example, the genealogy associated with leaf 2 is $G(2)=\{g_2,g_1\}$, while the genealogy of the internal node g_4 is $G(g_4)=\{g_4,g_3,g_1\}$. We use the notation $g_k=\mathfrak{g}(g_l)$ to express g_k as a parent node to g_l. For example, $g_3=\mathfrak{g}(g_4)$ in the same figure.

We also provide a more concrete example in the right panel of the figure where the leaf nodes correspond to the individual stock returns (labeled as Coke, Pepsi, IBM, Intel, and Amazon), and the internal nodes correspond to the latent factors (labeled as Beverage, Silicon Cost, Technology, and Oil Cost). This example underscores the complex interrelationship that

might govern common stock returns. For example, even though Coke and Intel belong to different industries, their price fluctuations may be correlated through more complex supply chain relationships.

A more elaborate example of a dendrogram based on a correlation matrix computed from $N = 100$ daily returns of liquid U.S. common stocks with $T = 1{,}260$ observations obtained from the period 1994 to 1999 is shown in the left panel of Figure C.3.2. Note that the dendrogram in the figure has been enriched with further information such that the leaves corresponding to each stock are colored according to the stock's industry assignment based on the Global Industry Classification Standard (GICS) sector codes; see Table B.2.1. From the figure, we can see the different scale levels where pairs (or groups) of stocks are merged, along with the value of the distance at which they are merged. Although the clusters at each level have no intrinsic meaning, we do observe compatibility between the clusters of stocks and the GICS assignments.

Finally, one could study the hierarchical structure of the same set of common stock returns that is not explained by the market factor.[1] The right panel in Figure C.3.2 is a dendrogram that was derived by applying cluster analysis to the panel of residuals obtained by regressing out the market factor from the common stock returns. As a result of isolating the average behavior of the stocks, the distances between any two stocks are now at a higher level than in the dendrogram on the left panel. However, the hierarchical organization of common stock returns remains visible.

3.2.4 Hierarchically Ordered Correlation Matrix

We can also use the output of the clustering assignments G to re-order the stock indices of the correlation matrix in order to further expose the grouped-based behavior in the covariations between the stocks. Figure C.3.3 illustrates two heatmaps of the sample correlation matrix of stocks returns accompanied by a color bar at the side that indicates the strength of the pairwise correlations. The heatmap from the left panel corresponds to the stocks whose indices are unordered, whereas the heatmap on the right panel, the stocks have been re-ordered according to the indices from the genealogy set G.

We can see that after sorting the stocks by their clustering assignments, some block structures representing the clusters appear to be visible along the diagonal of the matrix. Furthermore, within each block diagonal, there are also additional block structures along its diagonal; this nested block structure reflects the hierarchical organization present in the correlation matrix of stock returns.

3.2.5 The Regularized Covariance Matrix

While cluster analysis can be viewed as an exploratory exercise, it can also be used to build a regularization scheme for the correlation matrix that retains information about the hierarchies. In particular, given a set of indices $\mathcal{I}_{pq} = \{(i, j) : i \in C_p, j \in C_q\}$, Tola et al. (2008) proposed a

[1] Although it has been widely documented in the literature of empirical finance since Sharpe (1963) that there is a single factor, namely a market factor, in the panel of stock returns, the cluster analysis of stock return correlations by King (1966) provided empirical evidence that traditional industrial groupings also explain part of the historical variability in stock returns after the removal of the market factor. This finding is also supported by studies of Cohen and Pogue (1967), Elton and Gruber (1973), among others, dealing with multi-index portfolio selection models.

regularized correlation matrix, denoted as $\hat{\Omega}$, whose (i,j)-th element is given by

$$\hat{\omega}_{i,j} \coloneqq 1 - d(p,q) \quad \text{where} \quad (p,q) \in G, \quad (i,j) \in \mathcal{I}_{pq}. \tag{3.3}$$

It is immediate that $\hat{\Omega}$ is a symmetric matrix with ones on the diagonal.

With our internal node labeling, we can associate to each internal node g_k a correlation coefficient such that $\hat{\omega}_{g_k} = \hat{\omega}_{i,j}$. Assuming that $\hat{\omega}_{g_1} \geq 0$, the agglomeration algorithm implies that $\hat{\omega}_{g_k} \leq \hat{\omega}_{g_{k+1}}$ for all k. Intuitively, the element $\hat{\omega}_{i,j}$ is equal to the correlation value associated with the first node in the dendrogram in which the two elements merge together. It also corresponds to the simplest matrix that shares the same dendrogram as the original sample correlation matrix R. Thus, $\hat{\Omega}$ is a reduced matrix that retains the hierarchical structure of the original sample correlation matrix.

Note that the regularized correlation matrix $\hat{\Omega}$ has $N-1$ distinct correlation values since it is constructed from a genealogy set G, which has $N-1$ elements. Hence, it presents a significant reduction in the degrees of freedom over the sample correlation matrix. Figure C.3.4 illustrates the heatmap of the regularized correlation matrix $\hat{\Omega}$ based on the same set of stock returns from Section 3.2.3. It is apparent that the regularization procedure from Equation (3.3) smooths out the elements of the (sub-)matrices of the sample correlation matrix by means of averaging. In addition, the regularized correlation matrix in this example has a condition number of 104 versus 125 for the sample correlation matrix, therefore it is better conditioned.

A regularized covariance matrix $\hat{\Sigma}_x$ can be built by means of the sample variances and the regularized correlations:

$$\hat{\sigma}_{i,i} = s_{i,i}, \quad \text{and} \quad \hat{\sigma}_{i,j} = \hat{\omega}_{i,j}\sqrt{s_{i,i}s_{j,j}}. \tag{3.4}$$

Remark 5 (Positive definiteness). Tumminello et al. (2007) demonstrated that the regularized correlation matrix $\hat{\Omega}$ obtained from an average linkage clustering procedure is positive definite provided all the elements in the obtained regularized correlation matrix are positive. Since this condition may not be guaranteed in practice, we enforce a positive definiteness on the covariance by maintaining its positive principal components.

3.2.6 Factor Models

Another class of dimension reduction tool that is often compared with clustering is the factor model, which has a long history in finance and widespread use in practice. They are characterized by a K-dimensional factor $f_t \coloneqq (f_{t,1}, \ldots, f_{t,K})'$ representing the comovements, and a residual component $u_t \coloneqq (u_{t,1}, \ldots, u_{t,N})'$ representing the individual stock features. The idea is to decompose the stock returns for every $i = 1, \ldots, N$, into the following representation

$$x_{t,i} = \alpha_i + \beta_i' f_t + u_{t,i}, \tag{3.5}$$

where α_i is an intercept, $\beta_i \coloneqq (\beta_{1,i}, \ldots, \beta_{K,i})'$ is a K-dimensional vector of factor loadings such that $K < N$. The factors can either be observed or latent. Examples of observed factors are those from Fama and French (1992, 2015), while latent factors are obtained through factor analysis or principal component analysis (PCA) (Jolliffe, 2002). The factors can either be static or dynamic. Static factors can only have a contemporaneous effect on the returns, while dynamic

factors allow for lagged effects. In this paper, we shall restrict our attention to the static factor models.

If orthogonality between the common and residual components is assumed, then the covariance matrix implied by this model is given by

$$\Sigma_x = B'\Sigma_f B + \Sigma_u, \tag{3.6}$$

where $\Sigma_f \coloneqq \mathsf{Cov}(f_t)$ and $\Sigma_u \coloneqq \mathsf{Cov}(u_t)$, and B is a matrix whose ith column is the factor loading.

The classes of static factor models can be distinguished based on the nature of the idiosyncratic errors denoted as u_t, for $t = 1, \ldots, T$. In the so-called *exact* factor model (EFM), the factors and the idiosyncratic errors are independently and identically distributed in time and factors and idiosyncratic errors are uncorrelated. Thus, the idiosyncratic errors have no cross-sectional dependence, and so Σ_u is a diagonal covariance matrix. By an extension, Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986) proposed the *approximate* factor model (AFM), which relaxes the assumption on the idiosyncratic errors from an exact factor model and allows the idiosyncratic errors to have mild cross-sectional dependence.[2] Hence, Σ_u may not necessarily be a diagonal matrix.

To estimate the covariance in Equation (3.6), there are three matrices that require attention: B, Σ_f, and Σ_u. In latent factor models, the factors and the loadings are not identified, and so one can assume without loss of generality that $\Sigma_f = \mathbb{I}_K$.[3] The factor loading matrix B can be estimated through maximum likelihood or PCA. In observed factor models, the estimation can be performed by means of regression of stock i's returns on the factors. In doing so, this would provide us with estimates for the intercept $\hat{\alpha}_i$ the factor loadings $\hat{\beta}_i$, and hence the residuals $\hat{u}_t$. The covariance of factors can be estimated by the sample covariance matrix, which we denote as $\hat{\Sigma}_f$. Finally, in both latent or observed settings, the covariance of estimated residuals $\hat{\Sigma}_{\hat{u}}$ is determined based on the exact or approximate nature of the factor model. In either case, we write the covariance estimator of the returns as

$$\hat{\Sigma}_x = \hat{B}'\hat{\Sigma}_f\hat{B} + \hat{\Sigma}_{\hat{u}}. \tag{3.7}$$

In this paper, we are interested in the approximate factor model because it captures the idiosyncratic cross-sectional dependencies that are relevant to many financial applications. Indeed, the left panel of Figure C.3.5 shows that the sample correlations across residual components after subtracting the market component are in agreement with an approximate factor model. Furthermore, by applying cluster analysis to the residuals, both right panels of Figure C.3.2 and C.3.5 suggest that the hierarchical organization of the data is still preserved along with its compatibility with the GICS sectors. This observation is consistent with that of King (1966) who finds that the residual covariances after removing the market factor can be attributed to comovements within the industries.

[2] The term "mild" means that the idiosyncratic components can be averaged out.

[3] This is possible due to the rotational indeterminacy of the latent factor model, where one can always rotate the factors (and the loadings) so that their covariance matrix is an identity matrix.

3.2.7 Hierarchical Trees and Factor Models

Both tree-based clustering methods and latent factor models are intimately related. For example, a single-factor model can be thought of as a tree model where the root node is connected to all the leaf nodes. Such a link is useful as it helps us to dovetail hierarchical clustering with factor models.

Tumminello et al. (2007) showed that $\hat{\Omega}$ is a correlation matrix that is associated with a latent factor model that has a hierarchically nested structure given by

$$x_{t,i} = \sum_{k=1}^{K} \beta_{k,i} f_{t,k} + \eta_i u_{t,i}, \quad \text{with} \quad \eta_i \coloneqq \sqrt{1 - \sum_{k=1}^{K} \beta_{k,i}^2}. \tag{3.8}$$

The latent factors are associated with the internal nodes, and so we have $K = N - 1$ factors. For example, the nodes $g_1, \ldots, g_4$ in the binary tree of Figure C.3.1 represent factors that are potentially unobserved. This expression has a similar form to that in Equation (3.5) with the exception that $\alpha_i = 0$ and the residual is now scaled by η_i. In this setting, the index k of the factor model now indexes the internal nodes. Moreover, each latent factor can be thought of as returns to a portfolio of stocks from a given cluster.

In order to ensure that the correlation matrix of the hierarchically nested factor model (3.8) is $\hat{\Omega}$, the factor loadings $\beta_{k,i}$ have to be chosen according to

$$\begin{aligned} \beta_1 &= \sqrt{\hat{\omega}_{g_1}}, \\ \beta_k &= \sqrt{\hat{\omega}_{g_k} - \hat{\omega}_{\mathfrak{g}(g_k)}} \quad \text{for} \quad k = 2, \ldots, N-1, \end{aligned} \tag{3.9}$$

where, assuming $\hat{\omega}_{g_1} \geq 0$, all the factor loadings β_k are non-negative real numbers.

An illustration of this idea can be seen in the correlation matrix heatmap from Figure C.3.4. It shows the presence of a large cluster at the bottom right corner of the matrix, which can be associated with a dominant latent factor that captures the co-movements of stock returns. We can draw similar remarks for other visibly large-sized clusters.

3.3 Proposed Estimation Scheme

Building on the insight from Section 3.2.6, we propose an approximate factor model (3.7) for the covariance of stock returns that accounts for the market factor while applying the regularization methodology from Section 3.2.5 to the residual covariance matrix Σ_u. This allows us to exploit the hierarchical structure in the residuals while also reducing their intrinsic dimensionality. Figure C.3.6 shows the correlation matrix heatmap of the residualized stock returns that arise from this regularization procedure.

Notwithstanding this innovation, this operation may still result in a severely biased estimate of the covariance structure of stock returns due to structural misspecification. The reason is that the regularization procedure assigns a common value to all merged elements from the cluster pair, which can inadvertently suppress important variations from the sample correlation matrix. Therefore, we look to augment the sample information into the estimation process that is unexplained by our estimated approximate factor structure.

To this end, we borrow the preconditioning idea by Ledoit and Wolf (2017) to isolate the

hierarchically nested factor structure in the stock returns by defining the following returns matrix $\tilde{X} \coloneqq X \times \hat{\Sigma}_x^{-1/2}$. We then employ shrinkage to the preconditioned returns matrix $\tilde{X}$, in order to yield a shrunked covariance $\hat{\Sigma}_{\tilde{x}}$. For this effort, we rely on the cross-validation shrinkage estimator, EWA-CV from Tan and Zohren (2024) since it offers comparable performance to existing state-of-the-art shrinkage estimators while also providing researchers the option to emphasize more recent observations with an exponential decay rate parameter δ. Our hybrid estimator $\hat{\Sigma}$ is then obtained by reassembling the hierarchical structure through the following operation:

$$\hat{\Sigma} \coloneqq \hat{\Sigma}_x^{1/2} \times \hat{\Sigma}_{\tilde{x}} \times \hat{\Sigma}_x^{1/2}. \tag{3.10}$$

This sequence of operations preserves the cross-sectional structure of the data while simultaneously shrinking part of it that is unexplained by the hierarchical structure.

3.4 Empirical Analysis

3.4.1 Data and Portfolio Construction Rules

We obtain daily stock returns data from the University of Chicago's Center for Research in Security Prices (CRSP), starting at 01/01/1995 and ending at 12/31/2021. Our analysis is limited to stocks traded on the NYSE, AMEX, and NASDAQ stock exchanges. The size of the investment universe that we consider is $N \in \{100, 500, 1{,}000\}$. Further, we only study the universe of U.S. common stocks (CRSP share codes 10 and 11); this excludes preferred stocks, unit investment trusts, closed-end funds, real estate investment trusts, Americus Trusts, foreign stocks, and American Depositary Receipts. We also download the market factor from Kenneth French's website.

We define one "month" and "year" as 21 and 252 consecutive trading days, respectively. The portfolios are rebalanced on a monthly basis using a rolling window scheme where only past information is used to avoid a look-ahead bias. During the month, the number of shares is kept fixed in order to avoid incurring unnecessary transaction costs. The covariance matrices are estimated using an in-sample period of size $T = 1{,}260$. The out-of-sample period starts at 01/21/2000 and ends at 12/31/2021. This provides us with a total of 264 months (or 5544 trading days) of consecutive non-overlapping observations of length 21 days for which the portfolios are rebalanced.

To construct a well-defined investment universe on which we can estimate the covariances on, we use a similar procedure described in De Nard et al. (2021). For each rebalancing month $h = 0, 1, \ldots, 263$ (using zero-based indexing), we first select the stocks that have complete data over the past 1,260 days as well as over the subsequent 21 days. The backward and forward restrictions ensure that we have data to estimate our models and to evaluate out-of-sample. Then, we search for pairs of highly correlated stocks (that is, those with a sample correlation exceeding 0.95) and remove the stock with the lower volume in each pair. From this remaining set of stocks, we select the largest N stocks as measured by their market capitalization on the rebalancing month h to include in our investment universe.

3.4.2 Candidate Portfolios

To test the performance of different covariance matrix estimators in the context of portfolio selection, we consider two investment problems: (1) the global minimum variance (GMV) portfolio and (2) the Markowitz portfolio with a given signal. Both problems are described in Section 3.4.4 and 3.4.6, respectively. The GMV problem will be our principal focus in the main text because it is useful for isolating the estimation error stemming from the expected returns and allows us to assess the efficiency gains of the covariance matrix estimator.

In either setting, we employ the following candidate estimators as their inputs:

- SF: an exact factor model based on the market factor.
- NL: the nonlinear shrinkage estimator of Ledoit and Wolf (2022).
- EWA-CV: the cross-validation shrinkage estimator of Tan and Zohren (2024). The parameter δ controls the degree of exponential smoothing, that is the number of effective observations.
- HC: the hierarchical clustering-based regularized estimator of Tola et al. (2008).
- FNL: the approximate factor model estimator of De Nard et al. (2021) based on the market factor with nonlinear shrinkage applied to the residuals.
- FHC: an approximate factor model estimator based on the market factor with hierarchical clustering-based regularization applied to the residuals.
- NL-SF: the single-factor-preconditioned nonlinear shrinkage estimator of Ledoit and Wolf (2017).
- EWA-CV-FHC: our proposed FHC-preconditioned cross-validation shrinkage estimator.

Note that for the EWA-CV-based estimators, we choose a fixed exponential decay rate that is relatively high at $\delta = 0.997$, which makes them slightly closer to the uniformly weighted setting.

To keep our empirical analysis brief, we only consider seven static estimators, with NL-SF as the main benchmark in both experiments. This is because we are interested to test the hypothesis of whether preconditioning the data with an approximate factor model (modeled as the market and hierarchically embedded factor structure, FHC) would result in better out-of-sample properties with shrinkage than an exact factor model (modeled as the single market factor, SF) with shrinkage.

3.4.3 Evaluation Methodology

We evaluate the performance of the different portfolios using four main out-of-sample performance measures: the average out-of-sample returns (AV), the standard deviation of portfolio returns (SD), the information ratio (IR), and the Sharpe ratio (SR). For ease of interpretability, all performance measures are annualized with 252 trading days. The information ratio is computed with respect to the actual returns (as opposed to returns in excess of a given benchmark) since we believe it is more relevant in our context where the portfolios are formed solely on the basis of risky stocks. On the other hand, the Sharpe ratio is the information ratio with the risk-free rate as the benchmark. In addition, we also include the maximum drawdown (MDD), which measures the largest cumulative loss from peak to trough over the entire out-of-sample period.

We also report the information ratio and Sharpe ratio net transaction costs denoted as

$\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. In particular, we assume a constant transaction cost of 5 bps for all stocks, which is a reasonable approximation today. Even though the investment problems that we consider do not explicitly account for transaction costs at the portfolio pre-formation stage, it is nevertheless worthwhile to investigate if our methods survive transaction costs in an unconstrained setup.

Additionally, we report the following portfolio weight statistics averaged over the 264 trading months: turnover (TO), gross exposure (GE; computed as the sum of the absolute value of long and short positions), proportional leverage (PL; computed as the fraction of negative weights), minimum weight (MIN), maximum weight (MAX), the standard deviation of weights (SD), and mean absolute deviation from an equally-weighted portfolio (MAD). Note that these statistics are not our primary focus since our proposed method is not optimized to account for these measures. Nevertheless, they are provided to give a better overview of the different methods.

The primary metric for which we evaluate the performance of the GMV and Markowitz portfolios will be the out-of-sample standard deviation and information ratio, respectively. We also report the statistical significance for the differences of the SD and IR values between EWA-CV-FHC and NL-SF. The tests are based on the two-sided p-value of the prewhitened HAC_{PW} test described by Ledoit and Wolf (2011, Section 3.1) and Ledoit and Wolf (2008, Section 3.1), respectively. Since the purpose of this paper is to demonstrate that EWA-CV-FHC is an improvement over NL-SF, we confine our attention to the comparison of EWA-CV-FHC with NL-SF to avoid the issue of multiple testing.

3.4.4 Global Minimum Variance Portfolio

The global minimum variance (GMV) portfolio is expressed as

$$\begin{aligned} &\min_{w} w'\Sigma w \\ \text{such that} \quad &w'\mathbf{1} = 1, \end{aligned} \tag{3.11}$$

where $w = (w_1, \ldots, w_N)'$ is a vector of portfolio weights and $\mathbf{1}$ is an N-dimensional conformable vector of ones. Empirical studies of Haugen and Baker (1991) and Jagannathan and Ma (2003), among others, have found that the GMV portfolios enjoy favorable out-of-sample performance both in terms of risk and Sharpe ratio when compared with other benchmark portfolios.

In this setting, the solution for optimal weights has an explicit form given by

$$w = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \tag{3.12}$$

To yield a feasible portfolio $\hat{w}$ that can be computed from the returns data, we replace Σ with the covariance matrix estimator $\hat{\Sigma}$. We also include the equal-weighted portfolio promoted by DeMiguel et al. (2009) as another benchmark, denoted as $1/N$, since it is argued to be difficult to beat.

The out-of-sample performances are reported in Table B.2.2 and the results from the SD column are summarized as follows:

- The EWA-CV-FHC portfolio is the best overall performer for all N. Furthermore, given that it is a hybrid portfolio, this also implies that it outperforms its standalone counterparts, namely EWA-CV and FHC.

- The improvements of EWA-CV-FHC over NL-SF are 5.4% 5.5%, and 4% for $N = 100, 500,$ and 1,000, respectively; each is statistically significant at the 1% level.
- FHC has at least a 3.8% lower standard deviation than HC in all settings. This indicates that there are some performance benefits to be realized from keeping the market factor untouched and averaging out the correlation values in the residuals.
- Both HC and FHC consistently have lower standard deviations than SF. This provides some empirical support for the extra-market covariance modeled as a hierarchical structure.

Additionally, we find that EWA-CV-FHC outperforms NL-SF in terms of information ratio by 12.5% and 13.8% for $N = 500$ and 1,000, respectively. Once we have taken into account the transaction costs at the post-formation stage of the portfolio, we see a decline in the information ratio across all portfolios but the relative performance remains similar. A similar observation also holds for the Sharpe ratio. In terms of the maximum drawdown incurred by the portfolio, we see that EWA-CV-FHC suffers the least for $N = 500, 1{,}000$.

Table B.2.3 describes the distribution of the portfolio weights of the estimated portfolios. We see that EWA-CV-FHC has similar weight dispersion with other portfolios. However, it also experiences comparably high turnover and gross exposure.

3.4.5 Robustness Checks

In this section, we inspect whether the outperformance of our proposed estimator is robust to different revisions in the current empirical set-up. In particular, we will be interested in parameters dictating the (1) subsample period and (2) exponential decay rate.

3.4.5.1 Subsample Analysis

Here, we check if there are any peculiar subsample effects that may drive the performances of our proposed scheme. We divide the out-of-sample period into four equally-sized subsamples of 100 months (that is, roughly 6 trading years) each: (1) 1999 to 2005, (2) 2005 to 2011, (3) 2011 to 2016, and (4) 2016 to 2021. Then we perform the same procedure in each subsample. The results are provided in Table B.2.4.

We see that the ranking of performance across different portfolios is relatively consistent over time. Generally, EWA-CV-FHC is the best performer although there is a brief episode where it is not. For example, EWA-CV-FHC only trails NL-SF in the second place by 1% in Sample 2 for $N = 1{,}000$.

3.4.5.2 Sensitivity Analysis

In our base experiments, we parameterized the EWA-CV-based estimators with a decay rate of 0.997. We now repeat the same backtest exercise as in the previous section but focus only on EWA-CV-FHC for different decay rates $\delta \in \{0.999, 0.997, 0.995, 0.99, 0.980, 0.97, 0.96\}$ and $N \in \{100, 500, 1{,}000\}$. This allows us to examine if the performance of our estimator is sensitive to changes in the decay rates or if there are other choices that may lead to better results.

We see that our default value $\delta = 0.997$ tends to deliver the lowest standard deviation for $N = 500, 1{,}000$. Moreover, there is some advantage in tilting away from the uniformly-weighted configuration; the portfolio standard deviation reduces by more than 1% as δ decreases from 0.999 to 0.997. Given the magnitude of this decrease, this also means that the improvement of our estimator comes mostly from the preconditioning of the hierarchical factor structure. This

observation holds over the full out-of-sample period as well as across different subsamples.

3.4.6 Markowitz Portfolio with Signal

The Markowitz problem with a fully-invested constraint is expressed as

$$\text{such that} \quad \begin{aligned} &\min_w w'\Sigma w \\ &w'm \geq \mathcal{G}, \text{ and} \\ &w'\mathbf{1} = 1, \end{aligned} \tag{3.13}$$

where m is a predictive signal and $\mathcal{G}$ is an investor's minimum target expected return. We consider the momentum signal as there is substantial empirical evidence that documents this anomaly in the returns of individual stocks (Jegadeesh, 1990; Jegadeesh and Titman, 1993). A momentum signal is calculated by taking the geometric average of the past year's stock returns, excluding the most recent month's returns. Taking the current momentum values for all N stocks in the universe produces a cross-sectional predictive signal m.

The well-known solution to this optimization problem from Merton (1972) is given by

$$\begin{aligned} w &= \gamma_1 \Sigma^{-1}\mathbf{1} + \gamma_2 \Sigma^{-1} m, \\ \text{where } \gamma_1 &:= \frac{A_3 - \mathcal{G}A_2}{A_1A_3 - A_2^2}, \quad \text{and} \quad \gamma_2 := \frac{\mathcal{G}A_1 - A_2}{A_1A_3 - A_2^2}, \\ \text{with } A_1 &:= \mathbf{1}'\Sigma^{-1}\mathbf{1}, \quad A_2 := \mathbf{1}'\Sigma^{-1}m, \quad \text{and} \quad A_3 := m'\Sigma^{-1}m, \end{aligned} \tag{3.14}$$

A feasible portfolio $\hat{w}$ can be obtained by replacing the unknown covariance Σ with an estimator $\hat{\Sigma}$ in Equation (3.14). We also include an equally-weighted portfolio of the top quintile stocks sorted according to their momentum value, which is labeled as EW-TQ. Further, we set the target expected return $\mathcal{G}$ to be the average of the momentum values of the stocks from in this portfolio.

We report the out-of-sample performances in Table B.2.6 and summarize the results from the IR column as follows:

- EWA-CV-FHC exhibits the best performance for $N = 1{,}000$, while it is competitive with the best portfolios for $N = 100, 500$. This indicates that there is some benefit of the hybrid approach in large dimensions.
- The outperformance of EWA-CV-FHC over NL-SF for all N and is statistically significant at the 1% level for $N = 500, 1{,}000$.
- Both HC and FHC consistently have higher information ratios than SF. This provides some empirical support for the extra-market covariance modeled as a hierarchical structure.

In addition, we find that EWA-CV-FHC produces the highest out-of-sample information ratio and Sharpe ratio after incorporating transaction costs at the post-formation stage of the portfolio for $N = 1{,}000$. Further, Engle and Colacito (2006) suggests assessing the performance of a Markowitz portfolio in terms of the out-of-sample standard deviation. Based on this metric, we find that EWA-CV-FHC has the lowest value across different N. In terms of the maximum drawdown incurred by the portfolio, we see that the numbers from EWA-CV-FHC are always comparable with the lowest.

Table B.2.7 describes the distribution of the portfolio weights of the estimated portfolios.

We see that EWA-CV-FHC has similar weight dispersion with other portfolios. However, it also experiences comparably high turnover and gross exposure.

3.4.6.1 Robustness Checks: Subsample Analysis

Here, we check if there are any peculiar subsample effects that may drive the performances of our proposed scheme. We divide the out-of-sample period into four equally-sized subsamples of 100 months (that is, roughly 6 trading years) each: (1) 1999 to 2005, (2) 2005 to 2011, (3) 2011 to 2016, and (4) 2016 to 2021. Then we perform the same procedure in each subsample. The results are provided in Table B.2.8.

We see that the ranking of information ratios across different portfolios varies quite a bit over time. On the other hand, the ranking of the standard deviation is more stable over time; for example, EWA-CV-FHC is consistently the lowest. This analysis could suggest that our proposed estimator may be good at minimizing a portfolio variance but at the expense of giving up some upside gains in average returns.

3.4.6.2 Robustness Checks: Sensitivity Analysis

In our base experiments, we parameterized the EWA-CV-based estimators with a decay rate of 0.997. We now repeat the same backtest exercise as in the previous section but focus only on EWA-CV-FHC for different decay rates $\delta \in \{0.999, 0.997, 0.995, 0.99, 0.980, 0.97, 0.96\}$ and $N \in \{100, 500, 1{,}000\}$. This allows us to examine if the performance of our estimator is sensitive to changes in the decay rate specification or if there are other choices that may lead to better results.

Interestingly, for all sizes N, there is some benefit to deviating from our default decay rate configuration of $\delta = 0.997$ for subsamples 1,2, and 4. For example, for $N = 1{,}000$, $\delta = 0.99$ gives the highest information ratio.

3.5 Conclusion

This paper introduces a covariance estimator that (1) leverages upon the hierarchical organization of financial returns, and (2) shrinks part of the cross-section of the returns that is unexplained by this structure. The former is achieved through our approximate factor structure, FHC, where the factor is modeled through the market component, and the covariance of residuals is handled through the hierarchical clustering-based regularized estimator of Tola et al. (2008). The latter is taken care of by the cross-validation shrinkage estimator with exponential smoothing (EWA-CV) of Tan and Zohren (2024). We term our hybrid covariance estimator EWA-CV-FHC.

We demonstrated on historical data that there is an advantage in combining these two approaches in order to produce an estimator that not only outperforms their standalone counterparts but also the challenging hybrid single factor nonlinear shrinkage estimator of Ledoit and Wolf (2017). In terms of future work, we can adapt the techniques developed in Engle et al. (2019) to obtain a dynamic estimator that accounts for both the time-varying volatility and the hierarchical structure in stock returns. It will also be interesting to expand the current modeling effort by by incorporating firm characteristics, such as the book-to-market ratio and market value of equity, to better capture extra-market covariance; for example, see Rosenberg (1974).

Part III

Portfolio Selection

Chapter 4

Canonical Portfolios: Optimal Asset and Signal Combination

A version of this chapter has been published in the *Journal of Banking & Finance*

> Diversification is the only free lunch in finance.
>
> —Harry M. Markowitz

4.1 Introduction

The investment decisions of portfolio managers are often guided by return-predictive signals that reflect their expectations about future returns.[1] Traditionally, the process of constructing a portfolio conditional on the investor's information set is divided into two stages according to modern portfolio theory pioneered by Markowitz (1952). In the first stage, investors make inferences about the return-generating process by either plug-in estimation or subjective belief formation. In the second stage, optimal portfolio weights are formed using estimates from the first stage.

In this paper, we draw inspiration from Brandt and Santa-Clara (2006) and solve a dynamic portfolio selection problem with multiple return-predictive signals in a single stage. Our first contribution is to reformulate the authors' original investment problem into a tractable one that yields a dynamic portfolio policy that directly invests in each basis asset. This is achieved by modeling the joint distribution of the returns and signals as stationary Gaussian variables. Additionally, our approach provides a theoretical framework for optimal investment in the presence of cross-predictability among assets and signals, as well as correlations within them. Our proposed solution is straightforward to implement in practice as the standard static Markowitz solution and is well-suited to handle large-scale cross-sectional applications, such as the equity universe.

The second contribution of our work is to provide insights into our proposed solution by reorganizing the set of assets and accompanying signals into a set of uncorrelated managed

[1] The terminology "signals" is also often referred to as predictors, alphas, attributes, characteristics, state variables, instrumental variables, cross-sectional anomalies, etc. in the finance literature.

portfolios. Each managed portfolio invests in each asset in a proportion that scales with the size of the signals. As discussed in Firoozye and Koshiyama (2020, Section 4), this can be accomplished through the use of a powerful tool from multivariate analysis known as canonical correlation analysis (CCA), which was developed by Hotelling (1936) and has a deep economic interpretation, which we provide. CCA generalizes the principal component analysis (PCA) of Hotelling (1933) to two sets of random variables and their joint association. This makes it particularly relevant to our case, where we have multiple correlated assets and signals that are mutually linked by a joint correlation objective. CCA allows us to express any portfolio returns derived from our framework in terms of their exposures to uncorrelated sources of returns, by reweighting the original set of managed portfolio returns while maximally retaining their correlation. We refer to these special weights as *canonical portfolios* since they can be viewed as long or short positions across different assets and signals.[2]

There are two common approaches for solving the dynamic portfolio selection problem when conditioning information is available. The first approach involves specifying the joint conditional distribution of asset returns and then forming the optimal portfolio. Alternatively, Hansen and Richard (1987) proposed a second approach, which involves augmenting the basis assets to include conditional portfolios, and solve a simpler unconditional mean-variance problem. Research efforts that have built upon the latter approach to developing dynamic trading strategies that exploit time-varying investment opportunities include those by Ferson and Siegel (2001), Brandt and Santa-Clara (2006), and others. Of particular relevance to our study is Brandt and Santa-Clara (2006), in which the authors analyze parametric portfolio policies that are linear in the state variables and/or firm characteristics. Through this parameterization, they are able to recast a dynamic portfolio selection problem involving basis assets into a static portfolio selection problem in an asset space augmented with managed portfolios. The optimization of parameters in the static problem leads to a dynamic strategy that selects among managed portfolios. By directly focusing on the portfolio weights, their optimal portfolio policy accounts for time-variation in the entire return distribution, rather than time-variation in expected returns only.

In this study, we directly model the portfolio weights with a similar parameterization and adopt a joint Gaussian distribution to model the relationship between returns and signals. Our modeling approach allows us to make analytical progress in two main areas. First, it facilitates the derivation of a closed-form dynamic portfolio policy that is applicable even in large-dimensional settings. Second, it allows us to decompose the joint correlation structure of our optimal portfolio policy into linearly independent orthogonal portfolios using CCA, thereby providing valuable insights into the underlying factors that drive our portfolio's performance.

However, these innovations come with additional input requirements when compared to the conventional mean-variance framework, which solely relies on the covariance of the returns matrix. Specifically, we need to consider the covariance of the signals and the cross-covariances between the returns and signals. To render our theoretical framework practicable, we utilize regularization techniques for these large-dimensional objects to encourage stability in the out-

[2]The idea of recasting the asset universe into their orthogonal components through PCA can be traced back to Partovi and Caputo (2004) with further extensions and applications from Meucci (2009) and Avellaneda and Lee (2010), as well as from Kozak et al. (2018, 2020) in the context of risk pricing in a no near-arbitrage framework. Our work aims to broaden that horizon to *two* sets of variables by also considering the signal as an important input in the analysis of their joint association with returns.

of-sample results. In particular, we apply shrinkage to the covariances, a technique commonly employed in portfolio selection problems; see, for example, Frost and Savarino (1986) and Ledoit and Wolf (2004a,b, 2017). Additionally, we identify the leading few canonical portfolios that exhibit the highest predictability, discarding noisy and unstable ones. Through a series of backtesting simulations, we demonstrate that our proposed method consistently outperforms competitive benchmarks.

Our study extends the work of Firoozye and Koshiyama (2019) on the use of total least squares (TLS) (Golub and Van Loan, 1980) for optimally combining signals for a univariate return. Their original study focused on a novel objective function for return prediction. While most research analysts will seek to find a good forecast for future returns via methods such as ordinary least squares (OLS), or in the nonlinear context, via a large suite of machine learning-based methods, Firoozye and Koshiyama (2019, 2020) argue that if the goal is to maximize the Sharpe ratio of trading strategy, then this can be achieved through a linear combination of signals that maximizes the correlation between their combination and the returns. In the linear context, the solution to this problem comes out via TLS, which is an errors-in-variables formulation of regression. TLS has been well-studied primarily among numerical analysts and is used less formally on trading desks of investment banks and hedge funds, typically under the moniker of PCA regression. Given the close relation between TLS and CCA, our work extends this research to a multivariate context, which is more relevant for problems in portfolio selection.

The remainder of the paper is organized as follows. Section 4.2 gives the description of our dynamic portfolio selection problem. Section 4.3 provides the financial interpretation of our optimal portfolio policy with CCA and details our proposed estimation approach. Section 4.4 describes the empirical methodology and presents the results of the out-of-sample backtest experiments with Fama-French equity sorted portfolios. Section 4.5 concludes. Appendix A.1, B.3, and C.4 contain all the additional mathematical derivations, tables, and figures.

4.2 Setting the Stage

4.2.1 Notation

In this section, we introduce the notational convention for our analysis. Let the subscript i index the variables such that $i \in \{1, \ldots, N\}$, where N denotes the number of assets. The subscript t indexes the trading dates such that $t \in \{1, \ldots, T\}$, where T denotes the number of observations. The notation $\mathbb{E}_t[\cdot] \coloneqq \mathbb{E}[\cdot|\mathcal{F}_t]$ represents the mathematical expectation operator of a random vector, conditioned on the information set $\mathcal{F}_t$ available at time t. Furthermore, $\mathsf{Cov}(\cdot,\cdot)$ represents the covariance matrix between two random vectors, $\mathsf{Tr}(\cdot)$ represents the trace of a square matrix, and $\mathsf{Diag}(\cdot)$ represents a diagonal matrix of a given vector. Denote $\mathbb{I}_N$ to be an identity matrix of dimensions $N \times N$. For any N-dimensional vector z, its ℓ_2-norm is given by $\|z\|_2 \coloneqq \sqrt{z_1^2 + \ldots z_N^2}$, and for any real matrix Z, its Frobenius norm is $\|Z\|_{\mathrm{F}} \coloneqq \sqrt{\mathsf{Tr}(Z'Z)}$. Here, the symbol $(')$ denotes the transpose operator of a vector or matrix, the symbol $\coloneqq$ is a definition sign, and the symbol $=$ is the equal sign.

Let $r_{t,i}$ be the return for a risky asset i sampled at trading date t, stacked into a N-dimensional vector $r_t \coloneqq (r_{t,1}, \ldots, r_{t,N})'$. Also, let $x_{t,i}$ be an m-dimensional vector return-predictive signal at date t for asset i, stacked into a M-dimensional vector $x_t \coloneqq (x_{t,1}, \ldots, x_{t,N})'$,

where $M \coloneqq m \times N$ is the total number of signals. This means every asset is accompanied by at least one signal and so m is necessarily a positive integer. For example, with two assets and two signals per asset, we have $m = 2$ and $M = 4$. The signals are understood to proxy the future expected returns conditional on the investor's information set at time t.

We impose the following assumptions on both the returns and signals throughout our analysis.

Assumption 1 (Stationarity). The return sequence $\{r_{t+1}\}$ and the signal sequence $\{x_t\}$ exhibit stationarity and ergodicity, whose moments can be calculated by taking time-series averages.

Assumption 2 (Gaussianity). The returns and signals have multivariate distributions that are assumed to be Gaussian with zero expectations and covariances $\mathsf{Var}(r_{t+1}) \coloneqq \mathsf{Cov}(r_{t+1}, r_{t+1}) = \Sigma_r$ and $\mathsf{Var}(x_t) \coloneqq \mathsf{Cov}(x_t, x_t) = \Sigma_x$, and a cross-covariance $\mathsf{Cov}(r_{t+1}, x_t) = \Sigma_{rx} = \Sigma_{xr}'$. The column vector of stacked subsequent returns and signals $(r_{t+1}, x_t)'$ has a joint covariance matrix expressed in the following partitioned form:

$$\Sigma = \begin{pmatrix} \Sigma_r & \Sigma_{rx} \\ \Sigma_{rx}' & \Sigma_x \end{pmatrix}. \tag{4.1}$$

In our notation, we denote estimated quantities of their population analogs with a hat accent ($\hat{}$) on them. We use S_r, S_x, and S_{rx} to represent the sample covariances between the asset returns and signals, the sample covariances of the signals themselves, and the sample cross-covariance between the returns and signals, respectively. Specifically, S_r is computed as $T^{-1}\sum_{t=0}^{T-1}(r_{t+1} - \bar{r})(r_{t+1} - \bar{r})'$, where $\bar{r}$ is the sample mean of the returns. Similarly, S_x is computed as $T^{-1}\sum_{t=0}^{T-1}(x_t - \bar{x})(x_t - \bar{x})'$, where $\bar{x}$ is the sample mean of the signals. Finally, S_{rx} is computed as $T^{-1}\sum_{t=0}^{T-1}(r_{t+1} - \bar{r})(x_t - \bar{x})'$.

Due to the symmetric positive semidefinite property of the covariance matrices, they admit a spectral decomposition. For example, $\Sigma_r = P_r \mathsf{Diag}(d_{r,1}, \ldots, d_{r,N}) P_r'$, where $P_r \coloneqq [p_{r,1}, \ldots, p_{r,N}]$ is an orthogonal matrix ($P_r' P_r = P_r P_r' = \mathbb{I}_N$) whose columns contain the eigenvectors and $\mathsf{Diag}(d_{r,1}, \ldots, d_{r,N})$ is a diagonal matrix containing the eigenvalues. The eigenvalues are assumed to be sorted in ascending order. The matrix factor $\Sigma_r^{1/2}$ can be obtained by retaining the eigenvectors of Σ_r, but reassembling them with the square roots of the eigenvalues given by $(\sqrt{d_{r,1}}, \ldots, \sqrt{d_{r,N}})'$. An analogous notation applies for the decomposition of the covariance of signals Σ_x and its matrix of factors $\Sigma_x^{1/2}$.

4.2.2 Dynamic Strategies

We begin our analysis by considering the portfolio policies to be linear in the vector of signals at time t based on the following specification.

Assumption 3 (Linear portfolio policies). Let $w_t \coloneqq A' x_t$, where A is a $M \times N$ constant matrix of coefficients.[3]

The portfolio weights are conditional on the strength of the signals but the matrix A is a static object where column i maps the signal vector to the corresponding portfolio weight

[3]Note that the assumption that the optimal portfolio weights are linear functions of the signals is not entirely restrictive. Indeed, each of the signals in x_t can be composed of nonlinear functions of a more primitive set of underlying signals.

invested in basis asset i. From this parameterization, the dynamic portfolio selection problem that an investor faces at time t can be formulated in terms of a conditional mean-variance objective as follows

$$\max_{A} \mathbb{E}_t[x_t' A r_{t+1}] - \frac{\gamma}{2}\mathsf{Var}_t[x_t' A r_{t+1}], \tag{4.2}$$

where γ is an investor's risk aversion parameter.[4] This objective function underscores that the investor chooses to simultaneously allocate between the assets and signals in order to yield a trading strategy that optimizes the returns of the portfolio at time $t+1$, which we denote as r^w_{t+1}.

Brandt and Santa-Clara (2006) showed that this problem formulation is equivalent to an unconditional portfolio selection problem since the matrix of coefficients is constant through time and Assumption 1 holds. Thus, the matrix A that optimizes the investor's conditional problem at any given date is the same for all dates and hence, it also optimizes the investor's unconditional problem. Hence, we can drop the subscript t in (4.2) and consider the following revised problem instead[5]

$$\max_{A} \mathbb{E}[x_t' A r_{t+1}] - \frac{\gamma}{2}\mathsf{Var}[x_t' A r_{t+1}]. \tag{4.3}$$

If we further invoke Assumption 2, we can express this objective in terms of the unconditional second-moments with the following proposition.

Proposition 1. The objective function to the investor's problem (4.3) is given by

$$\max_{A} \mathsf{Tr}(A\Sigma_{rx}) - \frac{\gamma}{2}\left(\mathsf{Tr}(\Sigma_x A \Sigma_r A') + \mathsf{Tr}(\Sigma_{rx} A \Sigma_{rx} A)\right). \tag{4.4}$$

A closed-form expression for matrix A is provided in Theorem 2 from Appendix A.1.

