Essays on Factor Portfolios

Nikolaos Vasilas

Department of Accounting and Finance

Lancaster University

A thesis submitted in fulfillment of the requirements

for the degree of

Doctor of Philosophy in Finance

August 2024

Ἡμεῖς ἐσμέν ὅ,τι ἐπαναλαμβάνομεν. Ἐξαίρεσις οὖν οὐκ ἔστιν ἔργον, ἀλλὰ ἕξις.

We are what we repeatedly do. Excellence, then, is not an act, but a habit.

— Aristotle

Abstract

This thesis explores innovative approaches to factor investing by examining the dynamics of factor portfolios and introducing novel methodologies for constructing and utilizing characteristic-based equity factors. Chapter 1 addresses the predictability of factor portfolios within the context of factor timing. This is achieved by extending stock return predictability to a portfolio level and using various dimension reduction techniques in both the characteristics and returns space. The analysis demonstrates that factor portfolios are predictable based not only on their own but also on other characteristics, highlighting the significant potential for asset return prediction. This finding also suggests that different portfolios share similarities in terms of signal sources or underlying factors.

Chapter 2 introduces a new technique for constructing characteristic-based equity factors, termed "power sorting". This method leverages the non-linearities and asymmetries inherent in characteristic-return relationships while maintaining computational simplicity and avoiding excessive weighting. Empirical analysis shows that power sorting consistently delivers superior out-of-sample performance compared to traditional quantile sorting and other factor portfolio construction methods. The approach proves robust across different factors and time periods, with its effectiveness not attributable to increased turnover or tail risk. Moreover, power-sorted versions of well-known asset pricing factor models outperform their original counterparts.

Extending the power sorting methodology to the multivariate level, Chapter 3 investigates the evolution of portfolio dynamics when multiple characteristics are jointly considered. While Chapter 2 shows that, in a univariate context, individual characteristics drive portfolio performance primarily through the short side, the analysis in Chapter 3 reveals a shift in importance to the long side when characteristics are jointly analyzed. The multivariate power sorting approach achieves two key objectives: the development of multifactor strategies with significantly enhanced risk-adjusted performance, and the construction of a six-factor model that effectively spans the tangency portfolio.

Declaration

I affirm that this dissertation is entirely my own work, except where explicitly acknowledged. It contains original content and has not been submitted, in whole or in part, for any other degree or qualification at this university or elsewhere. I have not collaborated with others on this work, except as specified in the text and acknowledgments.

Nikolaos Vasilas

August 2024

Acknowledgements

The successful completion of this journey would not have been possible without the help and support of many people to whom I am deeply grateful.

I would like to express my heartfelt appreciation to my supervisors, Dr. Anastasios Kagkadis (University of Liverpool), Prof. Ingmar Nolte, and Prof. Sandra Nolte (both Lancaster University), for their invaluable feedback, unwavering support, and immensely helpful guidance throughout my PhD journey. I deeply appreciate their deep expertise in financial econometrics and asset pricing, which have significantly elevated the quality of my work. Our productive discussions have greatly benefited my research, and I hold the utmost respect for the passion, enthusiasm and patience they have shown in guiding me through my doctoral studies.

I would also like to extend my sincere thanks to my co-author, Dr. Harald Lohre (Robeco), for his invaluable mentorship, meticulous guidance, and keen attention to detail during our collaboration on the second and third chapters of my thesis. I am especially grateful for the opportunity to join him at Robeco, where I was able to further develop my ideas and elevate my research in a highly creative and stimulating environment.

I am grateful to Lancaster University Management School (LUMS) for significantly advancing my academic development throughout every stage of my higher education and for their generous financial support for my PhD studies.

I am also very thankful to the friends and colleagues I have encountered at LUMS and Robeco. Although there are too many to list, these past five years have been a highly rewarding and enjoyable part of my life. The numerous discussions and the support I received on theoretical and technical matters made my doctoral studies much more manageable. I would like to extend special thanks to Filip Bašić, Marco Cinquetti, Mustafa Berke Erdis, and Alexander Swade for their help, support, and friendship.

Finally, I would like to thank my family for their steady support and encouragement

throughout this journey. Their support has been incredibly important and I am deeply thankful for their presence and love.

Contents

Introduction **12**

1 Factor Timing with Portfolio Characteristics **15**

1.1 Introduction 16

1.2 Literature Review 20

1.2.1 Dimension reduction in asset pricing 20

1.2.2 Factor portfolio predictability 22

1.3 Methodology 24

1.3.1 General framework 24

1.3.2 Forecasting procedure 27

1.3.3 Benchmark models 31

1.4 Empirical Results 33

1.4.1 Data 33

1.4.2 Predictive performance 36

1.4.3 Investment performance 45

1.5 Alternative Approaches 52

1.6 Conclusion 56

2 Power Sorting **58**

2.1 Introduction 59

2.2 Power Sorting Methodology 65

2.2.1 Conventional long-short quantile factor portfolios 66

2.2.2 Power sorting 67
2.2.3 Power sorting and related literature 73
2.3 Optimal Power Sorting Portfolios 76
2.3.1 Characteristics and power thresholds 76
2.3.2 Estimation procedure 77
2.3.3 Power-sorted portfolios and concentration ratios 78
2.3.4 Power sorting versus conventional decile sorting 81
2.3.5 Dissecting long and short factor portfolio legs 84
2.3.6 Spanning regressions 86
2.3.7 Multi-factor portfolios 88
2.3.8 Asset pricing tests 92
2.4 Benchmarking and Robustness 94
2.4.1 Alternative Benchmarks 94
2.4.2 Robustness Tests 102
2.5 Conclusion 108

3 Multi-Factor Power Sorting **110**
3.1 Introduction 111
3.2 Literature Overview 114
3.3 Methodology 117
3.3.1 Univariate Power Sorting 117
3.3.2 Multivariate Power Sorting 118
3.4 Power Sorting Portfolios 119
3.4.1 Data 119
3.4.2 Univariate versus Multivariate Power Sorting 120
3.4.3 Multi-factor portfolios 122
3.5 Power Sorting-driven Factor Model 125
3.5.1 Iterative Factor Selection 125
3.5.2 Selected Factors 127
3.5.3 Comparative Analysis 132

3.6 Conclusion 135

Concluding remarks **137**

A Appendix **138**

A.1 List of characteristics 138

A.2 Supplementary Material to Chapter 1 145

A.2.1 Factor variation explained by the PC portfolios 145

A.2.2 Details of the dimension reduction and regularization techniques . . 146

A.2.3 Sources of variation in the PC characteristics 150

A.2.4 Details of the predictive models 158

A.2.5 Factor statistics 158

A.2.6 Interpreting the PC portfolios 161

A.2.7 Factor timing strategy constituents 164

A.3 Supplementary Material to Chapter 2 168

A.3.1 Supplementary Figures 168

A.3.2 Return-spread maximization objective 169

A.3.3 Multi-factor strategies for alternative benchmarks 170

A.4 Supplementary Material to Chapter 3 175

A.4.1 Squared Sharpe ratio from adding each factor in each iteration . . . 175

References 178

List of Tables

1.1 OOS R^2 for PC portfolios and total OOS R^2 across all anomalies 37
1.2 Modified Diebold-Mariano test for models that employ lasso and benchmark models 40
1.3 Percentage of correct sign identifications and average cross-sectional correlation 42
1.4 Portfolio evaluation measures for long-short strategies 48
1.5 Total OOS R^2 and average cross-sectional correlation of factor portfolio return forecasts based on various forecasting methods 53
1.6 Portfolio evaluation measures for long-short (LSS), time-series (TSFM), and cross-sectional (CSFM) strategies under the alternative specifications . 55

2.1 Average performance measures across factors for power and conventional long-short portfolios 81
2.2 Portfolio evaluation measures for multi-factor portfolios 90
2.3 Asset pricing models based power-sorted versus original factors 93
2.4 Power sorting versus rank sorting 94
2.5 Power sorting versus Parametric Portfolio Policy 97
2.6 Power sorting versus efficient sorting 101
2.7 Robustness with respect to different lookback windows 103
2.8 Robustness with respect to the maximum stock weight 104
2.9 Robustness with respect to concentration on mega-cap stocks 106
2.10 Robustness across different sub-periods 107

3.1 Multi-factor portfolios 124
3.2 Iterative Factor Selection 128
3.3 Overall Performance of Factor Models 133

A.1 Listing of firm characteristics used in the study, including the source and the exact definition. 139
A.2 Stylized example of estimating PC portfolio characteristics and returns using three anomalies 151
A.3 Detailed description of all the predictive models used in the study 159
A.4 Descriptive statistics of factor portfolios 160
A.5 Power portfolios under a return maximization objective 170
A.6 Portfolio evaluation measures for multi-factor power and rank portfolios . . 171
A.7 Portfolio evaluation measures for multi-factor power and PPP portfolios . 172
A.8 Portfolio evaluation measures for multi-factor power and efficient portfolios 174

List of Figures

1.1 Visual depiction of our modeling procedure 31

1.2 Average monthly returns of factor portfolios with 95% confidence intervals 35

1.3 OOS R^2 for individual anomalies under the characteristic-based models that employ lasso and benchmark models 39

1.4 Number of features for each PC portfolio under the different models 44

1.5 Cumulative return performance of factor timing strategies 50

2.1 Conditional monthly returns and conventional equal-weighted decile-sorted factor portfolio weights for six characteristics 61

2.2 The opportunity set of power sorting 70

2.3 Average estimated concentration ratios $\hat{\hat{p}}$ and $\hat{\hat{q}}$ of each factor 79

2.4 Conditional monthly returns and power factor portfolio weight function for six characteristic-based factor portfolios 80

2.5 Portfolio evaluation measures for conventional long-sort and power versions 83

2.6 Concentration ratios and excess returns 85

2.7 Spanning regression alphas 87

2.8 Average weight function for efficient sorting and power sorting 100

3.1 Average standardized powers of each factor for univariate versus multivariate optimization 121

3.2 Standardized powers for each factor in the power sorting factor model per iteration 131

A.1 Percentage of the variation explained by each PC of factor portfolio returns under PCA and RPPCA 145
A.2 Monthly cross-sectional skewness of 72 characteristics across factor portfolios 153
A.3 Monthly cross-sectional kurtosis of 72 characteristics across factor portfolios 154
A.4 Average monthly cross-sectional correlations across 72 characteristics . . . 154
A.5 Standard deviation of monthly cross-sectional correlations across 72 characteristics 155
A.6 Recursively estimated PC portfolio weights on 72 factor portfolios using PCA 156
A.7 Recursively estimated PC portfolio weights on 72 factor portfolios using RPPCA 157
A.8 Average R^2 across month from recursive regressions of each PC portfolio on each anomaly 163
A.9 R^2 across months from a recursive regression of each PC portfolio on each of the following anomalies: beta, bm, mom12m, rd_sale, and salerec 164
A.10 Percentage frequency of LSS constituents over the out-of-sample period . . 166
A.11 Upper Thresholds for q_t^{max} (a) and p_t^{max} (b) for the 85 characteristics using a Maximum Weight Constraint of (w^{ceil}) 2% 168
A.12 Cross-sectional stock size for original sample and sub-sample with five years of daily return data 169
A.13 Squared Sharpe ratio from adding each factor to the factor model of the previous iteration 176

Introduction

The empirical asset pricing literature suggests that firm characteristics contain valuable information about future stock returns. This relationship is typically examined by constructing portfolios sorted by these characteristics, commonly referred to as factor portfolios or simply factors. This approach was popularized by Fama and French (1992) and Jegadeesh and Titman (1993), and these factor portfolios are widely used by academics for various asset pricing applications, as well as by investment professionals seeking systematic exposure to rewarded factors. However, the effectiveness of the standard procedure in efficiently extracting the risk premium from these characteristics remains unclear. Furthermore, with the growing number of identified characteristics—often referred to as the "factor zoo" (Cochrane 2011)—it becomes increasingly challenging to discern which characteristics provide unique information and which are merely different manifestations of the same underlying return driver. Additionally, these portfolios can exhibit significant time variation in performance and are susceptible to crashes (e.g. Daniel and Moskowitz 2016). These challenges naturally call for solutions that involve improving upon the ad-hoc, static factor approach, either through dynamic factor allocation or by refining portfolio construction techniques.

The aim of this dissertation is to address these challenges and explore innovative approaches to factor investing. By examining the dynamics of factor portfolios and introducing novel methodologies for constructing and utilizing characteristic-based equity factors, this work seeks to enhance our understanding of the relevance of firm characteristics in predicting asset returns and solving optimal portfolio construction problems.

First, we investigate return predictability from observable characteristics at the portfolio level and predict factor portfolio returns using a collection of portfolio characteristics. In doing so, we employ various combinations of dimension reduction techniques that have gained popularity in recent asset pricing applications to reduce both the number of predictors and the portfolios to predict. We observe significant benefits from timing factor portfolio returns using observed characteristics, which surpass existing methods documented in the literature. Our analysis also highlights the importance of focusing on the main sources of variation in factor portfolio returns, which are ultimately the most predictable, allowing for the detection of robust predictive patterns across factors. This chapter also implicitly conducts an indirect comparative analysis across different dimension reduction techniques, offering a reference point for future applications.

Second, we address the limitations of the conventional quantile approach in capturing premiums across various characteristics. Specifically, we develop a simple, data-oriented procedure called power sorting, which allows for characteristic-specific treatment and captures asymmetries and non-linearities in returns, thereby providing a deeper understanding of the drivers of factor premia. In our empirical application, we demonstrate that power sorting consistently achieves superior out-of-sample performance compared to traditional quantile sorting and other factor portfolio construction methods, both at the single-factor and multi-factor levels. We also document several important findings, including the existence of asymmetric and non-linear patterns between characteristics and returns, and the critical role of the short side in driving portfolio performance at a univariate level.

Third, we explore how portfolio dynamics change when various characteristics are considered jointly, enabling the examination of potential interaction effects. We extend the power sorting methodology from Chapter 2 to a multivariate level by focusing on maximizing the squared Sharpe ratio of the collection of factors, rather than the Sharpe ratios of individual factors. This multivariate extension significantly improves the risk-adjusted performance of multi-factor strategies proposed in Chapter 2 and facilitates the construction of a six-factor model that spans the tangent portfolio, ultimately addressing which characteristics are most important. Importantly, our analysis in Chapter 3 reveals a shift in importance from the short side to the long side when characteristics are considered jointly. This suggests that while different characteristics on the short side capture similar effects, those on the long side provide more complementary information.

The remainder of this dissertation is structured as follows. Chapter 1 is based on the research paper titled *Factor Timing with Portfolio Characteristics,"* which has been published in the *Review of Asset Pricing Studies*. This chapter utilizes portfolio characteristics and various combinations of dimensionr reduction techniques to effectively predict factor portfolio performance and construct factor timing strategies. Chapter 2 is drawn from the research paper *Power Sorting,"* which won first place in the Chicago Quantitative Alliance 30th Annual Academic Competition. In this chapter, we develop a data-oriented power sorting procedure that directly models factor portfolio weights as a function of firm characteristics. Chapter 3 is based on the paper titled *"Multi-Factor Power Sorting,"* which extends the power sorting method to a multivariate setting and explores interesting asset pricing and investment applications. The final chapter concludes this dissertation, summarizing the key findings and implications.

Chapter 1

Factor Timing with Portfolio Characteristics

This project is a joint work with Anastasios Kagkadis, Ingmar Nolte, and Sandra Nolte. It is published in the *Review of Asset Pricing Studies* (Volume 14, Issue 1, pages 84–118). We would like to thank Jeffrey Pontiff (the Editor), an anonymous referee, David Buckle, Amit Goyal, Harald Lohre, Alejandro Lopez-Lira, Robert Macrae, Winfried Pohlmeier, as well as participants at the INQUIRE UK Spring 2022 Residential Seminar, the 11th International Conference of the Financial Engineering and Banking Society, the QFFE 2022 International Conference, the 2022 EFMA Annual Meeting, the 3rd Frontiers of Factor Investing Conference, and the 2022 FMA Annual Meeting for their valuable comments and suggestions.

1.1 Introduction

The asset pricing literature has long been shaped by the idea that observable firm characteristics convey information about the cross-section of expected stock returns. A common practice in the literature is to extract the risk premium associated with these characteristics by constructing long-short (LS) factor portfolios (Fama and French 1993). Such zero-investment, market-neutral portfolios have given rise to so-called factor investing. Yet, there are benefits over and above static factor investing. Studies such as Stambaugh, Yu, and Yuan (2012), Jacobs (2015), Akbas et al. (2016), and Keloharju, Linnainmaa, and Nyberg (2016) show that the performance of LS portfolios, and therefore the benefits from factor investing, are significantly time-varying. More importantly, such time variation in performance is not harmonious across portfolios, allowing for substantial investment gains from timing factor portfolio returns.[1] As such, from an investor's perspective, timing is important, and an active factor allocation is needed in order to capitalize on the fluctuations in LS portfolio returns.

In a factor timing context, several studies have emerged utilizing a variety of predictive signals as a way to improve upon static factor investing. Valuation ratios, investor sentiment, issuer-repurchaser spread, and technical indicators, such as factor momentum, are the most prominent examples, among others. In this paper, we create an optimal factor timing strategy, going over and above existing methods for predicting factor portfolio returns. In doing so, we extend the predictability of stock returns from observable firm characteristics to a portfolio level and predict factor portfolio returns using a collection of portfolio characteristics. Specifically, the characteristics used to sort stocks into portfolios are subsequently aggregated into portfolio characteristics and used as predictive variables to forecast future factor portfolio returns. The use of multiple characteristics to predict individual factor portfolio returns is motivated by the fact that many stocks coexist in different factor portfolio legs simultaneously.[2] Hence, it is sensible to assess the

[1] For example, Haddad, Kozak, and Santosh (2020) find that the loadings of a size portfolio on the optimal factor timing portfolio are procyclical while those of a momentum portfolio are countercyclical.

[2] For example, the stocks with the highest asset growth are also the ones with the lowest book-to-market

joint predictability that arises from characteristics at a portfolio level and examine the possibility that factor portfolios are predictable by characteristics other than their own. A comparison of the collective characteristic-based predictability against alternative sets of predictors documented in the literature highlights the joint importance of characteristics in explaining the dynamics of factor portfolios.

A key aspect of our methodology is the use of different dimension reduction techniques to reduce the dimensions of both sides of the predictability problem. In line with Haddad, Kozak, and Santosh (2020), we begin by reducing the number of forecasting targets, recognizing the underlying factor structure in factor portfolio returns. Instead of independently predicting individual anomalies, we focus our attention on the main sources of return variation by isolating the first five principal components (PCs). These PCs capture around 67% of the variation in factor portfolio returns (see Figure A.1 in the Appendix), allowing us to greatly reduce the dimensions of the problem at the expense of little return variation foregone. Since the dominant PCs capture common variation in the underlying risk premia, being able to accurately predict their performance leads to the detection of robust predictive patterns across individual anomalies.[3] In addition, PCs are not just statistical factors but have an investable interpretation as well. As each PC is a linear combination of the underlying variables, PC portfolios are portfolios of factor portfolios, meaning that their returns and characteristics are calculable. To construct the PC portfolios, we use conventional principal component analysis (PCA), as well as the risk premium PCA (RPPCA) proposed by Lettau and Pelger (2020a).[4] Unlike PCA, RPPCA utilizes information of the mean returns of the factor portfolios in addition to their variances and leads to the extraction of factors that may explain a smaller part of the time-series variation but are important in pricing the cross-section. The resulting PCs have higher Sharpe ratios and in our context help guide the forecasting study around

ratio, the highest return on assets, and the highest accruals (Cooper, Gulen, and Schill 2008).