It is analytically convenient to assume that $\mathsf{Tr}(\Sigma_{rx} A \Sigma_{rx} A) \approx 0$ in Proposition 1. Thus, we can consider the following objective function

$$\max_{A} \mathbb{E}[x_t' A r_{t+1}] - \frac{\gamma}{2}\mathbb{E}[x_t' A \Sigma_r A' x_t]. \tag{4.5}$$

This approximation offers the advantage of formulating the optimal portfolio policy solely based on unconditional moments, without any recourse to canonical correlation analysis, which is necessary to solve the investment problem (4.4). It is a reasonable approximation as long as the weighted sum contributions of squared expected returns from the dollar-neutral managed portfolios, represented as x_t multiplied by the return of each asset, are small and can be approximated as zero.[6] Notwithstanding this approximation, we have further validated both the exact and approximate solutions empirically in Section 4.4.

Proposition 2. The objective function to the investor's problem (4.5) is given by

$$\max_{A} \mathsf{Tr}(A\Sigma_{rx}) - \frac{\gamma}{2}\mathsf{Tr}(\Sigma_x A \Sigma_r A'). \tag{4.6}$$

[4] Other reformulations of this problem include the maximization of mean return or minimization of risk. Regardless of the choice, all three problems result in the same mean-variance trade-off.

[5] Note that Equations (4.2) and (4.3) are not generally the same when time-varying covariances are present; see for example, Ang and Bekaert (2002). However, we will defer the investigation of this setting to future research endeavors.

[6] For instance, if a portfolio managed on the basis of momentum signals has an average return of 5%, then the square of its return is 0.25%.

The solution to this problem is given by

$$A = \frac{1}{\gamma}\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}. \tag{4.7}$$

The weight w_t allocated to each asset conditional on the signal at time t is then

$$w_t = \frac{1}{\gamma}\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}x_t. \tag{4.8}$$

In the absence of any further constraints on w_t, this is the frictionless portfolio policy that invests directly in each asset, and the size of the portfolio is determined by the level of the investor's risk aversion. The dynamics of the portfolio is guided by the trajectory of the signal x_t at each period. We see that the portfolio takes into account the cross-sectional information from the asset returns, the signals as well as their cross-relationships. The cross-covariance matrix Σ_{rx} can be interpreted as a matrix consisting of managed portfolios that encapsulate the potential earnings opportunities that are available to an investor.[7] For this reason itself, we term this matrix a "conditional portfolios" matrix, which consists of returns scaled by their predictive signals.

We conclude this section by presenting two examples along with some remarks.

Example 1 (Univariate portfolio). Let us consider one return-accompanying signal and postulate that all covariances in Equation (4.8) are the identity matrix. In this case, the weight is proportional to the conditioning variable, that is, $w_t = \gamma^{-1}x_t$. This portfolio policy is interesting in its own right because it ignores the impact that the returns and signals can have on the covariances. Hence, it is essentially a univariate system in which the signals only forecast their associated asset. We shall henceforth refer to this conditional portfolio as the univariate factor (UNI).

Example 2 (Two-assets portfolio). Suppose that $N = 2$, $m = 1$, and $\gamma = 1$. If the asset returns and signals have unit variances, then the covariances are

$$\Sigma_r = \begin{pmatrix} 1 & \rho_r \\ \rho_r & 1 \end{pmatrix}, \quad \Sigma_x = \begin{pmatrix} 1 & \rho_x \\ \rho_x & 1 \end{pmatrix}, \quad \text{and} \quad \Sigma_{rx} = \begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{pmatrix}. \tag{4.9}$$

For simplicity, we also suppose that the signals are uncorrelated, that is, $\rho_x = 0$. Inserting these objects into the optimal matrix A, we can write the weights invested in the two basis assets as

$$w_t = \left(\frac{\xi_{11} - \rho_r\xi_{21}}{1-\rho_r^2}x_1 + \frac{\xi_{12} - \rho_r\xi_{22}}{1-\rho_r^2}x_2, \frac{\xi_{21} - \rho_r\xi_{11}}{1-\rho_r^2}x_1 + \frac{\xi_{22} - \rho_r\xi_{12}}{1-\rho_r^2}x_2 \right). \tag{4.10}$$

Moreover, the expected portfolio return is

$$\mathsf{Tr}(A\Sigma_{rx}) = \mathsf{Tr}(\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}\Sigma_{rx}') = \frac{\xi_{11}^2 + \xi_{12}^2 + \xi_{21}^2 + \xi_{22}^2 - 2\rho_r(\xi_{11}\xi_{21} + \xi_{12}\xi_{22})}{1-\rho_r^2}. \tag{4.11}$$

This expression is non-negative given $\rho_r \in [-1, 1]$ since $\xi_{11}^2 + \xi_{21}^2 \geq 2\rho_r\xi_{11}\xi_{21}$ and $\xi_{12}^2 + \xi_{22}^2 \geq$

[7] In Kelly et al. (2023), the authors term the cross-covariance matrix as the "prediction matrix," which is also another valid interpretation since the correlation between an asset return and some signal is one measure of signal-return predictability. Given that a cross-covariance matrix is not symmetric in general, the predictive strength of a signal i on asset j may be different from that of signal j on asset i.

$2\rho_r\xi_{12}\xi_{22}$ by the inequality of arithmetic and geometric means. Let us consider three cases. First, if $\rho_r = 0$, then the expected return is simply the squared expected returns of all conditional portfolios that exploit own- and cross-predictabilities. Second, if $\rho_r \in (-1, 0)$, then there is an additional positive contribution to the conditional portfolios that arises from diversifying into imperfectly negatively correlated assets. The position is leveraged according to a factor of $1/(1-\rho_r^2)$, which increases as $|\rho_r|$ approaches one from below. Third, if $\rho_r \in (0, 1)$, then the expected portfolio return is now reduced since the gains to diversification are limited by the extent to which the assets are positively correlated. Notwithstanding the diminished returns, the optimal strategy is also leveraged with a factor $1/(1-\rho_r^2)$ in order to increase the portfolio returns to meet its profit objective. A similar conclusion holds for the case $\rho_r = 0$ and $\rho_x \neq 0$.

Remark 6 (Conditional portfolio policies). We can gain further intuition into the role of conditioning that our problem posits. Consider the following conditional portfolio selection problem:

$$\max_{w_t} \mathbb{E}_t[w_t' r_{t+1}] - \frac{\gamma}{2}\mathsf{Var}_t[w_t' r_{t+1}]. \tag{4.12}$$

Since the weights are known at time t, the problem can be solved to give

$$w_t = \frac{1}{\gamma}\mathsf{Var}_t[r_{t+1}]^{-1}\mathbb{E}_t[r_{t+1}]. \tag{4.13}$$

Observe that this optimal portfolio is determined by the mean and variance of returns that incorporate all available information up to time t. Hence, restricting the weights to be linear functions of a set of signals also restricts the solution set, which may lead to a suboptimal solution. However, specifying the conditional means and variances of the returns is a notoriously difficult task. Our weights restriction simplifies these dual tasks by converting a dynamic problem into an equivalent static problem for which a conditional portfolio policy is available solely in terms of the unconditional moments of the assets and signals.

Remark 7 (Regression-based policies). Notice that the portfolio weights (4.8) can be seen as one that had been obtained from a two-stage process of forecasting and asset allocation. To see this, let us consider the following return-generating process:

$$r_{t+1} = B' x_t + \varepsilon_{t+1}, \tag{4.14}$$

where B is a matrix of coefficients and ε_t is an idiosyncratic error that is independent and identically distributed (i.i.d.). A simple forecasting approach that captures this relationship between both variables would be to perform an ordinary least squares (OLS) regression. In particular, one would first perform N independent univariate OLS regressions of the returns to assets on the signals in order to obtain a vector of cross-sectional predictive signals, and then insert them into (4.13) to arrive at the vector of allocations:

$$w_t = \frac{1}{\gamma}\Sigma_\varepsilon^{-1}\Sigma_{rx}\Sigma_x^{-1}x_t, \tag{4.15}$$

where $\Sigma_\varepsilon \coloneqq \mathsf{Cov}(\varepsilon_t) = \Sigma_r - \Sigma_{rx}\Sigma_x^{-1}\Sigma_{rx}'$ is the covariance matrix of the residuals. In fact, this portfolio is also optimal under our assumption of joint Gaussianity between the asset returns and signals. By invoking the standard properties of the multivariate Gaussian distribution, the conditional distribution of r_{t+1} given x_t is also multivariate Gaussian distributed with means

and variances given by

$$\mathbb{E}_t[r_{t+1}] = \Sigma_{rx}\Sigma_x^{-1}x_t, \tag{4.16}$$
$$\mathsf{Var}_t[r_{t+1}] = \Sigma_r - \Sigma_{rx}\Sigma_x^{-1}\Sigma_{rx}. \tag{4.17}$$

Inserting these expressions into (4.13) gives portfolio (4.15). In contrast to our portfolio (4.8), this regression-based portfolio depends on the covariance structure of the unexplained parts of the returns. Another key difference here is that our portfolio policy respects symmetry; exchanging the roles of the assets and signals yields identical returns to the portfolio. This symmetry property may be warranted in situations where the signals have an influence on the asset returns but at the same time, the reverse association is also economically plausible. It also implies that our approach is more closely related to total least squares (TLS) than it is to OLS since TLS is an errors-in-variables regression model that optimizes for the correlation of both variables, which is a symmetric metric.

4.2.3 Fully Invested Constraint

In quantitative equity investing, it is commonplace to impose economic constraints on the portfolio weights. A prominent example would be the fully-invested portfolio, which enforces a budget constraint so that the weights add to one. This equality constraint can be included in our optimization problem and given that it is linear in the matrix of coefficients A, the portfolio weights can be fortunately solved in closed form.

Proposition 3. We can formulate the optimal asset and signal combination problem with a fully-invested constraint as follows

$$\begin{aligned} &\max_A \mathsf{Tr}(A\Sigma_{rx}) - \frac{\gamma}{2}\mathsf{Tr}(\Sigma_x A \Sigma_r A') \\ &\text{subject to } \mathbf{1}'A'x_t = 1, \end{aligned} \tag{4.18}$$

where $\mathbf{1}$ denotes the conformable vector of ones of dimension N. The problem has the following analytical solution

$$w_t^{\text{FI}} = (1-\kappa)\frac{\Sigma_r^{-1}\mathbf{1}}{\mathbf{1}'\Sigma_r^{-1}\mathbf{1}} + \kappa\frac{\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}x_t}{\mathbf{1}'\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}x_t}, \tag{4.19}$$

where $\kappa := \gamma^{-1}\mathbf{1}'\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}x_t$ is a scalar value.

Proposition 3 shows that the portfolio weight is expressed as a convex linear combination of the global minimum variance portfolio and the optimal unconstrained portfolio policy that is renormalized to one. The latter portfolio resembles the so-called "tangency portfolio," which is the highest Sharpe ratio of a portfolio of risky assets from the standard mean-variance framework. The fully-invested portfolio w_t^{FI} has a similar form to the one obtained from the standard Markowitz paradigm found in Merton (1972).

4.2.4 Relation to Existing Literature

At this juncture, it is important to highlight the differences between our work from those of the existing literature. We first draw our attention to the fact that the objective function (4.5)

under our consideration is similar to that of Brandt and Santa-Clara (2006).[8] Suppose that the joint distribution of the returns and signals is Gaussian and the squared returns of conditional managed portfolios are relatively small and can be approximately zero.[9] Then we can express the optimal matrix of coefficients A from Equation (4.7) in vectorized form as

$$\begin{aligned}
\mathsf{vec}(A) &= \frac{1}{\gamma}\mathsf{vec}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}) && (4.20)\\
&= \frac{1}{\gamma}(\Sigma_x^{-1}\otimes\Sigma_r^{-1})\mathsf{vec}(\Sigma_{rx}') && (4.21)\\
&= \frac{1}{\gamma}(\Sigma_x\otimes\Sigma_r)^{-1}\mathsf{vec}(\Sigma_{rx}') && (4.22)\\
&\approx \frac{1}{\gamma}\mathbb{E}[(x_t x_t')\otimes(r_{t+1}r_{t+1}')]^{-1}\mathbb{E}[x_t\otimes r_{t+1}], && (4.23)
\end{aligned}$$

where $\mathsf{vec}(\cdot)$ is an operator that stacks the columns of matrix A into a vector, and $\otimes$ is the Kronecker product of two matrices. The fourth (approximate) equality invokes our sparsity assumptions so that the second-moment matrix of the conditional product returns $x_t\otimes r_{t+1}$ is a Kronecker product of two covariances. This corresponds to the solution that one obtains by rewriting the portfolio returns as $x_t'Ar_{t+1}=\mathsf{vec}(A)(x_t\otimes r_{t+1})$, and optimizing over $\mathsf{vec}(A)$ with respect to $x_t\otimes r_{t+1}$.

The authors considered this new optimization scheme to be an augmented asset space form of the mean-variance optimization. They had removed one step in the process of forecasting and allocation and combined them into one step, thus reducing noise in the allocation process. However, there are two issues with this approach. First, augmenting the asset space results in a severe expansion of the problem dimension. The unrestricted covariance matrix in (4.23) has $NM(NM+1)/2$ parameters, while our Kronecker product factorization in (4.22) has only $N(N+1)/2+M(M+1)/2$ parameters. Clearly, the former becomes less computationally tractable when the dimensions of both the asset and signal space are large. Second, there is less clarity on the structure of their optimal solution due to the complicated interactions in the covariance between the subsequent returns and signals. Notwithstanding these shortcomings, our efforts in recasting the optimal strategy (4.23) into (4.7) (or (A.5) in Theorem 2) help to amend their approach towards large-scale financial applications. Although this requires us to model the joint distribution of returns and signals, our solution lends itself nicely to the study of the joint correlation structure of the optimal matrix of coefficients A, as we will demonstrate in Section 4.3.3.

Second, our work is related to the classic portfolio selection problem of Markowitz (1952) in its static form. If we assume independent and identically distributed (i.i.d.) returns with constant moments and time-invariant weights $w_t\equiv w$, the conditional portfolio problem (4.12) simplifies to an unconditional one, which has the following well-known solution

$$w=\frac{1}{\gamma}\mathbb{E}[r_{t+1}r_{t+1}']^{-1}\mathbb{E}[r_{t+1}]. \tag{4.24}$$

[8] A subtle difference worth pointing out is that the authors model the portfolio policy to be *affine* in the state variables, and hence the allocations are also driven by the assets themselves. Here, we work in a slightly restricted setting where the portfolio policy is only driven by the conditional managed portfolios themselves.

[9] For example, if a portfolio managed on the basis of a momentum signal has an average return of 5%, then the square of its return is 0.25%.

This portfolio policy is an unconditional strategy because it does not use any information today and is typically implemented through plug-in estimation. In this paper, we focus on the more realistic setting of non-i.i.d. returns, which motivates our model parameterization and use of conditioning variables that affect the joint distribution of returns.

Finally, our work is also related to Kelly et al. (2023) on cross-predictability. In their work, the authors pursued a similar parameterization for the portfolio weights but optimize an objective function that is subject to a robust risk constraint that controls for the leverage of A. As a result, their optimal portfolio policy depends only on the cross-covariance matrix. When viewed in the context of our framework, the analysis is simplified and avoids the challenges of inverting two additional large-dimensional matrices. However, this departs from our paper on two fronts. First, we assume that the investors have mean-variance preferences in a similar spirit to the standard mean-variance problem.[10] Second, our application of canonical correlation analysis formally suggests that the cross-covariance between *decorrelated* asset returns and signals is the more appropriate object for investment analysis as opposed to the cross-covariance of the original asset returns and signals. This is because the latter contains non-trivial variations that are embedded in both financial variables that may obscure the inference.

4.3 Strategy Diversification

4.3.1 Canonical Correlation Analysis

The goal of canonical correlation analysis (CCA) is to perform dimension reduction on two different data sets that comprise a large number of interrelated variables, while simultaneously retaining as much of the correlation present in the two data sets. This can be achieved through a transformation of the original set of variables into a new set of mutually orthogonal paired variables, which are ranked so that the largest few retain most of the correlation present in all of the original variables.

There are several standard approaches to solving the CCA problem as outlined in Uurtio et al. (2017). Indeed, this may involve solving a standard eigenvalue problem (Hotelling, 1936), or a singular value decomposition (SVD) (Healy, 1957), or through a generalized eigenvalue problem (Bach and Jordan, 2002). We adopt the SVD approach in this brief exposition of the classical subject by emphasizing a sequence of two change of basis operations to arrive at a reduced problem whose financial interpretation we provide in Section 4.3.3.

Given that scale invariance is a property of correlation, the first step is to orthogonalize the random vectors r and x such that their covariances are the identity matrices. This can be achieved through the following linear transformations $\tilde{r} \coloneqq \Sigma_r^{-1/2} r$ and $\tilde{x} \coloneqq \Sigma_x^{-1/2} x$. The transformed objects are now uncorrelated and have a unit variance and their joint covariance matrix is given by

$$\begin{pmatrix} \mathbb{I}_N & \Sigma_r^{-1/2} \Sigma_{rx} \Sigma_x^{-1/2} \\ \Sigma_x^{-1/2} \Sigma_{rx}' \Sigma_r^{-1/2} & \mathbb{I}_M \end{pmatrix}. \tag{4.25}$$

[10] Levy and Markowitz (1979) and Markowitz (1991) argue that the mean-variance preferences serve as a reasonable approximation to other utility preferences in portfolio selection problems.

It is helpful to introduce the object

$$\Sigma_{\tilde{r}\tilde{x}} \coloneqq \mathbb{E}[\tilde{r}\tilde{x}'] = \Sigma_r^{-1/2}\Sigma_{rx}\Sigma_x^{-1/2}, \tag{4.26}$$

as the adjusted cross-correlation matrix between the transformed variables $\tilde{r}$ and $\tilde{x}$.[11]

The second step is to perform an SVD operation on the adjusted cross-covariance matrix $\Sigma_{\tilde{r}\tilde{x}}$. Let $((s_1,\ldots,s_N);(u_1,\ldots,u_N);(v_1,\ldots,v_N))$ denote a system of singular values and singular vectors of the adjusted cross-covariance matrix $\Sigma_{\tilde{r}\tilde{x}}$. We assume that the singular values s_i are sorted in increasing order. The canonical correlations correspond to the singular values. The canonical variates of r and x are defined as $u_1'\tilde{r},\ldots,u_N'\tilde{r}$ and $v_1'\tilde{x},\ldots,v_N'\tilde{x}$, respectively. As the singular values, s_i are sorted in increasing order, the canonical variate with the largest correlation is given by the pair $(u_N'\tilde{r}, v_N'\tilde{x})$ while the canonical variate with the smallest correlation is $(u_1'\tilde{r}, v_1'\tilde{x})$.

Finally, since the solutions are expressed on a different basis than our original problem, we have to translate them back to our original basis. By inverting the change of variables we made, the ith canonical variates can be written as $(u_i'\Sigma_r^{-1/2}r, v_i'\Sigma_x^{-1/2}x)$. This pair can be seen as a linear combination of the original assets and signals with coefficients given by $q_{r,i} \coloneqq \Sigma_r^{-1/2}u_i$ and $q_{x,i} \coloneqq \Sigma_x^{-1/2}v_i$, which are the so-called canonical directions. The application of change of basis operations simplifies the covariance structure considerably.

In order to implement the CCA in practice, a common approach is to replace the population second-moments Σ_r, Σ_x, and Σ_{rx}, with their sample counterparts S_r, S_x, and S_{rx} computed from random samples $r_1,\ldots,r_{T+1}$ and $x_1,\ldots,x_T$. Let $((\hat{s}_1,\ldots,\hat{s}_N);(\hat{u}_1,\ldots,\hat{u}_N);(\hat{v}_1,\ldots,\hat{v}_N))$ be a system of singular values (sorted in increasing order) and singular vectors of the sample analog of (4.26) denoted by $S_{\tilde{r}\tilde{x}}$. Then the sample canonical variates of r and x are given by $\hat{u}_1'\tilde{r},\ldots,\hat{u}_N'\tilde{r}$ and $\hat{v}_1'\tilde{x},\ldots,\hat{v}_N'\tilde{x}$, respectively. However, one should be skeptical of sample canonical correlations as they are not reflective of the population canonical correlations. Indeed, the true canonical correlation of the ith canonical variate pair $(\hat{u}_i'\tilde{r}, \hat{v}_i'\tilde{x})$ is $\hat{v}_i'\Sigma_{\tilde{r}\tilde{x}}'\hat{u}_i$, as opposed to $\hat{s}_i = \hat{v}_i'S_{\tilde{r}\tilde{x}}'\hat{u}_i$. This is because the sample canonical correlations are known to be inconsistent estimates of their population counterparts when the dimensionality of the problem for both variables is large relative to the number of observations.

To bring this point home, let us consider a specific case when the two variable sets r and x are independent and each set consists of i.i.d. Gaussian random variables. Figure C.4.1 shows the distribution of sample canonical correlations for various values of the limiting ratios $N/T \in \{0.003, 0.03, 0.3\}$ and $M/T \in \{0.006, 0.06, 0.6\}$. We obtained this figure by invoking the asymptotic formula derived by Wachter (1980, Theorem 3.1).[12] We observe that the sample canonical correlations are biased upwards away from zero and the severity of the bias worsens as N/T or M/T or both increases. This distributional shape is expected to hold when N, M and T are approximately large and it only depends on the ratio N/T and M/T; this is true regardless of any particular realization of the adjusted sample cross-covariance matrix. We attempt to

[11] Note that the matrix (4.26) is different from the *unadjusted* cross-covariance matrix Σ_{rx}. The adjusted cross-covariance matrix is the central object in CCA as it internalizes the variabilities from the asset returns and signals that might potentially obscure the relationships between both of the variables. Moreover, since the optimal matrix of coefficients for our problem (4.5) is indeed the adjusted cross-covariance matrix under an appropriate change of basis (see Section 4.3.3), it makes sense for us to focus on this object for our analysis.

[12] Wachter's result also holds under a less restrictive setting of i.i.d. data with finite second-moments; see, for example, Yang and Pan (2012, Theorem 1).

address these problems with regularization in Section 4.3.6.

4.3.2 Relation to PCA, PLS, and RRR

Having laid the groundwork for CCA, it is interesting to explore the relationship between CCA and a variety of linear dimension-reduction techniques that are commonly employed to reveal underlying economic structures. We shall restrict our attention to the principal component analysis (PCA), partial least squares (PLS), and reduced rank regression (RRR) due to their tight connection with one another and extensive applications in the finance and econometrics literature.

PCA, for instance, has been used to extract common factors from a covariance matrix of individual stocks among others by Roll and Ross (1980), Connor and Korajczyk (1986, 1988), and Lehmann and Modest (1988). In the context of describing risk compensation through estimation of the stochastic discount factor, Kozak et al. (2018, 2020) invokes no near-arbitrage arguments for the use of PCA to identify factor risk exposures that price the cross-section of expected returns of characteristic-managed portfolios. For PLS, one of the earliest applications to finance can be ascribed to Kelly and Pruitt (2013), where the authors used a three-pass regression filter (which is a special case of PLS) to forecast market return and cash flow growth. The RRR, alongside CCA, features prominently in the estimation of the celebrated vector-error-correction model (VECM) of Engle and Granger (1987) as a means to detect cointegration, which are economic relationships that cannot be obtained through standard regression approaches; see Johansen (1988, 1991, 1995).

Starting with PCA developed by Hotelling (1933), the method seeks to break down the variation of the data into mutually orthogonal components whose variances can be ranked from the smallest to the largest. In the context of a paired dataset consisting of asset returns with their associated signals, PCA focuses on searching for directional vectors that have the maximum variation in each variable separately. In contrast to CCA, however, it does not take into account any relationship between the two financial variables. In this paper, we would like to focus on how financial variables are related, not how much they vary. We summarize the key properties distinguishing PCA and CCA in Table B.3.1.

On the other hand, PLS was introduced by Wold (1975) as an econometric technique that aims to explain the relationship of a paired dataset by studying the (cross-)covariance. This can be understood by relaxing the variance constraints in the optimization problem of CCA, which results in a maximum (cross-)covariance problem. The solution(s) can be obtained via singular value decomposition of the cross-covariance matrix, Σ_{rx}, and the corresponding managed portfolios closely mimic that of Kelly et al. (2023). This differs from CCA where the cross-covariance matrix is normalized with respect to the covariances from both the r_{t+1} and the x_t variables.

Finally, the RRR proposed by Anderson and Rubin (1949, 1950) and Anderson (1951) can be understood through the regression framework (4.14) with the following error minimization problem:

$$\min_{B} \mathbb{E}[\|\Gamma^{1/2}(r_{t+1} - B'x_t)\|_2^2], \tag{4.27}$$

where Γ is a $N \times N$ weighting matrix, and the matrix B is subject to a rank constraint; see

also, Velu and Reinsel (2013). Expressing the objective function in terms of the unconditional moments, we have

$$\mathbb{E}\big[\|\Gamma^{1/2}(r_{t+1} - B'x_t)\|_2^2\big] = \mathbb{E}\big[\|\Gamma^{1/2}r_{t+1}\|_2^2\big] - 2\mathbb{E}\big[x_t'B\Gamma r_{t+1}\big] + \mathbb{E}\big[x_t'B\Gamma B'x_t\big] \tag{4.28}$$

$$= N - 2\mathsf{Tr}(B\Gamma\Sigma_{rx}) + \mathsf{Tr}(B\Gamma B'\Sigma_x) \tag{4.29}$$

$$= \|\Gamma^{1/2}B\Sigma_x^{1/2} - \Gamma^{1/2}\Sigma_{rx}\Sigma_x^{-1/2}\|_{\mathrm{F}}^2 + \text{constant}. \tag{4.30}$$

The solution to this rank-restricted minimization problem follows from the theorem of Eckart and Young (1936) and it relies on the singular value decomposition of $\Gamma^{1/2}\Sigma_{rx}\Sigma_x^{-1/2}$. The typical objective function of RRR occurs when $\Gamma = \mathbb{I}_N$. If additionally, $\Sigma_x = \mathbb{I}_M$, then RRR coincides with PLS. Finally, observe that if $\Gamma = \mathbb{I}_N$ and $r_{t+1} = x_t$, then the objective function coincides with that of PCA.

Interestingly, if $\Gamma = \Sigma_r^{-1}$, then the solution of RRR generalizes to CCA. This speaks to the difference between optimizing a regression-based objective and a correlation-based objective. Indeed, regression attempts to explain as much of the variation in the returns r_{t+1} as possible. However, if r_{t+1} is decorrelated and isotropic along all variance directions, then minimizing a regression loss function becomes equivalent to maximizing a correlation objective. Unlike conventional regression-based approaches, there is no notion of independent and dependent variables in CCA because the correlation metric is symmetric; that is, if we exchange the roles of asset returns and signals, CCA yields identical outcomes for the canonical directions and canonical correlations.

In summary, RRR, PLS, and CCA can all be viewed under the same framework, which is solving a cross-covariance maximization problem. Each of these linear dimensional reduction techniques is subject to different normalization schemes with CCA serving as the most general one that nests these methods.

4.3.3 Reformulation as Canonical Portfolios

The portfolio returns obtained with (4.7) is not particularly intuitive as the expression involves large-dimensional matrix inversion and multiplication operations. However, we can make progress by performing a CCA in order to decompose the portfolio selection problem into one that we can provide financial interpretation.

We start by expressing the portfolio returns in terms of their transformed objects as

$$x_t'Ar_{t+1} = \frac{1}{\gamma}x_t'\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}r_{t+1} = \frac{1}{\gamma}\tilde{x}_t'\Sigma_{\tilde{r}\tilde{x}}\tilde{r}_{t+1}, \tag{4.31}$$

Since the decorrelated objects $\tilde{r}_{t+1}$ and $\tilde{x}_t$ span the same universe as the original assets and signals, we shall term them "synthetic assets" and "synthetic signals." These newly defined objects have identity covariances, and so this is the cross-sectional version of risk parity.

From expression (4.31), we can see that the portfolio returns depend intricately on the synthetic asset returns and synthetic signals, which are coupled through their adjusted cross-covariances. The portfolio returns do not depend on whether we exchange the role of the asset returns of signals; what matters is how they are related to each other.

Our next step is to choose basis vectors such that $\Sigma_{\tilde{r}\tilde{x}}$ is diagonal. In particular, we perform an SVD operation such that all cross-relationships are eliminated in the new basis in order to

arrive at the following simplified expression for the portfolio returns:

$$x_t' A r_{t+1} = \frac{1}{\gamma} \sum_{i=1}^{N} s_i (v_i' \tilde{x}_t)(u_i' \tilde{r}_{t+1}). \tag{4.32}$$

Thus, we see that a generic strategy return can be viewed either as a combination of the original assets and signals that have been optimally blended with the matrix of coefficients A, or as a combination of N uncorrelated managed long-short portfolios weighted by their canonical correlations. The N long-short managed portfolios span the same space of investment opportunities as the original assets and signals.

Economically, we interpret the change of basis as an operation that reorganizes the set of N synthetic assets and M synthetic signals into a set of N uncorrelated managed portfolios. Each of these managed portfolios is expressed as a certain weighted combination of the assets and signals themselves. These particular combinations are determined by the singular vectors or canonical directions; in either case, we shall refer to them as *canonical portfolios* henceforth to reinforce the idea that these vectors are essentially managed portfolios that are loading onto their respective asset and signal variables. The higher the ith canonical correlation, the higher the return of the ith canonical portfolio. Said differently, the canonical portfolios are ordered in such a way that the highest canonical correlation corresponds to the *most linearly predictable portfolio*, and the second highest canonical correlation, the *second most linearly predictable portfolio*, and so on. This notion of predictability distinguishes canonical portfolios from other concepts of orthogonal managed portfolios such as those constructed by PCA.

Circling back to Equation (4.32), the source of the portfolio returns can be seen to be distributed across the canonical portfolios of $\Sigma_{\tilde{r}\tilde{x}}$. This reinforces the economic notion of diversification: having exposure to different uncorrelated canonical portfolios is akin to putting each of your eggs into different baskets. We see that the capital assigned to each canonical portfolio is proportional to its original returns r_{t+1} and signal x_t, and proportional to the correlation s_i. Put differently, we want to assign capital to managed portfolios that have high correlations but are also orthogonal to each other.

4.3.4 Canonical Portfolio Analysis

With our CCA decomposition, we can also gain some insight into how the canonical portfolios impact the returns of a portfolio. We start by working out the expected returns and variance of these canonical portfolios with the next proposition.

Proposition 4. Let $\pi_{t,i} \coloneqq s_i (v_i' \tilde{x}_t)(u_i' \tilde{r}_{t+1})$ be the return that the ith canonical portfolio generates. The expected value and variance return of each canonical portfolio are

$$\mathbb{E}[\pi_{t,i}] = s_i^2, \quad \text{and} \quad \mathsf{Var}[\pi_{t,i}] = s_i^2(1 + s_i^2). \tag{4.33}$$

Furthermore, the expected value of the optimal portfolio returns and its squared returns are given by

$$\mathbb{E}[x_t' A r_{t+1}] = \frac{1}{\gamma} \sum_{i=1}^{N} s_i^2, \quad \text{and} \quad \mathbb{E}[x_t' A \Sigma_r A' x_t] = \frac{1}{\gamma^2} \sum_{i=1}^{N} s_i^2. \tag{4.34}$$

We see that the expected portfolio returns are positive, and hence, the optimal portfolio is a profit-generating strategy that has economic value. Moreover, the expected portfolio returns can be expressed as the sum of the contributions from the squared canonical correlations; the larger the expected return of the ith canonical portfolio, the larger its contribution towards the overall returns.[13] This suggests that the squared canonical correlations can serve as a natural measure to rank the performance of the canonical portfolios.

Interestingly, the canonical correlations are also intimately connected to the Sharpe ratio of the returns associated with our portfolio policy.

Corollary 1. The Sharpe ratio of the ith canonical portfolio and the optimal portfolio is given by

$$\mathsf{SR}_i := \frac{\mathbb{E}[\pi_{t,i}]}{\sqrt{\mathsf{Var}[\pi_{t,i}]}} = \frac{s_i}{\sqrt{1+s_i^2}}, \quad \text{and} \quad \mathsf{SR} := \frac{\mathbb{E}[x_t' A r_{t+1}]}{\sqrt{\mathbb{E}[x_t' A \Sigma_r A' x_t]}} = \sqrt{\sum_{i=1}^{N} s_i^2}. \tag{4.35}$$

From Corollary 1, we see the squared canonical correlations also influence the portfolio's Sharpe ratio. We end this section with two simple examples to give further intuition into these results.

Example 3 (Equal canonical correlations). If we suppose $s_i = s$ for all i, then Equation (4.35) simplifies to $\mathsf{SR} = s\sqrt{N}$. In this special case, we arrive at the Fundamental Law of Active Management of Grinold (1989), where the so-called "information coefficient" is s and the effective number of statistically independent investments (also referred to as "breadth") is N. We see that the Sharpe ratio improves with the average canonical correlation and the number of canonical portfolios, with the latter exhibiting diminishing gains to returns due to the square-root scaling in its exponent. In the general case where the canonical correlations are not identical across all canonical portfolios, the canonical correlations are thought to be the information coefficient over clusters of signals.[14]

Example 4 (Bias of in-sample returns). Let $\hat{A} := \gamma^{-1} S_x^{-1} S_{rx}' S_r^{-1}$ is the optimal matrix of coefficients that replaces the population moments in A with sample moment-based estimates. For simplicity, suppose that $S_r = \Sigma_r$ and $S_x = \Sigma_x$. Since the Frobenius norm is a convex function, by Jensen's inequality, the expected portfolio return satisfies

$$\mathbb{E}[\mathsf{Tr}(\hat{A} S_{rx})] = \frac{1}{\gamma}\mathbb{E}[\mathrm{tr}\{(S_{\tilde{r}\tilde{x}}' S_{\tilde{r}\tilde{x}}\})] \geq \frac{1}{\gamma}\mathsf{Tr}(\mathbb{E}[S_{\tilde{r}\tilde{x}}]'\mathbb{E}[S_{\tilde{r}\tilde{x}}]) = \frac{1}{\gamma}\mathbb{E}[\mathrm{tr}\{(S_{\tilde{r}\tilde{x}}' \Sigma_{\tilde{r}\tilde{x}}\})], \tag{4.36}$$

where we used the fact that $\mathbb{E}[S_{rx}] = \Sigma_{rx}$ in the last equality due to the unbiasedness of the sample cross-covariance. Hence,

$$\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i^2] \geq \sum_{i=1}^{N} s_i^2 = \sum_{i=1}^{N} \mathbb{E}[\hat{s}_i s_i^\circ], \tag{4.37}$$

[13] The formulas (4.33) and (4.34) have some close similarities to a result in Kelly et al. (2023, Proposition 4) for the expected returns. One notable difference, however, is that our singular values are raised to the power of two. This is implicit in the fact that the expected return of each canonical portfolio is scaled by a leverage factor of s_i. On the other hand, this multiplicative factor is absent in their expression since leverage is explicitly controlled for in their problem formulation.

[14] We thank Attilio Meucci for this insight.

where $s_i^\circ := \hat{q}_{r,i}' \Sigma_{rx} \hat{q}_{x,i}$ is the out-of-sample "canonical correlation" associated with the ith estimated canonical portfolio $\hat{q}_{r,i}$ and $\hat{q}_{x,i}$.[15] Hence, on average, the in-sample returns are always optimistic but the out-of-sample evaluation disappoints. This is because both the in-sample singular values and singular vectors are estimated with a bias. Hence, in order to ensure that the in-sample and out-of-sample returns are more in sync, we have to shrink the in-sample singular values and align singular vectors closer to the truth.

4.3.5 Static and Dynamic Returns Decomposition

In our standing assumptions, we supposed that both the asset returns and signals are centered and have zero means. However, when evaluating the out-of-sample performance of the portfolio, the original non-demeaned returns and signals tend to be used instead. Moreover, the means of both financial variables may be non-negligible and contain useful information for describing the cross-predictability of future returns. In general, we can write the second-moment matrix of conditional portfolios using the following decomposition:

$$\Sigma_{rx} = \mathbb{E}[r_{t+1} x_t'] = \mathbb{E}[r_{t+1}]\mathbb{E}[x_t]' + \mathsf{Cov}(r_{t+1}, x_t). \tag{4.38}$$

The first component is a rank-one outer product matrix of the unconditional means in the asset returns and signals while the second component is the cross-covariance matrix of the demeaned returns and signals. In particular, the first component places emphasis on the cross-sectional differences driven by the unconditional levels since all time-series variabilities have been averaged out, while the second component captures time-series variations in the returns and signals that are expressed as the deviations from their unconditional means. Kelly et al. (2023) describe the first component as static bets and the second component as dynamic bets.

We can derive the expected returns of the optimal portfolio in this setting in the next proposition.

Proposition 5. If the expected value of the returns and signals are $\mathbb{E}[r_{t+1}] = \mu_r$ and $\mathbb{E}[x_t] = \mu_x$, then the expected return of a portfolio utilizing the second-moment matrix (4.38) is given by

$$\mathbb{E}[x_t' A r_{t+1}] = \frac{1}{\gamma}(\mu_r' \Sigma_r^{-1} \mu_r)(\mu_x' \Sigma_x^{-1} \mu_x) + \frac{1}{\gamma} \sum_{i=1}^{N} s_i^2. \tag{4.39}$$

Proposition 5 shows that there is an additional non-negative contribution to the portfolio returns, which is due to the (squared) maximum Sharpe ratio achievable from the assets. This observation aligns nicely with our interpretation that the sum of the squared canonical correlations is related to the squared Sharpe ratio of optimal portfolio returns from Corollary 1. That is, the expected portfolio returns can be attributed to the Sharpe ratios from both static and dynamic bets.

We shall let our portfolio exploit both investment opportunities that arise from the second-moment matrix of managed portfolio returns. This is possible within our framework since the assumption of zero-mean returns and signals strictly applies to the variance of the portfolio returns, and so only the covariances of the returns and signals have to be centered. However, we shall apply CCA to the cross-covariance matrix to fulfill its modeling

[15] In fact, a more precise relationship between the in-sample and out-of-sample portfolio returns have been established by Benaych-Georges et al. (2023, Proposition 2.9).

assumptions since otherwise, it can result in a generic top canonical portfolio that is mainly driven by the unconditional levels.[16]

4.3.6 Estimation

Our optimal portfolio policy requires the knowledge of the population covariances of the asset returns, the covariances of the signals, and cross-covariance between both variables. These objects are generally unknown to us, and so in order to render our framework to practice, we have to estimate them with real data. Given the large-dimensional nature of our problem, it is necessary to regularize the covariances of our portfolio to reduce the estimation errors. Unfortunately, simply maintaining the top canonical portfolio is not sufficient for empirical analysis. Indeed, if either the assets or the signals have rank T (which tends to be the case when $N \gg T$), then the canonical portfolios can take on any arbitrary values. Moreover, since the smallest amount of variability in each dataset gets rescaled to one, CCA can produce spurious outcomes.

The challenge of estimating the covariance matrix of financial covariances is well known amongst practitioners (Jobson and Korkie, 1980). A standard approach is to use the sample covariance matrix. However, when the dimensionality of the problem is large relative to the number of observations, estimation error of the sample covariance matrix can create issues for portfolio optimizers; they tend to place extreme bets on low-risk sample eigenvectors. In fact, this observation led Michaud (1989) to refer to mean-variance optimizers as "error maximization" schemes. There have been several approaches from practice to address this problem using methods from bootstrapping (Michaud and Michaud, 2008) to Bayesian estimators (Black and Litterman, 1992; Lai et al., 2011).

We obtain regularized covariance matrices by applying the linear shrinkage technology from Ledoit and Wolf (2004b). The covariances of returns will be estimated as $\hat{\Sigma}_r = (1-\delta_r)S_r + \delta_r N^{-1}\mathsf{Tr}(S_r)\mathbb{I}_N$, where the shrinkage intensity δ_r is determined based on an asymptotic formula. We also apply the same linear shrinkage technology for the covariances of signals $\hat{\Sigma}_x$ with a shrinkage intensity parameter δ_x. The choice of linear shrinkage of the covariances has a nice interpretation in our context in that varying the shrinkage intensities allows us to interpolate between the maximum covariance (PLS) and maximum correlation (CCA) problems. Shrinking towards maximum covariance helps to break the singularities by considering managed canonical portfolios that have better out-of-sample properties. In the extreme regularization setting with both $\hat{\Sigma}_r$ and $\hat{\Sigma}_x$ being identity matrices, our estimated canonical portfolios mimics the principal portfolios approach of Kelly et al. (2023).

With both of these estimated covariances at hand, we proceed to build a regularized adjusted cross-covariance matrix by pre-conditioning the sample covariances S_{rx} on both sides with the estimated matrix factors $\hat{\Sigma}_r^{1/2}$ and $\hat{\Sigma}_x^{1/2}$. One can then choose to maintain the top few canonical correlations of the regularized adjusted sample cross-covariance matrix defined as

$$\hat{\Sigma}_r^{-1/2} S_{rx} \hat{\Sigma}_x^{-1/2}, \tag{4.40}$$

and set the bottom ones to zero.[17] This thresholding operation has the effect of regularizing

[16] CCA is sometimes applied to non-demeaned variables. In such a case, this is more closely related to using a cosine similarity objective instead of a correlation objective.

[17] The regularized sample adjusted cross-covariance matrix (4.40) has a close similarity to that of Vinod (1976),

the problem as it reduces the effective number of parameters that we have to estimate by maintaining the top few canonical portfolios that are the most predictable.

4.4 Empirical Analysis

4.4.1 Data and Portfolio Construction Rules

For our empirical analysis, we download six datasets from Kenneth French's data library, which are characteristic-sorted long-short stock portfolios.[18] They are daily returns on portfolios of stocks sorted on the basis of size and book-to-market (FF), size and operating profitability (ME/OP), and size and investment (ME/INV), each of which is of universe size 25 and 100. In comparison to individual stocks, each return of a portfolio is an average return of a group of stocks sharing similar characteristics and so they are less subject to large variabilities due to idiosyncratic risks. Hence, they serve as useful test assets that may allow us to more easily harvest the predictability in stock returns.

Although some of these portfolio returns have been available since 1926, we conduct most of our analysis on the period from July 1963 to October 2022, for which most of the returns are available. For simplicity, we suppose that 21 consecutive days constitute one trading "month" and 252 consecutive days as one trading "year." We adopt a sequential updating scheme and rebalance the portfolio every "month" on a rolling walk-forward basis. To obtain a well-defined investment universe for which we can estimate the portfolios, we use the following rule. For each rebalancing date, we select test assets that have a complete return history over the most recent $T = 120$ months as well as a complete return "future" over the subsequent trading month. The backward and forward restrictions ensure that we have data to estimate our models and to evaluate out-of-sample. This provides us with 578 monthly out-of-sample returns, which covers the out-of-sample investment period from 09-09-1974 through 10-21-2022.

4.4.2 Signal Construction

As a demonstration of our method for a dynamic portfolio selection problem between equity-sorted portfolios, we start with a single conditioning variable as our signal. We consider the momentum signal as there is substantial empirical evidence that documents this anomaly in the returns of individual stocks (Jegadeesh, 1990; Jegadeesh and Titman, 1993), industries (Moskowitz and Grinblatt, 1999), and of size and value portfolios (Lewellen et al., 2010). This application is similar to Kelly et al. (2023) for reproducibility.

To construct a momentum signal, we compute for each asset the lagged one-month return defined as the simple average return over the previous 21 trading days. Similar to Asness et al. (2019), Freyberger et al. (2020), Kozak et al. (2020) and Kelly et al. (2023), we rank the momentum signals across the assets from 1 to N, dividing the ranks by the number of assets, and then centering the normalized ranks to map the signals into the range $[-0.5, 0.5]$.

where ridge regression was proposed as a means of regularizing the sample adjusted cross-covariance matrix. We resort to the class of linear shrinkage estimators due to their ability to also reduce the influence of large variance directions.

[18] The description of all portfolio construction can be found on Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. At the time of writing, these datasets were based on the 10-2022 CRSP database.