[3] Applying principal component analysis to a set of factor portfolio returns in order to achieve dimension reduction has recently gained a lot of attention in asset pricing. For example, Haddad, Kozak, and Santosh (2020) form PC portfolios by running PCA on a set of 50 anomalies and use their own book-to-market ratio to predict their performance.

[4] Henceforth, "PC portfolios" refers to the estimation of the principal component portfolios using either PCA or RPPCA.

factor portfolios with higher average returns.

We then proceed by compressing the predictive information from the characteristics of the PC portfolios. To achieve this, we not only rely on PCA, but employ methods that account for the covariance structure between predictors and forecasting targets, such as partial least squares (PLS) (Wold et al. 1984). Conventional PCA focuses on the variance within the predictors and can lead to components that mix return-relevant and -irrelevant variation. By using PLS we aim to capture only the variation in the characteristics that is relevant in predicting returns, potentially resulting in sparser and more accurate models.

After rotating characteristics in space using PCA or PLS, we either use the first characteristic component in standard predictive regressions or apply lasso on the whole set of characteristic components to identify the relevant subset of features.[5] The first case is used to investigate the predictability in the simplest case of a single predictive factor, while the use of lasso allows for successive components to be included in the surviving subset of predictors, with the importance of each characteristic component being assessed based on its contribution to minimizing the forecasting error rather than the magnitude of its eigenvalue. Our procedure is implemented recursively, and the optimal degree of coefficient shrinkage is identified separately for each PC portfolio based on a cross-validation step. This approach has two important implications. First, the number of factors can be different across PC portfolios, allowing for different sources of variation in factor portfolio returns to be approximated by models of different complexity. For instance, many characteristic components may be required to predict the first PC portfolio but only a few for the second. Second, allowing for different values for the level of coefficient shrinkage across time allows us to examine the time variation in the strength of the characteristic signal overall.

In our empirical analysis, we use a collection of 72 anomalies spanning the period from 1970 to 2019 and find that characteristics are particularly useful for factor timing

[5]The combination of lasso with PCA or PLS is particularly suitable in this case because the PC portfolios are by construction orthogonal.

purposes. We distinguish factor portfolio predictability in terms of exact predictive accuracy (comparing predicted with future realized returns) and ability to predict the cross-sectional dispersion in returns (differentiating winners from losers). The characteristic-based models that incorporate lasso are the most successful and consistently outperform existing methods in both terms as they deliver smaller forecasting errors and higher cross-sectional correlations between forecasted and realized returns. They also deliver average monthly returns of up to 1.47% and annualized Sharpe ratios of up to 0.73, while the best benchmark model delivers 1.06% and 0.55, respectively. Importantly, our factor timing strategies show no decay in return performance over time, although many individual anomalies have been found empirically to do so (McLean and Pontiff 2016).

In terms of the different methods used, the implications of using PCA or RPPCA to reduce the number of portfolios to predict are minimal. Yet, when it comes to reducing the number of predictors down to a single predictive factor, the dimension reduction technique matters. In particular, PCA delivers slightly better exact predictability but severely underperforms PLS in terms of ranking the anomaly portfolios successfully. Essentially, when a single-factor model is used, it is better to condense the information from the predictors using a tool that is specifically designed for forecasting purposes.

Nonetheless, the difference between PCA and PLS disappears when multiple characteristic components are considered in conjunction with lasso, suggesting that the exact rotation method of the predictors is less important once we account for the whole information set. After employing lasso, results improve uniformly across models reflecting the importance of accounting for further components and the benefits of regularization in dealing with overfitting. Furthermore, the cross-validation step reveals that the required number of features varies significantly across time for all the PC portfolios. This implies that characteristics work better in predicting returns in certain periods than others, which is expected given the time variation in factor risk premia. Our lasso-based factor timing strategies are flexible enough to downgrade (upgrade) information in the characteristics when their informativeness is low (high).

1.2 Literature Review

Our paper is related to several strands of the literature. Without attempting a full-scale review, we discuss briefly how we contribute to two main categories, namely studies that utilize dimension reduction techniques in the context of asset pricing and studies that explore factor portfolio predictability.

1.2.1 Dimension reduction in asset pricing

Machine learning has surfaced in recent years in various asset pricing applications due to the limitations of standard methodologies in a high-dimensional setting. Gu, Kelly, and Xiu (2020) compare various machine learning techniques in their effort to forecast stock returns using a large collection of stock characteristics. Similarly, numerous studies attempt to identify the extent to which characteristics are associated with expected returns by regularizing the cross-sectional regressions or the characteristic-based portfolio sorts used in the estimation of risk premia. For instance, DeMiguel et al. (2020), Freyberger, Neuhierl, and Weber (2020), and Feng, Giglio, and Xiu (2020) employ lasso regularization to create a stochastic discount factor (SDF) with sparse characteristic exposure. However, imposing sparsity in the number of return predictors under a lasso approach may not be a realistic assumption after all due to the diverse characteristic space (Kozak, Nagel, and Santosh 2020). Nevertheless, sparse models allow for a parsimonious representation of the cross-section of expected stock returns and an easier interpretation and link to economic theories. In our empirical application, we apply lasso on a set of characteristic PCs instead of raw characteristics. Hence, our approach still encourages a sparse factor structure, while allowing multiple characteristics to have an effect on expected factor portfolio returns through their exposure to the characteristic PCs.

Another strand of the literature applies PCA on a set of stock or portfolio returns to reduce their dimensions. Examples of PCA applications in asset pricing include Connor and Korajczyk 1988, who apply asymptotic PCA on asset returns to extract latent factors, and Kozak, Nagel, and Santosh (2018), who form a low-dimensional SDF using the first few PCs of anomaly returns. Kozak, Nagel, and Santosh (2020) also find that a low-dimensional specification in terms of PC portfolios is feasible due to the high degree of common variation in factor portfolio returns. In general, the use of PCA in this context is both economically and empirically motivated. Economically, the existence of arbitrageurs in the economy implies that near-arbitrage opportunities, meaning extremely high Sharpe ratios, are implausible to achieve. Hence, high Sharpe ratios associated with low eigenvalue PCs should make no contribution to explaining returns (Kozak, Nagel, and Santosh 2018).[6] Empirically, returns possess a spiked covariance structure, meaning the variance-covariance matrix is dominated by a small number of large eigenvalues, separated from the rest. Combining these facts implies that asset returns should be adequately explained by a small number of dominant PCs. We contribute to this literature by constructing PC portfolios of LS portfolios and examining their predictability.

Several recent studies also focus on modifying conventional PCA with the purpose of making it more suitable for asset pricing applications. Kelly, Pruitt, and Su (2019) propose a new method of instrumental principal components, allowing latent factor loadings to be time-varying and partially dependent on firm characteristics.[7] They find that only a small number of characteristic-based factors are important for identifying a successful latent factor model. Lettau and Pelger (2020a) augment standard PCA by a cross-sectional pricing error in order to extract factors that can simultaneously explain the time-series variation and the cross-section of asset returns, and Lettau and Pelger (2020b) demonstrate the superiority of the estimator compared to standard PCA on a set of 37 factor portfolios. Finally, Giglio and Xiu (2021) account for omitted factors in the estimation of risk premia by combining PCA with two-pass cross-sectional regressions. We exploit the recent advancements in the literature by also using the RPPCA of Lettau and Pelger (2020a) to extract factors from LS portfolio returns.

[6]Still, this argument does not explicate whether high eigenvalue PCs reflect risk or mispricing.

[7]The method is an extension of the projected PCA by Fan, Liao, and Wang (2016) and can be thought of as standard PCA on characteristic sorted portfolios.

1.2.2 Factor portfolio predictability

In a factor timing context, factor momentum has emerged as a mechanism to time factor portfolio returns. Early contributors to this literature include Grundy and Martin (2001), who document a momentum effect in the factor component of stock returns. The momentum effect in factor portfolio returns is strong and has its own distinctive behavior, different from that of stock momentum. For example, Arnott et al. (2021) and Gupta and Kelly (2019) find that the effect is the strongest at the one-month horizon, even though stocks exhibit reversals in such short intervals. Nonetheless, factor momentum captures the effect at its purest form as it subsumes stock, industry momentum, as well as momentum found in other well-diversified portfolios (Arnott et al. 2021). Furthermore, factor momentum is concentrated in the highest eigenvalue PCs of factor portfolio returns, which implies that momentum is intertwined with the covariance structure of factor portfolios (Ehsani and Linnainmaa 2022). Whether looking at PC portfolios or individual factors, factor momentum can accommodate factor timing simply by buying (selling) portfolios that have performed well (poorly) in the recent past or relative to their peers. Such strategies deliver strong return performance and are not susceptible to crashes, as stock momentum (Gupta and Kelly 2019). Nevertheless, using exactly the same investment rule, we show that characteristic-based forecasts provide superior information and result in more profitable investment strategies compared to factor momentum.

Outside factor momentum, numerous studies attempt to predict the performance of individual factor portfolios using a collection of potential predictors. Daniel and Moskowitz (2016) forecast stock momentum using market indicators and volatility proxies in an effort to explain momentum crashes. Similarly, Huang (2022) finds that the return spread between winners and losers negatively predicts stock momentum returns. Baba-Yara, Boons, and Tamoni (2021) analyze the ability of the value spread to forecast the returns of the value-minus-growth portfolio across asset classes. They find that the first principal component of the value spread captures most of the variation in expected value returns. In a similar manner, we also use the first principal component of multiple characteristics to predict PC portfolio returns, even though we examine the possibility that further characteristic components are required. In contrast to previous studies targeting only specific anomalies, we examine factor portfolio predictability across a large set of factor portfolios.

Other studies also examine the predictability of multiple portfolios at once, using either a single or multiple predictors. Asness et al. (2017) use the value spread to construct timing strategies for value, momentum, and betting-against-beta portfolios, though they observe little improvement upon a constant multi-style strategy. Greenwood and Hanson (2012) show that corporate share issuance can be used to forecast the performance of factor portfolios related to size and value. Stambaugh, Yu, and Yuan (2012) find that LS strategies appear to be stronger following periods of high investor sentiment. They find the sentiment effect to be concentrated on the short leg of anomalies, which they base on the short-sale impediments that results in relatively higher overpricing compared to underpricing. On a much larger scale, Jacobs (2015) confirms the findings of Stambaugh, Yu, and Yuan (2012) by examining the role of sentiment in a large set of 100 anomalies. Kelly and Pruitt (2013) forecast four sets of characteristic-sorted portfolios using the cross-section of book-to-market ratios and observe higher predictability at lower frequencies. Dichtl et al. (2019) attempt to predict 20 equity factors using fundamental and technical indicators. They distinguish between cross-sectional and time-series predictability, which results in factor-tilting and factor timing portfolio allocations, respectively. Haddad, Kozak, and Santosh (2020) construct PC portfolios by running PCA on the time series of 50 anomalies and find that the largest eigenvalue PCs are the most predictable by their own book-to-market ratio.

We extend Haddad, Kozak, and Santosh's (2020) framework by incorporating information across a large set of observable characteristics to predict a large set of factor portfolio returns. Furthermore, we allow the effect of characteristics to be independently identified for every PC portfolio, examining the possibility that different characteristics affect different sources of variation in factor portfolio returns.

1.3 Methodology

This section begins by setting out the general framework, followed by our forecasting procedure and the benchmark models employed. Section A.2.2 of the Appendix introduces the statistical methods used in this study and provides a comprehensive overview of their functional form and statistical properties.

1.3.1 General framework

The main objective is to predict a large set of factor portfolio returns using a large set of portfolio characteristics. Let R be a $(T \times N)$ matrix of N factor portfolio returns for T periods. Equivalently, let $R_{t,\cdot} = (R_{t,1}, \ldots, R_{t,N})$ be a $(1 \times N)$ vector of portfolio returns at time t and C_t, a $(N \times M)$ matrix of M characteristics for N factor portfolios at time t. The base case arises from a conditional version of Cochrane's (2011) framework for modeling returns as a function of characteristics:

$$R_{t+1,n} = a_{t,n} + \sum_{m=1}^{M} b_{t,n}^{m} C_{t,n}^{m} + \varepsilon_{t+1,n}, \tag{1.1}$$

where $a_{t,n}$ and $b_{t,n}^{m}$ denote the conditional alpha and beta at time t, and $\varepsilon_{t+1,n}$ is the pricing error at time $t+1$. Entertaining time variation in $b_{t,n}^{m}$ and $a_{t,n}$ due to changes in portfolio attributes is the essence of factor timing.[8] By combining different dimension reduction techniques, we essentially investigate the possibility that the conditional alphas and betas are a function of the covariance of returns, the covariance of the characteristics, or even the covariance of returns with the characteristics. The covariance of returns comes into play by focusing on the dominant components of factor portfolio returns instead of predicting each factor portfolio separately. More concretely, assuming a linear latent factor specification, excess asset returns can be expressed as:

[8]Cochrane (2011) uses the formulation in Equation (1.1) to model the returns of an individual stock in excess of the risk-free rate. In our setting, we model factor returns, i.e., the returns of a long portfolio in excess of the returns of a short portfolio. Getting from individual stocks to factor portfolios is straightforward, and hence we focus directly on the latter to simplify the exposition of our framework.

$$R_{t+1,.} = \mathrm{Z}_{t+1,.} W_t' + \Theta_{t+1}, \tag{1.2}$$

where $Z_{t+1,.} = (z_{t+1,1}, z_{t+1,2}, \ldots, z_{t+1,K})$ is a $(1 \times K)$ vector of factor returns with $K << N$, $W_t = (w_{t,1}, w_{t,2}, \ldots, w_{t,K})$ is a $(N \times K)$ matrix of factor loadings and Θ_{t+1} is a $(1 \times N)$ vector of idiosyncratic errors. The time dimension in this context arises by the recursive estimation of eigenvectors and principal components. The first term of the right-hand side reflects compensation for the exposure on systematic risk factors while the second term reflects asset-specific risk. Under the assumption that the factors and the errors are uncorrelated, the variance-covariance matrix of asset returns can be decomposed into a systematic and idiosyncratic part. A common practice is to estimate $Z_{t+1,.}$ and W_t directly, by applying PCA on the variance-covariance matrix of R and retaining the dominant components (e.g., Connor and Korajczyk [1986] and Kozak, Nagel, and Santosh [2018]). Provided that time variation in asset risk premia is driven by exposure to time-varying aggregate risk, being able to accurately predict the dominant components $Z_{t+1,.}$ allows us to form forecasts for individual anomalies through W_t. By only focusing on $Z_{t+1,.}$, we isolate common sources of predictability across factor portfolios and ignore spurious predictability associated with smaller PCs.

In order to forecast $Z_{t+1,.}$, we model PC portfolio returns as a function of observable characteristics. Specifically, lagged characteristics are used to predict next-period PC portfolio returns. The characteristics of the PC portfolios are computed by combining factor portfolio characteristics according to their weights given by the i^{th} eigenvector $w_{t,i}$. The cross-section of characteristics for the i^{th} $i = (1, \ldots, K)$ PC portfolio is calculated as $H_{t,i} = w_{t,i}' C_t$. Repeating the process for every t and every i results in a $(T \times M)$ matrix H_i of characteristics for each PC portfolio.

However, using raw characteristics as inputs in standard predictive regressions would be suboptimal due to high correlations and lack of predictive information for some of them. Therefore, we transform the characteristics of PC portfolios into scores by using PCA and PLS. This is achieved by multiplying the matrix of characteristics H_i with a matrix of eigenvectors, such as:

$$X_i = H_i Q_{t,i}, \tag{1.3}$$

where X_i is a $(T \times M)$ matrix of component scores, or linear combinations of the underlying characteristics, of the i^{th} PC portfolio. Similarly, $Q_{t,i} = (q^1_{t,i}, q^2_{t,i}, \ldots, q^M_{t,i})$ is an $(M \times M)$ matrix of eigenvectors estimated at time t and sorted by their corresponding eigenvalues. For PCA, $Q_{t,i}$ is estimated based on the eigenvalue decomposition of $Var(H_i)$, while for PLS it is based on the eigendecomposition of $cov(Z_{.,i}, H_i)$; more information on how to obtain $Q_{t,i}$ and X_i under the different methods is provided in the Appendix. Dominant PCA components capture most of the variation within the characteristics, while dominant PLS components capture most of the covariation between lagged characteristics and next-period returns. Making a connection to the conditional betas, the use of PCA for the characteristics introduces their covariance into the conditional beta function, while the use of PLS introduces the covariance of PC portfolio returns with their characteristics.

Next, we model PC portfolio returns using the characteristic components:

$$z_{t+1,i} = \beta^0_{t,i} + \sum_{m=1}^{M} \beta^m_{t,i} X^m_{t,i} + \epsilon_{t+1,i}, \tag{1.4}$$

where $X^m_{t,i}$ is the m^{th} characteristic component of the i^{th} PC portfolio at time t, and $z_{t+1,i}$ is the one-month-ahead return of the same portfolio. Equations (1.1) to (1.4) lead to:

$$R_{t+1,.} = \sum_{i=1}^{K} w'_{t,i}\beta^0_{t,i} + \sum_{i=1}^{K}\sum_{m=1}^{M} w'_{t,i}(\beta^m_{t,i} w'_{t,i} C_t q^m_{t,i}) + \eta_{t+1,.}, \tag{1.5}$$

where $\eta_{t+1,.}$ is a $(1 \times N)$ vector of composite errors capturing both the unexplained return variation from the characteristics as well as the variation from potentially omitting higher-order PC components. Equation (1.5) shows that $a_{t,n}$ and $b^m_{t,n}$ from Equation (1.1) end up being functions of the eigenvectors of the covariance of factor portfolio returns, $w_{t,i}$, the eigenvectors of the covariance of characteristics, $q^m_{t,i}$, and the betas, $\beta^m_{t,i}$, from regressing PC portfolio returns on their characteristic components. Specifically, multiplying the $(1 \times N)$ vector $w'_{t,i}$ with the $(N \times M)$ matrix C_t gives the cross-section of PC portfolio characteristics, while multiplying that product with the $(M \times 1)$ vector $q_{t,i}^m$ gives the m^{th} characteristic component of the i^{th} PC portfolio. Further multiplying that new product with $\beta_{t,i}^m$ and summing over M gives the value for the i^{th} PC. Finally, multiplication with $w'_{t,i}$ and summation over K generates a return vector for the whole cross-section of factor portfolio returns. Note that all the objects have time t subscripts since they can be estimated recursively.

1.3.2 Forecasting procedure

We use at least 20 years (240 months) of information to estimate the PC portfolios and their characteristics to then make return predictions at $t+1$. Our forecasts employ an expanding estimation window, with the estimation sample always starting at the beginning of the sample period and incorporating additional observations as they become available. PC portfolios are recursively reestimated at each point in time, using an updated $w_{t,i}$ with $i = 1, \ldots, K$ based on the in-sample variance-covariance matrix of factor portfolio returns.[9] Notice that PC portfolio characteristics H_i change not only because of the change in the underlying factor portfolio characteristics C_t, but because of the change in the weighting vectors $w_{t,i}$ as well. Overall, our approach is flexible enough to account for a potentially unstable correlation structure in the factor portfolio returns.

In a similar fashion, the matrix of characteristic components is obtained as follows: for PCA, which only utilizes information contained in the characteristics to extract the latent factors, characteristics up to t are used to estimate X_i. For PLS, which uses information in both characteristics and returns, characteristics up to $t-1$ and PC portfolio returns up to t are used to estimate X_i. The βs in Equation (1.4) are always estimated using returns up to t and values in X_i up to $t-1$. Values of X_i at t are then plugged into Equation (1.4) to obtain forecasts for each PC portfolio return at $t+1$. Hence, our forecasts are completely out of sample and do not suffer from any look-ahead bias.