This provides us with a set of dollar-neutral signals that are insensitive to outliers for which we further divide by the sum of their absolute values. This keeps the gross exposure (that is, the sum of the absolute amount of long and short positions) fixed since otherwise doubling the number of assets at any time t will result in signals that are two times more aggressive even though the investment opportunities remain the same. We assume \$1 of capital is invested to \$1 of long and short positions.

We collect the individual momentum values of the N assets to yield a predictive signal x_t for the subsequent monthly returns. Additionally, we impose a one-day buffer between the constructed signals and the subsequent returns to limit the effects of illiquidity from driving our results and to bring our backtest simulations closer to being tradeable practice.[19]

4.4.3 Candidate Portfolios

Given the time series panels of asset returns and signals, we consider the following portfolios in our study:

- **CP2**: Our proposed portfolio contruction methodology based on Equation (4.8).
- **MVO**: The mean-variance optimization portfolio of Markowitz (1952).
- **PP2**: The Principal Portfolios of Kelly et al. (2023).
- **UNI**: The univariate factor where the weights are the signals.

The competing portfolios have been chosen because they can be subsumed in our proposed method and hence, serve as natural benchmarks for us to determine where the contribution to any improved performance comes from. The suffix number attached to the portfolio labels CP and PP indicates the number of managed portfolios that we retain; for example, CP2 means that we choose to keep the leading two most predictable managed portfolios.[20]

The covariance of returns in CP2 and MVO is estimated with the linear shrinkage of Ledoit and Wolf (2004b). Additionally, we also apply linear shrinkage to the covariance of signals for CP2 but choose a high shrinkage intensity with value $\delta_x = 0.9$ without relying on the asymptotic formula from Ledoit and Wolf (2004b) for this purpose since it was developed for financial returns that are assumed to be independent and identically distributed.

Finally, we renormalize the estimated portfolios so that the sum of the absolute value of their weights equals one. This allows all the portfolios to be comparable in scale. It also implies that the gross exposure for all portfolios is one dollar by construction, that is, we apply one dollar of capital for one dollar of long and short positions. This is sensible for long-short equity hedge fund managers who face institutional constraints such as limits on gross exposure by their

[19]The challenge working with the equity portfolios from Kenneth French's website is that they contain small illiquid stocks. Moreover, given that we are using daily returns, which are close-to-close returns, asynchronous trading at the end of the day may arise; see, for example, Lo and MacKinlay (1990). Consequently, this may induce some lead-lag relationships among the stocks, which can lead to autocorrelation in the portfolio returns and spurious correlation estimates due to the Epps (1979) effect. The latter problem can be particularly acute given our use of multivariate techniques in the estimation of portfolios. Therefore, we employ a one-day buffer along with monthly returns to mitigate the effects that asynchronous or infrequent trading can have on the portfolio returns.

[20]We cross-sectionally demean the returns in the construction of the cross-covariance matrix for PP as suggested in Kelly et al. (2023) to focus on the cross-sectional differences. However, we choose to ignore otherwise for the other portfolios.

prime brokerage.[21]

4.4.4 Evaluation Methodology

To evaluate the performance of the different portfolios, we report three main out-of-sample performance measures: the average cumulative out-of-sample returns, the standard deviation of the out-of-sample returns, and the Sharpe ratio defined as the ratio of the average returns to the standard deviation of returns. For ease of interpretability, all performance measures are annualized with 12 trading "months." The Sharpe ratio is computed with respect to the actual returns (as opposed to returns in excess of a given benchmark) since we believe it is more relevant in our context where the portfolios are formed solely on the basis of risky assets.

We also report three additional performance measures based on the out-of-sample returns in excess of a 6-factor benchmark; that is, the 5 factors from Fama and French (2015) augmented with out-of-sample returns from UNI. We compute the Jensen's alpha, beta to the out-of-sample UNI returns, and information ratio from a 6-factor regression model. This is done for all portfolios except for UNI. The alpha and information ratio are annualized with 12 trading months. We also provide the t-statistics of the Sharpe ratio and information ratio, which are computed with approximate standard errors from Lo (2002).

Additionally, we report the following portfolio weight statistics averaged over the 578 trading months: turnover, proportional leverage (computed as the fraction of negative weights), the sum of negative weights, and the minimum and maximum weight. Note that these statistics are not our primary focus since our proposed method is not optimized to account for these measures. Nevertheless, they are provided to give a better overview of the different methods.

4.4.5 Application of Canonical Portfolio Analysis

We can use the results of Proposition 4 to identify the sources of portfolio returns by estimating the canonical portfolios that contribute most to its profitability. To this end, we will consider the FF25, ME/OP25, and ME/INV25 test assets for this purpose. The FF25 is widely studied amongst academics and it allows us to check if our method produces the expected results. Moreover, we know from Lewellen et al. (2010) that the size and value portfolio returns have a strong factor structure explained mostly by the three-factor model Fama and French (1993). Kozak et al. (2018) finds that retaining the first three principal components extracted from FF25, closely reproduces the Fama and French three-factor model. Thus, this observation could potentially be exploited in our method.

The left panels of Figure C.4.3 show the squared sample canonical correlations of the demeaned conditional portfolios matrix for the three test assets. It is also overlaid with the squared canonical correlations generated from randomly permuting the signals observations for each asset and repeating a similar exercise.[22] The squared canonical correlations are averaged over the rebalancing dates and ordered from the smallest to the largest. From this in-sample analysis, we can see that the leading two squared sample canonical correlations have values larger than their pseudo-random generated counterparts and that the top one "sticks

[21] The insights are similar if we renormalize the portfolios to achieve a target level of volatility or return.

[22] Randomly shuffling the time series helps generate a null distribution. It is a useful heuristic to determine the importance of each canonical correlation relative to a random benchmark compared to a formal statistical test of significance from Yang and Pan (2015).

out" and extends beyond the value of one.[23] Moreover, the spacing between the sample canonical correlations is wider at the top end of the spectrum and more uniform at the bottom end of the spectrum.

We contrast these findings with the right panels of Figure C.4.3, which shows the out-of-sample returns of each canonical portfolio computed as the product of the in-sample and out-of-sample canonical correlation, accompanied by their standard error. Not surprisingly, we find a performance deterioration in the out-of-sample performance due to the bias in the in-sample predictions. Nevertheless, there is some coherence between the in-sample predictions and the out-of-sample returns in that the top canonical portfolio possesses the most realized returns followed by the second one, while the bottom ones are close to zero.

Given the prominence of the leading canonical correlation, we plot its corresponding weights. The left panels of Figure C.4.4 show the top canonical portfolios (that is, the canonical directions) averaged over the rebalancing dates. To ensure the signs of the canonical portfolios are consistent across time, we flip the sign of the canonical portfolios at any given rebalancing date if its cosine similarity with the canonical portfolios obtained from the previous rebalancing date is negative. While there is not any clear pattern of trades that we can immediately discern, it nonetheless differs from other research findings vis-á-vis PCA; for example, Kelly et al. (2023) finds that the top principal component of a symmetrized cross-covariance matrix goes long (short) on big (small) equity portfolios, and long (short) on value (growth) portfolios. The right panels of Figure C.4.4 provide the final weights invested in each asset.

4.4.6 Empirical Results

Table B.3.2 summarizes the performance of the various portfolio methods for the different test assets. Restricting our attention to the FF25 column, we see that CP2 has an average return, which is lower than that of PP2 (2.20% versus 3.76%) but has much lower volatility (2.18% versus 5.85%).[24] Altogether, this translates into Sharpe ratio of 1.01 which is a 50% improvement over the closest competitor, PP2, at 0.66. The alpha of the CP2 is 1.84% and has a low beta to the univariate factor return of 0.15. Adjusting the alpha by the idiosyncratic volatility of 2.04% gives an information ratio of 0.9 for CP2. A similar conclusion holds for the ME/OP25 and ME/INV25 test assets, although we observe a deterioration in the Sharpe ratio for all portfolios in larger-sized test asset possibly due to estimation errors. CP2 also underperforms PP2 in the ME/INV100 dataset.

Table B.3.3 describes the distribution of the portfolio weights of the estimated portfolios. In the FF25 column, we see that CP2 has the lowest turnover. This is interesting given that we make no effort to control the trajectory of the weights. The average sum of negative weights in the CP2 is -0.496 and the average proportional leverage in CP2 is less than 0.5 indicating a slight tilt towards long positions. The weights are CP2 appear to be the most dispersed but do but are relatively not extreme. We draw a similar conclusion for the other test assets.

It would be interesting to investigate where the performance improvement of CP2 comes

[23] The values of (squared) regularized canonical correlations can exceed one unlike its sample-based analog, which is constrained to the interval (0,1).

[24] Note that the magnitudes of the returns for all portfolios appear to be low in comparison to those published in hedge fund return indices. This is expected since the gross exposure of our portfolios are all constrained to one. In practice, one would typically apply a leverage factor greater than one in order to magnify the returns.

from: is better at selecting assets that have historically performed well on average or at varying the positions in the assets dynamically? One way to discern between the two possible explanations is to decompose the portfolio returns into the following components:

$$\mathbb{E}[w_t' r_{t+1}] = \underbrace{\mathbb{E}[w_t]' \mathbb{E}[r_{t+1}]}_{\text{static}} + \underbrace{\mathsf{Cov}(r_{t+1}, w_t)}_{\text{dynamic}}. \tag{4.41}$$

Intuitively, static bets refer to the investor's ability to get the long-run allocations right. On the other hand, the covariance between the signal and subsequent returns refers to an investor's ability to tactically "time" the market, this is, to accurately identify movements in the assets and gain exposure to those assets accordingly.

Table B.3.4 reports the results of this decomposition, where we estimate the static and dynamic components of returns according to Equation (4.41) with their corresponding sample analogs. Panels A and B show that both the static and dynamic components contribute positively to the overall performance of the portfolio. Panel C of Table B.3.4 breaks down the share of total returns due to taking dynamic bets for CP2 and P2. It shows that on average the returns of CP2 and PP2 come from taking dynamic bets. Therefore, the performance of our proposed method does not come from its ability to long (short) more highly performing (underperforming) assets; it must emanate from its ability to tactically time the market.

Finally, we also isolate the portfolio returns due to the long and short legs of the trade. We write $r_t^w = l((r_t^w)_+ - (r_t^w)_-)$ where $(r_t^w)_+$ is the return on the long leg and $(r_t^w)_-$ is the return on the short leg, with weights of both trade legs normalized to sum to one. Here, l denotes the leverage of the long-short portfolio, but since our estimated portfolios have a unit gross exposure by construction, l is close to one. Panel D and E of Table B.3.4 present the statistics of the long and short legs, respectively. For CP2, the average return on the long leg is 8.01% and the short leg is 5.71%. This indicates that the profits of our portfolio come from the long side of the trades.

4.4.7 Robustness Checks

In this section, we inspect whether the outperformance of our proposed portfolio construction methodology is robust to different revisions in the current empirical set-up. In particular, we will be interested in results based on (1) subsample period, (2) forecast horizon (3) shrinkage in the signal covariance, (4) momentum lookback window, and (5) "approximate" versus "actual" portfolio policy.

4.4.7.1 Sub-Period Analysis

In this section, we check if there are any peculiar subsample effects that may drive the performances of our proposed scheme. We divide the out-of-sample period into four roughly equally-sized subsamples of 144 months (that is, 12 trading years) each: (1) 1986-09-25 to 1986, (2) 1986 to 1998, (3) 1998 to 2010, and (4) 2010 to 2022. Then we perform the same procedure in each subsample. The results are provided in Table B.3.5.

Generally, the performance of all portfolios appears stronger in the earlier periods of the sample but poorer in the recent decade. We see that the outperformance of CP2 over the competing portfolios is consistent over time for FF25 and ME/INV25. However, CP2 underperforms PP2 in (1) ME/INV100 for most of the subsamples, and (2) FF100,

ME/OP25, and ME/OP100 in the earlier subsample.

4.4.7.2 Forecast Horizon

To make a forecast of subsequent returns, we have used a horizon length of 21 trading days, which roughly corresponds to one month. Given our use of non-overlapping observations between the signals and subsequent returns, this implies that the data are sampled on a monthly basis, which also corresponds to the frequency with which we rebalance our portfolios. We now change the forecast horizon from 21 days to 1, 5, and 10 days. Each of these forecast horizons covers an investment period from (1) 01-06-1965 to 10-21-2022, (2) 12-06-1966 to 10-21-2022, and (3) 06-12-1969 to 10-21-2022; this provides us with 14550 daily, 2814 weekly, and 1347 fortnightly out-of-sample returns, respectively.

Table B.3.6 shows that the annualized Sharpe ratio of all portfolios tends to be better at shorter holding periods. Barring the potential side effects of illiquidity in daily returns, this suggests that the conditioning information becomes more relevant as the holding period decreases. This makes sense since the conditional portfolios become more reactive to changes in the states of the market. Overall, the ranking of the methods remains similar relative to Table B.3.2 with the exception of the 1-day horizon, where CP2 is the best performer for all test assets.

4.4.7.3 Shrinkage Intensity

We examine the effect of shrinkage on the covariance of signals in our proposed method. Our default choice in the analysis was $\delta_x = 0.9$. The annualized Sharpe ratios for CP2 corresponding to shrinkage values of $\delta_x \in \{0, 0.1, 0.5, 0.8, 1\}$ are presented in Table B.3.7 over different subsamples.

We see that there is no specific level of shrinkage that provides consistent outperformance for all test assets. This indicates that the shrinkage intensity δ_x is time-varying in nature. Shrinkage values greater than or equal to 0.5 appears to work well across different subsamples for larger-sized equity portfolios up until the recent decade, where there is some benefit of using more sample information from the signal correlations in datasets ME/OP100 and ME/INV100.

4.4.7.4 Momentum Lookback Window

The 21-day momentum signal that we used as a base case is perceived to be a relatively fast signal, reacting quickly to changes in market conditions. We now consider using momentum signals computed with a longer lookback window of sizes 42 days, 63 days, 84 days, 126 days, and 252 days. The remaining details remain similar.

Table B.3.8 demonstrates that CP2 generally outperforms the other portfolios in terms of annualized Sharpe ratio for different momentum signals and different test assets. Although we observe some performance degradation for all portfolios as we increase the lookback size from 21 days to 126 days, the ranking of the methods remains similar to Table B.3.2. There is, however, a significant performance gain in CP2 not seen in the other portfolios as we extend the momentum lookback size from 126 days to 252 days. This is comforting since momentum signals with longer lookback horizons tend to have lower turnover.

4.4.7.5 Approximate Versus Actual Solution

We have worked with the "approximated" mean-variance problem from Proposition 2 throughout this paper since it helped us to simplify the analysis. Given that we also have a

closed-form expression to the "full" mean-variance problem from Theorem 2 in Proposition 1, we now check if it offers any practical benefits over our approximate solution. The only difference between both solutions is in terms of how the canonical correlations enter into the reconstruction of the optimal portfolio policy. The full solution essentially nonlinearly adjusts the canonical correlations, while no adjustment takes place in the approximated solution. Figure C.4.2 shows that both solutions are similar for small canonical correlation values but for large canonical correlation values, the conservative behavior of the optimal strategy is reinforced by downweighting its influence. This makes sense because we are taking into account more terms that affect the risk profile of the strategy.

From Table B.3.9, we see that our approximated formula generally performs better than the full formula across different test assets. One reason for this observation is that the nonlinear adjustment of the canonical correlations may be too conservative relative to the unadjusted one. For example, a canonical correlation value of one gets reduced by half through the nonlinear adjustment. This holds mechanically irrespective of the data. Consequently, this behavior may inadvertently under-leverage the leading two canonical portfolios, which are the most profitable streams.

4.4.8 Extension: Two-Signal Case

Our canonical portfolios modeling framework is flexible enough to accommodate multiple signals. In this empirical exercise, we expand our signal vector to include two momentum signals of different lookback windows; one with a 21-day lookback window, and another with a 252-days lookback window. This expanded $2N$-dimensional signal vector will serve as input for both CP and PP. On the other hand, we assume that both MVO and UNI take in an equal-weighted average of the two momentum signals as inputs since these methods do not have an optimal way of blending different signals together in a single stage. The rest of the empirical setup remains unchanged.

From Table B.3.10, we see that the Sharpe ratio of all portfolios, with the exception of MVO, generally improves in comparison to the one signal case. More importantly, the outperformance of CP2 over competing methods continues to hold up for different test assets.

Turning our attention to Table B.3.11, we see that the turnover is reduced for all portfolios compared to the base case. This observation can be attributed to the inclusion of the signal with a long lookback window, which tends to have a lower turnover than one with a short lookback window. This is appealing since we can expect the performance after factoring in transaction costs to be better than using a single conditioning variable. We also observe lower proportional leverage throughout the portfolios indicating a tilt toward long positions. The weights also appear to be less dispersed than in the base case.

4.5 Conclusion

In this paper, we provide a novel framework for portfolio managers and academics to conceptualize the optimal asset and signal combination problem with canonical correlation analysis (CCA). Our contribution can be summarized as follows. First, we recast the original investment problem of Brandt and Santa-Clara (2006) into a tractable one that allows us to derive an optimal portfolio policy that is applicable to large cross-sectional financial

applications. Our portfolio policy is able to ingest multiple return-predictive signals, and account for cross-predictability and correlations in both the returns and signals. All of these properties are achieved by solving the portfolio selection problem in a single stage and assuming Gaussianity on the joint distribution of the assets and signals. Second, we attempted to lift the veil of complexity from our large-dimensional investment problem through a novel application of CCA. In particular, CCA breaks down the correlations of all of the asset returns and signals into independent long–short managed portfolios, which we term as *canonical portfolios*. Each of the canonical portfolios can be ranked from the one with the smallest correlation to the one with the highest; the canonical portfolios with the highest predictable returns get scaled up the most.

Having established the theoretical contents of our method, we bring it to the empirical test. We ran backtest simulations on Fama-French sorted equity portfolios with a momentum signal, and we have found that our proposed method generally outperforms natural benchmarks. The performance of our method further improves when the analysis is extended to two momentum signals of different lookback windows. These results are made possible by introducing regularization techniques to overcome the instabilities in the estimation and exploiting the most predictable dimensions of the data.

In terms of future work, our proposed modeling framework is not set in stone and is flexible enough to accommodate further improvements. Some research ideas include introducing nonlinearities or regime-switching into the modeling process, and relaxing the Gaussianity assumption on the joint return distribution. Another interesting avenue to explore would be to apply these ideas to develop a test for asset pricing models or to form portfolios on different asset classes such as individual stocks portfolios, fixed-income portfolios, currency portfolios, and so forth. Last but not least, recasting the portfolio selection problem into a CCA framework enables researchers to leverage the insights and techniques from the rich literature of CCA that has been expanded by developments in machine learning.

Chapter 5

Dynamic Portfolio Selection under Transaction Cost and Signal Decay

> The government has a strategy; the people have a counter-strategy.
>
> Ancient Chinese Proverb;
> —Thomas J. Sargent, *Rational Expectations and Inflation* (2013)

5.1 Introduction

Portfolio managers often rely on a diverse set of return-predictive signals to capture investment opportunities that vary over time.[1] Trading signals that are temporary react quickly to short-term market fluctuations but have a short lifespan, while permanent ones have a longer lifespan and can potentially provide more reliable information about long-term market trends. To combine these signals into a single predictor that informs investment decisions, there is a standard approach based on modern portfolio theory pioneered by Markowitz (1952). This approach is built on the foundation of myopic investors with mean-variance preferences who construct their portfolios at the start of the investment period and hold them until the end of the period. Any change to the investment opportunity set will lead to an adjustment in the mean-variance efficient portfolio, without considering transaction costs.

However, in practice, trading is costly due to market frictions. These costs can either be explicit or implicit (Keim and Madhavan, 1998). Explicit costs include commissions, brokerage fees, and taxes, while implicit costs include bid-ask spread, price impact, and opportunity costs. For large trades, price impact, which refers to the difference between the executed price and what the price would have been in the absence of the transaction, tends to be the dominant source of transaction costs. Price impact can be both temporary and permanent, with temporary impact affecting a single transaction and permanent impact affecting every future transaction. To mitigate transaction costs, it is common practice for large institutional investors to periodically

[1] The terminology "signals" is also often referred to as predictors, alphas, attributes, characteristics, state variables, instrumental variables, cross-sectional anomalies, etc. in the finance literature.

adjust their portfolio on a quarterly or annual basis. However, doing so may also limit the potential gains from investment opportunities that require more frequent rebalancing. Hence, active managers who are forward-looking and prioritize cost efficiency must carefully balance the trade-offs involved in reducing transaction costs and capitalizing on profitable near-term investment opportunities.

This paper studies a multiperiod portfolio selection problem that considers both transaction costs and time-varying investment opportunities while also being scalable to a large investment universe. The investment problem involves minimizing a loss function, which comprises the accumulation of costs incurred over the lifetime of the investment that arises from deviations from a benchmark index[2] or target portfolio, and the trading costs. The former cost is referred to as a *tracking error*, which is widely used in the investment management industry in lieu of postulating a utility function for wealth, which can be difficult to specify in practice. The latter cost is modeled as a combination of temporary and permanent price impact that are linear in the size of the trades. Although empirical and theoretical studies, such as Hasbrouck (1991), Hausman et al. (1992), Keim and Madhavan (1996), Lillo et al. (2003), Almgren et al. (2005), and Engle et al. (2012), have argued that price impact is a positive and concave function of the trade, our cost specification renders the investor's problem to be analytically tractable. This allows us to derive a closed-form expression for the optimal portfolio policy, which involves smoothly adjusting toward a forward-looking portfolio (represented as a composite sum of current and expected future exogenously given target portfolios) at a rate determined by the investor's time, risk, and liquidity preferences.

As an application of our methodology, we also derived analytical formulas for the steady-state performance of the optimal portfolio policy in relation to the structural parameters of the model. Our analysis shows that investment portfolios reach a steady-state over time that is systematically predictable, in line with the findings of Sneddon (2008). This insight can provide portfolio managers with a useful tool to understand the turnover rate and Sharpe ratio of their investment strategies. By using these formulas, portfolio managers can determine the degree of flexibility they have to adjust the scale of their strategies, which is crucial for optimizing their performance.

Multiperiod portfolio selection has been a subject of extensive research in the academic literature, dating back to the early works of Samuelson (1969) and Merton (1969, 1971). These papers study dynamic portfolio selection models that can accommodate time-varying investment opportunities and general objective functions. However, incorporating transaction costs into the framework has posed challenges in solving the standard multiperiod investment problem. Indeed, although many papers have developed analytical frameworks for scenarios with multiple assets, they have largely focused on numerical examples involving only a small number of risky assets. Previous works by Balduzzi and Lynch (1999, 2000) have examined the effects of proportional transactions with return predictability on utility costs and optimal rebalancing rules for a single-risky asset by discretizing the state space of the dynamic program. This approach was extended in Lynch and Tan (2010) to include two risky assets but at a significant computational expense. To address this challenge, Brown and Smith (2011) proposed heuristic trading strategies for the multiple assets case and employed upper bounds based on duality

[2]Typical examples include the 60/40 portfolio of stocks and bonds, or the 40/40/20 portfolio of stocks, bonds, and cash.

theory to assess the effectiveness of the proposed heuristics.

Our paper shares a close connection with Gârleanu and Pedersen (2013, 2016) as we adopt a similar analytically tractable framework that considers a quadratic criterium. Indeed, this allows us to arrive at a closed-form expression for the optimal dynamic portfolio policy that invests in multiple assets, which can serve as an approximation to more intricate trading models and provides us with a readily understandable description of how trades impact prices. However, our work also differs in a few key aspects. First, the authors assume mean-variance preferences for investors, whereas we consider a tracking error problem, which is more relevant for most portfolio managers. Nevertheless, our approach leads to their main results under certain simplifying assumptions. Second, we solve the multiperiod investment problem by employing stochastic discrete-time calculus of variations and solving the prediction problem with the Wiener-Kolmogorov method (Whittle, 1983), whereas the authors adopt a recursive approach and solve the control problem using dynamic programming. Our approach offers distinct advantages over dynamic programming procedures in terms of more streamlined derivations and the explicit provision of investment decision rules. Finally, our paper aims to extend the results from our recent work in Firoozye et al. (2023), which studies dynamic portfolio policies that are parameterized linearly in the signals in the absence of transaction costs.

Our approach to solving the dynamic portfolio selection problem is based on classical variational methods that take inspiration from Hansen and Sargent (1980), Sargent (1987), and Ljungqvist and Sargent (2004). These procedures are widely used in macroeconomics in solving dynamic models of "rational expectations" (Lucas, 1976; Muth, 1961), which posits that the subjective beliefs held by each agent regarding future outcomes coincide with the actual probability distributions, conditional on the agent's information set. A prominent example includes models of dynamic stochastic general equilibrium (DSGE), which are widely used by central banks, government agencies, and academic economists to guide monetary and fiscal policy decisions. One of the essential characteristics of rational expectations is that it considers the forward-looking behavior of rational agents, a feature that is also fundamental in our investor's optimal portfolio policy rule.

Our paper is also related to the literature on optimal trade execution; see *inter alia* Perold (1988), Bertsimas and Lo (1998), Almgren and Chriss (2001), Engle and Ferstenberg (2007), Obizhaeva and Wang (2013). This strand of research investigates methods for scheduling large blocks of buy or sell orders into a sequence of smaller orders. As emphasized by Engle and Ferstenberg (2007), the optimal execution problem involves a trade-off between the cost of immediate execution and the risk of adverse price movements. While executing a trade as a single large block results in higher costs, dividing it into smaller blocks and gradually executing them reduces the cost of immediacy but at the expense of greater variability in the average execution price. Striking an appropriate balance among these costs naturally leads to a risk-to-reward trade-off frontier that closely aligns with the principles of modern portfolio theory.

The rest of the paper is organized as follows. Section 5.2 introduces our multiperiod portfolio selection framework and presents our analytical results. Section 5.3 explores modeling extensions that consider permanent price impact and portfolio growth. Section 5.4 studies the transmission effects of an exogenous disturbance on the portfolio over time. Section 5.5 provides analytical expressions for the performance of the optimal portfolio in steady-state. Section 5.6 concludes. Appendix A.2, B.4, C.5 contain all the mathematical derivations, tables, and figures.

5.2 Modeling Framework and Approach

5.2.1 Notation

In this section, we introduce the notational convention for our analysis. Let the subscript i index the assets such that $i \in \{1, \ldots, N\}$, where N denotes the number of assets; the superscript j indexes the return-predictive signals such that $j \in \{1, \ldots, m\}$, where m denotes the number of signal for each asset; and the subscript t indexes the trading dates such that $t \in \{1, \ldots, T\}$, where T denotes the number of observations. For any N-dimensional vector z, its ℓ_{p}-norm is given by $\|z\|_{\mathsf{p}} \coloneqq (|z_1|^{\mathsf{p}} + \ldots + |z_N|^{\mathsf{p}})^{1/\mathsf{p}}$. Denote $\mathbb{I}_N$ as an identity matrix of dimensions $N \times N$, and $\mathbf{1}_N$ as a vector of dimensions N. In addition, $\mathsf{Tr}(\cdot)$ represents the trace of a square matrix and $\mathsf{Diag}(\cdot)$ represents the diagonal matrix of a given vector. The symbol $(')$ denotes the transpose operator of a vector or matrix, the symbol $\coloneqq$ is a definition sign, and the symbol $=$ is the equal sign.

Let $p_{t,i}$ be the price of a risky asset i sampled at the trading date t, stacked into a vector $p_t \coloneqq (p_{t,1}, \ldots, p_{t,N})'$. The security returns at time t in excess of the risk-free rate, denoted as r_t, is defined as $r_t \coloneqq p_t - (1 + r^{\mathrm{f}})p_{t-1}$, where r^{f} represents the risk-free rate. Also, let $x^j_{t,i}$ be the jth signal for asset i at date t, stacked into a M-dimensional vector $x_t \coloneqq (x^1_{t,1}, \ldots, x^1_{t,N}, \ldots x^m_{t,1}, \ldots, x^m_{t,N})'$, where $M \coloneqq m \times N$ is the total number of signals.[3] This means every asset is accompanied by at least one signal and so m is necessarily a positive integer. For example, with two assets and two signals per asset, we have $m = 2$ and $M = 4$. The signals $\{x_t\}$ are assumed to be adapted to the information sets $\{\mathcal{F}_t\}$ and are understood to proxy the future expected returns, conditional on the investor's information set at time t. We impose the following assumption on both the returns and signals throughout our analysis.

Assumption 4 (Stationarity). The security returns sequence $\{r_{t+1}\}$ and the signal sequence $\{x_t\}$ exhibit stationarity and ergodicity, whose moments can be calculated by taking time series averages.

The notation $\mathbb{E}[\cdot]$ represents the mathematical expectation operator of a random vector and $\mathsf{Cov}(\cdot, \cdot)$ represents the covariance matrix between two random vectors. Their conditional counterparts, given the information set $\mathcal{F}_t$ available at time t–including the model's structure and all previous and current realizations of both exogenous and endogenous processes–are denoted as $\mathbb{E}_t[\cdot] \coloneqq \mathbb{E}[\cdot|\mathcal{F}_t]$ and $\mathsf{Cov}_t(\cdot, \cdot) \coloneqq \mathsf{Cov}(\cdot, \cdot|\mathcal{F}_t)$, respectively. The returns and signals have covariances denoted by $\mathsf{Var}(r_{t+1}) \coloneqq \mathsf{Cov}(r_{t+1}, r_{t+1}) = \Sigma_r$ and $\mathsf{Var}(x_t) \coloneqq \mathsf{Cov}(x_t, x_t) = \Sigma_x$, and a cross-covariance $\mathsf{Cov}(r_{t+1}, x_t) = \Sigma_{rx} = \Sigma'_{xr}$. Due to the symmetric positive semidefinite property of the covariance matrices, they admit a spectral decomposition. A matrix factor denoted by $\Sigma_r^{1/2}$ can be obtained by retaining the eigenvectors of the matrix and reassembling them with the square roots of its eigenvalues.

5.2.2 The Investor's Problem

The investor in our problem has to select a planned sequence of trade decisions $\{w_t\}_{t=0}^{\infty}$, which is a dimensionless amount of shares, at each time t in an infinite-horizon setting. We

[3] The signals x_t can incorporate information about the unconditional mean excess returns of each asset. This can be achieved by assigning a value of 1 for the signal of each asset for all time.

assume that the investor has access to unrestricted borrowing and there is no uncertainty in the execution of the trades, which means that they are limited to market orders. Let us denote W_t to be the investor's wealth and its change during period $t-1$ and t as $\Delta W_t \coloneqq W_t - W_{t-1}$, where Δ is the first-difference operator. The investor's excess wealth change at each period is given by $\Delta W_{t+1} - r^{\mathrm{f}} W_t = w_t' r_{t+1}$. This represents the excess gross dollar profits and losses that an investor would be subject to before transaction costs.

In addition, we assume that at each time t, there is a target level of the portfolio, w_t^*, called the "frictionless" or "idealized" portfolio, which is an exogenous stochastic process representing a hypothetical portfolio that would prevail in an identical market that is not subject to transaction costs. The difference between the portfolio w_t and the target portfolio w_t^* creates a tracking error in their returns, represented as $(w_t - w_t^*)' r_{t+1}$. The tracking error variance of a portfolio is thus, defined by the following bilinear form

$$\mathsf{TEV}(w_t) \coloneqq \frac{\gamma}{2} \mathsf{Var}_t[(w_t - w_t^*)' r_{t+1}] = \frac{\gamma}{2} (w_t - w_t^*)' \Sigma_{r,t} (w_t - w_t^*), \tag{5.1}$$

where $\gamma > 0$ is a constant that reflects the investor's *absolute* risk aversion towards deviations from the target portfolios.[4] If $w_t^* = 0$ for all t, then the tracking error variance collapses to the variance of a portfolio. The expression (5.1) assumes that the squared expected returns of the tracking error portfolio are approximately zero; this is a reasonable assumption which also helps to resolve the time-inconsistency issue in the intertemporal optimization problem attributed to the variance term.

We also consider the transaction cost that the investor has to pay in each period. This can be modeled based on the following quadratic specification from

$$\mathsf{TC}(\Delta w_t) \coloneqq \frac{1}{2} \Delta w_t' \Lambda_t \Delta w_t, \tag{5.2}$$

where Δw_t is the difference between the desired position and the previous day's position. Previous studies, such as Heaton and Lucas (1996), Grinold (2006), Engle and Ferstenberg (2007), and Gârleanu and Pedersen (2013, 2016), have examined similar cost structures. The quadratic nature of it means that if an investor acquires an asset twice the original amount of shares, this excess demand impacts the price by a factor of two, which results in a total transaction cost that is four times the original dollar amount.

The matrix Λ_t determines the transaction costs incurred when Δw_t shares are traded during the rebalancing of the portfolio at time t. Without loss of generality, we assume that Λ_t is symmetric.[5] It reflects the temporary price impact created by trading across different assets and can be seen as a multivariate version of Kyle (1985), which is often referred to as "Kyle's lambda." The (i,j)-th element of Λ_t represents the price impact on asset i caused by trading asset j at time t. This means that when an investor trades Δw_t shares, the average price is impacted linearly in it by an amount of $\Lambda_t \Delta w_t$ dollars. This yields a total transaction cost that is proportional to the product of the number of shares traded and the dollar price impact.

Moving forward, we will impose the following assumption on the conditional covariance matrix of both the returns and the transaction cost matrix throughout our analysis for

[4] Here, γ is measured in units of inverse dollars. This means the investor is willing to tolerate an additional tracking error variance of $1\$^2$ if it translates to a dollar reduction of $\gamma\$$ in our expected transaction cost.

[5] Indeed, the quadratic form of the transaction cost term always nullifies its antisymmetric component.

tractability purposes.

Assumption 5. The conditional covariance of the returns and the transaction cost matrix are time invariant.

Equipped with two competing economic objectives, our investor seeks to minimize the following discounted present value intertemporal optimization problem.

$$\lim_{T\to\infty} \min_{\{w_t\}_{t=0}^{T}} \mathbb{E}_0 \left[\sum_{t=0}^{T} \frac{(1-\rho)^{t+1}}{2} (w_t - w_t^*)' \gamma \Sigma_r (w_t - w_t^*) + \frac{(1-\rho)^t}{2} \Delta w_t' \Lambda \Delta w_t \right], \tag{5.3}$$

subject to w_{-1} given, where $\rho \in (0,1)$.[6] The investor is assumed to be infinitely lived to simplify the analysis; however, this assumption is not restrictive since we can modulate the effective investment horizon by varying the investor's rate of time preference, thereby adjusting the relative importance of the near future and the distant future. According to the objective function (5.3), the adjustment of the portfolio results in costs, which can be incurred due to deviations from the target portfolio and changes in the level of adjustment. This problem can be seen as an instance of the classic rational expectations policy model with costly adjustments from Hansen and Sargent (1980).[7]

Optimization of the portfolio over multiple periods is equivalent to optimal tracking of a sequence of target portfolios. Essentially, if the portfolio deviates considerably from the target portfolio, then it may be reasonable for the investor to pay the cost of the transaction to reduce the tracking error. In a setting with no transaction costs, it is always optimal to hold the target portfolio. However, it is not optimal to adjust the portfolio completely to the target portfolio since the marginal decrease in tracking error may not sufficiently outweigh the increase in transaction costs.

We end this section with two remarks.

Remark 8 (Expected utility maximization). It is natural to ask whether our investment problem formulation is consistent with the notion of expected utility maximization. The extent to which it does depends on the specification of the target portfolio. Consider the Markowitz portfolio $w_t^* = (\gamma\Sigma_r)^{-1}\mu_t$, where $\mu_t := \mathbb{E}_t[r_{t+1}]$ represents the expected future returns. In this context, our approach yields optimal portfolio policies for investors aiming to maximize the expected total of discounted utilities, with a mean-variance utility over changes in wealth. The mean-variance utility, at any time, can be formulated as follows:

$$w_t'\mu_t - \frac{\gamma}{2} w_t' \Sigma_r w_t = w_t^{*\prime} \mu_t - \frac{\gamma}{2} w_t^{*\prime} \Sigma_r w_t^* - \mathsf{TEV}(w_t). \tag{5.4}$$

Thus, the goal of maximizing the expected discounted sum of utilities minus transaction costs is the same as minimizing the discounted sum of tracking errors plus transaction costs. This minimization problem is precisely in line with our stated objective.

Remark 9 (Absolute and relative price changes). For short investment horizons, it is reasonable to base our modeling framework on changes in wealth derived from absolute price

[6] It is worth noting that transaction costs are incurred at the time of rebalancing, whereas the returns are realized at the end of the investment period. This difference in timing explains why different discount rates are used to calculate the present values of the tracking error variance versus the transaction cost.

[7] A similar objective function is also presented in the rational expectations model of Kennan (1979), within the context of analyzing the relationship between employment and output in U.S. manufacturing industries.

changes rather than relative (percentage) price changes. According to Mei et al. (2016), price changes tend to remain relatively stable in periods of one year. However, for long investment horizons or highly volatile markets, it might be more suitable to take into account compounded returns to wealth over time, along with relative price changes; see, for example, Collin-Dufresne et al. (2020) and Jensen et al. (2023).

5.2.3 The Euler Equation

We now proceed to minimize the total loss of our investor's problem (5.3) by means of first-order conditions.

Proposition 6. The optimal portfolio policies $w_0, w_1, \ldots, w_{T-1}$, satisfy the following system of second-order linear expectational difference equations

$$w_t = \Psi_1 w_t^* + \Psi_2 w_{t-1} + \Psi_3 \mathbb{E}_t[w_{t+1}], \tag{5.5}$$

and the following transversality condition

$$\lim_{T\to\infty} \bar{\rho}^T \mathbb{E}_0[\bar{\rho}\gamma\Sigma_r(w_T - w_T^*) + \Lambda(w_T - w_{T-1})] = 0. \tag{5.6}$$

The matrix coefficients are given by

$$\begin{aligned} \Psi_1 &\coloneqq \bar{\rho}[\bar{\rho}\Omega + (1+\bar{\rho})\mathbb{I}_N]^{-1}\Omega \\ \Psi_2 &\coloneqq [\bar{\rho}\Omega + (1+\bar{\rho})\mathbb{I}_N]^{-1} \\ \Psi_3 &\coloneqq \bar{\rho}[\bar{\rho}\Omega + (1+\bar{\rho})\mathbb{I}_N]^{-1}, \end{aligned} \tag{5.7}$$

where $\Omega \coloneqq \gamma\Lambda^{-1}\Sigma_r$ and $\bar{\rho} \coloneqq 1-\rho$ for notational convenience.[8]

The dynamic optimality conditions (5.5) are referred to as Euler equations, which relate the investor's current trading decision to the past, present and future. Intuitively, because investors are forward-looking, face transaction costs, and have forecasts multiple periods ahead, they base their investment decisions on the expected future behavior of the forecasts while also relying on past decisions as an anchor. The extent to which the investor trades off between these portfolios depends on the structural parameters of the optimization problem such as the investor's time, liquidity, and risk preferences.

The Euler equations are a necessary condition for optimality, but they are not sufficient. In addition, we require the transversality condition (5.6) to hold, which means that the present value of an investor's expected future benefits from an investment converges to zero as time approaches infinity. Colloquially speaking, the investors do not leave any money on the table at the end of time since otherwise, they are not optimizing their investment. This is needed to rule out explosive growth of the expectations of future portfolios w_t^*, and is satisfied if the following assumption is maintained.

Assumption 6. We assume that the target portfolios $\{w_t^*\}_{t=0}^{\infty}$ is a N-dimensional stochastic process of mean exponential order less than $\bar{\rho}^{-1/2}$, so that $\lim_{j\to\infty} \mathbb{E}_t[\bar{\rho}^{j/2} w_{t+k}^*] = 0$ for all t. That is, a sequence $\{w_t^*\}$ is said to be of exponential order less than $\bar{\rho}^{-1/2}$ if there exist t_0

[8] The matrix Ω represents a risk matrix that is adjusted to account for price impact considerations. It measures the relative importance of cost in relation to tracking error and price impact.

and $K > 0$ such that for every $t > t_0$, $\|w_t^*\| \leq Kz$ for some $z \in [1, \bar{\rho}^{-1/2})$, where $\|\cdot\|$ is any well-defined norm.

The above assumption states that the target portfolio should not grow faster than the square of the reciprocal of $\bar{\rho}$ indefinitely (that is, it decays exponentially as t increases). We will impose this requirement on the target portfolio process throughout.

5.2.4 The Analytical Solution

The optimality condition outlined in Proposition 6 provides us with a means to characterize the investor's optimal portfolio policy, but it does not determine the current optimal portfolio. Indeed, an investment decision rule is one that expresses the current optimal portfolio as a function of variables known to the investor at time t. Fortunately, the Euler equations and the transversality condition are linear equations, so it is possible to derive an analytical expression for the current optimal portfolio policy.

It is instructive to begin our analytical efforts by imposing the following structure on the transaction cost matrix.

Assumption 7. Let $\Lambda = \lambda\Sigma_r$, where λ denotes a price impact coefficient.

This specification in which the transaction cost matrix is proportional to the covariance matrix of returns was also considered in Gârleanu and Pedersen (2013, 2016). The authors suggested that the transaction cost in this setting can be thought of as the compensation a dealer would receive for taking the other side of the trade. In other words, it reflects the risk that the dealer would be exposed to, with their risk aversion being represented by the price impact coefficient λ.

Proposition 7. Under Assumption 4-7, the optimal portfolio policy with temporary transaction costs is given by the following convex combination

$$w_t = \theta_1 w_{t-1} + (1 - \theta_1)\mathsf{aim}_t, \tag{5.8}$$

where

$$\mathsf{aim}_t = (1 - \theta_2^{-1}) \sum_{k=0}^{\infty} \theta_2^{-k} \mathbb{E}_t[w_{t+k}^*]. \tag{5.9}$$

The constants are given as follows:

$$\theta_1 = 1 - \frac{-(\omega\bar{\rho} + \rho) + \sqrt{(\omega\bar{\rho} + \rho)^2 + 4\omega\bar{\rho}^2}}{2\bar{\rho}} \qquad \text{and} \qquad \theta_2 = (\bar{\rho}\theta_1)^{-1}, \tag{5.10}$$

where $\omega := \gamma/\lambda$. The distributed lag weights satisfy $\theta_2 > \bar{\rho}^{-1}$ while the quantity $\theta_1 \in (0, 1)$ depends negatively on ω.

Proposition 7 states that the current optimal portfolio depends on the past portfolio and the entire future path of the target portfolio. That is, at each trading date the optimal portfolio partially adjusts itself towards a composite portfolio, denoted as aim, which is expressed as a weighted sum of present and future values of the w_t^*. Thus, the trading is said to be purged into the direction of future forecasts.

Equation (5.8) bears a close resemblance to an exponentially weighted average scheme, which smooths out the time-series variations in the portfolio dynamics by anchoring to the optimal portfolio from the previous period. The rate at which the optimal portfolio adjusts itself towards the aim portfolio is governed by the parameter θ_1. This parameter also determines the rate at which the optimal portfolio converges to a steady state when it initiates away from it. The lower (higher) θ_1 is, the faster (slower) the portfolio adjustment. Moreover, the quantity θ_1 depends on the model's structural parameters such as the investor's time preference ρ, risk aversion γ, and the price impact coefficient λ.

It is interesting to observe that the profile of the optimal portfolio trajectory is better represented by the ratio ω. This ratio can be interpreted as a risk aversion parameter adjusted for price impact. Thus, even though a higher λ might result in increased transaction costs, if it is accompanied by a greater level of risk aversion γ, it allows the investor to trade more frequently to capitalize on time-varying investment opportunities.

Remark 10 (Certainty equivalence). The solution (5.8) exhibits the certainty equivalence or separation property. This means that we would arrive at the same solution for w_t if we had optimized the criterion by replacing w_t^* with $\mathbb{E}_0[w_t^*]$ and dropping the expectation operator $\mathbb{E}_0[\cdot]$ outside the summation in the objective function (5.3). This means that investors treat uncertain outcomes as if they were certain, based on their expected values. As a result, the optimal portfolio w_t depends solely on the conditional mean of the discounted value of the target portfolio sequence and it is independent of the conditional variances of the target portfolios around their predicted values. This property allows for the separation of the investor's minimization problem into two parts: an optimization or control problem, and a prediction problem.

Remark 11 (Single-period optimization). In a single-period model, where investors are myopic, the optimization problem can be written as:

$$\min_{w_t} \frac{1}{2}(w_t - w_t^*)'\gamma\Sigma_r(w_t - w_t^*) + \frac{1}{2}\Delta w_t'\Lambda\Delta w_t \tag{5.11}$$

In this case, the solution is given by

$$w_t = \frac{1}{1+\omega}w_{t-1} + \frac{\omega}{1+\omega}w_t^*. \tag{5.12}$$

This solution is also expressed in terms of a convex combination between two portfolios, wherein investors adjust partially toward the optimal target portfolio on each trading date. However, the main feature of the solution (5.8) is the inclusion of a combined forward-looking total of target portfolios. This particular term is what sets our proposed portfolio policy apart from alternative approaches that rely on single-period optimization.

We now generalize the results in Proposition 7 to allow for a more general structure of the transaction cost matrix.

Proposition 8. The optimal portfolio policy with temporary transaction costs is given by

$$w_t = \Theta_1 w_{t-1} + (\mathbb{I}_N - \Theta_1)\mathsf{aim}_t, \tag{5.13}$$

where

$$\mathsf{aim}_t = (\mathbb{I}_N - \Theta_1)^{-1} \sum_{k=0}^{\infty} \Theta_2^{-(k+1)} \Omega \mathbb{E}_t[w^*_{t+k}]. \tag{5.14}$$

The coefficient matrices Θ_1 and Θ_2 are provided in Appendix A.2, and their eigenvalues lie inside and outside the unit circle, respectively.

The results presented so far hold for any idealized portfolio sequence $\{w^*_t\}_{t=0}^{\infty}$. At present, we do not make any assumptions about how the return-generating process depends on the signal or how the signals evolve over time.

Remark 12. If the target portfolio w^*_t is the Markowitz portfolio and the returns have a linear factor model representation, then the optimal portfolio policy coincides with that of Gârleanu and Pedersen (2013, Proposition 2).

5.2.5 The Prediction Problem

We are now left to specify a forecasting model for the signals so that we can construct a proxy for the forward-looking term. To complete the model, we make the following specification for the exogenous stochastic process $\{w^*_t\}_{t=0}^{\infty}$, as considered in works from Brandt and Santa-Clara (2006), Kelly et al. (2023) and Firoozye et al. (2023).

Assumption 8. The target portfolio w^*_t is parameterized linearly in the signal vector, that is $w^*_t = A'x_t$ for all t, where A is some $M \times N$ time-invariant policy matrix.

We now make the pessimistic assumption that the signals have decaying strength, following an autoregressive process.

Assumption 9. The signals follow a vector autoregressive (VAR) model of order one given by

$$\Delta x_{t+1} = -\Phi x_t + \varepsilon_{t+1}, \tag{5.15}$$

where Φ is a matrix conformable in the dimensions of x_t with eigenvalues that lie inside the unit circle, and ε_t is a noise term with variance $\mathsf{Var}_t(\varepsilon_{t+1}) = \Sigma_\varepsilon$.

Although this assumption may not apply to all signal types, it is a plausible one for a large class of signals, such as momentum and mean-reversion signals. This autoregressive structure implies that if an investor has a mean-reversion signal of, say, 5 bps (basis points) today, they anticipate that the same signal will remain valid tomorrow but subsequently decays to 2 bps.

Proposition 9. Under Assumptions 4, 6, 8, and 9, the aim portfolio has the following closed-form expression

$$\mathsf{aim}_t = (\mathbb{I}_N - \Theta_1)^{-1} \Theta^{\mathsf{aim}} x_t, \tag{5.16}$$

where the $N \times M$ matrix Θ^{aim} is given in Appendix A.2. If Assumption 7 holds in addition, the aim portfolio simplifies to the following expression

$$\mathsf{aim}_t = A' \left(\mathbb{I}_M - (1 - \theta_2)^{-1} \Phi \right)^{-1} x_t. \tag{5.17}$$

The fact that Φ appears in the law of motion of the investment decision process and the law of motion of the exogenous stochastic process is an example of a *cross-equation restriction*. Hansen and Sargent (1980, Section 7) refer to this observation as the "hallmark" of rational expectations models. In particular, the estimation of the investor's portfolio policy has to be performed jointly with the stochastic process model they posit, subject to the cross-equation restriction, which ensures that the investor's expectations are consistent with the model's predictions. Any changes to the law of motion governing x_t will change the function that describes how the policy w_t depends on both current and past values of x_t. This interdependence reflects the idea that investors form their expectations considering all relevant information up to time t, including the longevity of the signals.[9]

Example 5. The expressions for the aim portfolio can be reduced further by imposing a diagonal structure on Φ, meaning that each signal's decay only depends on itself. Specifically, let us consider $\Phi = \mathsf{Diag}(\phi \otimes \mathbf{1}_N)$, where $\phi \coloneqq (\phi_1, \ldots, \phi_m)$, with the condition that $|\phi_k| < 1$ holds true for all $k = 1, \ldots, m$. The parameter ϕ_k represents the speed of reversion of the kth signal, which has a half-life given by $\log(0.5)/\log(1-\phi_k)$.[10] The aim portfolio has the following form

$$\mathsf{aim}_t = A' \left(\frac{x_{t,1}}{1-(1-\theta_2)^{-1}\phi_1}, \cdots, \frac{x_{t,M}}{1-(1-\theta_2)^{-1}\phi_m} \right). \tag{5.18}$$

We see that the aim portfolio assigns less weight to each return-predicting signal. As a result, the size of the position is reduced and the relative significance of each signal has shifted towards more permanent ones. This result is intuitive as the portfolio responds more to permanent components of the signals than to the less-permanent ones.

5.3 Modeling Extensions

5.3.1 Portfolio Growth

In certain asset classes, such as equities, it is common to model the returns of a security rather than its price changes. To accommodate for this practice, we suppose that the excess returns of the risky assets from period $t-1$ to t is (with a slight abuse of notation) given by $r_t = p_t/p_{t-1} - (1+r^{\mathrm{f}})\mathbf{1}_N$. In this case, the investment problem is formulated in terms of dollar holdings and returns. In particular, the excess change in the investor's wealth is given by $\Delta W_{t+1} - r^{\mathrm{f}} W_t = h_t' r_{t+1}$, where $h_t \coloneqq W_t \tilde{h}_t$ represents the dollar holdings and $\tilde{h}_t$ denotes the fraction of the capital invested in each asset. As a result, the investor's wealth grows proportionally with the portfolio returns.

Given the dollar trade vector denoted as g_t, the investor's dollar holdings has the following

[9] In Hansen and Sargent (1980), the modeling emphasis is geared towards econometric identification in relation to the Lucas (1976) critique, applied to settings involving aggregate data at the level of a 'representative' agent (e.g., a firm or a household). The optimal decision rule for an agent is derived as a means for an econometrician to test the overidentifying cross-equation restrictions imposed by the rational expectations hypothesis in order judge if the theoretical model is consistent with observed data. While the primary focus of that study is to obtain accurate estimates of the structural parameters of the decision rule for a representative agent to conduct policy analysis, our focus is on the optimal decision rule itself.

[10] The half-life represents the duration over which half of the signal is anticipated to decay; the smaller the half-life, the more rapidly the signal decays.

dynamics:

$$h_t = \eta_t h_{t-1} + g_t, \tag{5.19}$$

where $\eta_t \coloneqq \mathsf{Diag}((1+r^{\mathrm{f}})\mathbf{1}_N + r_t)$ is a diagonal matrix whose entries contain the gross rate of return on each asset. The dollar holdings at time t can be attributed to both the dollar trade and an inherited grown position. The latter refers to the situation where if we invest \$100 in an asset, which then realizes a 10% return, our dollar value position would grow to \$110, yielding a higher value than the original \$100 investment. Thus, even if the number of shares allocated to each asset is held the same during the period $t-1$ and t, the dollar holdings may grow to new values, purely due to the price movements of the assets.

Based on this setup, we can write the investor's intertemporal optimization problem as

$$\lim_{T\to\infty} \min_{\{h_t\}_{t=0}^{T}} \mathbb{E}_0 \left[\sum_{t=0}^{T} \frac{(1-\rho)^{t+1}}{2} (h_t - h_t^*)' \gamma \Sigma_r (h_t - h_t^*) + \frac{(1-\rho)^t}{2} g_t' \Lambda g_t \right], \tag{5.20}$$

where h_{-1} is given. The tracking error is computed with respect to a dollar-denominated target portfolio h_t^*.[11] The investor selects a sequence of dollar holdings $\{h_t\}_{t=0}^{\infty}$ in order to minimize this criterion. At the end of period $t-1$, the investor already holds h_{t-1} dollars and decides to trade g_t dollars to adjust the dollar holdings to $\eta_t h_{t-1} + g_t$. This adjustment generates a linear price impact, expressed in terms of a relative change in price, resulting in an associated trading cost represented as $\frac{1}{2} g_t' \Lambda g_t$.

Unfortunately, the introduction of portfolio growth into the decision dynamics gives rise to certain complications. In particular, the investor's decision rules can no longer be modeled as a time-invariant linear difference equation since the model coefficients depend on conditional expectations of future portfolio growths. Consequently, methods similar to those presented in Hansen and Sargent (1980) cannot be directly used to solve this problem. For the purpose of this paper, it is in our interest to remain within the domain of time-invariant linear stochastic difference equation so that the model is analytically tractable. We can simplify by assuming that $\eta_t \equiv \eta$ is a constant quantity, leading us to the set of optimality conditions to fulfill under this specific setting.

Proposition 10. The optimal portfolio policies $h_0, h_1, \ldots, h_{T-1}$, satisfy the following system of second-order linear expectational difference equations

$$h_t = \breve{\Psi}_1 h_t^* + \breve{\Psi}_2 \eta h_{t-1} + \breve{\Psi}_3 \mathbb{E}_t[h_{t+1}], \tag{5.21}$$

and the following transversality condition

$$\lim_{T\to\infty} \bar{\rho}^T \mathbb{E}_0[\bar{\rho}\gamma\Sigma_r(h_T - h_T^*) + \Lambda(h_T - \eta h_{T-1})] = 0. \tag{5.22}$$

[11] Our setup differs from Jensen et al. (2023) in that they focus on portfolio weights or fractional holdings as the decision variable and model the investor's wealth growth at some stochastic rate.

The matrix coefficients are given by

$$\begin{aligned}\breve{\Psi}_1 &:= \bar{\rho}[\bar{\rho}\Omega + \mathbb{I}_N + \bar{\rho}\eta]^{-1}\Omega \\ \breve{\Psi}_2 &:= [\bar{\rho}\Omega + \mathbb{I}_N + \bar{\rho}\eta]^{-1} \\ \breve{\Psi}_3 &:= \bar{\rho}[\bar{\rho}\Omega + \mathbb{I}_N + \bar{\rho}\eta]^{-1},\end{aligned} \tag{5.23}$$

As a practical matter, we can further simplify the analysis by not taking a stance on the conditional expected returns and only considering the realization of portfolio growth at time t. That is, we can assume that expected future portfolio growths due to the risky and risk-free assets are negligible.[12] This enables us to derive an explicit expression for the optimal portfolio policy within this setting.

Proposition 11. Suppose we replace the portfolio growth process in Equation (5.19) with its truncated counterpart $\tilde{\eta}_u := \mathsf{Diag}((1+r^{\mathrm{f}})\mathbf{1}_N + r_u\chi_{\{u<t+1\}})$, where $\chi_{\{\cdot\}}$ is an indicator function that returns a value of 1 when the condition inside the braces is satisfied and 0 otherwise. The optimal portfolio policy is given by

$$h_t = \Theta_1\tilde{\eta}_t h_{t-1} + (\mathbb{I}_N - \Theta_1)\mathsf{aim}_t, \tag{5.24}$$

where

$$\mathsf{aim}_t = (\mathbb{I}_N - \Theta_1)^{-1}\sum_{k=0}^{\infty}\Theta_2^{-(k+1)}\Omega\mathbb{E}_t[h^*_{t+k}]. \tag{5.25}$$

The matrices of coefficients Θ_1 and Θ_2 are precisely those provided in Proposition 8.

We see that the optimal portfolio displays a structure similar to the one derived using price changes and share quantities but with two key distinctions: it is expressed in terms of empirically relevant units, and it adjusts from its previous realized growth position. It is reasonable for us to suggest that the optimal portfolio mirrors the previous structure, especially when considering short time horizons. However, this assumption may not be appropriate for entities such as family offices that typically have a long-term investment horizon and are often concerned with wealth preservation and risk mitigation.

5.4 Impulse Response Functions

5.4.1 Temporary Price Impact

It would be interesting to investigate the short-term and long-term effects of an exogenous shock on the dynamics of the signal with respect to the optimal portfolio policy. With knowledge of the policy rule for our investor, we can generate the model's dynamics following such a shock through the use of impulse response functions (IRFs), which are commonly employed in many dynamic economic analyses. This analysis will enable us to gain a better understanding of the optimal portfolio dynamics by examining the duration and magnitude of the responses.

Let us consider two assets where the first is exposed to a temporary signal and the second to a permanent one. This relationship implies that the policy matrix A is a 2×2 diagonal

[12] This assumption is reasonable when the conditional expected returns from time $t+1$ onwards are either small or difficult to reliably forecast.

matrix, which maps both signals to the corresponding positions on the assets by rescaling the first and second signals by 50 and 10 times their original values, respectively. The temporary signal exhibits decay with a half-life of 3 days, and the permanent signal decays with a half-life of 205 days.

Suppose that the investor initially holds an amount of one million shares in both assets. Additionally, the investor maintains an annualized discount rate of 2% and an absolute-risk aversion parameter of $\gamma = 10^{-8}$, which corresponds to a large investor who manages \$100 million in assets with a relative-risk aversion value of one. We set the transaction cost parameter to be $\lambda = 20$ bps per 1% of ADV, where ADV is the average daily volume assumed to be \$10 million. Trading occurs on a daily basis. If the annualized volatility of each asset is 20%, the corresponding daily volatility is $0.2/\sqrt{252} = 0.012$. To obtain its absolute value, we rescale this value by the price, resulting in a value of 1.2. A summary of all the model parameters based on this experimental setup described in Example 5 is provided in the third column of Table B.4.1.

In Figure C.5.1a, we illustrate the impulse response functions of our portfolio policy to a one-unit shock in the signal at time $t = 0$. We also superimpose the figure with the portfolio trajectories followed by a myopic investor who takes transaction costs into account. In both panels, we observe that our portfolio reacts more to the shocks in the short term compared to the myopic portfolio with transaction cost since it internalizes the future trajectories of the signals. Our portfolio achieves a higher peak position more rapidly than the myopic portfolio and subsequently reduces its position at a faster rate as well. When contrasting both panels, it can be observed that our optimal portfolio quickly reduces its position for the asset with the temporary signal compared to the permanent signal. The reason is that our portfolio attempts to weigh both signals based on their strengths and durations as well as their accruing investment benefits in relation to risk and trading costs.

We can also allow the policy matrix A to have off-diagonal entries, which represent the additional exposure or cross-predictability that the first asset has to the second signal and the second asset has to the first signal. These interdependencies are assigned exposure values of -5 and -1, respectively. Figure C.5.1b shows the updated portfolio responses. We observe that the optimal response for Asset 1 initially follows a similar pattern, but gradually transitions into negative territory. The reason is due to the increased emphasis placed on the permanent signal, to which it has a negative exposure. Further, we see that the qualitative shape of the response for Asset 2 is similar albeit in a smaller magnitude due to its negative exposure to temporary signal.

5.4.2 Temporary and Permanent Price Impact

We now examine the effect of temporary and permanent price impacts on the optimal portfolio. The parameters that we use in this study are summarized in the fourth column of Table B.4.1, which are also based on the experimental setup described in Example 5. These parameter values serve as the default configuration for our experiments, with the exception of R and q, which we vary in the numerical experiments.

In Figure **??**, it can be observed that low price resiliency ($R = 0.1$) is associated with a larger initial position for Asset 1 and a more gradual decline thereafter. This can be attributed to the fact that when prices are not expected to quickly revert to their pre-trade levels, the costs involved in entering and exiting a position shortly after are relatively smaller. As a result,

investors tend to capitalize on available investment opportunities by minimizing the risk of price movements caused by adverse shifts or sustained shifts resulting from trading. They achieve this by establishing their target position early on and bearing the upfront costs, thus avoiding permanent transaction costs associated with multiple cumulative trades. However, as the level of price resiliency increases, it becomes more economically advantageous for investors to stagger their trades, thereby reducing immediate transitory costs.

Figure **??** illustrates the case where $q = 0.5$, that is when there is an equal balance of temporary and permanent impact costs. Here, we see that the profiles of the trajectories are quite similar as before except that the response with low price resilience on Asset 1 is more subdued at the initial stage and has a shorter duration. This outcome is expected as a higher proportion of transitory costs motivates investors to stagger their trades and quickly reduce their exposure to the temporary signal.

5.5 Performance in Steady-State

5.5.1 Turnover

We can derive the long-run (steady-state) turnover of a portfolio of risky assets with one accompanying signal, $m = 1$, and only temporary price impact. The long-run turnover is defined as the limiting behavior of the ratio between the expected trade and the expected size of the portfolio position as time approaches infinity. This is possible if we impose the assumption of Gaussianity in both the returns and signals.

Assumption 10 (Gaussianity). The joint probability distribution of the asset returns and their associated signals is assumed to be Gaussian with zero expectations and a covariance matrix expressed in the following partitioned form:

$$\Sigma = \begin{pmatrix} \Sigma_r & \Sigma_{rx} \\ \Sigma_{rx}' & \Sigma_x \end{pmatrix}. \tag{5.26}$$

The assumption of Gaussianity is a convenient one as it allows us to make analytical progress with the following proposition.[13]

Proposition 12. Suppose that Assumptions 4–10 hold with $\Phi = \mathsf{Diag}(\phi \otimes \mathbf{1}_N)$, where $|\phi_k| < 1$ for $k = 1, \ldots, m$, and each asset has one signal. The steady-state turnover of a portfolio with temporary price impact is given by

$$\lim_{t\to\infty} \frac{\mathbb{E}[\|\Delta w_t\|_1]}{\mathbb{E}[\|w_t\|_1]} = \sqrt{\frac{2\phi(1-\theta_1)}{1+\theta_1(1-\phi)}}. \tag{5.27}$$

The steady-state turnover is increasing with respect to ϕ, while decreasing with respect to θ_1.

We see that the long-run turnover is zero when $\phi = 0$; this makes sense since if the signal follows a random walk (that is, the signal changes cannot be forecasted) then on average the future forecast is simply the current forecast. Additionally, the turnover is zero if $\theta_1 = 1$, which happens when we do not rebalance the portfolio at all at each trading date.

[13] Campbell and Viceira (2002, Chapter 2) provide a comparison between normal returns and lognormal returns. They argue that this distinction becomes less significant when considering shorter periods.

Figure C.5.2a illustrates the relationship of the portfolio's turnover to the signal's reversion speed, as well as the investor's liquidity-adjusted risk aversion. A lower value of ω leads to a higher θ_1 and hence lower turnover. This makes sense since if the portfolio size increases relative to the level of transaction costs, the investor's ability to capitalize on investment strategies with potentially high price impact diminishes. Therefore, the optimal strategy suggests reducing the size of trades to minimize the impact of transaction costs, thereby reducing the turnover.

To provide a concrete illustration, we present a numerical example. Using the scalar parameter values for ρ, λ, and γ from Table B.4.1, we find that $\theta_1 = 0.84$ from Equation (5.10), which corresponds to a rate of adjustment of 0.16. Additionally, if we consider a signal with a reversion speed of $\phi = 0.2$, then the steady-state turnover of the optimal portfolio is 19.4%.[14]

5.5.2 Sharpe Ratio

We can study the Sharpe ratio of the optimal portfolio, which takes transaction costs into consideration. To this end, let us consider a target portfolio given by $w_t^* = A' x_t$ where

$$A = \frac{1}{\gamma} \Sigma_x^{-1} \Sigma_{rx}' \Sigma_r^{-1}. \tag{5.28}$$

The policy matrix A was derived in Firoozye et al. (2023) for investors with mean-variance preferences and Gaussian distributed returns and signals. To simplify our analysis, we make the following assumption.

Assumption 11. The signals x_u and returns r_{t+1} are uncorrelated for $u < t$.

In other words, the signals are not informative of the future returns beyond their own time period. We are now ready to derive the Sharpe ratio of our dynamic portfolio policy at steady-state.

Proposition 13. Suppose that Assumptions 4–11 hold with $\Phi = \mathsf{Diag}(\phi \otimes \mathbf{1}_N)$, where $|\phi_j| < 1$ for $j = 1, \ldots, m$, and each asset has one signal. Also, consider the policy matrix A as defined in (5.28). The steady-state Sharpe ratio of the optimal portfolio with temporary price impact is given by

$$\lim_{t\to\infty} \frac{\mathbb{E}[w_t' r_{t+1} - \frac{1}{2}\Delta w_t' \Lambda \Delta w_t]}{\sqrt{\mathbb{E}[w_t' \Sigma_r w_t]}} = \left[1 - \frac{1}{\omega} \frac{\phi(1-\theta_1)^2}{(1-(1-\theta_2)^{-1}\phi)(1+\theta_1(1-\phi))}\right] \cdot \mathsf{SR} \tag{5.29}$$

$$\text{with } \mathsf{SR} \coloneqq \sqrt{\sum_{i=1}^{N} s_i^2}, \tag{5.30}$$

where s_i is the ith singular value of the adjusted cross-covariance matrix denoted by $\Sigma_r^{-1/2} \Sigma_{rx} \Sigma_x^{-1/2}$.

In Firoozye et al. (2023, Corollary 1), it was demonstrated that the square root of the sum of the squared singular values in Equation (5.30) corresponds to the Sharpe ratio of w_t^* in the

[14] Note that this value is a dimensionless quantity. If an investor trades Δw_t shares over a time period of $\Delta t = 1$ day, this means that the steady-state turnover rate becomes 19.4% per day. Thus, our proposed formula holds for any trading frequency.

absence of transaction costs. The singular values are also known as canonical correlations, which arise from applying canonical correlation analysis (CCA) to the paired dataset of asset returns r_{t+1} and signals x_t. Further, if we suppose that $s_i = s$ for all i, then the Sharpe ratio can be further simplified to $\mathsf{SR} = s\sqrt{N}$. This particular case corresponds to the Fundamental Law of Active Management (FLAM) introduced by Grinold (1989), where the information coefficient is represented by s, and the number of statistically independent investments (also known as breadth) is represented by N.

However, our previous results were based on dynamic portfolios that were not aware of transaction costs. By extending our analysis to Proposition 13, our result generalizes that of the FLAM and provides a more conservative estimate of the portfolio's Sharpe ratio due to the presence of transaction costs and signal persistence. In particular, the Sharpe ratio (SR) of the optimal portfolio policy in Equation (5.30) is adjusted by a multiplicative factor of less than one, which deflates its original value. This adjustment factor equals one when the signals are highly permanent ($\phi = 0$) or there are no portfolio adjustments ($\theta_1 = 1$). In such cases, the Sharpe ratio remains unaffected as there are no transaction costs.

Figure C.5.2b illustrates the relationship between the Sharpe ratio and the signal's reversion speed, as well as the investor's liquidity-adjusted risk aversion. A higher mean-reversion speed leads to a greater downward adjustment in the Sharpe ratio due to the higher turnover of such signals. Interestingly, a lower value of ω leads to a higher θ_1, and hence a smaller downward adjustment. In this case, the optimal strategy suggests reducing the size of trades to minimize the impact of transaction costs associated with a low ω. Consequently, the cost adjustment factor needed to correct the Sharpe ratio is reduced.

As a numerical example, suppose that the number of assets is $N = 1{,}000$ and the canonical correlations (or information coefficients) between the returns and signals have a common value of 0.02. Then the expected Sharpe ratio of the optimal portfolio in the absence of transaction costs is $0.02 \times \sqrt{1000} = 0.63$. Using the scalar parameter values for ρ, γ, and λ from Table B.4.1 with $\phi = 0.2$, the multiplicative factor to correct the Sharpe ratio is 0.95, yielding an adjusted Sharpe ratio of 0.6.

5.6 Conclusion

In this paper, we investigate a multiperiod portfolio selection problem formulated as a tracking error problem with transaction costs, which is designed to track the performance of a benchmark index or a target portfolio. To address this problem, we formulate the problem as a Hansen and Sargent (1980) dynamic linear rational expectations model. This model can be effectively solved through two steps: (1) factorize the characteristic polynomial associated with the non-stochastic component of the problem to derive the feedback and feedforward portions of the investment decision rule; (2) employ the Wiener-Kolmogorov formula to express the feedforward part based on the information available at the investment decision date. Our model is built upon a foundation that is grounded in the optimization behavior of forward-looking investors. By adopting this approach, we provide researchers with a transparent understanding of how portfolio positions are established and offer them a means to investigate the transmission mechanisms through which shocks or changes in signals affect the investment process.

With our analytical solution, we demonstrate that the optimal portfolio policy smoothly adjusts towards a forward-looking portfolio expressed as a composite sum of current and future

exogenously given target portfolios. The rate at which the optimal portfolio adjusts accounts for the investor's time, risk, and liquidity preferences. The target portfolio can take any form based on analysts' estimates. In this study, we consider target portfolios parameterized linearly in the signal vector, with particular attention given to those introduced in Firoozye et al. (2023), as it allows us to construct an empirical proxy for the forward-looking portfolio by specifying a forecasting model for the signals with different persistence properties.

We also derived analytical expressions for the performance of the optimal portfolio in the steady state. The steady-state regime of the optimal portfolio is particularly relevant for active management, as it captures an aspect of the investment process that portfolio managers eventually experience. Furthermore, it is independent of its initial conditions and its expected performance can be predicted. Our analytical predictions for the long-run behavior of the turnover and Sharpe ratio (expressed in terms of the structural parameters of the model) can serve as a convenient tool for portfolio managers and practitioners, enabling them to perform back-of-the-envelope calculations to determine the scalability and ex-ante profitability of their trading strategy.

Part IV

Conclusion

Chapter 6

Concluding Remarks

> The difficulty lies not so much in developing new ideas as in escaping from old ones.
>
> —John M. Keynes, *The General Theory of Employment, Interest and Money* (1936)

This thesis attempts to contribute to the development of Modern Portfolio Theory by introducing innovative methods for estimating risk and constructing dynamic investment portfolios by building on established works from the financial economics and econometrics literature while also remaining faithful to its core tenets. While this thesis has made some noteworthy contributions to the field, it is important to acknowledge that there are several intriguing questions and research directions that could not be addressed within its scope but are nonetheless worth exploring.

Chapter 2 provides a novel nonparametric solution to the estimation of a covariance of stock returns that is measured with an exponentially weighted scheme by means of cross-validation. Future research endeavors should focus on establishing the theoretical properties of the proposed estimator, such as deriving its asymptotic properties, to further bolster its reliability. A valuable point of reference is Abadir et al. (2014, Propositions 3 and 4), where the authors demonstrate that the precision criteria for their estimator are asymptotically optimized. Additionally, adapting the arguments presented in Abadir et al. (2014, Proposition 1) and Lam (2016, Corollary 4) to the estimator proposed in this chapter would help rigorously demonstrate the conditions under which it is a positive-definite matrix.

The estimator introduced in Chapter 2 exhibits rotation-equivariance, thereby neglecting certain empirical cross-sectional structures inherent in financial returns. In an effort to overcome this limitation, Chapter 3 attempts to build upon this innovation by leveraging the hierarchically nested factor structure often found in stock returns. This chapter demonstrates that there are gains to performance when integrated with the previous method. While hierarchical clustering provides an elegant means to capture co-movements relying solely on stock returns data, further improvements can be achieved by incorporating firm-level characteristics to better capture priced or unpriced factors. Moreover, it is important to recognize that correlations are not constant and can vary depending on market conditions. As discussed in Ang and Chen (2002), the concept of 'asymmetric correlations' in the U.S. equity

market suggests that correlations between stocks and the overall market tend to be significantly higher during market downturns compared to market upswings. The implication for portfolio selection and risk management is that if stocks were to decline together during market downturns, the benefits of diversification may be overestimated by those who overlook the increase in downside correlations. Thus, accounting for this phenomenon can potentially refine the estimates of risk.

Chapter 4 presents a solution to the single-period dynamic portfolio selection problem that takes into account cross-predictability between asset returns and their return-predictive signals, as well as their respective covariances. To gain a deeper understanding of the joint correlation structure of the solution, Canonical Correlation Analysis (CCA) is utilized. However, standard CCA relies on the assumptions of a multivariate normal distribution and linear relationships between the variable sets. To address these limitations, we can modify the framework to accommodate nonlinear relationships. This can be accomplished by applying nonlinear transformations to certain original variables prior to the implementation of the CCA algorithm, thereby allowing CCA to capture more complex associations between the variables, and deliver better out-of-sample properties.

Further extensions to this modeling framework may involve incorporating *state-variable hedging* mechanisms, which consider how changing economic conditions or state variables affect asset prices (Cochrane, 2022). It is known that many hedge funds, active managers, and financial institutions heavily rely on mean-variance optimizers with strategies based on the belief that expected returns, variances, and covariances change significantly over time. However, they frequently neglect state-variable risk in practice. To illustrate the importance of addressing this oversight, consider a stock investor during the 2008 financial crisis using traditional mean-variance analysis. Such an analysis would advise selling stocks when volatilities surge and prices fall. This scenario underscores the limitations of single-period mean-variance analysis—which assumes constant volatility—in assessing *long-term* investment goals. Indeed, this analysis fails to consider the long-term perspective and the potential for increased expected returns following price declines. Thus, it is clear that long-term investors require state-variable hedging strategies that can account for changes in volatility and other relevant factors due to fluctuations in state variables.

Chapter 5 attempts to extend the ideas presented in Chapter 4 to include market frictions such as transaction costs. In particular, the transient portion of transaction costs in my multi-period framework is modeled as quadratic adjustment cost, which corresponds to a linear price impact. However, it is well-documented that price impact exhibits concavity for large trade volumes (Hasbrouck, 1991). Thus, the quadratic cost specification may overestimate the true costs of executing trades in those situations. Consequently, investors may adopt overly conservative trading strategies to mitigate the perceived high costs, which can lead to missed investment opportunities and suboptimal portfolio performance. Although obtaining an exact closed-form solution within this context is unlikely, it may be possible to derive an approximate analytical solution that can offer valuable insights into the behavior of the optimal portfolio policy.

Moving on to the broader context, Chapter 5 studies an investment problem that belongs to a class of dynamic rational expectations models that exhibits a property known as *certainty equivalence*. This concept can be attributed to Simon (1956) and Theil (1957). Essentially, in environments where certainty equivalence holds, economic agents can treat uncertain future

outcomes as if they were certain of achieving its average value when making decisions. The immediate implication of certainty equivalence for investors is that their decision-making relies solely on the average values of exogenous shocks, while disregarding additional information contained in the distribution of these shocks, such as higher-order moments. Such environments are conducive for economic analysis because they facilitate the computation of decision rules analytically. Looking ahead, an intriguing area for further research involves examining how deviations from certainty equivalence could impact economic analysis. This exploration would necessitate moving beyond the conventional linear-quadratic framework; for example, delving into other utility functions that could introduce additional sources of saving or hedging motives due to changes in the curvature of the investor's marginal utility.

In closing, the problem of selecting a portfolio is a long-standing one in finance. Over the years, various approaches have been introduced to tackle this problem, including those contributions presented in this thesis. Despite these efforts, a clear consensus around a single econometric or statistical methodology has yet to be established. Each approach has its own merits and shortcomings, rendering them suitable for specific situations while less ideal for others. Therefore, forging a methodology that demonstrates empirical validity, scalability, comprehensibility, along with sound theoretical and economic underpinnings, remains a key challenge in this research endeavor.

The foundational contributions of Harry Markowitz serve as a potent reminder that constructing an investment portfolio is not solely a quantitative or engineering exercise but a delicate balance of "*statistical techniques and the judgment of practical men*" (Markowitz, 1952, page 91). This insightful perspective underscores that the practice of portfolio selection is both a science and an art. In other words, while econometric or statistical techniques offer a scientific means to analyzing and optimizing portfolios based on historical data, they also require the subjective assessments of researchers, investors, and advisers who draw upon their experience, intuition, and expertise to make informed investment decisions. This highlights the essential human element in the art of portfolio selection.

Appendix A

Mathematical Derivations

A.1 Derivations for Chapter 4

Proof of Proposition 1. We provide the proof of this proposition in two parts.
Part 1. We drop the time subscript t for brevity. From the cyclic property of the trace operator and linearity of the expectation, the expected value of the portfolio returns is

$$\mathbb{E}[x'Ar] = \mathbb{E}[\mathsf{Tr}(Arx')] = \mathsf{Tr}(A\Sigma_{rx}). \tag{A.1}$$

In order to analyze the variance of the portfolio returns, we appeal to the following one-dimensional from Isserlis (1918) or Wick (1950), which expresses the higher moments of centered multivariate Gaussian vectors in terms of its second-moments.

Theorem 1. Let z_1, z_2, z_3, and z_4 be jointly Gaussian random variables with mean zero. Then we have the following results:

$$\begin{aligned}
\mathbb{E}[z_1] &= 0 \\
\mathbb{E}[z_1 z_2] &= \mathsf{Cov}(z_1, z_2) \\
\mathbb{E}[z_1 z_2 z_3] &= 0 \\
\mathbb{E}[z_1 z_2 z_3 z_4] &= \mathbb{E}[z_1 z_2]\mathbb{E}[z_3 z_4] + \mathbb{E}[z_1 z_3]\mathbb{E}[z_2 z_4] + \mathbb{E}[z_1 z_4]\mathbb{E}[z_2 z_3].
\end{aligned} \tag{A.2}$$

The last equation of (A.2) takes all partitions of size two of the four variables, which gives us three separate terms. By recasting fourth-order terms using the covariance terms, we have

$$\begin{aligned}
\mathsf{Var}[x'Ar] &= \mathbb{E}[(x'Ar)^2] - \mathbb{E}[x'Ar]^2 \\
&= \sum_{i,j,k,l} \mathbb{E}[A_{i,j}A_{k,l}x_i r_j x_k r_l] - \sum_{i,j,k,l} \mathbb{E}[A_{i,j}x_i r_j]\mathbb{E}[A_{k,l}x_k r_l] \\
&= \sum_{i,j,k,l} A_{i,j}A_{k,l}\mathbb{E}[x_i r_j]\mathbb{E}[x_k r_l] + \sum_{i,j,k,l} A_{i,j}A_{k,l}\mathbb{E}[x_i x_k]\mathbb{E}[r_j r_l] \\
&\quad + \sum_{i,j,k,l} A_{i,j}A_{k,l}\mathbb{E}[x_i r_l]\mathbb{E}[x_k r_j] - \sum_{i,j,k,l} A_{i,j}A_{k,l}\mathbb{E}[x_i r_j]\mathbb{E}[x_k r_l] \\
&= \sum_{i,j,k,l} A_{i,j}A_{k,l}\mathbb{E}[x_i x_k]\mathbb{E}[r_j r_l] + \sum_{i,j,k,l} A_{i,j}A_{k,l}\mathbb{E}[x_i r_l]\mathbb{E}[x_k r_j] \\
&= \sum_{i,j,k,l} \mathbb{E}[x_k x_i]A_{i,j}\mathbb{E}[r_j r_l]A_{k,l} + \sum_{i,j,k,l} \mathbb{E}[r_l x_i]A_{i,j}\mathbb{E}[r_j x_k]A_{k,l},
\end{aligned} \tag{A.3}$$

where $A_{i,j}$ is the (i,j)-th entry of A. Reverting back to matrix notation, we have

$$\mathsf{Var}[x'Ar] = \mathsf{Tr}(\Sigma_x A \Sigma_r A') + \mathsf{Tr}(\Sigma_{rx} A \Sigma_{rx} A). \tag{A.4}$$

Part 2. In order to solve the full mean-variance objective function (4.4), we use the tools from CCA that we have laid the ground in Section 4.3.1. To assist in our derivation, we expand our notation and let Q_r and Q_x be matrices whose columns contain the so-called ith canonical directions $q_{r,i} \coloneqq \Sigma_r^{-1/2} u_i$ and $q_{x,i} \coloneqq \Sigma_x^{-1/2} v_i$, respectively, and $D \coloneqq \mathsf{Diag}(s_1, s_2, \ldots, s_N)$ be a diagonal matrix containing the canonical correlations. We now state the main theorem:

Theorem 2. Suppose r_{t+1} and x_t are jointly Gaussian random variables. Then the objective function (4.2) is maximized at

$$A = Q_x \mathsf{Diag}(L_1, \ldots, L_N) Q_r', \tag{A.5}$$

where diagonal matrix consist of the following elements

$$L_i = \frac{1}{\gamma} \frac{s_i}{1 + s_i^2}, \tag{A.6}$$

for $i = 1, \ldots, N$.

Without loss of generality, let us reparameterize the matrix of coefficients in terms of the canonical directions as $A = Q_x L Q_r'$ where L is a variable matrix that we now have to optimize on. The expected value of the portfolio returns at time t can be written as

$$\begin{aligned} \mathbb{E}[x_t' A r_{t+1}] &= \mathsf{Tr}(A \Sigma_{rx}) && \text{(A.7)} \\ &= \mathsf{Tr}(LD), && \text{(A.8)} \end{aligned}$$

and the variance as

$$\begin{aligned} \mathsf{Var}[x_t' A r_{t+1}] &= \mathsf{Tr}(\Sigma_x A \Sigma_r A') + \mathsf{Tr}(\Sigma_{rx} A \Sigma_{rx} A) && \text{(A.9)} \\ &= \mathsf{Tr}(LL') + \mathsf{Tr}(DLDL). && \text{(A.10)} \end{aligned}$$

Putting all together, the objective function (4.4) can be written as:

$$\max_L \mathsf{Tr}(LD) - \frac{\gamma}{2}(\mathsf{Tr}(LL') - \mathsf{Tr}(DLDL)). \tag{A.11}$$

Using the rules for matrix derivatives from Lütkepohl (1997), we take first-order conditions with respect to the matrix L to get

$$D - \gamma(L + DL'D) = 0. \tag{A.12}$$

Since the matrix D is diagonal and thus symmetric, we necessarily have the following relationship

$$\frac{1}{\gamma} D = L + DL'D = L' + DLD. \tag{A.13}$$

Let $L^a \coloneqq L - L'$, which is an anti-symmetric matrix (that is, $L^{a'} = -L^a$) and has diagonal

elements zero. Then we have

$$L^a = DL^aD. \tag{A.14}$$

If we restrict our attention to the off-diagonal elements of (A.14), we see that $L^A_{i,j} = s_i s_j L^A_{i,j}$ for all $i \neq j$, so $(1 - s_i s_j)L^A_{i,j} = 0$. Therefore, either $s_i s_j = 1$ for all $i \neq j$ or $L^A_{i,j} = 0$. Consequently, L^a is identically zero and hence, L must be a symmetric matrix.

The benefit of variable matrix L being symmetric is that the condition (A.12) can now be written as

$$L + DLD = \frac{1}{\gamma}D. \tag{A.15}$$

If we focus on off-diagonal elements, we see that $(1 + s_i s_j)L_{i,j} = 0$ for all $i \neq j$. But we also know that the canonical correlations satisfy the following ordering $0 \leq s_1 \leq \ldots \leq s_N \leq 1$, and so it is impossible that $s_i s_j = -1$. Thus, $L_{i,j} = 0$ for all $i \neq j$, and so L must be a diagonal matrix. Thus, optimizing over the elements $L_{ii} \equiv L_i$ for $i = 1, \ldots, N$, boils down to maximizing the following univariate problems

$$\max_{L_1,\ldots,L_N} \sum_{i=1}^{N} L_i s_i - \frac{\gamma}{2}(L_i^2 + s_i^2 L_i^2). \tag{A.16}$$

Our problem can now be easily solved to give us

$$L_i = \frac{1}{\gamma}\frac{s_i}{1 + s_i^2}. \tag{A.17}$$

Hence, the diagonal elements are a nonlinear function of the canonical correlations. The optimal matrix A in (A.5) is effectively the optimal scaling of the canonical portfolios of the asset returns and signals. □

Proof of Proposition 2. Applying a change-of-variables $\tilde{A} \coloneqq \Sigma_x^{1/2} A \Sigma_r^{1/2}$, we have

$$\max_{\tilde{A}} \mathsf{Tr}(\tilde{A}\Sigma'_{\tilde{r}\tilde{x}}) - \frac{\gamma}{2}\mathsf{Tr}(\tilde{A}\tilde{A}'). \tag{A.18}$$

The problem (A.18) can be solved to yield

$$\tilde{A} = \frac{1}{\gamma}\Sigma_x^{-1/2}\Sigma'_{rx}\Sigma_r^{-1/2}. \tag{A.19}$$

Since the solution is expressed in a different basis, we rescale it back to the original assets through the following operation

$$A = \Sigma_x^{-1/2}\tilde{A}\Sigma_r^{-1/2} = \frac{1}{\gamma}\Sigma_x^{-1}\Sigma'_{rx}\Sigma_r^{-1}. \tag{A.20}$$

As a result of this approximation, we now have

$$L_i \approx \frac{1}{\gamma}s_i. \tag{A.21}$$

Figure C.4.2 contrasts the optimal singular value adjustment in (A.6) to the linear

approximation (A.21) for $\gamma = 1$. We can see that for small values of s_i, the optimal adjustment is approximately linear while for large values of s_i, the optimal adjustment deflates the values of the canonical correlations. □

Proof of Proposition 3. We start by writing down the expression for the Lagrangian

$$\mathsf{Tr}(A\Sigma_{rx}) - \frac{\gamma}{2}\mathsf{Tr}(\Sigma_x A \Sigma_r A') + \lambda(1 - \mathbf{1}'A'x_t), \tag{A.22}$$

$\lambda > 0$ is the Lagrange multiplier. Using the change-of-variables $\tilde{A} \coloneqq \Sigma_x^{1/2} A \Sigma_r^{1/2}$, we have

$$\mathsf{Tr}(\tilde{A}\Sigma_{\tilde{r}\tilde{x}}) - \frac{\gamma}{2}\mathsf{Tr}(\tilde{A}\tilde{A}') + \lambda(1 - \mathbf{1}'\Sigma_r^{-1/2}\tilde{A}'\Sigma_x^{-1/2}x_t). \tag{A.23}$$

Performing the first-order conditions

$$[\text{FOC } \tilde{A}]: \ \Sigma'_{\tilde{r}\tilde{x}} - \gamma\tilde{A} - \lambda\Sigma_x^{-1/2}x_t\mathbf{1}'\Sigma_r^{-1/2} = 0, \tag{A.24}$$

$$[\text{FOC } \lambda]: \ 1 - \mathbf{1}'w_t = 0. \tag{A.25}$$

By reverting the change-of-variables we have made and using the fact that the portfolio policies are linear the signals, the portfolio weights is then

$$w_t^{\text{FI}} = \frac{1}{\gamma}\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}x_t + \lambda(x_t'\Sigma_x^{-1}x_t)\mathbf{1}'\Sigma_r^{-1}. \tag{A.26}$$

Solving for λ gives

$$\lambda = \frac{1-\kappa}{(x_t'\Sigma_x^{-1}x_t)(\mathbf{1}'\Sigma_r^{-1}\mathbf{1})}, \tag{A.27}$$

where $\kappa \coloneqq \gamma^{-1}\mathbf{1}'\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}x_t$. Inserting the λ into (A.26) and rearranging the expression gives us the optimal portfolio (4.19). □

Proof of Proposition 4. If r_{t+1} and x_t are zero-mean Gaussian random variables, then their projections on to the ith canonical portfolios given by $u_i'\tilde{r}_{t+1}$ and $v_i'\tilde{x}_t$ are also Gaussian random variables with mean zero and unit variances and a joint correlation s_i. For the expected return of the ith canonical portfolio, we have

$$\mathbb{E}[\pi_{t,i}] = s_i v_i'\Sigma'_{\tilde{r}\tilde{x}}u_i = s_i^2. \tag{A.28}$$

Furthermore, the variance of the ith canonical portfolio is

$$\mathsf{Var}[\pi_{t,i}] = \mathbb{E}[s_i^2(v_i'\tilde{x}_t)^2(u_i'\tilde{r}_{t+1})^2] - s_i^4 \tag{A.29}$$

$$= s_i^2(1 + 2s_i^2) - s_i^4 \tag{A.30}$$

$$= s_i^2(1 + s_i^2). \tag{A.31}$$

The second equality follows from the second-moment-based result of the product of two correlated Gaussian variables from Haldane (1942, Section 6). Substituting the expression of the optimal policy A into the definition of the expected value and variance of the portfolio

returns, we have:

$$\mathbb{E}[x_t' A r_{t+1}] = \frac{1}{\gamma}\mathsf{Tr}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}\Sigma_{rx}) = \frac{1}{\gamma}\sum_{i=1}^{N} s_i^2 \tag{A.32}$$

$$\mathbb{E}[x_t' A \Sigma_r A' x_t] = \frac{1}{\gamma^2}\mathsf{Tr}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}\Sigma_{rx}) = \frac{1}{\gamma^2}\sum_{i=1}^{N} s_i^2. \tag{A.33}$$

The proof concludes. □

Proof of Corollary 1. This follows immediately Proposition 4 and the definition of the Sharpe ratio. □

Proof of Proposition 5. Substituting the second-moment matrix of managed portfolio returns (4.38) into the policy matrix A from (4.7), the expected returns based on this matrix of coefficients is then

$$\mathbb{E}[x_t' A r_{t+1}] = \frac{1}{\gamma}\mathsf{Tr}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}\Sigma_{rx}) = \frac{1}{\gamma}(\mu_r'\Sigma_r^{-1}\mu_r)(\mu_x'\Sigma_x^{-1}\mu_x) + \frac{1}{\gamma}\sum_{i=1}^{N} s_i^2, \tag{A.34}$$

and the claim follows. □

A.2 Derivations for Chapter 5

A.2.1 Proof of Proposition 6

Differentiating objective function (5.3) with respect to w_t for $t = 0, 1, \ldots, T-1$ gives[1]

$$\bar{\rho}\gamma\Sigma_r(w_t - w_t^*) + \Lambda(w_t - w_{t-1}) + \bar{\rho}\Lambda(w_t - \mathbb{E}_t[w_{t+1}]) = 0. \tag{A.35}$$

Differentiating with respect to the last term w_T and taking the limit of the resulting equation as $T \to \infty$ gives us the transversality condition:

$$\lim_{T\to\infty} \bar{\rho}^T \mathbb{E}_0[\bar{\rho}\gamma\Sigma_r(w_T - w_T^*) + \Lambda(w_T - w_{T-1})] = 0. \tag{A.36}$$

Rearranging the terms in Equation (A.35) yields

$$w_t = \Psi_1 w_t^* + \Psi_2 w_{t-1} + \Psi_3 \mathbb{E}_t[w_{t+1}], \tag{A.37}$$

where the matrix coefficients are given in Equation (5.7) where they satisfy $\Psi_1 + \Psi_2 + \Psi_3 = \mathbb{I}_N$.