Another subtle but important detail is the cross-sectional standardization of C_t to

[9]When RPPCA is used for the left-hand side, we use a constant value of $\gamma = 10$ for the weight on the cross-sectional pricing error.

account for the difference in the scale of the characteristics. Running raw PCA or PLS on C_t would tilt the PCs toward the larger characteristics, as those will have significantly higher variance. For this reason, we standardize the matrix of factor portfolio characteristics C_t cross-sectionally before calculating H_i and ultimately X_i by subtracting the cross-sectional characteristic mean and dividing by the cross-sectional characteristic standard deviation at each time t. Apart from ensuring a reasonable covariance matrix for the characteristics, such an approach allows us to focus on the cross-sectional differences in the data. As long as factor portfolio characteristics coincide with factor portfolio returns in cross-sectional terms, PC portfolio characteristics should coincide with returns across time, as they are both linear combinations of the cross-section, and thus making a predictive regression approach sensible.[10]

The first decision being made is on the optimal number of factors in Equation (1.2). Specifying the optimal number of PCs is ultimately an empirical question, as it depends on the underlying factor structure. Bai and Ng (2002), Onatski (2010), and Haddad, Kozak, and Santosh (2020) all develop critical value thresholds for determining the number of factors. We follow a simple approach and focus on the first five PCs, as they capture about 67% of the variation in factor portfolio returns. Selecting the first five PC portfolios is also consistent with similar studies performing PCA on a set of factor portfolios—for example, Haddad, Kozak, and Santosh (2020) and Lettau and Pelger (2020b). Hence, let $Z_{t,5} = (z_{t,1}, z_{t,2}, \ldots, z_{t,5})$ and $W_{t,5} = (w_{t,1}, w_{t,2}, \ldots, w_{t,5})$ be the set of the five largest PC portfolios and eigenvectors.

The second decision to be made is on how to estimate $\beta^m_{t,i}$ in Equation (1.4). Here, we examine two different cases, one that imposes sparsity and one that is data-driven. In the first case, we only use the first characteristic component of each PC portfolio (i.e., the first column of X_i) in standard bivariate predictive regressions. Although this is the sparsest specification possible, multiple characteristics can have an effect on PC portfolio returns through their weights on the first characteristic PC. As an alternative, we apply

[10]In Section A.2.3 of the Appendix, we provide a detailed description of the standardization approach and explain the drivers of variation in the PC characteristics across time.

lasso on the whole set of characteristic components for each PC portfolio (i.e., the whole matrix X_i) to identify a subset that is useful for our forecasting objective. Hence, βs in Equation (1.4) for the first case are obtained through OLS for a single predictive factor ($m = 1$), and in the second case the βs are obtained through lasso for $m = 1, \ldots, M$.

When performing lasso the optimal amount of coefficient shrinkage is selected by conducting cross-validation on a rolling basis. In particular, before every forecasting step we separate the in-sample period into a training and a validation sample. The training sample is used to estimate the PC portfolios and characteristic PCs, and the validation sample is used to identify the degree of model complexity that delivers reliable out-of-sample performance.[11] At the start, the training sample is used to forecast the first period in the validation sample subject to a geometric sequence of shrinkage values.[12] The actual value of the forecasted data point is then used as part of the next training set to forecast the subsequent point in the validation sample. After repeating this procedure for every period in the validation sample, we pick the level of shrinkage that minimizes the mean-squared error. We then reestimate the PC portfolios and characteristic PCs using the whole in-sample period (training and validation) and apply lasso using the fixed value for the shrinkage parameter to estimate $\beta_{t,i}^m$ and predict PC portfolios at $t+1$. Depending on the magnitude of the shrinkage, our approach examines the possibility that none of the characteristic components are relevant in predicting PC portfolio returns, in which case returns forecasts shrink down to a constant term.

As already discussed, lasso is applied separately on each PC portfolio, meaning that the number of features can be different across PC portfolios. Essentially, our method allows for different sources of variation in factor portfolio returns to be approximated by models of different complexity, examining the possibility that characteristic importance varies across the main sources of return variation. Furthermore, since lasso is applied iteratively, the number of features can also vary across time for each PC portfolio de-

[11]The validation sample covers the last five years (60 months) of the in-sample period, while the training sample increases by one at each forecasting step.

[12]The sequence of shrinkage values is strictly positive and terminates at a value for which all coefficients are equal to zero.

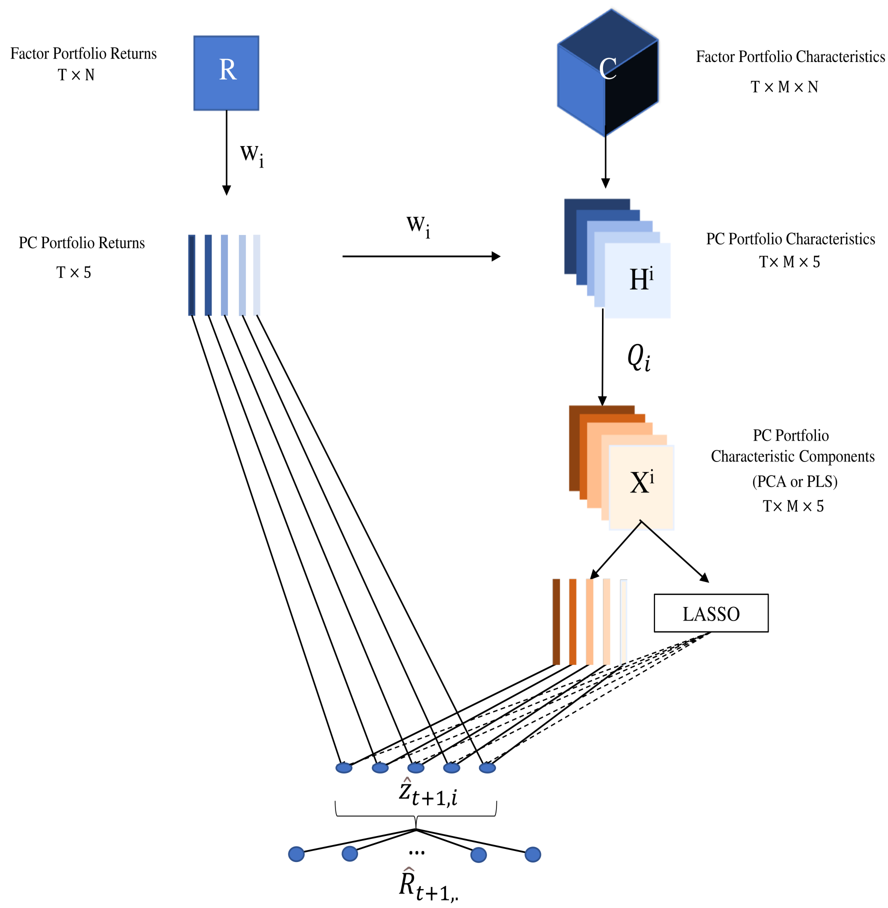
pending on how strong the characteristic signal has been in the recent past. Lastly, it is important to highlight that lasso can select low eigenvalue characteristic PCs, as long as they contribute to minimizing the forecasting error in the validation period.

To summarize, we attempt to regularize both the left-hand-side (LHS) and the right-hand side (RHS) of the predictability problem by combining different dimension reduction techniques. Regularization in the number of forecasting targets is achieved with the use of PCA or RPPCA and in the number of predictors with the use of PCA or PLS, resulting in four base models that we define as PCA, RPPCA, PCA-PLS, and RPPCA-PLS.[13] Figure 1.1 provides a visual depiction of our procedure that can be summarized in the following steps:

1. Reduce a set of factor portfolios to their first five components using PCA or RPPCA.
2. Estimate the characteristics of the PC portfolios using their loadings from the first step.
3. Rotate PC portfolio characteristics using either PCA or PLS.
4. Either select the first characteristic PC or apply lasso on the whole set of characteristic PCs of each PC portfolio.
5. Produce separate forecasts for each PC portfolio using the selected number of features.
6. Expand these forecasts to individual factor portfolios using their loadings on each PC portfolio.

[13] All models are estimated using either single or multiple predictors (via lasso), resulting in a total of eight forecasting models. Panel A of Table A.3 includes the listing of the four main models, which are estimated using either a single predictor or multiple predictors in combination with lasso.

Figure 1.1: Visual depiction of our modeling procedure. The figure presents the process of forecasting factor portfolio returns using their portfolio characteristics. PC portfolios are calculated as linear combinations of factor portfolios. The same weighting vectors are used to decompose the three-dimensional set of characteristics into five independent matrices of characteristics (one for each PC portfolio). The matrices of predictors are transformed to components, and either the first component is retained or lasso is applied on the whole set of components to pick those that are the most informative. Individual forecasts for each PC portfolio are produced, and those forecasts are aggregated into factor portfolio return forecasts using the weighting vectors that were used to aggregate factor portfolios into PC portfolios.



1.3.3 Benchmark models

To examine whether characteristic-based models provide superior information compared to different approaches, we employ alternative information sets to predict factor portfolio returns. Panel B of Table A.3 includes a list of all the benchmark models used. In a general setting, we form the baseline benchmark models following the methodological framework proposed by the original authors. Section 1.5 modifies the original models in various ways in order to examine the robustness of our results.

2.3.1 Factor momentum. The first benchmark is the one-month momentum strategy (1mMOM), which forms the momentum signal based on a look-back window of one month. Essentially, the return at time t becomes the prediction for the return at time $t+1$. The second benchmark is the 12-month momentum strategy (12mMOM), which forms the momentum signal based on a look-back window of 12 months. In this case, the prediction for the return at time $t+1$ is the average monthly return of the previous 12 months. In order to improve consistency across characteristic and momentum models, in Section 1.5 we also apply both momentum strategies to the PC portfolios and then extend the forecasts to individual anomalies as in Equation (1.5). Hence, we also examine the possibility of a stronger momentum effect on the main sources of variation of factor portfolio returns.[14]

2.3.2 Valuation ratios. As a third benchmark, we use only the book-to-market ratio of factor portfolios as a return predictor. Specifically, we follow Haddad, Kozak, and Santosh (2020) in predicting the first five PCs by their own book-to-market ratio and then extending the forecasts to individual anomalies. In order to keep things consistent with our framework, we estimate the PC portfolios recursively rather than using the first half of the sample. In Section 1.5, we simultaneously use the book-to-market ratio of all dominant PC portfolios in combination with lasso as an alternative to the baseline model.