A.2.2 Proof of Proposition 7

To solve the Euler equations subject to the transversality condition and the given initial portfolio w_{-1}, we appeal to the variational methods from Sargent (1987, Chapter 9). We begin by rewriting the second-order difference equation (5.5) in terms of lag and forward operators[2], denoted as L and F, respectively:

$$\mathbb{E}_t\left[\left(F^2 - \frac{1}{\psi_3}F + \frac{\psi_2}{\psi_3}\right) L w_t\right] = -\frac{\psi_1}{\psi_3} w_t^*, \tag{A.38}$$

where

$$\psi_1 := \frac{\bar{\rho}\gamma}{\bar{\rho}\gamma + \lambda + \bar{\rho}\lambda}, \quad \psi_2 := \frac{\lambda}{\bar{\rho}\gamma + \lambda + \bar{\rho}\lambda}, \quad \psi_3 := \frac{\bar{\rho}\lambda}{\bar{\rho}\gamma + \lambda + \bar{\rho}\lambda}, \tag{A.39}$$

are the scalar counterparts of the matrix coefficients from (5.7). The characteristic equation is given by

$$F^2 - \frac{1}{\psi_3}F + \frac{\psi_2}{\psi_3} = 0. \tag{A.40}$$

In terms of the forward operator F, the characteristic equation can be factorized as

$$(F - \theta_1)(F - \theta_2) = F^2 - (\theta_1 + \theta_2)F + \theta_1\theta_2. \tag{A.41}$$

[1] Although the expectation in Equation (A.35) should be taken with respect to $\mathbb{E}_0[\cdot]$, we can simplify the expression by applying the law of iterated expectations and restricting it to the information set $\mathcal{F}_t$ for $t = 1, \ldots, T$, since $\mathbb{E}_0[\mathbb{E}_t[\cdot]] = \mathbb{E}_0[\cdot]$.

[2] The lag and forward operators on any stochastic process $\{z_t\}$ are defined by $L^k z_t = z_{t-k}$, and $F^k z_t = z_{t+k}$, respectively, for some positive integer $k > 0$.

This implies the following restrictions on the roots

$$\theta_1\theta_2 = \frac{\psi_2}{\psi_3}, \qquad \text{and} \qquad \theta_1 + \theta_2 = \frac{1}{\psi_3}. \tag{A.42}$$

Since the sum and product of the roots, θ_1 and θ_2, are positive, it follows that both roots are positive. The first equality in (A.42) establishes that $\theta_1 = \bar{\rho}^{-1}\theta_2^{-1}$, so that the two roots come in $\bar{\rho}$-reciprocal pairs. Hence, $\bar{\rho}\theta_1 + \theta_1^{-1} = \bar{\rho}\psi_3^{-1}$. If we consider the convex function $\mathsf{F}(\theta) := \bar{\rho}\theta + \theta^{-1}$, it has a minimum at $\theta = \bar{\rho}^{-1/2}$. This implies that $0 < \theta_1 < \bar{\rho}^{-1/2} < \theta_2$, taking θ_1 as the smallest root without loss of generality.

The two roots of the characteristic polynomial are given by:

$$\theta_1 = \frac{1 - \sqrt{1 - 4\psi_2\psi_3}}{2\psi_3} \qquad \text{and} \qquad \theta_2 = \frac{1 + \sqrt{1 - 4\psi_2\psi_3}}{2\psi_3}. \tag{A.43}$$

The roots are positive, real, and distinct due to the condition $1 - 4\psi_2\psi_3 > 0$, which holds based on our assumptions about the signs and magnitudes of the structural parameters ρ, γ, and λ. Furthermore, we can establish the inequalities $\theta_1 < 1 < \bar{\rho}^{-1} < \theta_2$. To demonstrate this, consider the relationship $\bar{\rho}\psi_3^{-1} = \bar{\rho}\omega + 1 + \bar{\rho}$. Observing that $\bar{\rho}\psi_3^{-1} > 1 + \bar{\rho}$, we conclude that the solutions to the equation $\mathsf{F}(\theta) = \bar{\rho}\psi_3^{-1}$ must satisfy $\theta_1 < 1$ and $\bar{\rho}^{-1} < \theta_2$.[3] The constants in Equation (5.10) can be arrived by means of substituting the coefficients (A.39) into the expressions (A.43).

Having achieved this factorization, we write the difference equation from (A.38) as

$$\mathbb{E}_t\left[(F - \theta_1)(F - \theta_2)Lw_t\right] = -\frac{\psi_1}{\psi_3}w_t^*, \tag{A.44}$$

or, after premultiplying both sides by the forward inverse $(F - \theta_2)^{-1}$ to satisfy the transversality condition, which gives us

$$\mathbb{E}_t\left[(F - \theta_1)Lw_t\right] = -\frac{\psi_1}{\psi_3}\mathbb{E}_t\left[\frac{1}{F - \theta_2}w_t^*\right]. \tag{A.45}$$

The unique solution of the Euler equations that satisfies the transversality condition is given by

$$w_t = \theta_1 w_{t-1} + \frac{\psi_1}{\psi_3\theta_2}\sum_{k=0}^{\infty}\theta_2^{-k}\mathbb{E}_t[w_{t+k}^*]. \tag{A.46}$$

Since $\theta_2 > \bar{\rho}^{-1/2}$, Assumption 6 is sufficient to ensure that the geometric sums of future w_t^* on Equation (A.46) converges.

Finally, a more intuitive re-formulation of the equation can be obtained by making use of the fact that $\psi_1(\psi_3\theta_2)^{-1} = (1 - \theta_1)(1 - \theta_2^{-1})$, which comes from using the restrictions in (A.42) and the fact that the sum of coefficients is unity. By substituting this expression into (A.46), we get

$$w_t = \theta_1 w_{t-1} + (1 - \theta_1)(1 - \theta_2^{-1})\sum_{k=0}^{\infty}\theta_2^{-k}\mathbb{E}_t[w_{t+k}^*]. \tag{A.47}$$

[3] This has to hold since the roots to the equation $\mathsf{F}(\theta) = 1 + \bar{\rho}$ are such that $\theta_1 = 1$ and $\theta_2 = \bar{\rho}$.

The proof concludes.

A.2.3 Proof of Proposition 8

We now generalize the arguments presented in Proposition 7 and rewrite the difference equation in terms of the matrix coefficients provided in Equation (5.7):

$$\mathbb{E}_t\left[\left(F^2 - \Psi_3^{-1}F + \Psi_3^{-1}\Psi_2\right) Lw_t\right] = -\Psi_3^{-1}\Psi_1 w_t^*. \tag{A.48}$$

The characteristic equation is given by the following matrix quadratic equation:

$$Q_1F^2 + Q_2F + Q_3 = 0. \tag{A.49}$$

where $Q_1 \coloneqq \mathbb{I}_N$, $Q_2 \coloneqq -\Psi_3^{-1}$, and $Q_3 \coloneqq \Psi_3^{-1}\Psi_2$. Since $Q_3 = \bar{\rho}^{-1}\mathbb{I}_N$ is a non-singular matrix, there exist two matrices Θ_1 and Θ_2 that satifies the following factorization:

$$(F - \Theta_1)(F - \Theta_2) = F^2 - (\Theta_1 + \Theta_2)F + \Theta_1\Theta_2, \tag{A.50}$$

so that we require

$$\Theta_1\Theta_2 = \Psi_3^{-1}\Psi_2, \qquad \text{and} \qquad \Theta_1 + \Theta_2 = \Psi_3^{-1}. \tag{A.51}$$

Since quadratic equation (A.49) has matrix coefficients, the usual formula for the roots of a scalar quadratic equation does not apply unless the following conditions hold: (i) $Q_1 = \mathbb{I}_N$, (ii) Q_2 commute with Q_3, and (iii) the "square root" of $Q_2^2 - 4Q_3$ exists. In our setting, the first two conditions can be easily verified. The last condition can be achieved if we write

$$(\Psi_3^{-2} - 4\Psi_3^{-1}\Psi_2)^{1/2} = \Lambda^{-1/2}(\Lambda^{1/2}Q_4\Lambda^{1/2})^{1/2}\Lambda^{1/2}, \tag{A.52}$$

where $Q_4 \coloneqq (\Psi_3^{-2} - 4\Psi_3^{-1}\Psi_2)\Lambda^{-1}$ is a positive-definite symmetric matrix. Hence, we can express the square root of $Q_2^2 - 4Q_3$ in a way that involves only taking the square root of matrices that are both symmetric and positive semi-definite.

We can write the roots to the matrix quadratic equation as

$$\Theta_1 = \frac{1}{2}[\Psi_3^{-1} - (\Psi_3^{-2} - 4\Psi_3^{-1}\Psi_2)^{1/2}] \qquad \text{and} \qquad \Theta_2 = \frac{1}{2}[\Psi_3^{-1} + (\Psi_3^{-2} - 4\Psi_3^{-1}\Psi_2)^{1/2}], \tag{A.53}$$

It can be shown that the eigenvalues of Θ_1 lie inside the unit circle. To see this, note that the function $\mathsf{G}(z) \coloneqq 0.5(z - (z^2 - 4\bar{\rho}^{-1})^{1/2}) < 1$ for all $z \geq 2\bar{\rho}^{-1/2}$. Thus, it suffices to show that $\Psi_3^{-1} - 2\bar{\rho}^{-1/2}\mathbb{I}_N$ is positive definite, that is, all eigenvalues are positive. Since $\Omega = \gamma\Lambda^{-1/2}(\Lambda^{-1/2}\Sigma_r\Lambda^{-1/2})\Lambda^{1/2}$, the matrix Ω is similar to the positive definite symmetric matrix $\gamma\Lambda^{-1/2}\Sigma_r\Lambda^{-1/2}$; that is, both matrices share the same positive eigenvalues. Since both Ω and the identity matrix trivially commute and $1 + \bar{\rho}^{-1} \geq 2\bar{\rho}^{-1/2}$ for $\rho \in (0, 1)$, this implies that the eigenvalues of Ψ_3^{-1} are greater than $2\bar{\rho}^{-1/2}$. Hence, all the eigenvalues of Θ_1 are inside the unit circle. A similar argument can be made to show that the eigenvalues of Θ_2 are outside the unit circle.

We can now proceed using similar arguments from Sargent (1987) by solving the stable root backward and the unstable root forward in order to satisfy the transversality condition and

produce a saddle-path solution

$$\mathbb{E}_t\left[(F-\Theta_1)Lw_t\right] = -\mathbb{E}_t\left[(F-\Theta_2)^{-1}\Psi_3^{-1}\Psi_1 w_t^*\right]. \tag{A.54}$$

In doing so, the Euler equations have a unique solution, which can be expressed as:

$$w_t = \Theta_1 w_{t-1} + \sum_{k=0}^{\infty}\Theta_2^{-(k+1)}\Psi_3^{-1}\Psi_1\mathbb{E}_t[w_{t+k}^*] \tag{A.55}$$

$$= \Theta_1 w_{t-1} + \sum_{k=0}^{\infty}\Theta_2^{-(k+1)}\Omega\mathbb{E}_t[w_{t+k}^*], \tag{A.56}$$

since $\Psi_3^{-1}\Psi_1 = \Omega$. This expression can be easily reformulated in terms of the expression for the aim portfolio given in Equation (5.14).

A.2.4 Proof of Proposition 9

Substituting the law of motion of the signals into the expression for the aim portfolio, we have

$$\mathsf{aim}_t = (\mathbb{I}_N - \Theta_1)^{-1}\sum_{k=0}^{\infty}\Theta_2^{-(k+1)}\Omega\mathbb{E}_t[w_{t+k}^*] \tag{A.57}$$

$$= (\mathbb{I}_N - \Theta_1)^{-1}\sum_{k=0}^{\infty}\Theta_2^{-(k+1)}\Omega A'\mathbb{E}_t[x_{t+k}] \tag{A.58}$$

$$= (\mathbb{I}_N - \Theta_1)^{-1}\left[\sum_{k=0}^{\infty}\Theta_2^{-(k+1)}\Omega A'(\mathbb{I}_M-\Phi)^k\right]x_t. \tag{A.59}$$

The third equality follows from the Wiener-Kolmogorov formula for k-step ahead linear least squares predictions of first-order autoregressive processes.

Let us denote the expression in the square brackets of Equation (A.59) as

$$\Theta^{\mathsf{aim}} := \sum_{k=0}^{\infty}\Theta_2^{-(k+1)}\Omega A'(\mathbb{I}_M-\Phi)^k. \tag{A.60}$$

The infinite sum in Θ^{aim} will converge to some matrix. We can write an expression for Θ^{aim} as follows:

$$\mathsf{vec}\left(\Theta^{\mathsf{aim}}\right) = \sum_{k=0}^{\infty}((\mathbb{I}_M-\Phi')\otimes\Theta_2^{-1})^k\mathsf{vec}(\Theta_2^{-1}\Omega A') \tag{A.61}$$

$$= (\mathbb{I}_{NM} - (\mathbb{I}_M-\Phi')\otimes\Theta_2^{-1})^{-1}\mathsf{vec}(\Theta_2^{-1}\Omega A'), \tag{A.62}$$

where the function $\mathsf{vec}(\Theta^{\mathsf{aim}})$ stacks the columns of matrix Θ^{aim} on top of each other to form a vector, and the symbol $\otimes$ represents the Kronecker product of two matrices. The first equality follows from the linearity of the $\mathsf{vec}(\cdot)$ operator, its mixed-product property, and the property that $\mathsf{vec}(Z_1Z_2Z_3) = (Z_3\otimes Z_1)\mathsf{vec}(Z_2)$ for any generic conformable matrices Z_1, Z_2, and Z_3.

Finally, if Assumption 7 holds, the aim portfolio at time t in Equation (A.59) simplifies to

$$\mathsf{aim}_t = A'(1-\theta_2^{-1})\sum_{k=0}^{\infty}\theta_2^{-k}(\mathbb{I}_M-\Phi)^k x_t \tag{A.63}$$

$$= A'\left(\mathbb{I}_M-(1-\theta_2)^{-1}\Phi\right)^{-1}x_t. \tag{A.64}$$

The proof concludes.

A.2.5 Proof of Proposition 10

Apply a similar argument to Proposition 6.

A.2.6 Proof of Proposition 11

If $\eta_u = 1$ for $u > t$, then the Euler equation (5.21) reduces to

$$h_t = \Psi_1 h_t^* + \Psi_2\eta_t h_{t-1} + \Psi_3\mathbb{E}_t[h_{t+1}], \tag{A.65}$$

where the matrices of coefficients are precisely those provided in Proposition 6 (since the future portfolio growths of all securities are identically equal to one.) To solve this problem, we use the method of undetermined coefficients. To this end, let us postulate that the solution has the following form:

$$h_t = \breve{\Theta}_1\eta_t h_{t-1} + \sum_{k=0}^{\infty}\breve{\Theta}_2^{-k}\breve{\Theta}_3\mathbb{E}_t[h_{t+k}^*], \tag{A.66}$$

where $\breve{\Theta}_1$, $\breve{\Theta}_2$ and $\breve{\Theta}_3$ are matrices that we seek to obtain. Shifting this equation one step forward, we have the following

$$h_{t+1} = \breve{\Theta}_1 h_t + \sum_{k=0}^{\infty}\breve{\Theta}_2^{-k}\breve{\Theta}_3\mathbb{E}_t[h_{t+k+1}^*]. \tag{A.67}$$

Substituting Equation (A.67) into Equation (A.65) gives

$$h_t = \Psi_1 h_t^* + \Psi_2\eta_t h_{t-1} + \Psi_3\left(\breve{\Theta}_1 h_t + \sum_{k=0}^{\infty}\breve{\Theta}_2^{-k}\breve{\Theta}_3\mathbb{E}_t[h_{t+k+1}^*]\right) \tag{A.68}$$

$$(\Psi_3^{-1}-\breve{\Theta}_1)h_t = \Psi_3^{-1}\Psi_2\eta_t h_{t-1} + \Psi_3^{-1}\Psi_1 h_t^* + \sum_{k=0}^{\infty}\breve{\Theta}_2^{-k}\breve{\Theta}_3\mathbb{E}_t[h_{t+1+j}^*] \tag{A.69}$$

Re-arranging the last equation, we have

$$\begin{aligned} h_t = {} & (\Psi_3^{-1}-\breve{\Theta}_1)^{-1}\Psi_3^{-1}\Psi_2\eta_t h_{t-1} + (\Psi_3^{-1}-\breve{\Theta}_1)^{-1}\Psi_3^{-1}\Psi_1 h_t^* \\ & + (\Psi_3^{-1}-\breve{\Theta}_1)^{-1}\sum_{k=0}^{\infty}\breve{\Theta}_2^{-k}\breve{\Theta}_3\mathbb{E}_t[h_{t+1+j}^*]. \end{aligned} \tag{A.70}$$

Equating (A.66) and (A.70) together implies the following restrictions on the matrices:

$$\breve{\Theta}_1 = (\Psi_3^{-1} - \breve{\Theta}_1)^{-1}\Psi_3^{-1}\Psi_2 \tag{A.71}$$
$$\breve{\Theta}_2 = \Psi_3^{-1} - \breve{\Theta}_1 \tag{A.72}$$
$$\breve{\Theta}_3 = (\Psi_3^{-1} - \breve{\Theta}_1)^{-1}\Psi_3^{-1}\Psi_1. \tag{A.73}$$

One can verify that these restrictions are the same as those from the proof in Proposition 8.

A.2.7 Proof of Proposition 12

We consider the expectations of the following processes defined as $\zeta_1 \coloneqq \mathbb{E}[\|\mathsf{aim}_t\|_2^2]$, $\zeta_2 \coloneqq \mathbb{E}[\mathsf{aim}_t' w_t]$, and $\zeta_3 \coloneqq \mathbb{E}[\|w_t\|_2^2]$. Given our assumption on stationarity, in a steady state, these variables remain constant over time indicating that the system has achieved equilibrium and their expectations do not vary.

By rewriting the expectation of the quadratic transaction costs, we have:

$$\begin{aligned}
\mathbb{E}[\|\Delta w_t\|_2^2] &= (1-\theta_1)^2\mathbb{E}[\|\mathsf{aim}_t - w_{t-1}\|_2^2] && \text{(A.74)}\\
&= (1-\theta_1)^2\mathbb{E}[\mathsf{aim}_t'\mathsf{aim}_t - 2\mathsf{aim}_t' w_{t-1} + w_{t-1}' w_{t-1}] && \text{(A.75)}\\
&= (1-\theta_1)^2(\zeta_1 - 2(1-\phi)\zeta_2 + \zeta_3). && \text{(A.76)}
\end{aligned}$$

The third equality uses the iterate law of expectations. If we take the discrete derivative of ζ_2, we have

$$\begin{aligned}
\Delta\zeta_2 &= \mathbb{E}[\Delta(\mathsf{aim}_t' w_t)]\\
&= \mathbb{E}[\mathsf{aim}_t'\Delta w_t + w_{t-1}'\Delta\mathsf{aim}_t]\\
&= \mathbb{E}[(1-\theta_1)(\mathsf{aim}_t'\mathsf{aim}_t - \mathsf{aim}_t' w_{t-1}) - \phi w_{t-1}'\mathsf{aim}_{t-1}]\\
&= (1-\theta_1)(\zeta_1 - (1-\phi)\zeta_2) - \phi\zeta_2.
\end{aligned} \tag{A.77}$$

The third equality uses Equations (5.8). At steady-state, $\Delta\zeta_2 = 0$, and therefore ζ_2 can be solved to give

$$\zeta_1 = \left(1 - \phi + \frac{\phi}{1-\theta_1}\right)\zeta_2. \tag{A.78}$$

We take the discrete derivative of the expectation of the risk component:

$$\begin{aligned}
\Delta\zeta_3 &= \mathbb{E}[\Delta\|w_t\|_2^2]\\
&= \mathbb{E}[w_t'\Delta w_t + w_{t-1}'\Delta w_t]\\
&= (1-\theta_1)\mathbb{E}[w_t'\mathsf{aim}_t + w_{t-1}'\mathsf{aim}_t - w_{t-1}' w_t - \|w_{t-1}\|_2^2]\\
&= (1-\theta_1)[\zeta_2 + (1-\phi)\zeta_2 - \theta_1\zeta_3 - (1-\theta_1)(1-\phi)\zeta_2 - \zeta_3].
\end{aligned} \tag{A.79}$$

The third and fourth inequality use Equation (5.8) and the law of iterated conditional expectations. At steady-state, we have $\Delta\zeta_3 = 0$, and so ζ_3 can be solved to give

$$\zeta_3 = \left(1 - \frac{\phi\theta_1}{1+\theta_1}\right)\zeta_2. \tag{A.80}$$

Given that x_t follows a Gaussian distribution with an unconditional mean of zero, it follows

that w_t is similarly Gaussian distributed with an unconditional mean of zero. Geary (1935) demonstrated that the mean absolute deviation of a Gaussian distributed random variable is equal to its standard deviation multiplied by $\sqrt{2/\pi}$. Consequently, this leads to the expressions $\mathbb{E}[\|w_t\|_1]^2 = (2N^2/\pi)\mathbb{E}[\|w_t\|_2^2]$, and $\mathbb{E}[\|\Delta w_t\|_1]^2 = (2N^2/\pi)\mathbb{E}[\|\Delta w_t\|_2^2]$. We can write the expression for the steady-state turnover as

$$\lim_{t\to\infty} \frac{\mathbb{E}[\|\Delta w_t\|_1]}{\mathbb{E}[\|w_t\|_1]} = \lim_{t\to\infty} \sqrt{\frac{\mathbb{E}[\|\Delta w_t\|_2^2]}{\mathbb{E}[\|w_t\|_2^2]}} \tag{A.81}$$

$$= \sqrt{\frac{(1-\theta_1)^2(\zeta_1 - 2(1-\phi)\zeta_2 + \zeta_3)}{\zeta_3}} \tag{A.82}$$

$$= (1-\theta_1)\sqrt{\left[\left(1-\phi+\frac{\phi}{1-\theta_1}\right) - 2(1-\phi)\right]\left(1-\frac{\phi\theta_1}{1+\theta_1}\right)^{-1} + 1} \tag{A.83}$$

$$= \sqrt{\frac{2\phi(1-\theta_1)}{1-\phi\theta_1+\theta_1}}. \tag{A.84}$$

By differentiating the expression on the right-hand side of Equation (A.84) with respect to θ_1, we have

$$-\frac{\phi^{3/2}}{\sqrt{2-2\theta_1}(\theta_1(\phi-1)+1)^{3/2}}. \tag{A.85}$$

This expression can be verified to be negative for $\phi, \theta_1 \in (0,1)$. Furthermore, differentiating the right-hand side of Equation (A.84) with respect to ϕ, we have

$$\frac{(1-\theta_1)^{3/2}}{\sqrt{2\phi}(\theta_1(\phi-1)+1)^{3/2}}. \tag{A.86}$$

This expression can be verified to be positive for $\phi, \theta_1 \in (0,1)$.

A.2.8 Proof of Proposition 13

Due to Assumption 11, $\mathbb{E}[w'_{t-1}r_{t+1}] = 0$. Using Equation (5.8), we have

$$\mathbb{E}[w'_t r_{t+1}] = \theta_1 \mathbb{E}[w'_{t-1}r_{t+1}] + (1-\theta_1)\mathbb{E}[\mathsf{aim}'_t r_{t+1}] \tag{A.87}$$

$$= (1-\theta_1)\mathbb{E}[\mathsf{aim}'_t r_{t+1}]. \tag{A.88}$$

Substituting the expression for the aim portfolio (5.18) and the optimal policy matrix A in (4.7), the expected returns of the portfolio is

$$\mathbb{E}[w'_t r_{t+1}] = \frac{1-\theta_1}{1-(1-\theta_2)^{-1}\phi}\mathbb{E}[x'_t A r_{t+1}] \tag{A.89}$$

$$= \frac{1-\theta_1}{1-(1-\theta_2)^{-1}\phi}\mathsf{Tr}(A\Sigma_{rx}) \tag{A.90}$$

$$= \frac{1}{\gamma}\frac{1-\theta_1}{1-(1-\theta_2)^{-1}\phi}\mathsf{Tr}(\Sigma_x^{-1}\Sigma'_{rx}\Sigma_r^{-1}\Sigma_{rx}). \tag{A.91}$$

Moreover, the variance (or more precisely, the second moment) of the portfolio returns is given by[4]

$$\mathbb{E}[w_t'\Sigma_r w_t] = \left(\frac{1-\theta_1}{1-(1-\theta_2)^{-1}\phi}\right)^2 \mathsf{Tr}(A\Sigma_r A'\Sigma_x) \tag{A.92}$$

$$= \frac{1}{\gamma^2}\left(\frac{1-\theta_1}{1-(1-\theta_2)^{-1}\phi}\right)^2 \mathsf{Tr}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}\Sigma_{rx}). \tag{A.93}$$

Thus, the ratio of the expected returns to the standard deviation of the portfolio is

$$\frac{\mathbb{E}[w_t' r_{t+1}]}{\sqrt{\mathbb{E}[w_t'\Sigma_r w_t]}} = \sqrt{\mathsf{Tr}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}\Sigma_{rx})}. \tag{A.94}$$

Finally, to get the ratio of the expected turnover to the standard deviation of the portfolio, we can apply a similar reasoning as in Proposition 12 to the processes defined as $\tilde{\zeta}_1 \coloneqq \mathbb{E}[\mathsf{aim}_t'\Sigma_r \mathsf{aim}_t]$, $\tilde{\zeta}_2 \coloneqq \mathbb{E}[\mathsf{aim}_t'\Sigma_r w_t]$, and $\tilde{\zeta}_3 \coloneqq \mathbb{E}[w_t'\Sigma_r w_t]$ (this gives similar results since the matrix Σ_r is time-invariant.) In doing so, we have

$$\lim_{t\to\infty} \frac{1}{2}\frac{\mathbb{E}[\Delta w_t'\Lambda\Delta w_t]}{\sqrt{\mathbb{E}[w_t'\Sigma_r w_t]}} = \lim_{t\to\infty} \frac{\lambda}{2}\sqrt{\mathbb{E}[w_t'\Sigma_r w_t]}\cdot\frac{\mathbb{E}[\Delta w_t'\Sigma_r\Delta w_t]}{\mathbb{E}[w_t'\Sigma_r w_t]} \tag{A.95}$$

$$= \frac{\lambda}{2}\left(\frac{1}{\gamma}\frac{1-\theta_1}{1-(1-\theta_2)^{-1}\phi}\sqrt{\mathsf{Tr}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}\Sigma_{rx})}\right)\frac{2\phi(1-\theta_1)}{1-\phi\theta_1+\theta_1} \tag{A.96}$$

$$= \frac{\lambda}{\gamma}\frac{\phi(1-\theta_1)^2}{(1-(1-\theta_2)^{-1}\phi)(1+\theta_1(1-\phi))}\sqrt{\mathsf{Tr}(\Sigma_x^{-1}\Sigma_{rx}'\Sigma_r^{-1}\Sigma_{rx})}. \tag{A.97}$$

The second equality follows from the derived expressions for the optimal portfolio's risk and turnover.

[4] This assumes that the squared expected returns of the portfolio are negligible and can be approximated by zero.

Appendix B Tables

B.1 Tables for Chapter 2

Table B.1.1 – Performance Summary

This table shows the annualized out-of-sample performances (in percentages) of global minimum variance portfolios of various covariance estimators. The performances are the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The most competitive value in the SD column is highlighted in **bold**. For the EWA-SC and EWA-CV portfolios, an asterisk indicates a significant outperformance of one portfolio versus the other in terms of SD: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
$1/N$	12.20	18.08	0.67	0.51	0.67	0.51	54.48
SC	9.60	11.86	0.81	0.55	0.79	0.53	38.11
EWA-SC	8.45	11.37	0.74	0.47	0.71	0.44	37.66
LS	9.77	11.79	0.83	0.57	0.81	0.55	36.48
NL	9.67	11.75	0.82	0.56	0.80	0.54	36.96
EWA-CV	9.00	**11.17***	0.81	0.53	0.78	0.51	36.03
AFM1-NL	9.66	11.76	0.82	0.56	0.80	0.54	36.90
AFM1-EWA-CV	9.28	11.57	0.80	0.54	0.78	0.51	34.84
Panel B: Universe Size, $N = 500$							
$1/N$	12.63	17.68	0.71	0.54	0.71	0.54	55.47
SC	9.13	8.80	1.04	0.69	0.94	0.59	34.24
EWA-SC	8.48	9.08	0.93	0.60	0.80	0.46	39.77
LS	9.43	8.39	1.12	0.76	1.05	0.68	32.38
NL	8.99	7.94	1.13	0.75	1.08	0.70	30.15
EWA-CV	8.77	**7.58*****	1.16	0.75	1.11	0.70	25.54
AFM1-NL	8.93	7.94	1.13	0.74	1.08	0.69	30.54
AFM1-EWA-CV	8.86	7.78	1.14	0.75	1.09	0.70	27.47
Panel C: Universe Size, $N = 1{,}000$							
$1/N$	12.83	17.40	0.74	0.56	0.74	0.56	55.04
SC	6.92	9.93	0.70	0.39	0.37	0.07	34.47
EWA-SC	6.91	10.46	0.66	0.37	0.32	0.03	35.83
LS	8.10	7.11	1.14	0.71	0.99	0.56	33.56
NL	9.01	6.51	1.38	0.92	1.31	0.84	27.81
EWA-CV	9.46	**6.39*****	1.48	1.00	1.42	0.94	25.63
AFM1-NL	8.96	6.51	1.38	0.91	1.30	0.83	27.82
AFM1-EWA-CV	9.56	6.51	1.47	1.00	1.41	0.94	27.11

Table B.1.2 – Weight Statistics

This table shows the average monthly weight statistics of global minimum variance portfolios of various covariance estimators. The statistics are the turnover (TO), the gross exposure (GE), the proportional leverage (PL), the minimum weight (MIN), the maximum weight (MAX), the standard deviation of weights (SD), and the mean absolute deviation from an equally-weighted portfolio (MAD). Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The summary statistics are obtained from an average over the 400 portfolio formation periods.

	TO	GE	PL	MIN	MAX	SD	MAD
Panel A: Universe Size, $N = 100$							
$1/N$	0.064	1.000	0.000	0.010	0.010	0.000	0.000
SC	0.626	2.956	0.442	−0.083	0.253	0.045	0.030
EWA-SC	0.882	3.164	0.442	−0.094	0.249	0.047	0.032
LS	0.572	2.772	0.429	−0.075	0.193	0.040	0.028
NL	0.548	2.713	0.428	−0.072	0.217	0.040	0.027
EWA-CV	0.663	2.613	0.410	−0.069	0.180	0.037	0.026
AFM1-NL	0.552	2.727	0.429	−0.073	0.221	0.041	0.027
AFM1-EWA-CV	0.643	2.715	0.416	−0.074	0.206	0.039	0.027
Panel B: Universe Size, $N = 500$							
$1/N$	0.052	1.000	0.000	0.002	0.002	0.000	0.000
SC	1.670	5.191	0.475	−0.070	0.255	0.021	0.010
EWA-SC	2.322	5.935	0.476	−0.081	0.276	0.023	0.012
LS	1.284	4.475	0.459	−0.036	0.138	0.015	0.009
NL	0.802	3.399	0.437	−0.046	0.201	0.015	0.007
EWA-CV	0.812	3.141	0.424	−0.044	0.191	0.014	0.006
AFM1-NL	0.811	3.425	0.439	−0.045	0.203	0.015	0.007
AFM1-EWA-CV	0.815	3.223	0.427	−0.043	0.196	0.014	0.006
Panel C: Universe Size, $N = 1{,}000$							
$1/N$	0.048	1.000	0.000	0.001	0.001	0.000	0.000
SC	5.693	9.734	0.486	−0.087	0.235	0.017	0.010
EWA-SC	6.368	10.453	0.486	−0.089	0.240	0.018	0.010
LS	2.029	5.521	0.465	−0.023	0.077	0.008	0.006
NL	1.025	3.517	0.426	−0.037	0.128	0.007	0.003
EWA-CV	0.845	3.113	0.407	−0.024	0.089	0.006	0.003
AFM1-NL	1.037	3.536	0.427	−0.036	0.129	0.007	0.004
AFM1-EWA-CV	0.857	3.185	0.410	−0.024	0.091	0.006	0.003

Table B.1.3 – Subsample Analysis

This table shows the annualized out-of-sample standard deviation and information ratio of global minimum variance portfolios of various covariance estimators over different subsamples. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The subsamples are divided equally into Sample 1 (1986 through 1994), Sample 2 (1994 through 2003), Sample 3 (2003 through 2011), and Sample 4 (2011 through 2019). The most competitive value in the SD column is highlighted in **bold**.

	Sample 1		Sample 2		Sample 3		Sample 4	
	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: Universe Size, $N = 100$								
$1/N$	16.62	0.71	19.72	0.77	21.25	0.36	13.84	1.02
SC	11.74	0.74	14.00	0.54	11.54	0.97	9.81	1.12
EWA-SC	**10.91**	0.83	13.38	0.43	11.41	0.82	9.46	1.02
LS	11.67	0.75	13.94	0.54	11.44	1.00	9.74	1.16
NL	11.66	0.76	13.86	0.52	11.44	0.99	9.68	1.17
EWA-CV	11.09	0.81	**13.07**	0.45	**11.00**	0.91	**9.19**	1.20
AFM1-NL	11.68	0.76	13.86	0.52	11.45	0.98	9.68	1.16
AFM1-EWA-CV	11.40	0.79	13.77	0.42	11.17	1.01	9.57	1.16
Panel B: Universe Size, $N = 500$								
$1/N$	14.80	0.81	17.86	0.76	22.75	0.50	13.96	0.97
SC	9.02	1.11	9.59	1.08	10.32	0.84	5.49	1.38
EWA-SC	8.90	1.03	10.28	1.07	10.76	0.66	5.43	1.22
LS	8.52	1.14	9.16	1.11	9.77	0.89	5.49	1.66
NL	8.01	1.14	8.92	0.97	9.20	1.08	4.89	1.68
EWA-CV	**7.41**	1.17	**8.73**	0.97	**8.83**	1.13	**4.54**	1.75
AFM1-NL	8.00	1.13	8.92	0.96	9.19	1.08	4.90	1.66
AFM1-EWA-CV	7.68	1.16	9.01	0.91	8.96	1.17	4.64	1.70
Panel C: Universe Size, $N = 1{,}000$								
$1/N$	13.46	0.89	16.90	0.77	23.13	0.55	14.49	0.94
SC	9.75	0.67	9.71	0.92	12.21	0.56	7.48	0.72
EWA-SC	10.37	0.62	10.45	0.75	12.76	0.67	7.64	0.63
LS	6.70	1.08	6.80	1.47	8.83	0.75	5.75	1.48
NL	6.08	1.32	6.73	1.43	8.12	1.15	**4.61**	1.96
EWA-CV	**5.84**	1.32	**6.69**	1.45	**7.90**	1.22	4.72	2.29
AFM1-NL	6.08	1.32	6.72	1.42	8.12	1.16	4.61	1.93
AFM1-EWA-CV	5.97	1.33	6.81	1.37	8.02	1.25	4.79	2.28

Table B.1.4 – Sensitivity Analysis

This table shows the annualized out-of-sample standard deviation (in percentages) and information ratio of global minimum variance portfolios estimated with EWA-SC and EWA-CV for different decay rates across different subsamples. Panels A-B and C-D display the results for investment universe sizes $N = 100$ and $N = 1{,}000$, respectively. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The subsamples are divided equally into Sample 1 (1986 through 1994), Sample 2 (1994 through 2003), Sample 3 (2003 through 2011), and Sample 4 (2011 through 2019).