2.3.3 Issuer-repurchaser spread. Following Greenwood and Hanson (2012), we estimate the issuer-repurchaser spread of each portfolio and use it to predict next-period factor portfolio returns. The issuer-repurchaser spread is defined as the average characteristic decile difference between issuers and repurchasers. Repurchasers are defined as firms that have reduced their shares outstanding by more than 0.5% during the fiscal year, and issuers are firms that have increased their shares outstanding by more than 10% during the fiscal year. The metric can take values from -9 to 9, with low values implying that issuers are located in the low leg and repurchasers in the high leg of each factor portfolio (and vice versa). In Section 1.5, we generalize this approach by considering the

[14]For instance, Ehsani and Linnainmaa (2022) observe that momentum is highly concentrated among the first five PC portfolios.

issuer-repurchaser spreads of the PC portfolios.

2.3.4 Investor sentiment. We explore the role of investor sentiment in predicting factor portfolio returns. Stambaugh, Yu, and Yuan (2012) and Jacobs (2015) find that anomaly performance is stronger following periods of high sentiment. To examine the effect of sentiment, we use the investor sentiment index of Baker and Wurgler (2006), which captures the common component in five sentiment proxies, with each proxy being orthogonalized with respect to six macroeconomic indicators. Specifically, next-period factor portfolio returns are regressed on the lagged values of the index, and forecasts for individual anomalies are formed based on a standard regression setting. In Section 1.5, we also form forecasts for individual PC portfolios and employ lasso to examine potential time variability in the sentiment signal.

2.3.5 Historical sample mean. Finally, we use the in-sample average of factor portfolio returns as a forecast for the next period, as in Campbell and Thompson (2008). Such a simple nonparametric technique utilizes information in the returns only, allowing us to examine the incremental effect of sophisticated statistical techniques and different information sets.

1.4 Empirical Results

1.4.1 Data

We replicate a large set of 72 characteristics, also considered by Green, Hand, and Zhang (2017). The characteristics are calculated using data from the Center of Research in Securities Pricing (CRSP) and Compustat. Our dataset covers the 50-year period from January 1970 to December 2019. The stock universe includes common stocks listed on NYSE, AMEX, and NASDAQ that have a record of month-end market capitalization on CRSP and a nonmissing and non-negative common value of equity on Compustat. Additional information about the characteristics, including origination and characteristic description, can be found in Table A.1 of the Appendix.

For every month in our sample, stock returns at month t are matched against their respective characteristics at month $t-1$. For accounting data, we allow at least six months to pass from the firms' fiscal year end before they become available and at least four months to pass for quarterly data. We also winsorize characteristics cross-sectionally at a 99% confidence level to account for extreme outliers. Finally, to isolate the effect of microcaps, we remove stocks with price below $5 at the portfolio formation period and use NYSE breakpoints to split stocks into deciles, following Fama and French (2008). These adjustments help us robustify our inferences, since many anomalies have been found to work better on small stocks (Fama and French 2008).

We then move to the construction of the factor portfolios. For each anomaly, we first group stocks into value-weighted deciles based on their characteristic exposure in the previous month and then go long decile 10 and short decile 1,[15] even if the characteristic is negatively related to future returns. Such an approach requires no ex ante information about the relationship between characteristics and returns, and results in the highest dispersion in factor portfolio returns. Furthermore, given that factor timing strategies can take long and short positions on factors, the sign of factor portfolio returns is irrelevant.[16] Similarly to computing factor portfolio returns, the characteristics of factor portfolios can be computed by value-weighting characteristics of stocks within each decile portfolio and then subtracting the value of the bottom from the top decile. Notice that the portfolio constructed based on a particular characteristic sort will also have the highest characteristic score by construction.[17]

Figure 1.2 displays the average monthly returns of the factor portfolios together

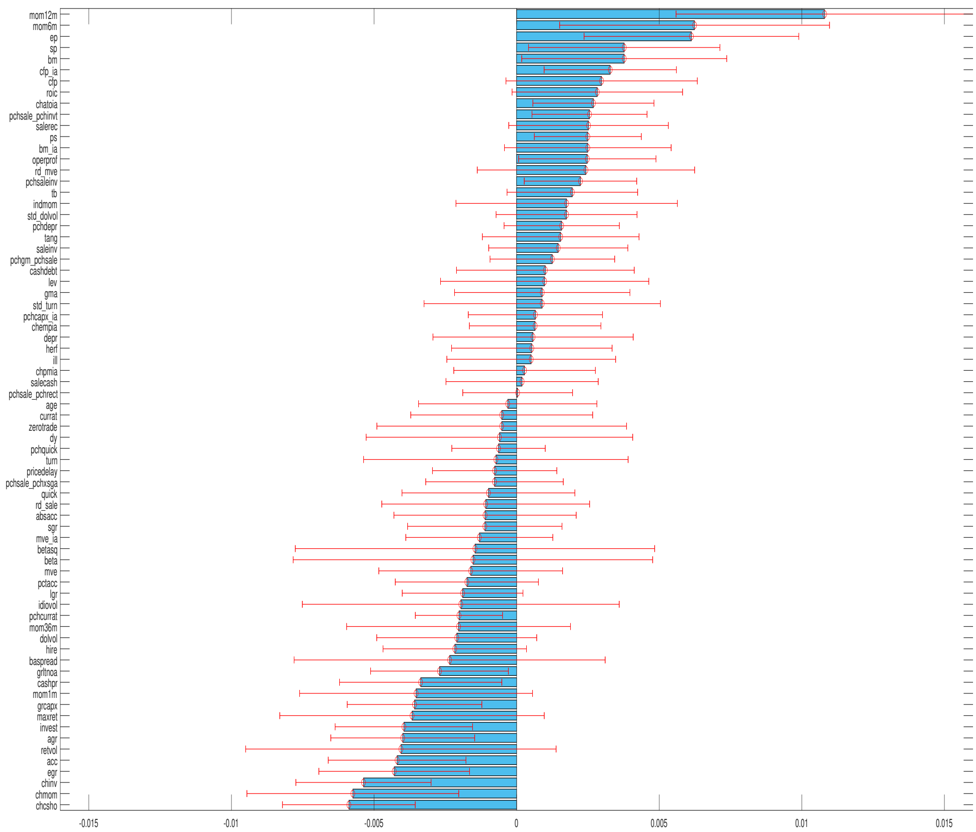
[15]In the early years of the sample period, there are few characteristics, such as characteristics based on research and development expenses, which do not have enough variation in order to form 10 separate portfolios. To account for this, we allow the number of quantiles to be less than 10 for months in which the required number of cutoff points is not reached. In other words, LS portfolio returns are calculated as long as there are at least two different values for the same characteristic in a particular month.

[16]Hence, strategies with a negative risk premium, such as asset growth, should on average be allocated in the short side of our factor timing portfolio.

[17]For example, the momentum portfolio will always have the highest momentum score compared to all the other factor portfolios.

with the 95% confidence intervals. Out of all the factor portfolios, 12-month momentum (mom12m) has the highest average return, followed by 6-month momentum (mom6m). Out of the 72 portfolios, only 22 have significant average returns, confirming a high degree of redundancy among the documented factors (Hou, Xue, and Zhang 2020). When we focus on the out-of-sample period only, this number goes down to 10, reflecting the decay in the performance of the anomalies over time (McLean and Pontiff 2016). Further descriptive statistics for the factor portfolios can be found in Table A.4 in the Appendix.

Figure 1.2: Average monthly returns of factor portfolios with 95% confidence intervals for the period January 1970 to December 2019.



As already discussed, we proceed by constructing recursively five PC portfolios—that is, linear combinations of the 72 factor portfolios using either PCA or RPPCA. These PC portfolios are by construction affected by all factor portfolios in a time-varying fashion; as a result, at a first glance they might look as if they do not have any economic interpretation. In order to tackle this, we recursively regress each PC portfolio return on each of the 72 anomalies and estimate the monthly time series of R^2 values for each anomaly. The analysis, which is detailed in Section A.2.6 of the Appendix, shows that the constructed PC portfolios have in fact a quite clear economic interpretation. For example, the first PC portfolio (based on PCA) loads heavily on volatility characteristics, the second one loads more on value characteristics, while the third one is driven mostly by momentum characteristics. Moreover, despite the recursive construction procedure of the PC portfolios, these economic relations are very stable over time.

1.4.2 Predictive performance

We examine the out-of-sample performance of our predictive models using standard forecast evaluation measures and a monthly holding period as in Campbell and Thompson (2008). We use an in-sample window of at least 240 months, with the initial in-sample period covering the period January 1970 to December 1989 and forecasts being obtained out-of-sample for the period January 1990 to December 2019. As a first indication of the out-of-sample fit of our models, we estimate the out-of-sample R^2 for each individual PC portfolio as:

$$\text{OOS } R^2 = 1 - \frac{\sum_{t=240}^{T-1} \left(z_{i,t+1} - \widehat{z}_{i,t+1}\right)^2}{\sum_{t=240}^{T-1} \left(z_{i,t+1} - \bar{z}_{i,t+1}\right)^2}, \tag{1.6}$$

where $\widehat{z}_{i,t+1}$ is the PC portfolio return forecast at time $t+1$ and $\bar{z}_{i,t+1}$ is the average PC portfolio return using information up to period t. We also estimate a total OOS R^2, which pools squared errors across factor portfolios and across time:

$$\text{Total OOS } R^2 = 1 - \frac{\sum_{i=1}^{N} \sum_{t=240}^{T-1} \left(R_{i,t+1} - \widehat{R}_{i,t+1}\right)^2}{\sum_{i=1}^{N} \sum_{t=240}^{T-1} \left(R_{i,t+1} - \bar{R}_{i,t+1}\right)^2}. \tag{1.7}$$

Total OOS R^2 assesses the predictive ability of each model under a grand panel framework and therefore is a bulk measure of the accuracy of the model-based predictions of future factor portfolio returns. Table 1.1 presents the OOS R^2 results for individual PC portfolios, as well as the total OOS R^2 under the various models. Apropos panel A, characteristic-based models with one predictive factor deliver negative OOS R^2, with only the second PC portfolio being predictable. With regards to the different dimension reduction techniques used, models that use PCA for the RHS outperform their PLS counterparts in terms of total OOS R^2, although someone would expect the opposite given that PCA factors capture variation among returns-related and unrelated variables.