	Full Sample		Sample 1		Sample 2		Sample 3		Sample 4	
	SD	IR	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: EWA-SC, Universe Size: $N = 100$										
$\beta = 0.999$	11.56	0.80	11.30	0.78	13.62	0.51	11.36	0.93	9.61	1.08
$\beta = 0.997$	11.37	0.74	10.91	0.83	13.38	0.43	11.41	0.82	9.46	1.02
$\beta = 0.995$	11.55	0.68	10.94	0.83	13.53	0.34	11.77	0.70	9.61	1.01
$\beta = 0.992$	12.03	0.60	11.24	0.76	14.03	0.22	12.41	0.59	10.07	0.99
$\beta = 0.990$	12.38	0.55	11.48	0.70	14.44	0.15	12.81	0.55	10.41	0.96
$\beta = 0.985$	13.27	0.44	12.11	0.57	15.57	−0.00	13.73	0.47	11.26	0.90
Panel B: EWA-CV, Universe Size: $N = 100$										
$\beta = 0.999$	11.43	0.82	11.30	0.81	13.46	0.48	11.17	0.98	9.44	1.16
$\beta = 0.997$	11.17	0.81	11.09	0.81	13.07	0.45	11.00	0.91	9.19	1.20
$\beta = 0.995$	11.15	0.80	11.20	0.78	12.94	0.44	11.00	0.86	9.13	1.28
$\beta = 0.992$	11.23	0.80	11.43	0.75	12.90	0.40	11.11	0.83	9.15	1.40
$\beta = 0.990$	11.28	0.80	11.53	0.72	12.95	0.40	11.15	0.83	9.16	1.45
$\beta = 0.985$	11.40	0.78	11.72	0.64	13.03	0.42	11.28	0.81	9.24	1.49
Panel C: EWA-SC, Universe Size: $N = 1{,}000$										
$\beta = 0.999$	9.97	0.69	9.83	0.67	9.80	0.87	12.25	0.58	7.41	0.71
$\beta = 0.997$	10.46	0.66	10.37	0.62	10.45	0.75	12.76	0.67	7.64	0.63
$\beta = 0.995$	11.35	0.63	11.23	0.51	11.47	0.65	13.76	0.75	8.28	0.58
$\beta = 0.992$	12.94	0.57	12.73	0.34	13.13	0.57	15.63	0.81	9.54	0.54
$\beta = 0.990$	14.00	0.54	13.75	0.26	14.22	0.53	16.86	0.82	10.41	0.53
$\beta = 0.985$	16.46	0.47	16.15	0.13	16.83	0.42	19.66	0.76	12.39	0.53
Panel D: EWA-CV, Universe Size: $N = 1{,}000$										
$\beta = 0.999$	6.45	1.48	5.92	1.35	6.68	1.45	7.99	1.22	4.77	2.23
$\beta = 0.997$	6.39	1.48	5.84	1.32	6.69	1.45	7.90	1.22	4.72	2.29
$\beta = 0.995$	6.48	1.46	5.91	1.25	6.83	1.40	7.95	1.24	4.83	2.26
$\beta = 0.992$	6.69	1.40	6.15	1.13	7.04	1.34	8.10	1.25	5.11	2.14
$\beta = 0.990$	6.82	1.37	6.32	1.08	7.16	1.32	8.19	1.24	5.28	2.09
$\beta = 0.985$	7.13	1.31	6.71	0.97	7.45	1.28	8.45	1.18	5.62	2.05

Table B.1.5 – No Short-Sale Constraint

This table shows the annualized out-of-sample performances (in percentages) of global minimum variance portfolios of various covariance estimators. Short sales are prohibited due to a lower bound of zero on all portfolio weights. The performances are the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The most competitive SD value is highlighted in **bold**. For the EWA-SC and EWA-CV portfolios, an asterisk indicates a significant outperformance of one portfolio versus the other in terms of SD: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
$1/N$	12.20	18.08	0.67	0.51	0.67	0.51	54.48
SC	9.83	12.61	0.78	0.54	0.77	0.53	40.81
EWA-SC	9.46	12.04	0.79	0.53	0.78	0.52	38.25
LS	9.86	12.60	0.78	0.54	0.78	0.53	40.35
NL	9.80	12.60	0.78	0.54	0.77	0.53	40.84
EWA-CV	9.51	**12.10**	0.79	0.53	0.78	0.53	38.65
AFM1-NL	9.80	12.61	0.78	0.54	0.77	0.53	40.92
AFM1-EWA-CV	9.78	12.57	0.78	0.54	0.77	0.53	38.60
Panel B: Universe Size, $N = 500$							
$1/N$	12.63	17.68	0.71	0.54	0.71	0.54	55.47
SC	10.39	9.86	1.05	0.74	1.05	0.74	39.10
EWA-SC	10.22	**9.23**	1.11	0.78	1.09	0.76	38.97
LS	10.46	9.87	1.06	0.75	1.05	0.74	38.72
NL	10.13	9.91	1.02	0.71	1.02	0.71	38.98
EWA-CV	9.77	9.33	1.05	0.72	1.04	0.71	37.92
AFM1-NL	10.12	9.92	1.02	0.71	1.01	0.71	39.06
AFM1-EWA-CV	9.80	9.83	1.00	0.69	0.99	0.68	38.00
Panel C: Universe Size, $N = 1{,}000$							
$1/N$	12.83	17.40	0.74	0.56	0.74	0.56	55.04
SC	8.72	7.76	1.12	0.73	1.11	0.72	31.69
EWA-SC	8.22	**7.32**	1.12	0.71	1.11	0.69	34.00
LS	9.12	7.85	1.16	0.77	1.15	0.76	32.66
NL	9.51	8.27	1.15	0.78	1.14	0.77	34.34
EWA-CV	9.26	7.82	1.18	0.79	1.17	0.78	33.64
AFM1-NL	9.49	8.27	1.15	0.78	1.14	0.77	34.52
AFM1-EWA-CV	9.26	8.36	1.11	0.74	1.10	0.73	35.96

Table B.1.6 – Devolatization

This table shows the annualized out-of-sample performances (in percentages) of global minimum variance portfolios for various estimators. The performances are the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The most competitive value in the SD column is highlighted in **bold**.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
VOL	11.88	15.41	0.77	0.57	0.76	0.56	45.63
DCC-NL	9.18	11.69	0.78	0.52	0.67	0.41	38.91
EWA-CV	10.63	11.57	0.92	0.65	0.82	0.56	40.00
AFM1-DCC-NL	9.02	**11.52**	0.78	0.52	0.72	0.45	37.03
AFM1-EWA-CV	9.35	11.60	0.81	0.54	0.74	0.47	32.39
Panel B: Universe Size, $N = 500$							
VOL	11.47	13.68	0.84	0.62	0.82	0.60	47.56
DCC-NL	9.10	7.48	1.22	0.81	1.05	0.65	27.89
EWA-CV	9.25	7.19	1.29	0.86	1.15	0.73	30.13
AFM1-DCC-NL	9.22	7.27	1.27	0.85	1.14	0.72	29.83
AFM1-EWA-CV	9.49	**7.13**	1.33	0.90	1.21	0.78	28.32
Panel C: Universe Size, $N = 1{,}000$							
VOL	11.33	12.91	0.88	0.64	0.86	0.63	48.20
DCC-NL	8.20	5.95	1.38	0.87	1.20	0.69	29.78
EWA-CV	8.30	5.77	1.44	0.91	1.29	0.77	33.10
AFM1-DCC-NL	8.22	5.50	1.49	0.94	1.34	0.79	24.23
AFM1-EWA-CV	8.45	**5.44**	1.55	0.99	1.41	0.85	27.02

Table B.1.7 – Performance Summary

This table shows the annualized out-of-sample performances (in percentages) of Markowitz portfolios of various covariance estimators. The performances are the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The most competitive value in the IR column is highlighted in **bold**. For the EWA-SC and EWA-CV portfolios, an asterisk indicates a significant outperformance of one portfolio versus the other in terms of IR: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
EW-TQ	15.62	22.30	0.70	0.56	0.69	0.55	56.71
SC	12.19	14.47	0.84	0.63	0.80	0.59	36.83
EWA-SC	10.98	14.01	0.78	0.57	0.73	0.51	41.63
LS	12.40	14.41	0.86	0.65	0.82	0.61	36.60
NL	12.46	14.42	0.86	0.65	0.83	0.61	37.25
EWA-CV	11.98	13.85	**0.87****	0.64	0.82	0.60	41.32
AFM1-NL	12.47	14.42	0.86	0.65	0.83	0.61	36.98
AFM1-EWA-CV	12.11	14.19	0.85	0.64	0.81	0.60	40.41
Panel B: Universe Size, $N = 500$							
EW-TQ	14.69	21.44	0.69	0.54	0.67	0.53	58.63
SC	11.28	10.36	1.09	0.79	0.97	0.67	33.83
EWA-SC	9.96	10.88	0.92	0.64	0.76	0.48	36.06
LS	11.84	9.99	1.18	0.88	1.09	0.78	32.23
NL	12.00	9.67	1.24	0.93	1.17	0.86	29.03
EWA-CV	12.16	9.24	**1.32*****	0.99	1.24	0.91	23.39
AFM1-NL	11.94	9.67	1.24	0.92	1.17	0.85	29.44
AFM1-EWA-CV	12.14	9.46	1.28	0.96	1.21	0.89	25.80
Panel C: Universe Size, $N = 1{,}000$							
EW-TQ	14.63	20.87	0.70	0.55	0.69	0.54	58.56
SC	7.55	11.88	0.64	0.38	0.30	0.04	34.37
EWA-SC	6.84	14.59	0.47	0.26	0.10	−0.10	38.35
LS	10.06	8.53	1.18	0.82	1.01	0.65	30.96
NL	12.01	8.00	1.50	1.12	1.41	1.03	25.29
EWA-CV	12.70	7.80	**1.63*****	1.24	1.55	1.16	23.19
AFM1-NL	11.96	7.99	1.50	1.11	1.40	1.02	25.37
AFM1-EWA-CV	12.72	7.93	1.60	1.22	1.52	1.14	24.82

Table B.1.8 – Weight Statistics

This table shows the average monthly weight statistics of Markowitz portfolios of various covariance estimators. The statistics are the turnover (TO), the gross exposure (GE), the proportional leverage (PL), the minimum weight (MIN), the maximum weight (MAX), the standard deviation of weights (SD), and the mean absolute deviation from an equally-weighted portfolio (MAD). Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The summary statistics are obtained from an average over the 400 portfolio formation periods.

	TO	GE	PL	MIN	MAX	SD	MAD
Panel A: Universe Size, $N = 100$							
EW-TQ	0.533	1.000	0.000	0.000	0.050	0.020	0.016
SC	1.231	3.465	0.444	−0.101	0.242	0.050	0.035
EWA-SC	1.542	3.774	0.448	−0.119	0.236	0.053	0.038
LS	1.162	3.289	0.434	−0.093	0.198	0.045	0.033
NL	1.126	3.200	0.432	−0.089	0.207	0.044	0.032
EWA-CV	1.250	3.155	0.426	−0.090	0.172	0.042	0.031
AFM1-NL	1.127	3.211	0.433	−0.090	0.210	0.045	0.032
AFM1-EWA-CV	1.198	3.184	0.427	−0.086	0.182	0.043	0.032
Panel B: Universe Size, $N = 500$							
EW-TQ	0.509	1.000	0.000	0.000	0.010	0.004	0.003
SC	2.380	6.037	0.474	−0.087	0.253	0.022	0.012
EWA-SC	3.205	7.237	0.474	−0.147	0.275	0.026	0.014
LS	1.896	5.134	0.464	−0.043	0.137	0.016	0.010
NL	1.321	3.953	0.443	−0.051	0.196	0.016	0.008
EWA-CV	1.320	3.745	0.430	−0.055	0.187	0.015	0.007
AFM1-NL	1.328	3.977	0.444	−0.049	0.198	0.016	0.008
AFM1-EWA-CV	1.303	3.782	0.433	−0.053	0.190	0.015	0.007
Panel C: Universe Size, $N = 1{,}000$							
EW-TQ	0.494	1.000	0.000	0.000	0.005	0.002	0.002
SC	7.178	11.847	0.484	−0.122	0.217	0.019	0.012
EWA-SC	9.467	14.992	0.485	−0.194	0.237	0.024	0.015
LS	2.663	6.347	0.467	−0.027	0.075	0.009	0.006
NL	1.430	3.978	0.431	−0.038	0.123	0.008	0.004
EWA-CV	1.268	3.616	0.416	−0.025	0.086	0.006	0.004
AFM1-NL	1.439	3.996	0.432	−0.037	0.124	0.008	0.004
AFM1-EWA-CV	1.265	3.660	0.418	−0.026	0.087	0.006	0.004

Table B.1.9 – Subsample Analysis

This table shows the annualized out-of-sample standard deviation and information ratio of Markowitz portfolios of various covariance estimators over different subsamples. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The subsamples are divided equally into Sample 1 (1986 through 1994), Sample 2 (1994 through 2003), Sample 3 (2003 through 2011), and Sample 4 (2011 through 2019). The most competitive value in the IR column is highlighted in **bold**.

	Sample 1		Sample 2		Sample 3		Sample 4	
	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: Universe Size, $N = 100$								
EW-TQ	19.46	0.79	28.18	**0.90**	23.43	0.28	16.33	0.91
SC	14.00	0.82	18.28	0.58	13.19	1.01	11.58	1.14
EWA-SC	13.42	**0.91**	17.76	0.49	12.94	0.92	11.08	1.00
LS	13.94	0.82	18.24	0.60	13.11	1.03	11.51	1.18
NL	14.02	0.83	18.17	0.60	13.15	1.03	11.48	**1.19**
EWA-CV	13.69	0.87	17.42	0.61	12.60	1.04	10.87	1.13
AFM1-NL	14.03	0.83	18.16	0.61	13.17	1.03	11.49	1.18
AFM1-EWA-CV	13.82	0.86	18.11	0.58	12.68	**1.09**	11.23	1.09
Panel B: Universe Size, $N = 500$								
EW-TQ	17.88	0.88	25.14	0.80	25.28	0.39	15.78	0.83
SC	10.38	1.31	11.46	1.01	11.69	0.93	7.32	1.23
EWA-SC	10.50	1.12	12.61	0.90	12.22	0.83	7.42	0.89
LS	9.93	1.36	11.04	1.07	11.21	1.01	7.32	**1.46**
NL	9.71	1.38	10.82	1.08	10.72	1.21	6.95	1.43
EWA-CV	9.11	**1.44**	10.63	**1.21**	10.13	1.35	6.56	1.36
AFM1-NL	9.69	1.37	10.80	1.08	10.72	1.21	6.95	1.41
AFM1-EWA-CV	9.40	1.42	10.98	1.14	10.24	**1.36**	6.63	1.33
Panel C: Universe Size, $N = 1{,}000$								
EW-TQ	17.06	0.95	23.81	0.81	24.99	0.40	16.15	0.80
SC	11.45	0.71	12.35	0.82	13.79	0.52	9.53	0.51
EWA-SC	13.37	0.39	17.00	0.46	15.93	0.57	11.41	0.45
LS	8.06	1.33	8.57	1.49	9.98	0.74	7.29	1.28
NL	7.54	1.66	8.43	1.59	9.35	1.26	6.38	1.62
EWA-CV	7.28	**1.69**	8.36	**1.71**	8.95	1.41	6.34	**1.83**
AFM1-NL	7.53	1.66	8.41	1.58	9.35	1.26	6.38	1.60
AFM1-EWA-CV	7.46	1.68	8.54	1.62	9.06	**1.42**	6.38	1.83

Table B.1.10 – Sensitivity Analysis

This table shows the annualized out-of-sample standard deviation and information ratio (in percentages) of Markowitz portfolios estimated with EWA-SC and EWA-CV based on decay rates $\beta \in \{0.999, 0.997, 0.995, 0.992, 0.990, 0.985\}$ over the different subsamples. Panels A-B and C-D display the results for investment universe sizes $N = 100$ and $N = 1{,}000$, respectively. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The subsamples are divided equally into Sample 1 (1986 through 1994), Sample 2 (1994 through 2003), Sample 3 (2003 through 2011), and Sample 4 (2011 through 2019).

	Full Sample		Sample 1		Sample 2		Sample 3		Sample 4	
	SD	IR	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: EWA-SC, Universe Size: $N = 100$										
$\beta = 0.999$	14.16	0.83	13.63	0.86	17.91	0.55	12.96	1.00	11.31	1.09
$\beta = 0.997$	14.01	0.78	13.42	0.91	17.76	0.49	12.94	0.92	11.08	1.00
$\beta = 0.995$	14.27	0.74	13.67	0.92	18.03	0.42	13.30	0.84	11.23	0.96
$\beta = 0.992$	14.93	0.68	14.30	0.91	18.79	0.32	14.00	0.75	11.75	0.91
$\beta = 0.990$	15.41	0.64	14.77	0.90	19.40	0.25	14.46	0.71	12.12	0.87
$\beta = 0.985$	16.68	0.55	15.99	0.86	21.07	0.11	15.53	0.65	13.12	0.80
Panel B: EWA-CV, Universe Size: $N = 100$										
$\beta = 0.999$	14.10	0.87	13.73	0.87	17.77	0.60	12.86	1.06	11.18	1.16
$\beta = 0.997$	13.85	0.87	13.69	0.87	17.42	0.61	12.60	1.04	10.87	1.13
$\beta = 0.995$	13.90	0.87	13.99	0.86	17.36	0.64	12.59	1.02	10.83	1.16
$\beta = 0.992$	14.05	0.89	14.39	0.87	17.43	0.67	12.72	1.01	10.84	1.22
$\beta = 0.990$	14.15	0.90	14.59	0.85	17.54	0.69	12.77	1.01	10.86	1.26
$\beta = 0.985$	14.31	0.91	14.84	0.81	17.70	0.75	12.92	1.00	10.93	1.30
Panel C: EWA-SC, Universe Size: $N = 1{,}000$										
$\beta = 0.999$	12.32	0.59	11.76	0.63	13.16	0.71	14.16	0.52	9.79	0.50
$\beta = 0.997$	14.59	0.47	13.37	0.39	17.00	0.46	15.93	0.57	11.41	0.45
$\beta = 0.995$	17.80	0.38	15.84	0.18	21.93	0.31	18.42	0.64	13.99	0.40
$\beta = 0.992$	21.39	0.37	19.02	0.10	26.83	0.27	21.40	0.73	17.05	0.41
$\beta = 0.990$	23.24	0.39	20.80	0.11	29.22	0.29	22.99	0.77	18.58	0.44
$\beta = 0.985$	27.44	0.44	24.71	0.16	34.75	0.33	26.65	0.80	22.00	0.51
Panel D: EWA-CV, Universe Size: $N = 1{,}000$										
$\beta = 0.999$	7.89	1.60	7.35	1.71	8.37	1.65	9.16	1.36	6.43	1.82
$\beta = 0.997$	7.80	1.63	7.28	1.69	8.36	1.71	8.95	1.41	6.34	1.83
$\beta = 0.995$	7.88	1.63	7.41	1.62	8.50	1.69	8.96	1.48	6.42	1.82
$\beta = 0.992$	8.10	1.59	7.76	1.52	8.74	1.64	9.07	1.52	6.64	1.76
$\beta = 0.990$	8.25	1.57	7.99	1.46	8.90	1.61	9.16	1.52	6.76	1.74
$\beta = 0.985$	8.61	1.52	8.50	1.34	9.30	1.56	9.40	1.49	7.02	1.75

Table B.1.11 – No Short-Sale Constraint

This table shows the annualized out-of-sample performances (in percentages) of Markowitz portfolios of various covariance estimators. Short sales are prohibited due to a lower bound of zero on all portfolio weights. The performances are the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The most competitive IR value is highlighted in **bold**. For the EWA-SC and EWA-CV portfolios, an asterisk indicates a significant outperformance of one portfolio versus the other in terms of IR: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
EW-TQ	15.62	22.30	0.70	0.56	0.69	0.55	56.71
SC	13.44	18.33	0.73	0.57	0.71	0.55	48.35
EWA-SC	12.93	17.82	0.73	0.55	0.70	0.53	47.02
LS	13.44	18.32	0.73	0.57	0.71	0.55	47.94
NL	13.37	18.33	0.73	0.56	0.71	0.54	48.22
EWA-CV	13.01	17.85	0.73	0.56	0.71	0.54	47.62
AFM1-NL	13.39	18.33	0.73	0.56	0.71	0.54	48.23
AFM1-EWA-CV	13.48	18.19	**0.74**	0.57	0.72	0.55	46.30
Panel B: Universe Size, $N = 500$							
EW-TQ	14.69	21.44	0.69	0.54	0.67	0.53	58.63
SC	13.29	15.02	0.88	0.68	0.86	0.66	38.68
EWA-SC	14.26	14.28	1.00	0.79	0.97	0.76	40.11
LS	13.34	14.99	0.89	0.69	0.87	0.66	38.61
NL	13.59	14.94	0.91	0.71	0.89	0.68	39.26
EWA-CV	14.38	14.20	**1.01**	0.80	0.99	0.77	37.75
AFM1-NL	13.58	14.94	0.91	0.70	0.89	0.68	39.36
AFM1-EWA-CV	13.95	14.78	0.94	0.74	0.92	0.71	36.40
Panel C: Universe Size, $N = 1{,}000$							
EW-TQ	14.63	20.87	0.70	0.55	0.69	0.54	58.56
SC	12.60	13.25	0.95	0.72	0.92	0.69	33.25
EWA-SC	12.49	12.56	0.99	0.75	0.96	0.72	35.00
LS	12.76	13.25	0.96	0.73	0.94	0.71	34.29
NL	13.24	13.37	0.99	0.76	0.97	0.74	35.68
EWA-CV	12.75	12.69	**1.01**	0.76	0.98	0.74	36.83
AFM1-NL	13.24	13.38	0.99	0.76	0.97	0.74	35.73
AFM1-EWA-CV	12.96	13.24	0.98	0.75	0.95	0.72	36.48

Table B.1.12 – Devolatization

This table shows the annualized out-of-sample performances (in percentages) of Markowitz portfolios for various estimators. The performances are the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 09/04/1986 up to 12/31/2019. The most competitive value in the IR column is highlighted in **bold**.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
VOL	13.85	18.79	0.74	0.57	0.71	0.55	50.92
DCC-NL	12.41	14.51	0.86	0.65	0.74	0.53	37.29
EWA-CV	13.62	14.45	**0.94**	0.73	0.85	0.64	44.01
AFM1-DCC-NL	12.16	14.16	0.86	0.64	0.79	0.57	34.91
AFM1-EWA-CV	12.45	14.34	0.87	0.66	0.79	0.58	31.73
Panel B: Universe Size, $N = 500$							
VOL	13.68	16.00	0.85	0.66	0.82	0.63	42.52
DCC-NL	11.63	9.07	1.28	0.95	1.12	0.78	24.36
EWA-CV	11.90	8.82	1.35	1.00	1.21	0.87	29.30
AFM1-DCC-NL	12.18	8.83	1.38	1.03	1.24	0.89	28.00
AFM1-EWA-CV	12.42	8.72	**1.42**	1.08	1.29	0.94	26.16
Panel C: Universe Size, $N = 1{,}000$							
VOL	13.28	14.74	0.90	0.69	0.87	0.66	43.16
DCC-NL	10.37	7.19	1.44	1.02	1.26	0.83	25.86
EWA-CV	10.64	7.09	1.50	1.07	1.35	0.92	30.29
AFM1-DCC-NL	11.09	6.88	1.61	1.17	1.44	1.00	21.41
AFM1-EWA-CV	11.15	6.85	**1.63**	1.18	1.46	1.01	23.39

B.2 Tables for Chapter 3

Table B.2.1 – GICS color code

GICS sectors as provided by Morgan Stanley Capital International (MSCI) and its associated color code. Note that a color code has not been assigned to the real estate sector since all of our empirical analysis is restricted to the universe of U.S.-denominated common stocks.

Sector	Sector Description	Color Code
10	Energy	Brown
15	Materials	Dark Green
20	Industrials	Blue
25	Consumer Discretionary	Red
30	Consumer Staples	Green
35	Health Care	Cyan
40	Financials	Fuchsia
45	Information Technology	Yellow
50	Communication Services	Cornflower Blue
55	Utilities	Pink
60	Real Estate	Not Available

Table B.2.2 – Performance Summary

This table shows the annualized out-of-sample performances (in percentages) of global minimum variance portfolios of various covariance estimators. The columns display the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 01/01/1995 up to 12/31/2021. The most competitive value in the SD column is highlighted in **bold**. For the NL-SF and EWA-CV-FHC portfolios, an asterisk indicates a significant outperformance of one portfolio versus the other in terms of SD: *** denotes significance at the 1% level; ** denotes significance at the 5% level; and * denotes significance at the 10% level.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
$1/N$	9.83	19.61	0.50	0.42	0.50	0.42	54.16
SF	8.87	16.63	0.53	0.44	0.53	0.44	32.03
HC	8.35	13.61	0.61	0.50	0.59	0.48	32.09
NL	8.20	13.42	0.61	0.50	0.59	0.48	38.22
EWA-CV	7.10	12.96	0.55	0.43	0.52	0.41	35.76
FHC	7.20	13.09	0.55	0.44	0.53	0.41	32.30
FNL	8.18	13.42	0.61	0.50	0.59	0.48	38.23
NL-SF	8.28	13.43	0.62	0.51	0.60	0.49	37.93
EWA-CV-FHC	7.48	**12.68*****	0.59	0.47	0.56	0.45	33.41
Panel B: Universe Size, $N = 500$							
$1/N$	12.51	20.47	0.61	0.54	0.61	0.54	54.97
SF	10.94	16.51	0.66	0.57	0.66	0.57	41.14
HC	9.94	12.02	0.83	0.70	0.78	0.66	35.44
NL	8.32	10.42	0.80	0.65	0.75	0.60	30.06
EWA-CV	8.53	9.98	0.85	0.70	0.80	0.65	27.33
FHC	10.33	11.29	0.91	0.78	0.87	0.74	29.79
FNL	8.34	10.42	0.80	0.66	0.75	0.61	30.29
NL-SF	8.99	10.19	0.88	0.74	0.84	0.69	29.01
EWA-CV-FHC	9.49	**9.62*****	0.99	0.83	0.93	0.77	26.21
Panel C: Universe Size, $N = 1{,}000$							
$1/N$	13.29	21.60	0.62	0.55	0.62	0.55	55.98
SF	11.19	15.76	0.71	0.62	0.71	0.61	43.66
HC	10.74	11.01	0.98	0.84	0.93	0.79	34.94
NL	9.36	9.46	0.99	0.83	0.93	0.77	31.41
EWA-CV	9.43	9.25	1.02	0.86	0.97	0.81	30.23
FHC	10.57	10.45	1.01	0.87	0.96	0.82	33.34
FNL	9.38	9.46	0.99	0.83	0.93	0.77	31.56
NL-SF	9.63	8.86	1.09	0.92	1.04	0.87	27.29
EWA-CV-FHC	10.53	**8.50*****	1.24	1.06	1.17	0.99	25.35

Table B.2.3 – Weight Statistics

This table shows the average monthly weight statistics of global minimum variance portfolios of various covariance estimators. The columns display the turnover (TO), the gross exposure (GE), the proportional leverage (PL), the minimum weight (MIN), the maximum weight (MAX), the standard deviation of weights (SD), and the mean absolute deviation from an equally-weighted portfolio (MAD). Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The summary statistics are obtained from an average over the 264 portfolio formation periods.

	TO	GE	PL	MIN	MAX	SD	MAD
Panel A: Universe Size, $N = 100$							
$1/N$	0.061	1.000	0.000	0.010	0.010	0.000	0.000
SF	0.234	2.203	0.432	−0.042	0.100	0.029	0.022
HC	0.735	2.443	0.471	−0.042	0.150	0.034	0.025
NL	0.564	2.853	0.407	−0.084	0.158	0.039	0.028
EWA-CV	0.687	2.766	0.407	−0.079	0.150	0.037	0.027
FHC	0.643	2.754	0.443	−0.079	0.160	0.039	0.028
FNL	0.565	2.861	0.407	−0.084	0.158	0.039	0.028
NL-SF	0.525	2.780	0.409	−0.085	0.161	0.039	0.028
EWA-CV-FHC	0.695	2.866	0.433	−0.083	0.176	0.041	0.029
Panel B: Universe Size, $N = 500$							
$1/N$	0.050	1.000	0.000	0.002	0.002	0.000	0.000
SF	0.179	2.370	0.454	−0.012	0.030	0.007	0.005
HC	1.020	3.158	0.512	−0.019	0.070	0.009	0.007
NL	1.088	4.689	0.446	−0.036	0.061	0.012	0.009
EWA-CV	1.044	4.063	0.428	−0.029	0.048	0.010	0.008
FHC	0.887	3.517	0.465	−0.031	0.069	0.010	0.007
FNL	1.100	4.729	0.447	−0.036	0.062	0.012	0.009
NL-SF	0.879	4.522	0.464	−0.038	0.084	0.013	0.009
EWA-CV-FHC	1.115	4.400	0.464	−0.040	0.089	0.013	0.009
Panel C: Universe Size, $N = 1{,}000$							
$1/N$	0.046	1.000	0.000	0.001	0.001	0.000	0.000
SF	0.164	2.440	0.474	−0.007	0.020	0.004	0.003
HC	0.996	3.097	0.523	−0.010	0.047	0.005	0.003
NL	1.142	4.787	0.446	−0.018	0.031	0.006	0.005
EWA-CV	0.982	4.111	0.427	−0.015	0.025	0.005	0.004
FHC	0.936	3.584	0.490	−0.017	0.047	0.006	0.004
FNL	1.160	4.834	0.447	−0.018	0.031	0.006	0.005
NL-SF	0.843	4.666	0.478	−0.021	0.059	0.007	0.005
EWA-CV-FHC	1.147	4.483	0.487	−0.024	0.061	0.007	0.005

Table B.2.4 – Subsample Analysis

This table shows the annualized out-of-sample standard deviation and information ratio of global minimum variance portfolios of various covariance estimators over different subsamples. The columns display the subsample periods divided equally into Sample 1 (1999 through 2005), Sample 2 (2005 through 2011), Sample 3 (2011 through 2016), and Sample 4 (2016 through 2021). Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 01/01/1995 up to 12/31/2021. The most competitive value in the SD column is highlighted in **bold**.

	Subsample 1		Subsample 2		Subsample 3		Subsample 4	
	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: Universe Size, $N = 100$								
$1/N$	19.82	0.13	23.56	0.27	15.76	0.80	18.50	0.97
SF	18.89	0.40	15.55	0.42	12.15	1.54	18.99	0.13
HC	14.02	0.18	13.23	0.31	10.79	1.58	15.89	0.60
NL	13.87	0.35	13.00	0.59	10.46	1.22	15.81	0.47
EWA-CV	13.03	0.24	12.71	0.46	9.87	1.31	15.60	0.41
FHC	12.96	0.25	12.91	0.45	10.75	1.50	15.34	0.24
FNL	13.87	0.35	13.00	0.59	10.46	1.22	15.82	0.46
NL-SF	13.94	0.35	12.98	0.61	10.46	1.19	15.78	0.50
EWA-CV-FHC	**12.90**	0.29	**12.40**	0.51	**10.09**	1.42	**14.88**	0.36
Panel B: Universe Size, $N = 500$								
$1/N$	19.31	0.53	25.34	0.36	16.75	0.77	19.53	0.92
SF	16.19	0.91	15.85	0.25	12.30	1.84	20.62	0.11
HC	10.46	0.90	12.17	0.11	10.10	1.97	14.76	0.61
NL	10.05	0.52	10.49	0.56	8.72	1.74	12.12	0.57
EWA-CV	9.65	0.65	10.44	0.50	8.06	1.93	11.46	0.61
FHC	9.82	1.23	11.18	0.20	9.36	2.11	14.16	0.51
FNL	10.06	0.55	10.49	0.56	8.71	1.75	12.12	0.56
NL-SF	9.55	0.68	10.20	0.65	8.58	1.72	12.09	0.66
EWA-CV-FHC	**8.76**	0.98	**10.15**	0.52	**7.83**	2.09	**11.34**	0.69
Panel C: Universe Size, $N = 1{,}000$								
$1/N$	19.53	0.65	26.45	0.40	18.01	0.70	21.49	0.82
SF	14.22	1.23	15.35	0.21	12.19	1.89	20.15	0.05
HC	9.15	1.36	11.42	0.24	9.31	2.00	13.55	0.67
NL	8.95	1.04	9.88	0.54	7.91	1.82	10.85	0.78
EWA-CV	8.69	1.20	9.83	0.52	7.45	1.91	10.71	0.74
FHC	8.71	1.56	10.54	0.35	8.72	2.22	13.16	0.42
FNL	8.96	1.06	9.87	0.54	7.90	1.80	10.84	0.77
NL-SF	7.93	1.34	**9.20**	0.70	7.48	1.85	10.50	0.72
EWA-CV-FHC	**7.28**	1.62	9.28	0.74	**6.94**	2.34	**10.07**	0.71

Table B.2.5 – Sensitivity Analysis

This table shows the annualized out-of-sample standard deviation (in percentages) and information ratio of global minimum variance portfolios estimated with EWA-CV-FHC for different decay rates across different subsamples. The columns display the full out-of-sample period and the subsample periods divided equally into Sample 1 (1999 through 2000), Sample 2 (2005 through 2010), Sample 3 (2010 through 2016), and Sample 4 (2016 through 2021). Panels A-C display the results for investment universe sizes $N = 100, 500,$ and 1,000, respectively. The daily out-of-sample returns are from 01/01/1995 up to 12/31/2021.

	Full Sample		Subperiod 1		Subperiod 2		Subperiod 3		Subperiod 4	
	SD	IR	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: EWA-CV-FHC, Universe Size, $N = 100$										
$\delta = 0.999$	12.91	0.60	13.10	0.35	12.56	0.55	10.21	1.33	15.29	0.37
$\delta = 0.997$	12.68	0.59	12.90	0.30	12.40	0.51	10.09	1.43	14.88	0.36
$\delta = 0.995$	12.58	0.59	12.78	0.23	12.38	0.47	10.04	1.52	14.69	0.39
$\delta = 0.990$	12.51	0.58	12.61	0.15	12.36	0.44	10.13	1.53	14.55	0.43
$\delta = 0.980$	12.48	0.59	12.61	0.11	12.34	0.41	10.35	1.53	14.30	0.51
$\delta = 0.970$	12.51	0.60	12.65	0.12	12.37	0.42	10.44	1.54	14.26	0.51
$\delta = 0.960$	12.53	0.58	12.68	0.11	12.42	0.40	10.46	1.53	14.28	0.48
Panel B: EWA-CV-FHC, Universe Size, $N = 500$										
$\delta = 0.999$	9.82	0.93	8.93	0.93	10.18	0.50	8.17	1.91	11.65	0.63
$\delta = 0.997$	9.62	0.99	8.77	0.99	10.15	0.51	7.83	2.09	11.34	0.69
$\delta = 0.995$	9.65	1.04	8.77	1.05	10.16	0.47	7.80	2.20	11.46	0.78
$\delta = 0.990$	9.84	1.08	8.95	1.14	10.26	0.47	8.03	2.29	11.71	0.78
$\delta = 0.980$	10.06	1.08	9.14	1.17	10.38	0.42	8.34	2.22	11.98	0.82
$\delta = 0.970$	10.20	1.06	9.24	1.18	10.50	0.38	8.48	2.21	12.17	0.79
$\delta = 0.960$	10.26	1.04	9.31	1.17	10.52	0.35	8.56	2.20	12.27	0.77
Panel C: EWA-CV-FHC, Universe Size, $N = 1{,}000$										
$\delta = 0.999$	8.61	1.17	7.34	1.53	9.27	0.70	7.22	2.18	10.21	0.67
$\delta = 0.997$	8.50	1.24	7.28	1.64	9.28	0.74	6.94	2.33	10.08	0.71
$\delta = 0.995$	8.61	1.27	7.41	1.68	9.38	0.68	6.97	2.43	10.25	0.78
$\delta = 0.990$	8.86	1.28	7.71	1.67	9.54	0.66	7.29	2.49	10.51	0.77
$\delta = 0.980$	9.09	1.24	7.91	1.68	9.71	0.60	7.60	2.37	10.74	0.74
$\delta = 0.970$	9.21	1.20	8.01	1.65	9.77	0.54	7.74	2.33	10.93	0.71
$\delta = 0.960$	9.28	1.19	8.06	1.65	9.82	0.51	7.83	2.31	11.02	0.70

Table B.2.6 – Performance Summary

This table shows the annualized out-of-sample performances (in percentages) of Markowitz portfolios of various covariance estimators. The columns display the average returns (AV), the standard deviation of returns (SD), the information ratio (IR), the Sharpe ratio (SR), and the maximum drawdown (MDD). The information ratio and Sharpe ratio net of transaction costs of 5 basis points are also reported and denoted as $\widetilde{\text{IR}}$ and $\widetilde{\text{SR}}$, respectively. Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 01/01/1995 up to 12/31/2021. The most competitive value in the IR column is highlighted in **bold**. For the NL-SF and EWA-CV-FHC portfolios, an asterisk indicates a significant outperformance of one portfolio versus the other in terms of IR: *** denotes significance at the 1% level; ** denotes significance at the 5% level; and * denotes significance at the 10% level.

	AV	SD	IR	SR	$\widetilde{\text{IR}}$	$\widetilde{\text{SR}}$	MDD
Panel A: Universe Size, $N = 100$							
EW-TQ	11.52	23.45	0.49	0.43	0.48	0.42	52.06
SF	10.55	18.92	0.56	0.48	0.53	0.45	40.13
HC	10.79	15.75	**0.69**	0.59	0.64	0.55	30.42
NL	10.42	15.52	0.67	0.58	0.63	0.54	37.52
EWA-CV	9.64	15.08	0.64	0.54	0.59	0.49	34.91
FHC	9.28	15.46	0.60	0.50	0.56	0.46	31.30
FNL	10.41	15.52	0.67	0.57	0.63	0.54	37.53
NL-SF	10.47	15.54	0.67	0.58	0.64	0.54	37.06
EWA-CV-FHC	9.74	14.84	0.66	0.56	0.61	0.51	31.73
Panel B: Universe Size, $N = 500$							
EW-TQ	12.25	23.56	0.52	0.46	0.51	0.45	56.62
SF	12.78	18.54	0.69	0.61	0.66	0.58	50.01
HC	11.70	13.31	0.88	0.77	0.82	0.70	34.39
NL	9.38	11.96	0.78	0.66	0.72	0.59	30.12
EWA-CV	9.75	11.55	0.84	0.71	0.78	0.65	28.53
FHC	11.80	12.87	**0.92**	0.80	0.86	0.74	34.20
FNL	9.40	11.96	0.79	0.66	0.72	0.59	30.16
NL-SF	9.85	11.70	0.84	0.71	0.78	0.65	30.65
EWA-CV-FHC	10.12	11.09	0.91***	0.78	0.83	0.70	29.49
Panel C: Universe Size, $N = 1{,}000$							
EW-TQ	12.83	24.56	0.52	0.46	0.51	0.45	56.24
SF	12.63	17.80	0.71	0.63	0.68	0.60	51.87
HC	12.23	12.22	1.00	0.88	0.93	0.81	34.33
NL	10.14	10.83	0.94	0.80	0.86	0.72	29.45
EWA-CV	10.47	10.61	0.99	0.85	0.92	0.78	27.51
FHC	11.76	11.86	0.99	0.87	0.92	0.80	36.88
FNL	10.15	10.82	0.94	0.80	0.86	0.72	29.59
NL-SF	9.99	10.29	0.97	0.83	0.90	0.75	27.35
EWA-CV-FHC	10.73	9.81	**1.09*****	0.94	1.00	0.84	27.94

Table B.2.7 – Weight Statistics

This table shows the average monthly weight statistics of Markowitz portfolios of various covariance estimators. The columns display the turnover (TO), the gross exposure (GE), the proportional leverage (PL), the minimum weight (MIN), the maximum weight (MAX), the standard deviation of weights (SD), and the mean absolute deviation from an equally-weighted portfolio (MAD). Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The summary statistics are obtained from an average over the 264 portfolio formation periods.

	TO	GE	PL	MIN	MAX	SD	MAD
Panel A: Universe Size, $N = 100$							
EW-TQ	0.551	1.000	0.000	0.000	0.050	0.020	0.016
SF	0.893	2.714	0.424	−0.066	0.116	0.035	0.027
HC	1.308	3.019	0.460	−0.067	0.161	0.041	0.031
NL	1.185	3.385	0.414	−0.103	0.171	0.045	0.033
EWA-CV	1.321	3.347	0.420	−0.097	0.160	0.044	0.033
FHC	1.274	3.332	0.444	−0.101	0.167	0.045	0.033
FNL	1.186	3.390	0.415	−0.103	0.172	0.045	0.033
NL-SF	1.152	3.316	0.417	−0.103	0.175	0.045	0.033
EWA-CV-FHC	1.348	3.466	0.440	−0.103	0.184	0.047	0.035
Panel B: Universe Size, $N = 500$							
EW-TQ	0.520	1.000	0.000	0.000	0.010	0.004	0.003
SF	0.852	2.897	0.447	−0.018	0.034	0.008	0.006
HC	1.543	3.723	0.499	−0.024	0.073	0.011	0.008
NL	1.594	5.123	0.448	−0.039	0.063	0.013	0.010
EWA-CV	1.557	4.537	0.436	−0.032	0.051	0.012	0.009
FHC	1.453	4.057	0.466	−0.034	0.070	0.012	0.008
FNL	1.604	5.161	0.449	−0.039	0.064	0.013	0.010
NL-SF	1.443	4.970	0.466	−0.042	0.087	0.014	0.010
EWA-CV-FHC	1.723	4.948	0.469	−0.044	0.091	0.014	0.010
Panel C: Universe Size, $N = 1{,}000$							
EW-TQ	0.513	1.000	0.000	0.000	0.005	0.002	0.002
SF	0.826	2.950	0.461	−0.010	0.023	0.004	0.003
HC	1.514	3.668	0.508	−0.014	0.049	0.005	0.004
NL	1.578	5.191	0.448	−0.020	0.032	0.007	0.005
EWA-CV	1.445	4.559	0.432	−0.017	0.026	0.006	0.004
FHC	1.498	4.142	0.484	−0.019	0.049	0.006	0.004
FNL	1.595	5.238	0.448	−0.020	0.032	0.007	0.005
NL-SF	1.396	5.127	0.477	−0.023	0.060	0.007	0.005
EWA-CV-FHC	1.746	5.056	0.487	−0.026	0.062	0.008	0.005

Table B.2.8 – Subsample Analysis

This table shows the annualized out-of-sample standard deviation and information ratio of Markowitz portfolios of various covariance estimators over different subsamples. The columns display the subsample periods divided equally into Sample 1 (1986 through 1994), Sample 2 (1994 through 2003), Sample 3 (2003 through 2011), and Sample 4 (2011 through 2019). Panels A-C display the results for investment universe sizes $N \in \{100, 500, 1{,}000\}$. The daily out-of-sample returns are from 01/01/1995 up to 12/31/2021. The most competitive value in the IR column is highlighted in **bold**.

	Subsample 1		Subsample 2		Subsample 3		Subsample 4	
	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: Universe Size, $N = 100$								
EW-TQ	27.50	0.21	24.64	0.31	17.85	0.77	22.80	**0.86**
SF	20.40	0.26	18.17	0.56	14.10	**1.59**	22.08	0.19
HC	17.11	0.14	14.34	0.53	12.38	1.52	18.47	0.75
NL	17.14	0.29	14.53	0.70	12.39	1.26	17.48	0.62
EWA-CV	16.56	0.19	14.12	0.69	11.70	1.35	17.29	0.56
FHC	16.95	0.26	14.20	0.57	12.49	1.47	17.65	0.34
FNL	17.15	0.28	14.53	0.70	12.39	1.26	17.48	0.61
NL-SF	17.21	**0.29**	14.52	**0.70**	12.39	1.24	17.48	0.64
EWA-CV-FHC	16.52	0.26	13.67	0.67	11.89	1.45	16.74	0.48
Panel B: Universe Size, $N = 500$								
EW-TQ	26.04	0.50	26.57	0.23	18.33	0.75	22.43	**0.74**
SF	18.02	0.76	18.35	0.50	13.63	1.86	22.97	0.12
HC	11.53	0.66	12.78	0.49	11.26	1.99	16.89	0.62
NL	11.88	0.26	11.81	0.73	9.91	1.74	13.90	0.61
EWA-CV	11.73	0.46	11.54	0.72	9.29	1.81	13.27	0.64
FHC	11.71	**0.89**	12.00	0.54	10.83	**2.04**	16.25	0.51
FNL	11.87	0.28	11.82	0.73	9.91	1.75	13.89	0.60
NL-SF	11.27	0.35	11.48	**0.80**	9.88	1.72	13.82	0.67
EWA-CV-FHC	10.41	0.65	11.04	0.69	9.22	1.97	13.28	0.60
Panel C: Universe Size, $N = 1{,}000$								
EW-TQ	26.57	0.52	27.52	0.26	19.87	0.68	23.60	0.74
SF	16.64	1.04	17.46	0.43	13.62	1.84	22.35	0.03
HC	10.17	1.10	12.00	0.47	10.56	1.99	15.42	0.72
NL	10.79	0.73	10.85	0.60	9.09	1.79	12.33	**0.80**
EWA-CV	10.55	0.91	10.67	0.60	8.67	1.86	12.23	0.79
FHC	10.24	**1.23**	11.27	0.55	10.29	2.06	14.97	0.46
FNL	10.79	0.75	10.84	0.60	9.08	1.78	12.31	0.79
NL-SF	9.76	0.89	10.23	0.70	8.81	1.76	12.09	0.71
EWA-CV-FHC	8.71	1.17	10.04	**0.75**	8.34	**2.14**	11.78	0.62

Table B.2.9 – Sensitivity Analysis

This table shows the annualized out-of-sample standard deviation (in percentages) and information ratio of Markowitz portfolios estimated with EWA-CV-FHC for different decay rates across different subsamples. The columns display the full sample period and the subsample periods divided equally into Sample 1 (1986 through 1994), Sample 2 (1994 through 2003), Sample 3 (2003 through 2011), and Sample 4 (2011 through 2019). Panels A-C display the results for investment universe sizes $N = 100, 500$, and 1,000, respectively. The daily out-of-sample returns are from 01/01/1995 up to 12/31/2021.