Table 1.1: OOS R^2 for PC portfolios and total OOS R^2 across all anomalies for the period January 1990 to December 2019 in percentage terms. Panel A displays results using a single latent factor to predict PC portfolio returns. Panel B shows the results where the optimal number of factors is selected by applying lasso on the set of latent factors. Panel C displays results for the benchmark models.

	PC1	PC2	PC3	PC4	PC5	Total
A. Single factor						
PCA	−1.00	0.24	−1.33	0.36	0.28	−0.55
PCA-PLS	−3.29	0.76	−2.89	0.60	−0.95	−1.55
RPPCA	−0.75	0.90	−0.13	−1.15	−0.74	−0.38
RPPCA-PLS	−3.00	1.80	−3.28	−3.94	−1.60	−1.54
B. Time-varying number of factors using lasso						
PCA	2.71	2.85	1.23	3.24	−0.78	1.46
PCA-PLS	1.70	5.34	1.78	3.87	3.75	1.53
RPPCA	0.80	0.60	1.71	0.41	2.64	0.52
RPPCA-PLS	1.17	6.97	0.92	0.20	2.30	1.17
C. Benchmark models						
1mMOM						−88.98
12mMOM						−6.61
PCA-BM	0.40	3.47	−0.20	0.89	−0.62	0.43
IR spread						0.13
Sentiment						0.42

Moving to panel B, the combination of dimension reduction techniques with lasso significantly improves results for all models delivering positive total OOS R^2s. The predictive performance improves almost uniformly across all PCs, highlighting the importance of accounting for further characteristic components and the benefits of regularization on out-of-sample performance. The use of lasso in particular allows the models to underweight (overweight) information in the characteristics in periods where the characteristic signal diminishes (becomes stronger).[18] Overall, results in panel B confirm that imposing

[18]It is also important to highlight that lasso may select characteristic components other than the first,

a sparse or constant factor structure may not be a realistic assumption in the context of asset return prediction.

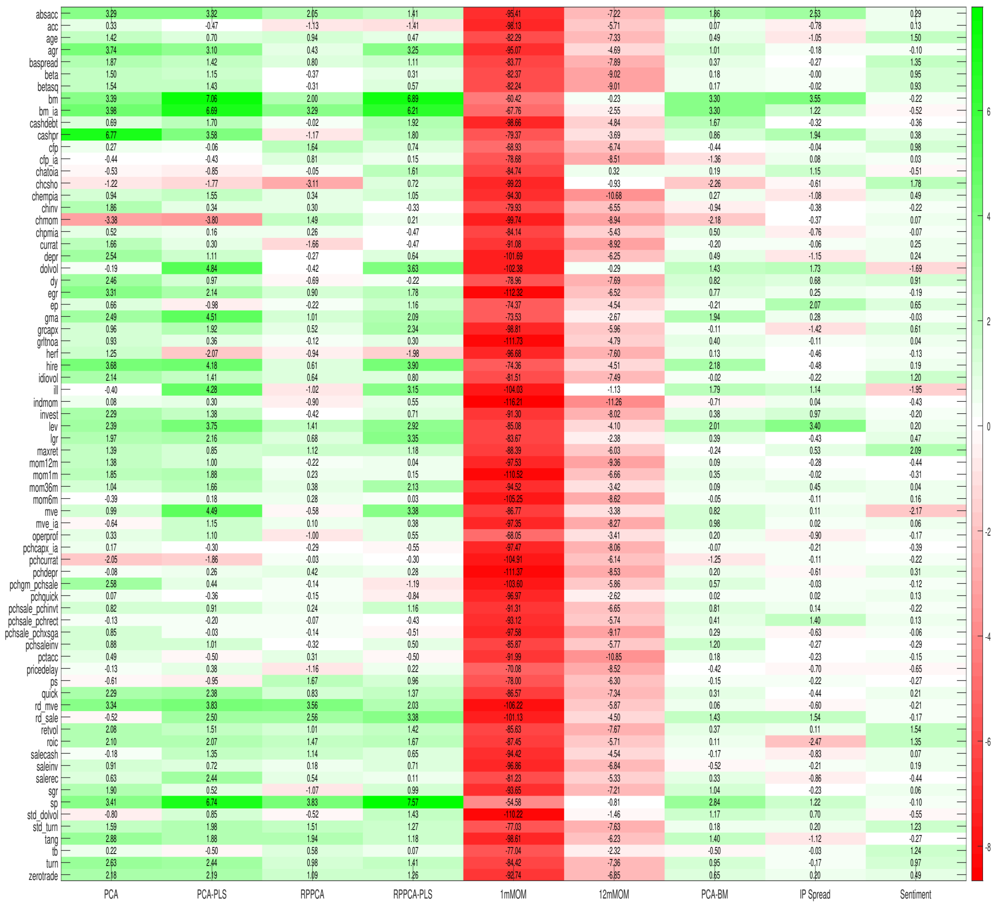
Finally, panel C displays the total OOS R^2 for the benchmark models.[19] With regards to factor momentum, previous month returns provide unreliable forecasts in exact terms, as implied by the highly negative total OOS R^2. When returns are averaged over the past 12 months, results improve significantly, although the total OOS R^2 remains on the negative side. Conversely, models based on the book-to-market ratio, issuer-repurchaser spread, and investor sentiment deliver positive total OOS R^2, though they still fall behind the characteristic-based models that employ lasso.

Ultimately, we are interested in the predictability of individual factor portfolios based on PC portfolio forecasts. As a measure of individual factor portfolio predictability, we estimate the individual OOS R^2 for all anomalies under the different models. Apropos Figure 1.3, expanding PC portfolio return forecasts to individual anomalies reveals predictive patterns in a robust and systematic way. In line with Haddad, Kozak, and Santosh (2020), we observe substantial anomaly predictability and find many predominant anomalies, such as value (bm) and sales-to-price ratio (sp) to be highly predictable by observed characteristics. However, almost all characteristic-based models fail to predict anomalies that are based on a % change in accounting variables, such as % change in sales minus % change receivables (pchsale_pchrect) and % change in the current ratio (pchcurrat) among others, located in the lower half of the heat map. These portfolios have returns statistically indistinguishable from zero and low covariance with the rest of the anomaly universe. As a result, they do not load heavily on the first five components, and their performance is not adequately captured by PC portfolio forecasts. With regard to the benchmark models, only factor momentum results in high forecasting errors and therefore negative OOS R^2 for almost all anomalies. The remaining benchmark models perform sufficiently well, delivering positive OOS R^2 for the majority of the anomaly universe.

potentially resulting in considerably different forecasts compared to the single-factor case.

[19]The historical sample mean is not included as it has a zero total OOS R^2 by construction.

Figure 1.3: OOS R^2 for individual anomalies under the characteristic-based models that employ lasso and benchmark models (historical sample mean is inferred by the R^2 metric). Negative values (in red) show lack of exact predictive ability, while positive values (in green) show exact predictive ability of the underlying model for a given factor portfolio.



Whereas OOS R^2 accommodates a general quantitative comparison of the predictive performance of the various models, it is also important to assess the statistical significance of the differences among model forecasts. To make pairwise comparisons of the out-of-sample predictive accuracy, we use the modified Diebold and Mariano (DM) test by Gu, Kelly, and Xiu (2020), which compares the cross-sectional average error differential between two models. The DM test statistic between two models (1) and (2) is defined as $DM_{1,2} = \bar{d}_{1,2}/\hat{\sigma}_{\bar{d}_{1,2}}$, where $\bar{d}_{1,2}$ and $\hat{\sigma}_{\bar{d}_{1,2}}$ are the mean and standard deviation of the error differential, defined as:

$$d_{1,2;t+1} = \frac{1}{N}\sum_{n=1}^{N}\left(\left(\hat{e}^{(1)}_{n,t+1}\right)^2 - \left(\hat{e}^{(2)}_{n,t+1}\right)^2\right), \tag{1.8}$$

where $\left(\widehat{e}_{n,t+1}^{(1)}\right)^2$ and $\left(\widehat{e}_{n,t+1}^{(2)}\right)^2$ denote the prediction error of factor portfolio return n at time $t+1$ under models (1) and (2), respectively.

Table 1.2: Modified Diebold-Mariano test for models that employ lasso and benchmark models. The table displays the modified DM statistic that compares the predictive performance of the column model with the row model. A positive value indicates that the column model outperforms the row model. The asterisks indicate statistical significance at a 10% (single), 5% (double), and 1% (triple) level.

	PCA	PCA-PLS	RPPCA	RPPCA-PLS	1mMOM	12mMOM	PCA-BM	IR spread	Sentiment
PCA-PLS	−0.10								
RPPCA	1.35*	1.26							
RPPCA-PLS	0.34	0.61	−0.93						
1mMOM	7.05***	7.19***	6.94***	7.17***					
12mMOM	2.73***	2.99***	2.40***	2.93***	−6.73				
PCA-BM	1.17	1.06	0.13	0.78	−6.88	−2.21			
IR spread	1.40*	1.32*	0.56	1.25	−6.85	−2.32	0.39		
Sentiment	1.16	1.08	0.19	0.94	−6.72	−2.30	0.02	−0.74	
Historical sample mean	1.75**	1.56*	0.89	1.35*	−6.83	−2.08	1.65*	0.31	0.94

Table 1.2 reports the results from the DM test for pairwise comparisons between the different models. For conciseness, we consider only the characteristic-based models that employ lasso as they outperform the single-factor models in terms of total OOS R^2. A positive value for the DM test statistic indicates that the column model outperforms the row model, and the asterisks indicate statistical significance at a 10% (single), 5% (double), and 1% (triple) level, respectively. We observe that the characteristic-based models provide significantly higher predictive accuracy than the factor momentum models and the historical sample mean model, even though the results are less strong in the latter case. In contrast, the higher predictive accuracy compared to the other benchmark model is not translated into statistical significance.