	Full Sample		Subperiod 1		Subperiod 2		Subperiod 3		Subperiod 4	
	SD	IR	SD	IR	SD	IR	SD	IR	SD	IR
Panel A: EWA-CV-FHC, Universe Size, $N = 100$										
$\delta = 0.999$	15.06	0.65	16.61	0.28	13.94	0.69	12.05	1.35	17.10	0.49
$\delta = 0.997$	14.84	0.66	16.52	0.27	13.67	0.67	11.89	1.45	16.74	0.48
$\delta = 0.995$	14.74	0.66	16.41	0.25	13.63	0.62	11.80	1.55	16.58	0.49
$\delta = 0.990$	14.65	0.66	16.28	0.22	13.59	0.58	11.78	1.58	16.42	0.54
$\delta = 0.980$	14.68	0.67	16.34	0.20	13.65	0.58	12.00	1.57	16.26	0.58
$\delta = 0.970$	14.78	0.66	16.47	0.19	13.68	0.60	12.13	1.55	16.36	0.55
$\delta = 0.960$	14.83	0.64	16.50	0.18	13.72	0.58	12.18	1.53	16.47	0.51
Panel B: EWA-CV-FHC, Universe Size, $N = 500$										
$\delta = 0.999$	11.37	0.85	10.59	0.56	11.22	0.66	9.59	1.82	13.67	0.58
$\delta = 0.997$	11.09	0.91	10.41	0.65	11.04	0.69	9.22	1.97	13.28	0.60
$\delta = 0.995$	11.06	0.97	10.41	0.72	10.97	0.66	9.16	2.08	13.30	0.67
$\delta = 0.990$	11.21	1.03	10.59	0.86	11.05	0.65	9.33	2.18	13.45	0.69
$\delta = 0.980$	11.44	1.05	10.81	0.92	11.28	0.66	9.70	2.13	13.61	0.73
$\delta = 0.970$	11.61	1.03	10.92	0.92	11.41	0.64	9.89	2.11	13.84	0.71
$\delta = 0.960$	11.72	1.02	11.00	0.89	11.45	0.61	10.02	2.10	14.02	0.70
Panel C: EWA-CV-FHC, Universe Size, $N = 1{,}000$										
$\delta = 0.999$	10.03	1.04	8.91	1.06	10.15	0.71	8.71	2.00	12.00	0.62
$\delta = 0.997$	9.81	1.09	8.71	1.18	10.04	0.75	8.34	2.14	11.78	0.62
$\delta = 0.995$	9.86	1.13	8.80	1.26	10.05	0.70	8.32	2.24	11.88	0.67
$\delta = 0.990$	10.04	1.17	9.08	1.30	10.15	0.69	8.60	2.31	12.00	0.70
$\delta = 0.980$	10.28	1.16	9.32	1.32	10.39	0.68	8.96	2.21	12.14	0.72
$\delta = 0.970$	10.44	1.14	9.42	1.30	10.49	0.64	9.15	2.17	12.38	0.70
$\delta = 0.960$	10.55	1.13	9.48	1.30	10.57	0.63	9.28	2.15	12.56	0.69

B.3 Tables for Chapter 4

Table B.3.1 – Summary of PCA and CCA

This table summarizes the key distinguishing features of PCA and CCA applied to a paired dataset of asset returns r_{t+1} and signals x_t. For CCA (PCA), it displays (1) the correlation (variance) maximization objective, (2) the equivalent error minimization objective, (3) the ith canonical (principal) direction, and (4) the ith correlation (variance) of the corresponding canonical (principal) direction.

PCA	(1) Maximize Variance	$\max_{\substack{q_r, q_x \\ \|q_r\|_2 = \|q_x\|_2 = 1}} \mathsf{Var}(q_r' r_{t+1}) + \mathsf{Var}(q_x' x_t)$
	(2) Minimize Error	$\min_{\|q_r\|_2 = \|q_x\|_2 = 1} \mathbb{E}\left[\left\|r_{t+1} - q_r q_r' r_{t+1}\right\|_2^2\right] + \mathbb{E}\left[\left\|x_t - q_x q_x' x_t\right\|_2^2\right]$
	(3) Principal Direction	$q_{r,i}, q_{x,i} = \operatorname{argmax}_{\substack{\|q_r\|_2 = \|q_x\|_2 = 1 \\ q_r' q_{r,j} = 0,\ j=1,\ldots,i-1 \\ q_x' q_{x,j} = 0,\ j=1,\ldots,i-1}} \mathsf{Var}(q_r' r_{t+1}) + \mathsf{Var}(q_x' x_t)$
	(4) Principal Variance	$d_{r,i} = \mathsf{Var}(q_{r,i}' r_{t+1}), \quad d_{x,i} = \mathsf{Var}(q_{x,i}' x_t)$
CCA	(1) Maximize Correlation	$\max_{\mathsf{Var}(q_r' r_{t+1}) = \mathsf{Var}(q_x' x_t) = 1} \mathsf{Cov}(q_r' r_{t+1}, q_x' x_t)$
	(2) Minimize Error	$\min_{\mathsf{Var}(q_r' r_{t+1}) = \mathsf{Var}(q_x' x_t) = 1} \mathbb{E}\left[\left\|q_r' r_{t+1} - q_x' x_t\right\|_2^2\right]$
	(3) Canonical Direction	$q_{r,i}, q_{x,i} = \operatorname{argmax}_{\substack{q_r' \Sigma_r q_r = q_x' \Sigma_x q_x = 1 \\ q_r' q_{r,j} = 0,\ j=1,\ldots,i-1 \\ q_x' q_{x,j} = 0,\ j=1,\ldots,i-1}} \mathsf{Cov}(q_r' r_{t+1}, q_x' x_t)$
	(4) Canonical Correlation	$s_i = \dfrac{\mathsf{Cov}(q_{r,i}' r_{t+1}, q_{x,i}' x_t)}{\sqrt{\mathsf{Var}(q_{r,i}' r_{t+1})}\sqrt{\mathsf{Var}(q_{x,i}' x_t)}}$

Table B.3.2 – Performance Summary

This table shows the out-of-sample performances (in percentages) of various portfolio selection methods using a 21-day momentum signal. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). Panels A-G display the average returns, the Sharpe ratio of returns, the alpha with respect to the Fama-French 5-factor regression augmented with the univariate factor (UNI), the beta of the simple factor, the idiosyncratic volatility, and the information ratio. The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Average Returns (%)						
CP2	2.196	1.334	1.741	0.953	1.809	0.782
PP2	3.762	3.462	3.806	3.490	3.377	3.432
MVO	0.381	0.188	0.358	−0.097	0.365	0.097
UNI	2.268	1.787	2.261	1.477	2.141	1.473
Panel B. Standard Deviation of Returns (%)						
CP2	2.175	1.366	2.048	1.260	2.057	1.248
PP2	5.853	5.690	4.912	4.931	4.642	4.433
MVO	1.857	1.366	1.685	1.301	1.595	1.261
UNI	5.280	4.814	4.173	3.663	4.384	3.970
Panel C. Sharpe Ratio						
CP2	1.010 (6.872)	0.977 (6.655)	0.854 (5.837)	0.759 (5.208)	0.883 (6.028)	0.631 (4.341)
PP2	0.662 (4.554)	0.628 (4.325)	0.786 (5.389)	0.722 (4.954)	0.740 (5.077)	0.785 (5.377)
MVO	0.214 (1.485)	0.145 (1.004)	0.220 (1.528)	−0.068 (−0.473)	0.236 (1.639)	0.083 (0.577)
UNI	0.452 (3.125)	0.393 (2.719)	0.557 (3.843)	0.419 (2.899)	0.506 (3.493)	0.389 (2.690)
Panel D. Alpha (%)						
CP2	1.844	1.189	1.116	0.823	1.340	0.683
PP2	2.252	2.240	2.183	2.347	2.069	2.565
MVO	−0.210	−0.097	−0.342	−0.391	−0.208	−0.214
Panel E. Beta to Simple Factor						
CP2	0.146	0.072	0.259	0.086	0.172	0.054
PP2	0.684	0.695	0.703	0.734	0.583	0.583
MVO	0.256	0.168	0.300	0.201	0.260	0.188
Panel F. Idiosyncratic Volatility (%)						
CP2	2.041	1.321	1.749	1.219	1.910	1.228
PP2	4.583	4.583	3.951	4.157	3.889	3.778
MVO	1.281	1.102	1.136	1.077	1.119	1.018
Panel G. Information Ratio						
CP2	0.903 (6.170)	0.900 (6.149)	0.638 (4.393)	0.675 (4.642)	0.701 (4.819)	0.556 (3.835)
PP2	0.491 (3.396)	0.489 (3.379)	0.553 (3.810)	0.565 (3.892)	0.532 (3.671)	0.679 (4.668)
MVO	−0.164 (−1.137)	−0.088 (−0.609)	−0.301 (−2.087)	−0.363 (−2.512)	−0.186 (−1.292)	−0.210 (−1.456)

Table B.3.3 – Weight Statistics

This table shows the average monthly weight statistics of various portfolio selection methods estimated using a 21-day momentum signal. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). Panels A-F display the turnover, the proportional leverage, the sum of negative weights in the portfolio, and the minimum and maximum portfolio weights. The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Turnover						
CP2	1.158	1.202	1.179	1.225	1.165	1.217
PP2	1.149	1.148	1.167	1.154	1.173	1.162
MVO	1.377	1.424	1.378	1.416	1.381	1.428
UNI	1.252	1.282	1.261	1.295	1.263	1.298
Panel B. Proportional Leverage						
CP2	0.493	0.497	0.499	0.494	0.498	0.495
PP2	0.496	0.497	0.493	0.497	0.500	0.502
MVO	0.499	0.499	0.504	0.497	0.502	0.500
UNI	0.489	0.500	0.489	0.500	0.489	0.500
Panel C. Sum of Negative Weights						
CP2	−0.496	−0.497	−0.498	−0.498	−0.497	−0.498
PP2	−0.500	−0.500	−0.500	−0.500	−0.500	−0.500
MVO	−0.500	−0.500	−0.501	−0.501	−0.500	−0.500
UNI	−0.500	−0.500	−0.500	−0.500	−0.500	−0.500
Panel D. Minimum Weight						
CP2	−0.106	−0.033	−0.106	−0.033	−0.103	−0.032
PP2	−0.102	−0.034	−0.105	−0.037	−0.104	−0.033
MVO	−0.097	−0.030	−0.094	−0.030	−0.094	−0.029
UNI	−0.077	−0.020	−0.077	−0.020	−0.077	−0.020
Panel E. Maximum Weight						
CP2	0.104	0.033	0.103	0.032	0.103	0.032
PP2	0.099	0.033	0.100	0.036	0.103	0.033
MVO	0.098	0.030	0.095	0.029	0.096	0.030
UNI	0.077	0.020	0.077	0.020	0.077	0.020
Panel F. Standard Deviation of Weights						
CP2	0.051	0.013	0.051	0.013	0.051	0.013
PP2	0.051	0.013	0.052	0.013	0.052	0.013
MVO	0.050	0.012	0.050	0.012	0.050	0.012
UNI	0.047	0.012	0.047	0.012	0.047	0.012

Table B.3.4 – Portfolio Returns Decomposition

This table shows the average returns (in percentages) of various portfolio selection methods due to static/dynamic trades and long/short trades. Panels A-C display the results for the dynamic and static trades as well as the share of returns due to dynamic trades. Panels D-E display the results for the long and short trades. The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Static Returns (%)						
CP2	0.184	0.204	0.182	0.159	0.178	0.138
PP2	0.043	0.032	0.209	0.165	0.081	0.026
MVO	0.118	0.159	0.084	0.103	0.142	0.153
UNI	0.168	0.170	0.190	0.166	0.157	0.158
Panel B. Dynamic Returns (%)						
CP2	2.008	1.127	1.561	0.791	1.640	0.657
PP2	3.718	3.429	3.614	3.349	3.302	3.426
MVO	0.263	0.029	0.268	−0.191	0.217	−0.067
UNI	2.097	1.614	2.059	1.303	1.988	1.316
Panel C. Share of Dynamic Returns (%)						
CP2	91.466	84.510	89.383	83.179	90.064	82.603
PP2	98.802	99.069	94.401	95.229	97.558	99.237
MVO	68.917	15.496	75.939	217.015	60.439	−77.498
UNI	92.467	90.326	91.399	88.562	92.652	89.249
Panel D. Long Leg Returns (%)						
CP2	8.012	7.621	7.666	7.505	8.021	7.725
PP2	8.893	8.532	8.664	8.506	8.716	8.924
MVO	7.506	7.306	7.248	7.197	7.570	7.457
UNI	8.057	7.916	7.940	7.822	8.255	8.015
Panel E. Short Leg Returns (%)						
CP2	5.714	6.297	5.785	6.522	6.011	6.848
PP2	4.764	4.747	4.543	4.697	5.081	5.240
MVO	6.939	6.979	6.637	7.152	7.001	7.224
UNI	5.476	5.861	5.442	6.163	5.866	6.347

Table B.3.5 – Subsample Analysis

This table shows the annualized out-of-sample Sharpe ratio of various portfolio selection methods over different subsamples. Panels A-D display the results for Sample 1 (1974 through 1986), Sample 2 (1986 through 1998), Sample 3 (1998 through 2010), and Sample 4 (2010 through 2022). The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Sample Period 1974 to 1986						
CP2	1.254 (4.208)	0.842 (2.874)	1.094 (3.685)	0.401 (1.378)	1.324 (4.411)	0.632 (2.163)
PP2	1.107 (3.741)	1.182 (3.979)	1.264 (4.224)	1.086 (3.660)	1.184 (3.973)	1.222 (4.093)
MVO	0.016 (0.056)	−0.152 (−0.527)	0.358 (1.232)	−0.168 (−0.580)	0.412 (1.419)	0.205 (0.707)
UNI	0.935 (3.181)	0.762 (2.607)	1.130 (3.802)	0.814 (2.773)	1.020 (3.446)	0.838 (2.850)
Panel B. Sharpe Ratio, Sample Period 1986 to 1998						
CP2	1.549 (5.117)	1.577 (5.199)	1.232 (4.124)	1.146 (3.854)	1.224 (4.100)	1.069 (3.604)
PP2	1.051 (3.560)	1.185 (3.989)	1.269 (4.239)	1.159 (3.892)	1.222 (4.094)	1.216 (4.074)
MVO	0.589 (2.025)	0.283 (0.977)	0.344 (1.185)	−0.125 (−0.431)	0.470 (1.616)	0.039 (0.134)
UNI	0.806 (2.755)	0.720 (2.466)	0.652 (2.230)	0.405 (1.393)	0.623 (2.134)	0.380 (1.307)
Panel C. Sharpe Ratio, Sample Period 1998 to 2010						
CP2	0.883 (3.010)	1.263 (4.237)	0.909 (3.086)	1.126 (3.788)	1.002 (3.390)	0.894 (3.034)
PP2	0.408 (1.407)	0.389 (1.342)	0.807 (2.747)	1.011 (3.418)	0.693 (2.369)	0.790 (2.692)
MVO	0.283 (0.977)	0.351 (1.214)	0.201 (0.693)	−0.109 (−0.378)	0.161 (0.557)	0.121 (0.416)
UNI	0.330 (1.140)	0.329 (1.137)	0.467 (1.606)	0.431 (1.483)	0.399 (1.373)	0.334 (1.151)
Panel D. Sharpe Ratio, Sample Period 2010 to 2022						
CP2	0.448 (1.539)	0.389 (1.339)	0.238 (0.819)	0.322 (1.105)	0.117 (0.401)	−0.174 (−0.599)
PP2	0.350 (1.205)	0.149 (0.515)	0.010 (0.034)	−0.213 (−0.731)	−0.136 (−0.468)	−0.054 (−0.185)
MVO	−0.062 (−0.212)	−0.006 (−0.021)	0.054 (0.186)	0.038 (0.132)	−0.041 (−0.140)	−0.160 (−0.552)
UNI	−0.011 (−0.037)	−0.031 (−0.108)	0.147 (0.505)	0.120 (0.412)	0.104 (0.358)	0.095 (0.327)

Table B.3.6 – Forecast Horizon

This table shows the annualized out-of-sample Sharpe ratio of various portfolio selection methods for different rebalancing frequencies. Panels A-C display the results for 1 day, 5 days, and 10 days. The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. Panel A is based on 14550 daily out-of-sample returns from 01-06-1965 until 10-21-2022, Panel B is based on 2814 weekly out-of-sample returns from 12-06-1966 until 10-21-2022, Panel C is based on 1347 fortnightly out-of-sample returns from 06-12-1969 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Horizon: 1 day						
CP2	2.104 (15.920)	1.301 (9.869)	1.853 (14.031)	0.993 (7.535)	1.811 (13.717)	1.091 (8.280)
PP2	1.226 (9.299)	1.070 (8.119)	1.155 (8.767)	0.796 (6.042)	1.336 (10.133)	0.957 (7.266)
MVO	1.089 (8.264)	−0.381 (−2.896)	0.786 (5.971)	−0.727 (−5.518)	0.816 (6.195)	−0.685 (−5.201)
UNI	1.147 (8.702)	0.660 (5.016)	0.981 (7.444)	0.320 (2.434)	1.012 (7.679)	0.364 (2.762)
Panel B. Sharpe Ratio, Horizon: 5 days						
CP2	1.527 (11.279)	1.187 (8.810)	1.372 (10.156)	0.848 (6.313)	1.383 (10.233)	1.074 (7.980)
PP2	1.002 (7.448)	1.033 (7.676)	1.040 (7.728)	0.934 (6.946)	1.207 (8.953)	1.078 (8.004)
MVO	0.555 (4.140)	−0.147 (−1.096)	0.415 (3.097)	−0.449 (−3.351)	0.486 (3.627)	−0.179 (−1.336)
UNI	0.971 (7.222)	0.681 (5.075)	0.861 (6.410)	0.453 (3.378)	0.853 (6.350)	0.465 (3.469)
Panel C. Sharpe Ratio, Horizon: 10 days						
CP2	1.302 (9.356)	1.025 (7.416)	1.382 (9.917)	1.033 (7.469)	1.287 (9.255)	0.989 (7.160)
PP2	0.925 (6.702)	0.952 (6.892)	0.950 (6.882)	0.827 (6.005)	0.864 (6.266)	0.914 (6.627)
MVO	0.422 (3.078)	0.154 (1.123)	0.365 (2.666)	−0.154 (−1.125)	0.459 (3.351)	0.028 (0.202)
UNI	0.789 (5.727)	0.686 (4.993)	0.861 (6.245)	0.626 (4.555)	0.778 (5.650)	0.569 (4.142)

Table B.3.7 – Shrinkage Intensity

This table shows the annualized out-of-sample Sharpe Ratio of Canonical Portfolio with two components for different shrinkage intensities in the estimation of the covariance of the signals: $\delta_x \in \{0, 0.2, 0.5, 1\}$. Panels A-E display the results for the full sample period, Sample 1 (1974 through 1986), Sample 2 (1986 through 1998), Sample 3 (1998 through 2010), and Sample 4 (2010 through 2022). The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Full Sample Period						
CP2 ($\delta_x = 1$)	0.986 (6.707)	0.939 (6.400)	0.842 (5.752)	0.634 (4.359)	0.849 (5.798)	0.588 (4.047)
CP2 ($\delta_x = 0.8$)	0.994 (6.764)	0.995 (6.766)	0.848 (5.791)	0.851 (5.817)	0.875 (5.970)	0.661 (4.545)
CP2 ($\delta_x = 0.5$)	0.973 (6.624)	0.849 (5.805)	0.836 (5.712)	0.914 (6.232)	0.869 (5.935)	0.728 (4.994)
CP2 ($\delta_x = 0.2$)	0.981 (6.679)	0.491 (3.393)	0.721 (4.949)	0.548 (3.778)	0.898 (6.125)	0.536 (3.694)
Panel B. Sharpe Ratio, Subsample Period 1974 to 1986						
CP2 ($\delta_x = 1$)	1.178 (3.969)	0.966 (3.284)	1.054 (3.558)	0.396 (1.364)	1.222 (4.092)	0.431 (1.482)
CP2 ($\delta_x = 0.8$)	1.315 (4.401)	0.757 (2.593)	1.103 (3.714)	0.495 (1.701)	1.384 (4.598)	0.752 (2.566)
CP2 ($\delta_x = 0.5$)	1.276 (4.278)	0.485 (1.672)	1.205 (4.040)	0.479 (1.645)	1.513 (4.990)	0.852 (2.896)
CP2 ($\delta_x = 0.2$)	1.280 (4.291)	0.059 (0.203)	0.937 (3.179)	−0.036 (−0.123)	1.580 (5.192)	0.486 (1.670)
Panel C. Sharpe Ratio, Subsample Period 1986 to 1998						
CP2 ($\delta_x = 1$)	1.465 (4.863)	1.406 (4.680)	1.134 (3.814)	0.728 (2.485)	1.159 (3.892)	1.043 (3.521)
CP2 ($\delta_x = 0.8$)	1.599 (5.267)	1.676 (5.493)	1.303 (4.348)	1.358 (4.518)	1.243 (4.158)	1.081 (3.644)
CP2 ($\delta_x = 0.5$)	1.703 (5.571)	1.548 (5.113)	1.288 (4.300)	1.489 (4.917)	1.061 (3.580)	1.147 (3.854)
CP2 ($\delta_x = 0.2$)	1.939 (6.245)	1.244 (4.176)	1.200 (4.023)	1.036 (3.500)	0.986 (3.336)	0.783 (2.669)
Panel D. Sharpe Ratio, Subsample Period 1998 to 2010						
CP2 ($\delta_x = 1$)	0.883 (3.009)	1.156 (3.896)	0.958 (3.244)	1.209 (4.052)	1.065 (3.594)	1.067 (3.599)
CP2 ($\delta_x = 0.8$)	0.852 (2.907)	1.295 (4.338)	0.869 (2.954)	1.025 (3.463)	0.912 (3.096)	0.798 (2.720)
CP2 ($\delta_x = 0.5$)	0.677 (2.322)	1.132 (3.819)	0.715 (2.441)	0.749 (2.556)	0.724 (2.473)	0.641 (2.193)
CP2 ($\delta_x = 0.2$)	0.489 (1.687)	0.448 (1.547)	0.769 (2.621)	0.285 (0.980)	0.812 (2.765)	0.591 (2.025)
Panel E. Sharpe Ratio, Subsample Period 2010 to 2022						
CP2 ($\delta_x = 1$)	0.487 (1.671)	0.376 (1.295)	0.228 (0.782)	0.121 (0.416)	0.087 (0.298)	−0.308 (−1.059)
CP2 ($\delta_x = 0.8$)	0.360 (1.240)	0.392 (1.348)	0.233 (0.802)	0.516 (1.764)	0.113 (0.390)	−0.023 (−0.078)
CP2 ($\delta_x = 0.5$)	0.437 (1.501)	0.272 (0.937)	0.376 (1.288)	0.984 (3.318)	0.330 (1.133)	0.338 (1.160)
CP2 ($\delta_x = 0.2$)	0.524 (1.799)	0.221 (0.762)	0.143 (0.490)	0.912 (3.083)	0.421 (1.442)	0.261 (0.898)

Table B.3.8 – Momentum Lookback Window

This table shows the annualized out-of-sample Sharpe Ratio of various portfolio selection methods based on momentum signals of different lookback window sizes. Panels A-E display the results for 42 days, 63 days, 84 days, 126 days, and 252 days. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Lookback Window Size: 42						
CP2	0.887 (6.054)	1.006 (6.837)	0.863 (5.895)	0.740 (5.074)	0.909 (6.198)	0.707 (4.853)
PP2	0.649 (4.465)	0.624 (4.298)	0.668 (4.587)	0.545 (3.759)	0.780 (5.342)	0.688 (4.728)
MVO	0.238 (1.647)	0.199 (1.380)	0.208 (1.444)	−0.051 (−0.355)	0.212 (1.468)	0.110 (0.761)
UNI	0.438 (3.031)	0.366 (2.532)	0.502 (3.462)	0.398 (2.748)	0.432 (2.982)	0.327 (2.265)
Panel B. Sharpe Ratio, Lookback Window Size: 63						
CP2	0.665 (4.575)	0.982 (6.681)	0.871 (5.948)	0.787 (5.389)	0.901 (6.143)	0.909 (6.199)
PP2	0.612 (4.216)	0.575 (3.962)	0.716 (4.913)	0.663 (4.555)	0.674 (4.627)	0.652 (4.479)
MVO	0.328 (2.273)	0.288 (1.995)	0.185 (1.284)	−0.195 (−1.354)	0.253 (1.753)	0.177 (1.229)
UNI	0.420 (2.901)	0.366 (2.536)	0.479 (3.308)	0.336 (2.325)	0.393 (2.716)	0.286 (1.979)
Panel C. Sharpe Ratio, Lookback Window Size: 84						
CP2	0.625 (4.306)	0.908 (6.196)	0.674 (4.628)	0.543 (3.743)	0.616 (4.237)	0.854 (5.834)
PP2	0.508 (3.504)	0.453 (3.129)	0.409 (2.826)	0.482 (3.329)	0.465 (3.211)	0.502 (3.460)
MVO	0.264 (1.829)	0.305 (2.110)	0.161 (1.117)	−0.097 (−0.675)	0.214 (1.480)	0.313 (2.169)
UNI	0.277 (1.917)	0.273 (1.891)	0.340 (2.352)	0.259 (1.796)	0.248 (1.715)	0.242 (1.678)
Panel D. Sharpe Ratio, Lookback Window Size: 126						
CP2	0.670 (4.605)	0.885 (6.042)	0.545 (3.755)	0.486 (3.352)	0.607 (4.179)	0.769 (5.270)
PP2	0.459 (3.169)	0.428 (2.957)	0.442 (3.055)	0.561 (3.865)	0.359 (2.486)	0.275 (1.907)
MVO	0.314 (2.172)	0.379 (2.620)	0.117 (0.808)	−0.107 (−0.743)	0.337 (2.331)	0.257 (1.780)
UNI	0.355 (2.455)	0.311 (2.154)	0.346 (2.390)	0.276 (1.910)	0.295 (2.040)	0.254 (1.760)
Panel E. Sharpe Ratio, Lookback Window Size: 252						
CP2	1.008 (6.851)	1.200 (8.089)	0.805 (5.511)	0.722 (4.953)	0.949 (6.460)	1.129 (7.629)
PP2	0.585 (4.035)	0.511 (3.526)	0.534 (3.683)	0.530 (3.654)	0.511 (3.521)	0.330 (2.282)
MVO	0.501 (3.458)	0.578 (3.986)	0.362 (2.502)	0.043 (0.300)	0.469 (3.240)	0.391 (2.701)
UNI	0.437 (3.023)	0.386 (2.672)	0.524 (3.612)	0.449 (3.103)	0.441 (3.042)	0.383 (2.650)

Table B.3.9 – Approximate versus Actual Solution

This table shows the out-of-sample performances (in percentages) of the 'approximated' versus 'actual' portfolio policy using a 21-day momentum signal. Panels A-G display the results for the average returns, the Sharpe ratio of returns, the alpha with respect to the Fama-French 5-factor regression augmented with the univariate factor (UNI), the beta of the simple factor, the idiosyncratic volatility, and the information ratio. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Average Returns (%)						
CP2 (Approx)	2.199	1.341	1.731	0.957	1.797	0.777
CP2 (Full)	2.253	1.175	1.578	0.822	1.675	0.596
Panel B. Standard Deviation of Returns (%)						
CP2 (Approx)	2.177	1.366	2.049	1.260	2.057	1.248
CP2 (Full)	2.210	1.356	2.119	1.229	2.061	1.309
Panel C. Sharpe Ratio						
CP2 (Approx)	1.011 (6.872)	0.982 (6.685)	0.848 (5.797)	0.763 (5.226)	0.877 (5.985)	0.626 (4.308)
CP2 (Full)	1.020 (6.933)	0.868 (5.934)	0.750 (5.140)	0.672 (4.620)	0.817 (5.588)	0.460 (3.177)
Panel D. Alpha (%)						
CP2 (Approx)	1.849	1.197	1.109	0.826	1.334	0.679
CP2 (Full)	1.908	1.043	0.942	0.739	1.224	0.536
Panel E. Beta to Simple Factor						
CP2 (Approx)	0.146	0.071	0.259	0.086	0.172	0.053
CP2 (Full)	0.144	0.067	0.267	0.057	0.165	0.046
Panel F. Idiosyncratic Volatility (%)						
CP2 (Approx)	2.043	1.321	1.749	1.220	1.911	1.229
CP2 (Full)	2.083	1.315	1.807	1.208	1.921	1.297
Panel G. Information Ratio						
CP2 (Approx)	0.905 (6.176)	0.906 (6.183)	0.634 (4.360)	0.677 (4.651)	0.698 (4.792)	0.553 (3.809)
CP2 (Full)	0.916 (6.247)	0.793 (5.434)	0.522 (3.596)	0.612 (4.210)	0.637 (4.379)	0.414 (2.858)

Table B.3.10 – Performance Summary for Two Signals Case

This table shows the out-of-sample performances (in percentages) of various portfolio selection methods using two momentum signals with a lookback window of 21 days and 252 days. Panels A-G display the results for the average returns, the Sharpe ratio of returns, the alpha with respect to the Fama-French 5-factor regression augmented with the univariate factor (UNI), the beta of the simple factor, the idiosyncratic volatility, and the information ratio. The test assets are 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Average Returns (%)						
CP2	2.925	1.807	2.308	1.027	2.274	1.551
PP2	4.160	4.143	3.779	4.082	2.893	2.522
MVO	0.761	0.581	0.595	−0.032	0.746	0.386
UNI	2.882	2.316	2.743	2.047	2.431	1.838
Panel B. Standard Deviation of Returns (%)						
CP2	2.343	1.384	2.153	1.312	2.072	1.293
PP2	6.042	6.053	5.288	5.488	4.707	4.760
MVO	1.893	1.431	1.807	1.332	1.642	1.347
UNI	5.489	5.097	4.543	4.095	4.619	4.280
Panel C. Sharpe Ratio						
CP2	1.244 (8.365)	1.302 (8.735)	1.072 (7.266)	0.786 (5.387)	1.096 (7.426)	1.197 (8.073)
PP2	0.706 (4.849)	0.702 (4.823)	0.729 (5.005)	0.758 (5.197)	0.630 (4.338)	0.548 (3.777)
MVO	0.410 (2.836)	0.412 (2.851)	0.337 (2.336)	−0.017 (−0.120)	0.461 (3.184)	0.293 (2.028)
UNI	0.546 (3.766)	0.476 (3.285)	0.619 (4.263)	0.516 (3.561)	0.544 (3.751)	0.448 (3.093)
Panel D. Alpha (%)						
CP2	2.287	1.576	1.508	0.781	1.641	1.300
PP2	2.048	2.097	1.691	2.263	1.318	1.205
MVO	0.025	0.185	−0.195	−0.391	0.117	−0.003
Panel E. Beta to Simple Factor						
CP2	0.206	0.100	0.277	0.106	0.202	0.115
PP2	0.735	0.866	0.783	0.922	0.600	0.716
MVO	0.254	0.174	0.291	0.171	0.257	0.198
Panel F. Idiosyncratic Volatility (%)						
CP2	2.040	1.285	1.745	1.240	1.856	1.198
PP2	4.494	4.155	3.909	3.986	3.858	3.673
MVO	1.288	1.124	1.235	1.142	1.136	1.058
Panel G. Information Ratio						
CP2	1.121 (7.584)	1.226 (8.254)	0.864 (5.906)	0.630 (4.336)	0.884 (6.037)	1.084 (7.349)
PP2	0.456 (3.149)	0.505 (3.485)	0.433 (2.990)	0.568 (3.914)	0.342 (2.364)	0.328 (2.272)
MVO	0.019 (0.133)	0.164 (1.139)	−0.158 (−1.095)	−0.343 (−2.373)	0.103 (0.715)	−0.003 (−0.019)

Table B.3.11 – Weight Statistics for Two Signals Case

This table shows the average monthly weight statistics of various portfolio selection methods estimated using a 21-day momentum signal. Panels A-F display the results for the turnover, the proportional leverage, the sum of negative weights in the portfolio, and the minimum and maximum portfolio weights. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Turnover						
CP2	0.749	0.437	0.806	0.542	0.762	0.466
PP2	0.806	0.763	0.878	0.794	0.891	0.856
MVO	1.064	1.075	1.065	1.082	1.060	1.079
UNI	0.959	0.978	0.970	0.990	0.969	0.989
Panel B. Proportional Leverage						
CP2	0.477	0.473	0.480	0.488	0.470	0.486
PP2	0.478	0.483	0.472	0.473	0.477	0.484
MVO	0.506	0.499	0.509	0.503	0.497	0.497
UNI	0.485	0.498	0.488	0.497	0.485	0.498
Panel C. Sum of Negative Weights						
CP2	−0.483	−0.480	−0.486	−0.482	−0.481	−0.479
PP2	−0.500	−0.500	−0.500	−0.500	−0.500	−0.500
MVO	−0.499	−0.499	−0.500	−0.501	−0.498	−0.498
UNI	−0.500	−0.500	−0.500	−0.500	−0.500	−0.500
Panel D. Minimum Weight						
CP2	−0.121	−0.035	−0.113	−0.036	−0.114	−0.035
PP2	−0.104	−0.034	−0.108	−0.037	−0.104	−0.033
MVO	−0.099	−0.031	−0.096	−0.030	−0.097	−0.031
UNI	−0.090	−0.025	−0.089	−0.025	−0.089	−0.024
Panel E. Maximum Weight						
CP2	0.109	0.032	0.102	0.032	0.102	0.032
PP2	0.092	0.031	0.096	0.033	0.096	0.030
MVO	0.102	0.031	0.099	0.031	0.097	0.030
UNI	0.088	0.025	0.087	0.025	0.089	0.025
Panel F. Standard Deviation of Weights						
CP2	0.053	0.013	0.052	0.013	0.052	0.013
PP2	0.051	0.013	0.052	0.013	0.051	0.013
MVO	0.051	0.013	0.050	0.012	0.050	0.012
UNI	0.049	0.012	0.049	0.012	0.049	0.012

B.4 Tables for Chapter 5

Table B.4.1 – Parameter Values for Numerical Experiments

This table reports the default parameter value configurations used in the numerical experiments described in Section 5.4. The trading frequency is taken to be daily.

Parameter	Units	Configuration 1 Value	Configuration 2 Value
A	N/A	$\begin{bmatrix} 50 & 0 \\ 0 & 10 \end{bmatrix}$	$\begin{bmatrix} 50 & 0 \\ 0 & 10 \end{bmatrix}$
Φ	N/A	$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0 \\ 0 & 0.002 \end{bmatrix}$
Σ_r	$\$^2/\text{share}^2$	$\begin{bmatrix} 1.2 & 0 \\ 0 & 1.2 \end{bmatrix}$	$\begin{bmatrix} 1.2 & 0 \\ 0 & 1.2 \end{bmatrix}$
$\tilde{\Lambda}$	$\$/\text{share}^2$	N/A	$\lambda\Sigma_r$
w_{-1}	share	$\begin{bmatrix} 10^6 \\ 10^6 \end{bmatrix}$	$\begin{bmatrix} 10^6 \\ 10^6 \end{bmatrix}$
x_0	share	$\begin{bmatrix} 10^6 \\ 10^6 \end{bmatrix}$	$\begin{bmatrix} 10^6 \\ 10^6 \end{bmatrix}$
D_0	\$/share	N/A	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$
γ	1/\$	10^{-8}	10^{-8}
λ	$\$/\text{share}^2$	2×10^{-7}	2×10^{-7}
ρ	N/A	$1-(1-0.02)^{\frac{1}{252}}$	$1-(1-0.02)^{\frac{1}{252}}$
R	N/A	N/A	0.1
q	N/A	N/A	0.1

Appendix C Figures

C.1 Figures for Chapter 1

Figure C.1.1 – Mean-Variance Efficient Frontier

This figure illustrates the mean-variance efficient frontier concept in portfolio theory. The x-axis represents the standard deviation of an investment portfolio's returns, while the y-axis represents the expected return of the portfolio. Each black node represents a possible risk return combination that can be generated by a hypothetical basket of securities. The black solid curve is the efficient frontier, which represents the best return to risk trade-off comprising of risky securities. The gray dashed line traces the capital market line (CML), which extends the concept of the efficient frontier by incorporating a risk-free asset into the mix; its intercept corresponds to the return on the risk-free rate, and its slope–representing the Sharpe ratio–reflects the additional excess return per unit of risk achieved by moving up the line. The red node denotes the 'tangency' portfolio, which is the portfolio of risky securities with the highest Sharpe ratio.

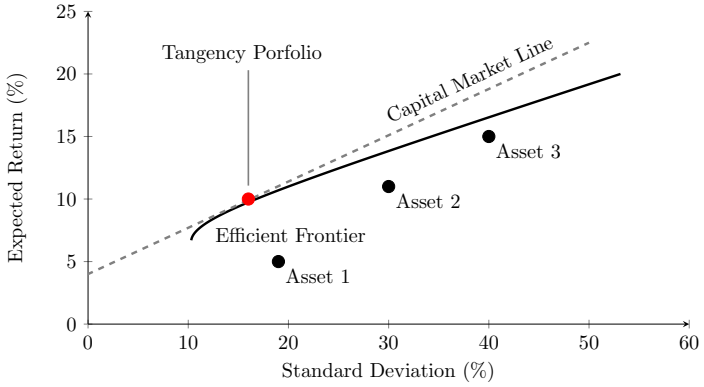
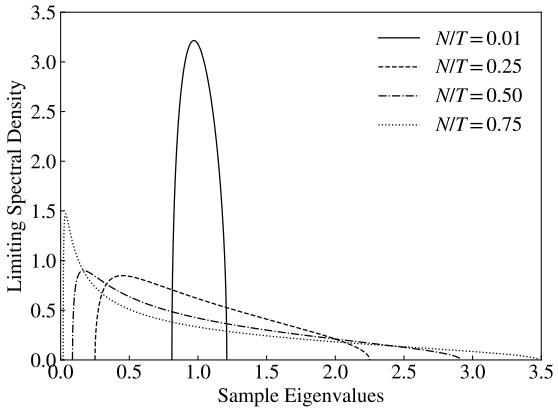


Figure C.1.2 – Marčhenko-Pastur Law

This figure displays the limiting spectral density of sample eigenvalues across various concentration ratios, denoted as N/T, under the specific condition where all population eigenvalues are located at one.



C.2 Figures for Chapter 2

Figure C.2.1 – Marčhenko-Pastur Law for the Exponentially-Weighted Sample Covariance Matrix

This figure the limiting spectral density of sample eigenvalues corresponding to an EWA random matrix with an identity population covariance for various values of $\tilde{c}$ with $c = 0.5$ (left panel) and various values of c with $\tilde{c} = 0.5$ (right panel).

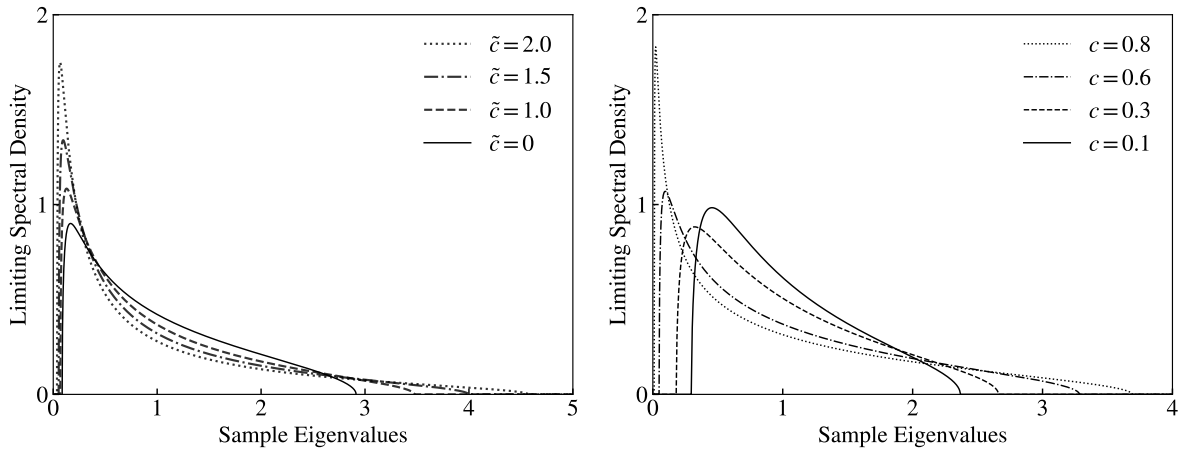


Figure C.2.2 – Schematic Overview of Five-Fold Cross-Validation

The figure displays a schematic diagram of a five-fold cross-validation scheme for estimating the vector of out-of-sample variances for the sample eigenvector portfolios. The light-shaded rectangles represent the training sets used to compute the sample eigenvectors, whereas the dark-shaded rectangles represent the testing sets used to estimate the vector of out-of-sample variances of the eigenvector portfolios.

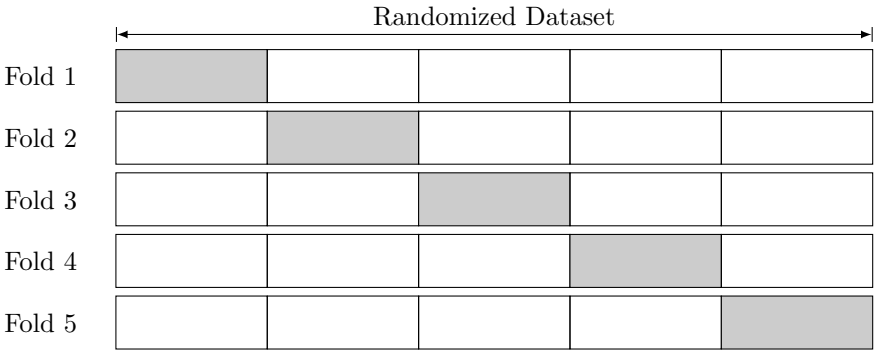
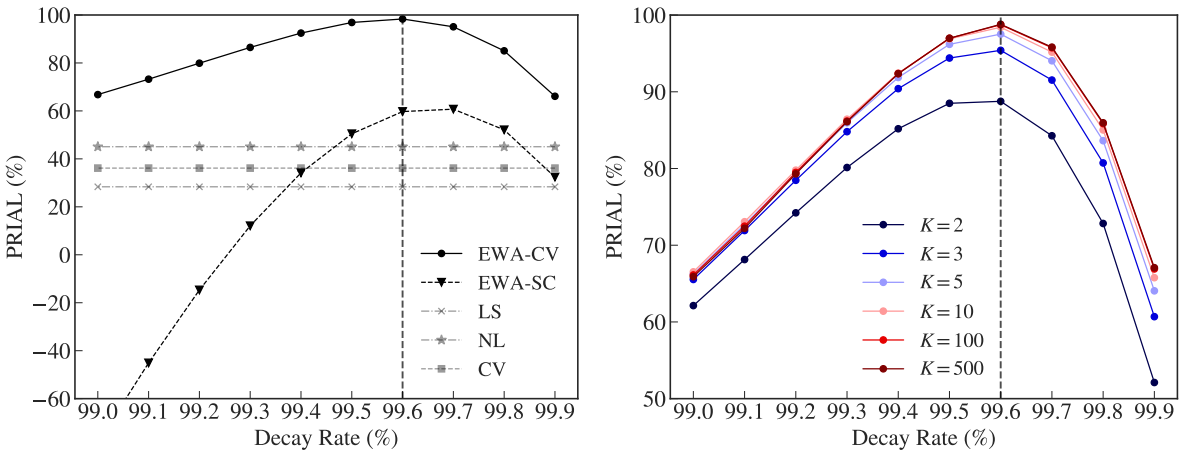


Figure C.2.3 – Out-of-Sample Risk Improvement of Different Estimators

The figures display the PRIAL of several candidate covariance estimators with respect to the standard sample covariance. The left panel compares uniformly-weighted and exponentially-weighted estimators over different decay rates β. The return generating process is stated in Section 2.5.1. The vertical dashed line at 0.996 corresponds to the intrinsic decay rate of population covariance. The average is taken over 100 repetitions. The right panel compares EWA-CV for the different number of folds K under the same setup.



C.3 Figures for Chapter 3

Figure C.3.1 – Binary Tree of Five Elements

Illustrative example of a binary tree associated with a system of $N = 5$ elements (corresponding to the leaves in the tree) with $N - 1 = 4$ nodes. Left panel: the numbers label the leaves and the variables $\{g_1, g_2, g_3, g_4\}$ label the nodes. Right panel: the stocks label the root nodes and the industry factors label the nodes.

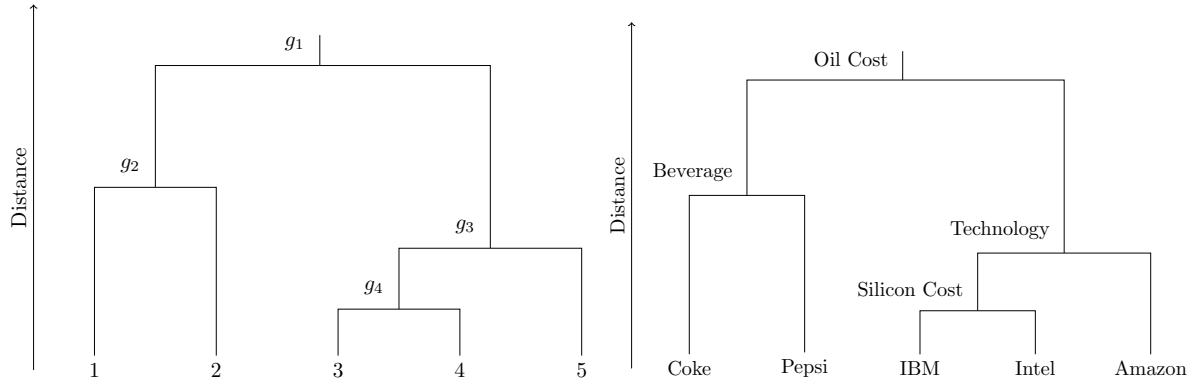


Figure C.3.2 – Dendrogram of Stocks Returns

This figure displays the dendrograms of the U.S. common stock returns (left panel) and residuals (right panel) obtained with an average linkage criterion over the period 1994 to 1999. Each leaf of the dendrogram represents a specific stock, with the color corresponding to the industry assignment defined by GICS; see Table B.2.1. The size of the investment universe is $N = 100$. The stocks in this sample are divided into the following sectors: 14 communication services, 9 consumer discretionary, 10 consumer staples, 4 energy, 18 financials, 12 health care, 6 industrials, 23 information technology, 3 materials, 1 utility.

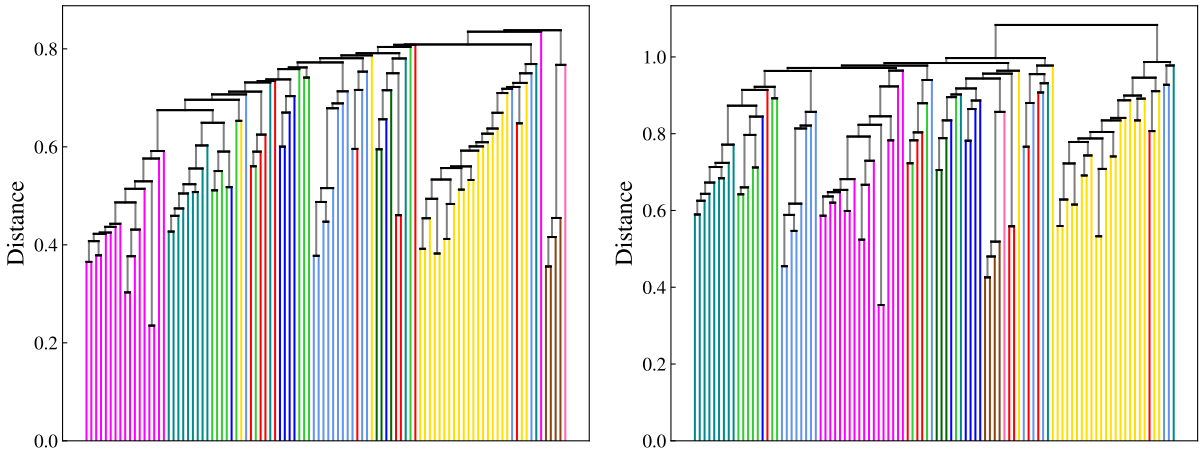


Figure C.3.3 – Heatmaps of Stock Returns Correlation Matrix

This figure displays the heatmap of the sample correlation matrix of U.S. common stock returns with $N = 100$ where the color code represents the correlation coefficient. Both x and y-axis cover the stock index. Left panel: Stock indices of the matrix are unordered. Right panel: Stock indices of the matrix are ordered according to the output of a hierarchical clustering average linkage procedure.

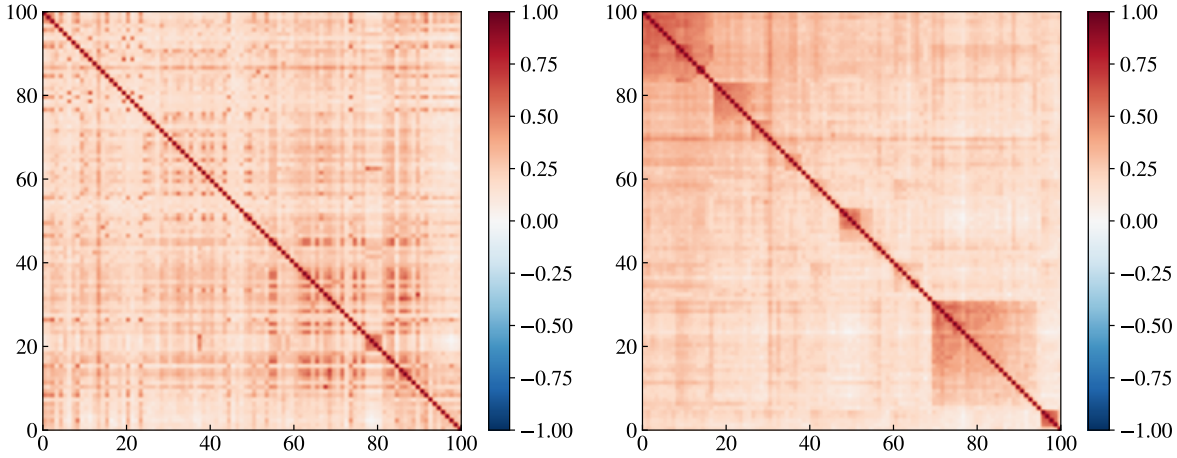


Figure C.3.4 – Heatmap of Regularized Stock Returns Correlation Matrix

This figure displays the heatmap of the regularized correlation matrix. The stock indices of the matrix are ordered according to the output of a hierarchical clustering average linkage procedure.

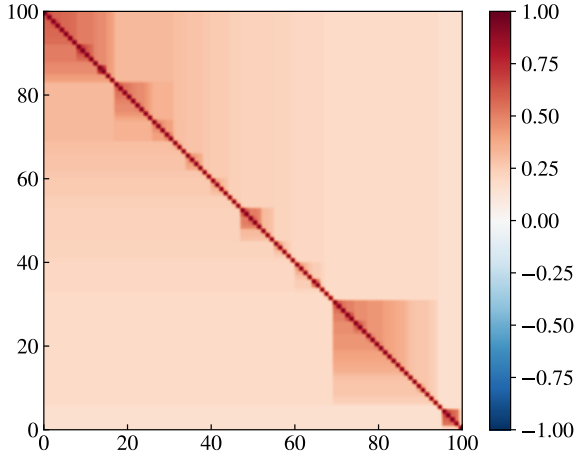


Figure C.3.5 – Heatmaps of Residual Stock Returns Correlation Matrix

This figure displays the heatmap of the sample correlation matrix of residualized stock returns with $N = 100$. Left panel: Stock indices of the matrix are unordered. Right panel: Stock indices of the matrix are ordered according to the output of a hierarchical clustering average linkage procedure.

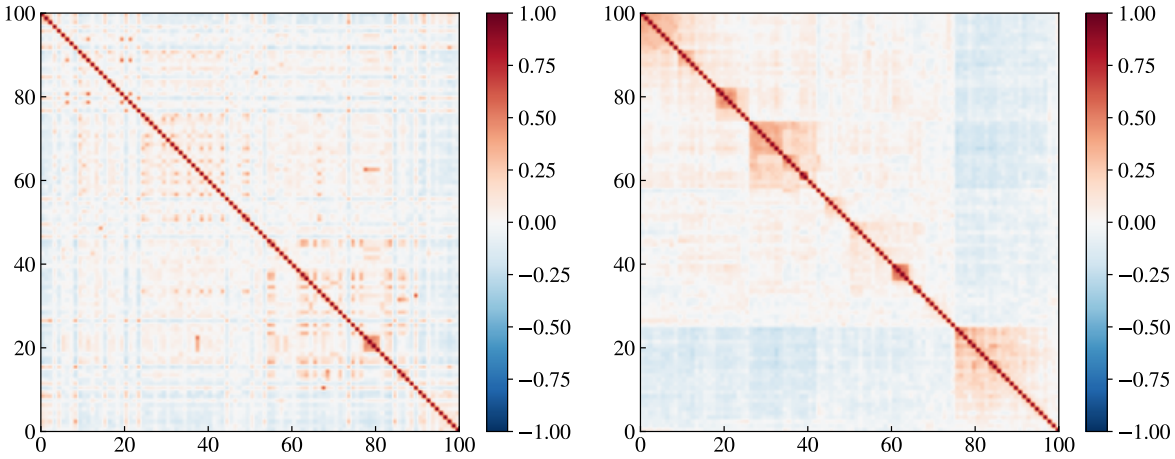
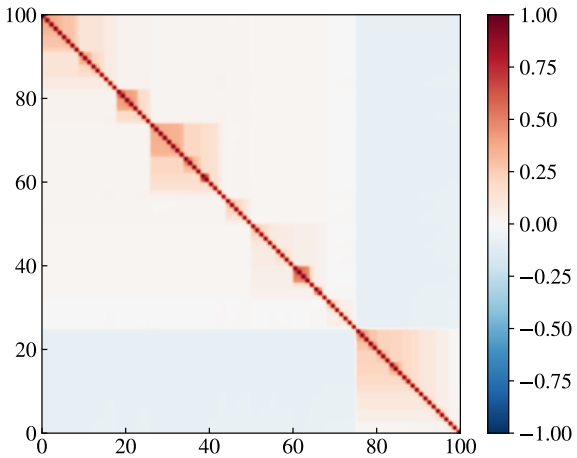


Figure C.3.6 – Heatmap of Regularized Residual Stock Returns Correlation Matrix

This figure displays the heatmap of the regularized residual correlation matrix. The stock indices of the matrix are ordered according to the output of a hierarchical clustering average linkage procedure.



C.4 Figures for Chapter 4

Figure C.4.1 – Wachter Law

This figure displays the limiting spectral density of the sample canonical correlations of Wachter (1980) for different ratios N/T and M/T. The population cross-covariance matrix is the zero matrix, and so all the population canonical correlation coefficients are equal to zero.

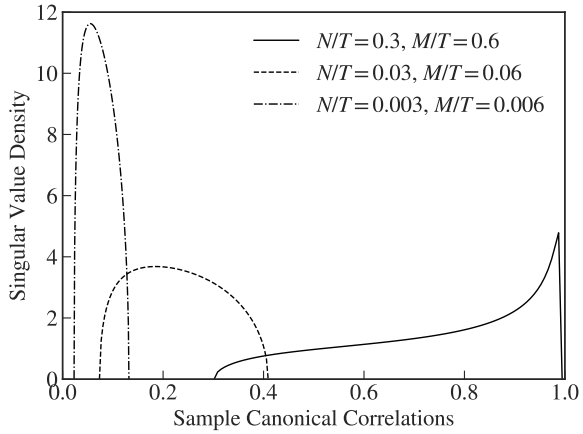


Figure C.4.2 – Optimal Singular Value Adjustment Versus Linear Approximation

This figure displays the optimal adjustment of the ith singular value versus an unadjusted ith singular value.

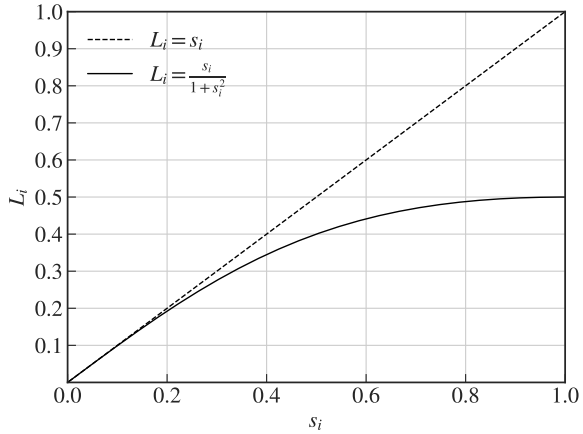
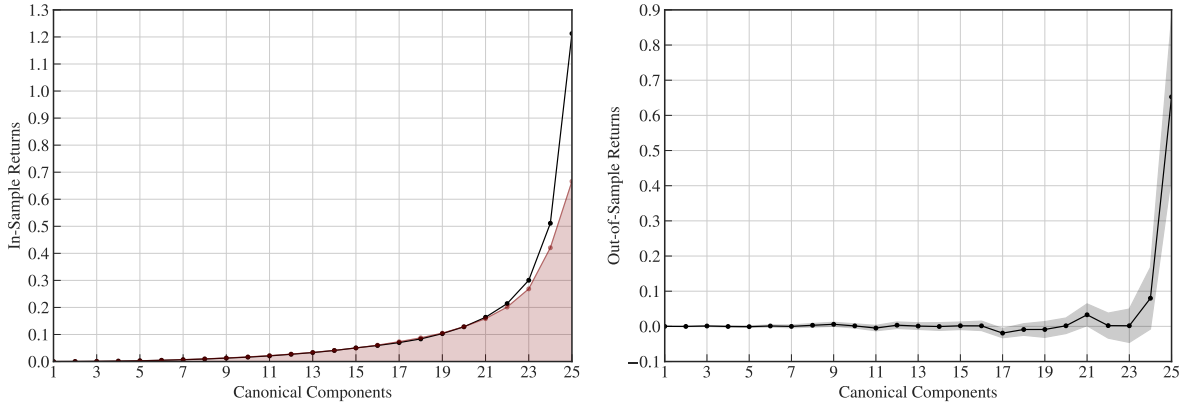


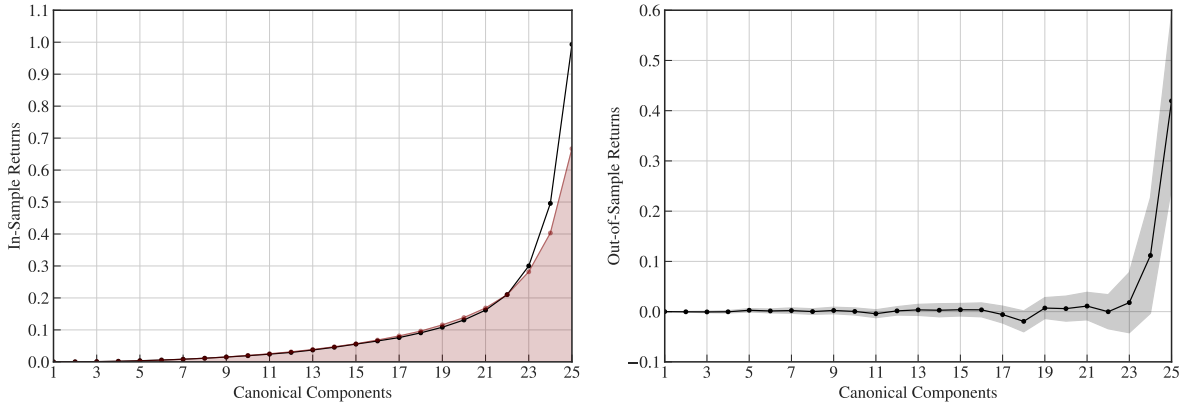
Figure C.4.3 – In-Sample and Out-of-Sample Canonical Portfolio Returns

This figure displays the in-sample and out-of-sample returns for each canonical portfolio. The left panel shows the squared in-sample regularized canonical correlations coefficients (black dots) overlaid with their pseudo-random generated counterparts (maroon dots). The right panel shows the out-of-sample returns of each canonical portfolio overlaid with a ±2 standard error band. Each dot corresponds to values that are averaged over the rebalancing dates. The estimates are obtained by using a 21-day momentum signal to predict subsequent monthly returns and 120 non-overlapping monthly returns. The test assets are the FF25 (top panel), ME/OP25 (middle panel), and MEINV25 (bottom panel). The out-of-sample period is covered from 09-09-1974 until 10-21-2022.

(a) 25 Size and Book-to-Market



(b) 25 Size and Operating Profitability



(c) 25 Size and Investment

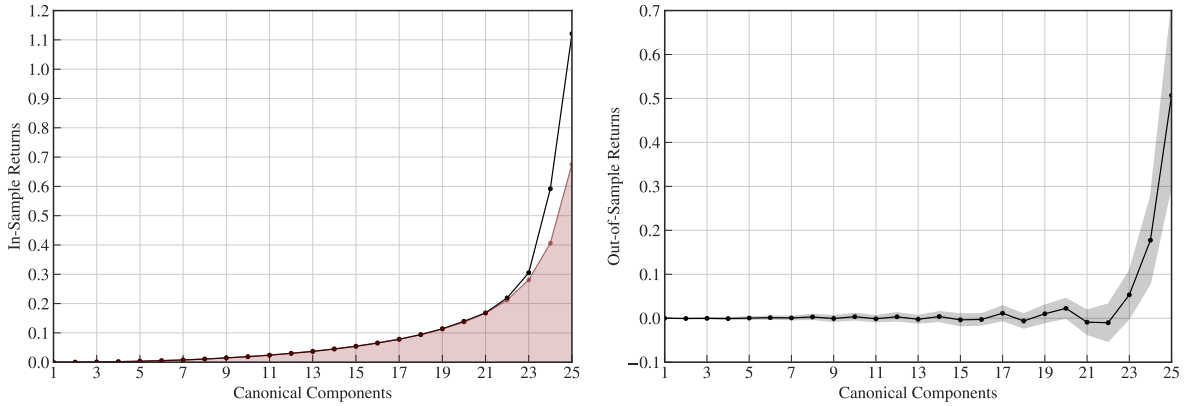
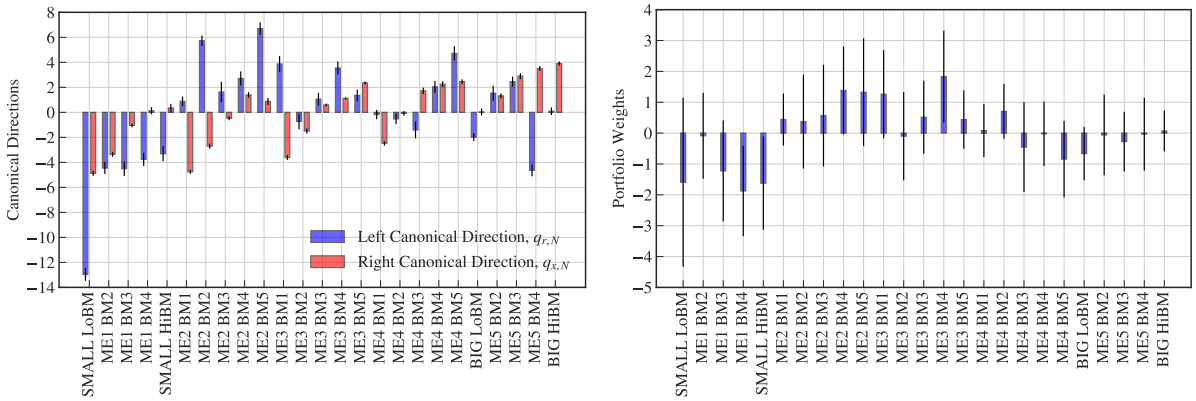


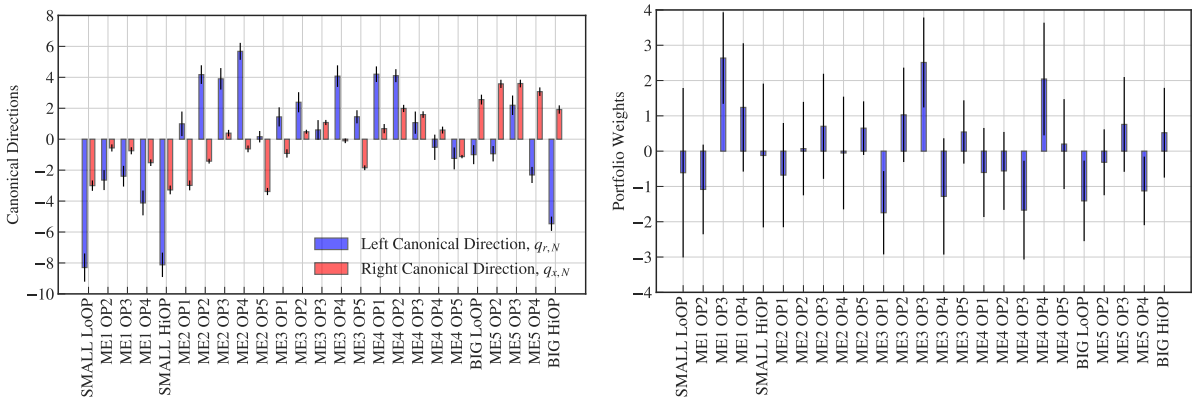
Figure C.4.4 – Weights for the Top Canonical Portfolio

This figure displays the weights of the leading canonical portfolio. The left panel shows the canonical 'weights' on the returns (blue) and signals (red). The right panel shows the final portfolio weights invested in each obtained from combining the returns and signals together through their canonical weights and scaling them up with the corresponding canonical correlation. The height of each bar corresponds to values that are averaged over the rebalancing dates, and they are overlaid with a ±2 standard error band. The estimates are obtained by using a 21-day momentum signal to predict subsequent monthly returns and 120 non-overlapping monthly returns. The test assets are the FF25 (top panel), ME/OP25 (middle panel), and MEINV25 (bottom panel). The out-of-sample period is covered from 09-09-1974 until 10-21-2022.

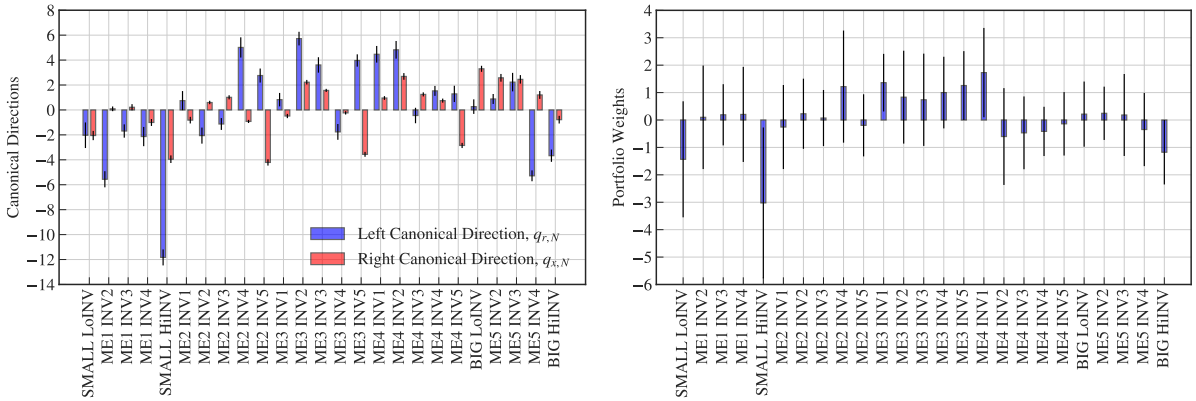
(a) 25 Size and Book-to-Market



(b) 25 Size and Operating Profitability



(c) 25 Size and Investment

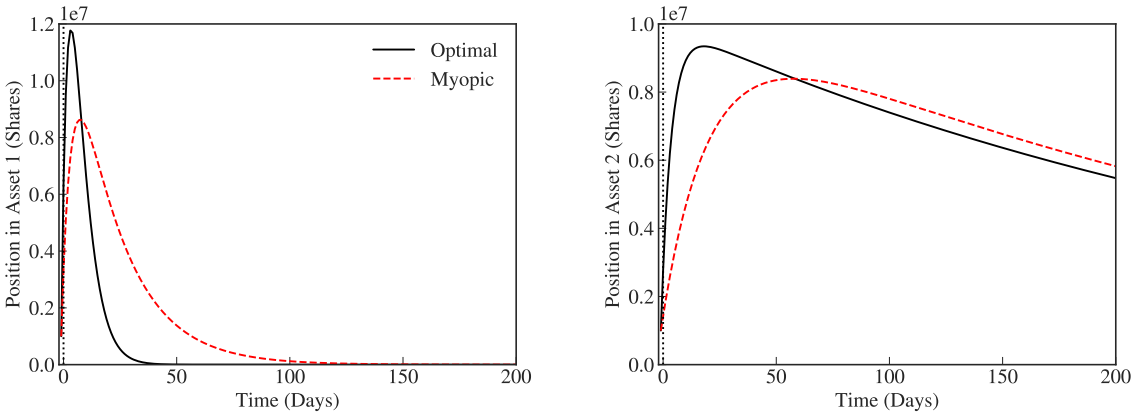


C.5 Figures for Chapter 5

Figure C.5.1 – Impulse Response to Shock under Temporary Price Impact

This figure displays the optimal portfolio response over time following a unit exogenous shock to the signals based on the parameter values of Configuration 1 from Table B.4.1. In numerical experiment 1 (top panels), Asset 1 is exposed to a temporary signal, while Asset 2 is exposed to a persistent signal. In numerical experiment 2 (bottom panels), each asset is also exposed to the signal from the other asset. The left panels display the responses in Asset 1, whereas the right panels display the responses in Asset 2. In both experiments, we analyze the trajectories of the 'Myopic' portfolio with transaction costs and our 'Optimal' dynamic portfolio.

(a) Numerical Experiment 1



(b) Numerical Experiment 2

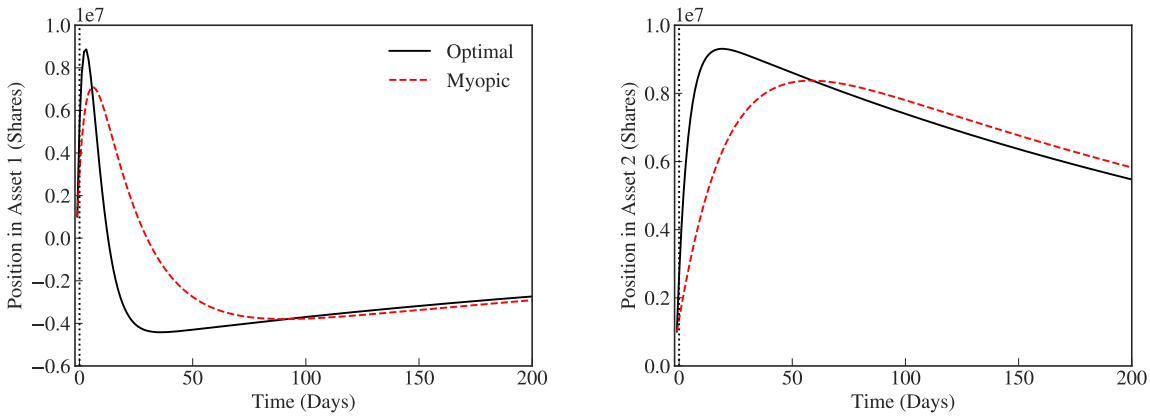
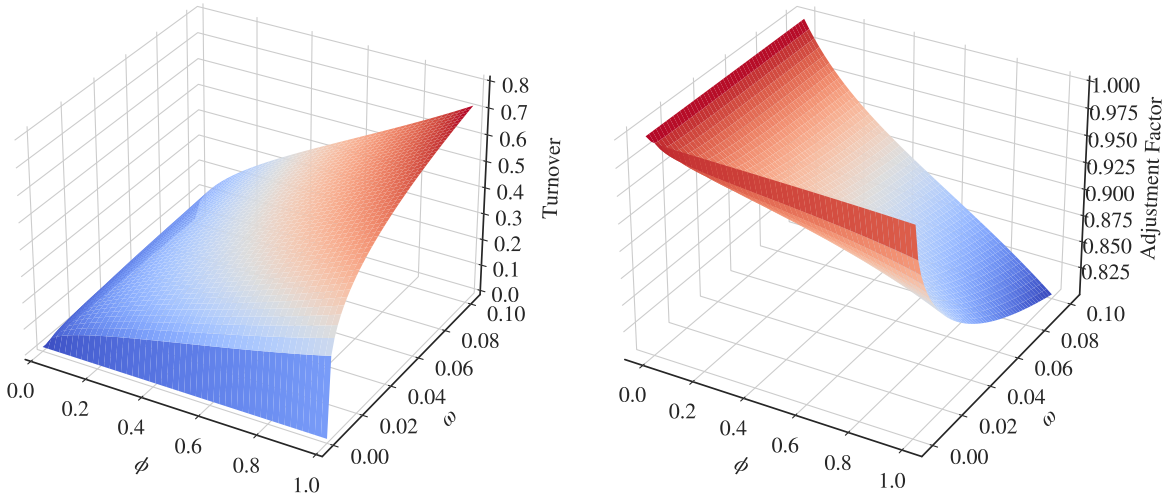


Figure C.5.2 – Steady-State Performance Analysis

This figure illustrates the relationship of the steady-state performance of the optimal portfolio to the signal's mean-reversion and the investor's liquidity-adjusted risk aversion. The left panel displays the turnover, while the right panel displays the adjustment factor used to correct the Sharpe ratio.



(a) Turnover

(b) Sharpe Ratio Adjustment Factor

References

Abadir, K. M., Distaso, W., and Žikeš, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2):165–180.

Ahn, D.-H., Conrad, J., and Dittmar, R. F. (2009). Basis assets. *Review of Financial Studies*, 22(12):5133–5174.

Aït-Sahalia, Y. and Brandt, M. W. (2001). Variable selection for portfolio choice. *Journal of Finance*, 56(4):1297–1351.

Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40.

Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18(7):58–62.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, pages 327–351.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20(1):46–63.

Anderson, T. W. and Rubin, H. (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, pages 570–582.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494.

Asness, C. S., Frazzini, A., and Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24(1):34–112.

Avellaneda, M. and Lee, J.-H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7):761–782.

Bach, F. R. and Jordan, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48.

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

Balduzzi, P. and Lynch, A. W. (2000). Predictability and transaction costs: The impact on rebalancing rules and behavior. *Journal of Finance*, 55(5):2285–2309.

Benaych-Georges, F., Bouchaud, J.-P., and Potters, M. (2023). Optimal cleaning for singular values of cross-covariance matrices. *Annals of Applied Probability*, 33(2):1295–1326.

Benzoni, L., Collin-Dufresne, P., and Goldstein, R. S. (2007). Portfolio choice over the life-cycle when the stock and labor markets are cointegrated. *Journal of Finance*, 62(5):2123–2167.

Bertsekas, D. P. (1976). *Dynamic Programming and Stochastic Control*. Academic Press, New York.

Bertsimas, D. and Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50.

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

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Annals of Statistics*, 36(1):199–227.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bouchaud, J.-P. and Potters, M. (2020). *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge University Press.

Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance*, 54(5):1609–1645.

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *Journal of Finance*, 61(5):2187–2217.

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, 22(9):3411–3447.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance*, 54(2):655–671.

Brown, D. B. and Smith, J. E. (2011). Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds. *Management Science*, 57(10):1752–1770.

Bun, J., Bouchaud, J.-P., and Potters, M. (2017). Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, 666:1–109.

Burda, Z. and Jarosz, A. (2022). Cleaning large-dimensional covariance matrices for correlated samples. *Physical Review E*, 105(3):034136.

Burda, Z., Jarosz, A., Nowak, M. A., Jurkiewicz, J., Papp, G., and Zahed, I. (2011). Applying free random variables to random matrix analysis of financial data. Part I: The Gaussian case. *Quantitative Finance*, 11(7):1103–1124.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Viceira, L. M. (1999). Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics*, 114(2):433–495.

Campbell, J. Y. and Viceira, L. M. (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investor*. Oxford University Press.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11.

Cocco, J. F., Gomes, F. J., and Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. *Review of Financial Studies*, 18(2):491–533.

Cochrane, J. H. (2001). *Asset Pricing*. Princeton University Press, Princeton.

Cochrane, J. H. (2022). Portfolios for long-term investors. *Review of Finance*, 26(1):1–42.

Cohen, K. J. and Pogue, J. A. (1967). An empirical evaluation of alternative portfolio-selection models. *Journal of Business*, 40(2):166–193.

Collin-Dufresne, P., Daniel, K., and Sağlam, M. (2020). Liquidity regimes and optimal dynamic asset allocation. *Journal of Financial Economics*, 136(2):379–406.

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3):373–394.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21(2):255–289.

Constantinides, G. M. (1983). Capital market equilibrium with personal tax. *Econometrica*, 51(3):611–636.

Constantinides, G. M. (1984). Optimal stock trading with personal taxes: Implications for prices and the abnormal january returns. *Journal of Financial Economics*, 13(1):65–89.

Davis, M. H. and Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4):676–713.

De Nard, G., Ledoit, O., and 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.

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

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419.

Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy*, 84(6):1161–1176.

Duffie, D. and Sun, T.-s. (1990). Transactions costs and portfolio choice in a discrete-continuous-time setting. *Journal of Economic Dynamics and Control*, 14(1):35–51.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, 36(6):2757–2790.

Elton, E. J. and Gruber, M. J. (1973). Estimating the dependence structure of share prices–implications for portfolio selection. *Journal of Finance*, 28(5):1203–1232.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24(2):238–253.

Engle, R. F. and Ferstenberg, R. (2007). Execution risk: It's the same as investment risk. *Journal of Portfolio Management*, 33(2):34–44.

Engle, R. F., Ferstenberg, R., and Russell, J. R. (2012). Measuring and modeling execution cost and risk. *Journal of Portfolio Management*, 38(2):14–28.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298.

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

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

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

Ferson, W. E. and Siegel, A. F. (2001). The efficient use of conditioning information in portfolios. *Journal of Finance*, 56(3):967–982.

Firoozye, N. and Koshiyama, A. (2019). Avoiding backtesting overfitting by covariance-penalties: An empirical investigation of the ordinary and total least squares cases. *Journal of Financial Data Science*, 1(4):63–83.

Firoozye, N. and Koshiyama, A. (2020). Optimal dynamic strategies on Gaussian returns. *Journal of Investment Strategies*, 9(1):23–53.

Firoozye, N., Tan, V., and Zohren, S. (2023). Canonical portfolios: Optimal asset and signal combination. *Journal of Banking & Finance*, 154:106952.

Fischer, S. (1977). Long-term contracts, rational expectations, and the optimal money supply rule. *Journal of Political Economy*, 85(1):191–205.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, 33(5):2326–2377.

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

Gârleanu, N. and Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *Journal of Finance*, 68(6):2309–2340.

Gârleanu, N. and Pedersen, L. H. (2016). Dynamic portfolio choice with frictions. *Journal of Economic Theory*, 165:487–516.

Geary, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, 27(3/4):310–332.

Golub, G. H. and Van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6):883–893.

Grinold, R. C. (1989). The fundamental law of active management. *Journal of Portfolio Management*, 15(3):30–37.

Grinold, R. C. (2006). A dynamic model of portfolio management. *Journal of Investment Management*, 4(2):5.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408.

Haff, L. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Annals of Statistics*, 8(3):586–597.

Haldane, J. (1942). Moments of the distributions of powers and products of normal variates. *Biometrika*, 32(3/4):226–242.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of Political economy*, 86(6):971–987.

Hansen, L. and Richard, S. F. (1987). The role of conditioning information in deducing testable. *Econometrica*, 55(3):587–613.

Hansen, L. P. and Sargent, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control*, 2:7–46.

Hansen, L. P. and Sargent, T. J. (1981). Linear rational expectations models for dynamically interrelated variables. In Lucas Jr., R. E. and Sargent, T. J., editors, *Rational Expectations and Econometric Practice*. University of Minnesota Press, Minneapolis, Minnesota.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207.

Haugen, R. A. and Baker, N. L. (1991). The efficient market inefficiency of capitalization–weighted stock portfolios. *Journal of Portfolio Management*, 17(3):35–40.

Hausman, J. A., Lo, A. W., and MacKinlay, A. C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31(3):319–379.

Healy, M. (1957). A rotation method for computing canonical correlations. *Mathematics of Computation*, 11(58):83–86.

Heaton, J. and Lucas, D. J. (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy*, 104(3):443–487.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4):321–377.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139.

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

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 361–379.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, 45(3):881–898.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Jensen, T. I., Kelly, B. T., Malamud, S., and Pedersen, L. H. (2023). Machine learning and the implementable efficient frontier. *Available at SSRN 4187217*.

Jing, B.-Y., Pan, G., Shao, Q.-M., and Zhou, W. (2010). Nonparametric estimate of spectral density functions of sample covariance matrices: A first step. *Annals of Statistics*, 38(6):3724.

Jobson, J. D. and Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer Verlag, New York.

J.P. Morgan/Reuters (1996). RiskMetrics - technical document. *New York: J.P. Morgan/Reuters*.

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

Keim, D. B. and Madhavan, A. (1996). The upstairs market for large-block transactions: Analysis and measurement of price effects. *Review of Financial Studies*, 9(1):1–36.

Keim, D. B. and Madhavan, A. (1998). The cost of institutional equity trades. *Financial Analysts Journal*, 54(4):50–69.

Kelly, B., Malamud, S., and Pedersen, L. H. (2023). Principal portfolios. *Journal of Finance*, 78(1):347–387.

Kelly, B. and Pruitt, S. (2013). Market expectations in the cross-section of present values. *Journal of Finance*, 68(5):1721–1756.

Kennan, J. (1979). The estimation of partial adjustment models with rational expectations. *Econometrica*, pages 1441–1455.

King, B. F. (1966). Market and industry factors in stock price behavior. *Journal of Business*, 39(1):139–190.

Kozak, S., Nagel, S., and Santosh, S. (2018). Interpreting factor models. *Journal of Finance*, 73(3):1183–1223.

Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.

Kydland, F. E. and Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica*, 50(6):1345–1370.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.

Lai, T. L., Xing, H., and Chen, Z. (2011). Mean–variance portfolio optimization when means and covariances are unknown. *Annals of Applied Statistics*, 5(2A):798–823.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Annals of Statistics*, 44(3):928–953.

Ledoit, O. and Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 151(1):233–264.

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

Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4):110–119.

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

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Ledoit, O. and Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, 2011(55):86–89.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139:360–384.

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

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

Ledoit, O. and Wolf, M. (2022). Quadratic shrinkage for large covariance matrices. *Bernoulli*, 28(3):1519–1547.

Lehmann, B. N. and Modest, D. M. (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, 21(2):213–254.

Leland, H. (2000). Optimal portfolio implementation with transactions costs and capital gains taxes. *Haas School of Business Technical Report*.

Levy, H. and Markowitz, H. M. (1979). Approximating expected utility by a function of mean and variance. *American Economic Review*, pages 308–317.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2):175–194.

Lillo, F., Farmer, J. D., and Mantegna, R. N. (2003). Master curve for price-impact function. *Nature*, 421(6919):129–130.

Ljungqvist, L. and Sargent, T. J. (2004). *Recursive Macroeconomic Theory, 2nd Edition*. The MIT Press.

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.

Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1-2):181–211.

Lucas, Jr., R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester Conference Series on Public Policy*, volume 1, pages 19–46. North-Holland.

Lütkepohl, H. (1997). Handbook of matrices. *Computational Statistics and Data Analysis*, 2(25):243.

Lynch, A. W. and Tan, S. (2010). Multiple risky assets, transaction costs, and return predictability: Allocation rules and implications for U.S. investors. *Journal of Financial and Quantitative Analysis*, 45(4):1015–1053.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, 36:394–394.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197.

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

Markowitz, H. M. (1991). Foundations of portfolio theory. *Journal of Finance*, 46(2):469–477.

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

Mei, X., DeMiguel, V., and Nogales, F. J. (2016). Multiperiod portfolio optimization with multiple risky assets and general transaction costs. *Journal of Banking & Finance*, 69:108–120.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, pages 247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7(4):1851–1872.

Meucci, A. (2009). Managing diversification. *Risk*, pages 74–79.

Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.

Michaud, R. O. and Michaud, R. O. (2008). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press.

Moskowitz, T. J. and Grinblatt, M. (1999). Do industries explain momentum? *Journal of Finance*, 54(4):1249–1290.

Mossin, J. (1968). Optimal multiperiod portfolio policies. *Journal of Business*, 41(2):215–229.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3):315–335.

Obizhaeva, A. A. and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32.

Pakel, C., Shephard, N., Sheppard, K., and Engle, R. F. (2021). Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics*, 39(3):652–668.

Partovi, M. H. and Caputo, M. (2004). Principal portfolios: Recasting the efficient frontier. *Economics Bulletin*, 7(3):1–10.

Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management*, 14(3):4.

Roll, R. and Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, 35(5):1073–1103.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis*, 9(2):263–274.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–60.

Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, pages 239–246.

Sargent, T. (1988). Linear optimal control, filtering, and rational expectations. *Federal Reserve Bank of Minneapolis, Working Paper 224*.

Sargent, T. J. (1978). Estimation of dynamic labor demand schedules under rational expectations. *Journal of Political Economy*, 86(6):1009–1044.

Sargent, T. J. (1987). *Macroeconomic Theory*. 2nd ed. New York: Academic Press.

Sargent, T. J. and Wallace, N. (1973). The stability of models of money and growth with perfect foresight. *Econometrica*, 41(6):1043–1048.

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

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Sheena, Y. and Takemura, A. (1992). Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss. *Journal of Multivariate Analysis*, 41(1):117–131.

Simon, H. A. (1956). Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica*, pages 74–81.

Sneddon, L. (2008). The tortoise and the hare: Portfolio dynamics for active managers. *Journal of Investing*, 17(4):106–111.

Stein, C. (1956a). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 197–206. University of California Press.

Stein, C. (1956b). Some problems in multivariate analysis, Part I. Technical Report 6, Department of Statistics, Stanford University.

Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

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

Stieltjes, T. J. (1894). Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse 1re Série*, 8(4):J1–J122.

Tan, V. and Zohren, S. (2024). Estimation of large financial covariances: A cross-validation approach. *Journal of Portfolio Management, Forthcoming*.

Taylor, J. B. (1979). Staggered wage setting in a macro model. *American Economic Review*, 69(2):108–113.

Theil, H. (1957). A note on certainty equivalence in dynamic planning. *Econometrica*, pages 346–349.

Tirole, J. (1982). On the possibility of speculation under rational expectations. *Econometrica*, 50(5):1163–1181.

Tola, V., Lillo, F., Gallegati, M., and Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32(1):235–258.

Tumminello, M., Lillo, F., and Mantegna, R. N. (2007). Hierarchically nested factor model from multivariate data. *Europhysics Letters*, 78(3):30006.

Uurtio, V., Monteiro, J. M., Kandola, J., Shawe-Taylor, J., Fernandez-Reyes, D., and Rousu, J. (2017). A tutorial on canonical correlation methods. *ACM Computing Surveys (CSUR)*, 50(6):1–33.

Velu, R. and Reinsel, G. C. (2013). *Multivariate reduced-rank regression: Theory and applications*, volume 136. Springer Science & Business Media.

Viceira, L. M. (2001). Optimal portfolio choice for long-horizon investors with nontradable labor income. *Journal of Finance*, 56(2):433–470.

Vinod, H. D. (1976). Canonical ridge and econometrics of joint production. *Journal of Econometrics*, 4(2):147–166.

Voiculescu, D. (1985). Symmetries of some reduced free product C^*-algebras. In *Operator algebras and their connections with topology and ergodic theory*, pages 556–588. Springer.

Voiculescu, D. (1991). Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1):201–220.

Wachter, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *Annals of Statistics*, pages 937–957.

Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press, Minneapolis.

Wick, G. (1950). The evaluation of the collision matrix. *Physical Review*, 80(2):268–272.

Wold, H. (1975). Soft modelling by latent variables: The non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability*, 12(S1):117–142.

Yang, Y. and Pan, G. (2012). The convergence of the empirical distribution of canonical correlation coefficients. *Electronic Journal of Probability*, 17:1–13.

Yang, Y. and Pan, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Annals of Statistics*, 43(2):467–500.

Zhou, G. (2008). On the fundamental law of active portfolio management: What happens if our estimates are wrong? *Journal of Portfolio Management*, 34(4):26–33.

Zumbach, G. (2007). The RiskMetrics 2006 methodology. *Technical Report, RiskMetrics Group*.