Nevertheless, predicting anomaly returns is of interest as long as it accommodates the construction of a profitable investment strategy. Specifically, in asset pricing the focus of interest is not so much on obtaining accurate predictions for individual returns, but rather on constructing portfolios with good risk-return properties (Nagel 2021). Put differently, we are more interested in predicting cross-sectional differences in returns than in predicting individual returns in exact terms. In that sense, total OOS R^2 is just a distance measure that does not reflect whether models can distinguish strong from weak performers. Consider, for example, a stylized hypothetical scenario with three factor portfolios and a forecasting period of only one month. If the realized returns of the portfolios are 3%, 2%, and 1%, the estimated historical samples means are 0%, 1%, and 2%, and the model-implied predictions are 6%, 5%, and 4%, respectively, then the predictive model will end up having a very negative OOS R^2 (–145.45%) even though it will be able to rank the portfolios perfectly. Consequently, models that yield higher total OOS R^2 do not necessarily yield better portfolios in terms of average returns or Sharpe ratios. This argument explains, for example, why the one-month factor momentum has been found empirically to be particularly profitable even though our results show that it has a very negative total OOS R^2. The disconnect between OOS R^2 and investment performance is discussed in detail both theoretically and empirically in Kelly, Malamud, and Zhou (forthcoming).

Given that predictive accuracy in relative terms might be more important than predictive accuracy in exact terms, we proceed by exploring two alternative measures—namely, the percentage of times that the sign of future factor portfolio returns is identified correctly and the average cross-sectional correlation between forecasted and realized returns. The former measure examines the ability of the models to predict the direction of individual factor portfolio returns, and the latter measure examines whether model-based forecasts capture the cross-sectional dispersion in factor portfolio returns. Table 1.3 presents the results.

When considering the single-factor predictive models in panel A, we observe that the models that use PLS for the RHS are far superior to the models that use PCA for the RHS, even though Table 1.1 shows that they exhibit worse OOS R^2. In order to understand this discrepancy, we can take the example of the PCA-PLS model. This model has a forecasting error that is lower than that of the historical sample mean model in 54% of

Table 1.3: Percentage of correct sign identifications and average cross-sectional correlation. Panel A displays results using a single latent factor to predict PC portfolio returns. Panel B shows the results where the optimal number of factors is selected by applying lasso on the set of latent factors. Panel C displays results for the different benchmark models.

	Proportion of correct sign (%)	Average cross-sectional correlation (%)
A. Single factor		
PCA	49.90	0.70
PCA-PLS	52.74	9.12
RPPCA	50.40	2.31
RPPCA-PLS	52.75	9.17
B. Time-varying number of factors using lasso		
PCA	52.19	7.83
PCA-PLS	51.95	9.04
RPPCA	52.23	7.74
RPPCA-PLS	52.24	8.51
C. Benchmark models		
1mMOM	51.88	5.98
12mMOM	52.40	5.43
PCA-BM	51.45	5.81
IR spread	51.22	4.26
Sentiment	51.27	4.71
Historical sample mean	50.50	3.24

the times. In those cases, it exhibits an OOS R^2 of 12.24% and an average cross-sectional correlation of 41%. In the remaining 46% of the cases, it exhibits an OOS R^2 of –17.17% and an average cross-sectional correlation of –28%. This means that, while the model's low overall OOS R^2 is driven by some large forecasting errors, its high overall cross-sectional correlation is due to the fact that in the majority of the cases it is particularly informative for the ranking of next-period portfolio returns. Importantly, panel B reveals that accounting for further components under a lasso approach harmonizes the performance across all four characteristic-based models. Finally, panel C shows that the benchmark models display slightly lower proportions of correct sign and markedly lower average cross-sectional correlations compared to the models in panel B.[20] Overall, results confirm that characteristic-based models can better distinguish anomaly performance compared to alternative approaches.

Recall from Section 1.3.2 that our predictive approach entails cross-sectional standardization of each characteristic in each month. Therefore, given the success of the approach, a natural question that arises is what is the source of variation in the characteristics of the factor portfolios that leads to predictability. This issue is discussed in detail in Section A.2.3 of the Appendix. We show that the main source of time variation comes from the higher moments of the cross-sectional distribution of the characteristics. This is intuitive given that the literature with respect to stock return predictability already establishes that the predictive power of several characteristics is closely related to their non-normal distribution. For example, Cooper, Gulen, and Schill (2008) show that asset growth is highly positively skewed, and accordingly, its predictive power is mainly driven by the high- rather than the low-asset growth stocks. Another source of variation comes from the time-varying correlations across the different characteristics. For example, it is possible that for a given month the correlation between stock momentum and value is high and hence the standardized momentum score of the respective factor portfolios is similar, while in another month the correlation might be low and hence the momentum score of the respective factor portfolios will be completely different. In the latter case, there is additional information content that can be exploited.[21]

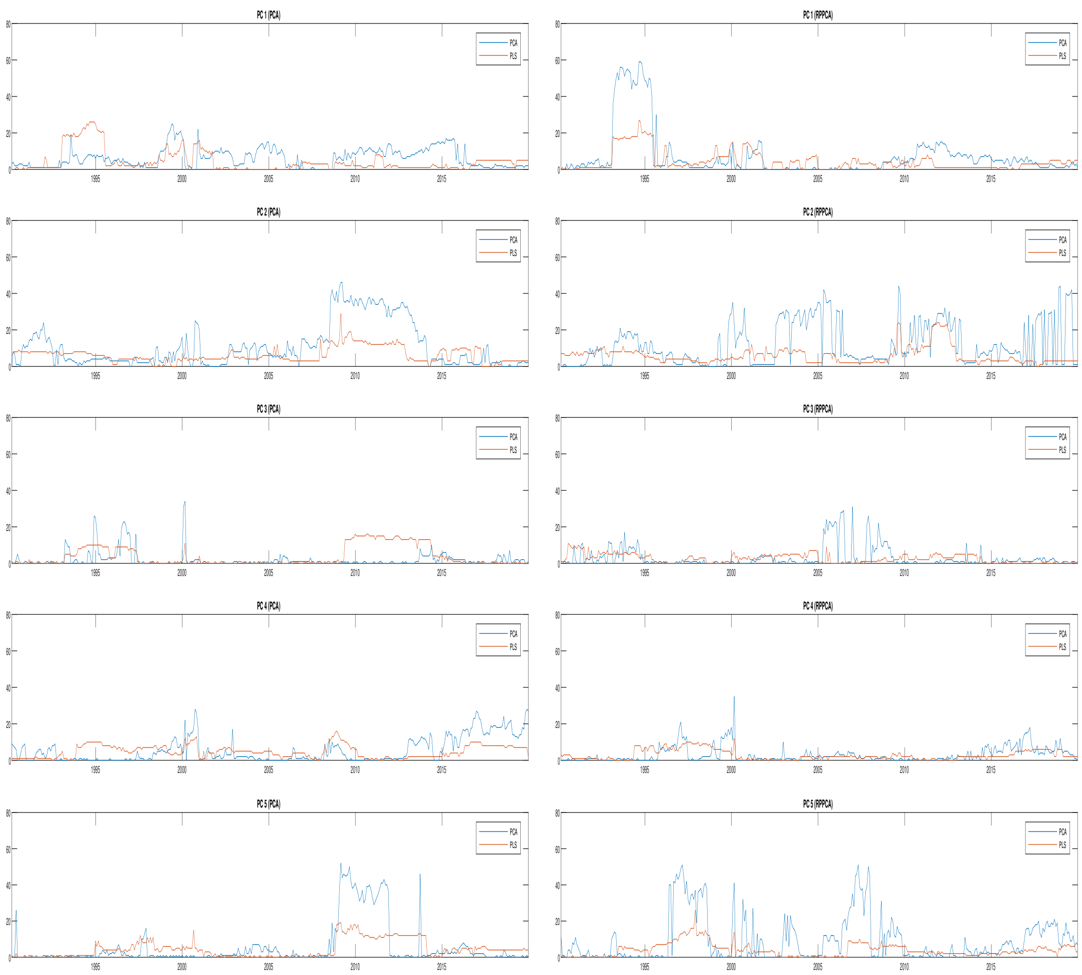
Finally, we examine the implications of applying lasso on the sets of characteristic components in terms of model complexity. Our approach allows for the number of features to vary across factor portfolios and across time, enabling us to see when the character-

[20] It is noteworthy that, similar to the case of the PLS single-factor models, the factor momentum models perform reasonably well despite their negative OOS R^2 values. In fact, the one-month factor momentum has a forecasting error that is lower than that of the historical sample mean model in only 20% of the times. In those cases, it exhibits an OOS R^2 of 38.93% and an average cross-sectional correlation of 60%. In the remaining 80% of the cases, it exhibits an OOS R^2 of –153.32% and an average cross-sectional correlation of –7%. This means that the good overall performance of the one-month factor momentum is driven by only a small subsample of observations during which it can predict future factor portfolio returns particularly well in terms of both exact and relative terms.

[21] Obviously, another source of variation stems from the recursively estimated weighting vector $w_{i,t}$. However, we show that this vector remains relatively stable across time.

istic signal is strong and when it diminishes. Figure 1.4 displays the number of nonzero coefficients for the characteristic PCs of each anomaly PC portfolio when PCA and PLS are used for the RHS in the out-of-sample period. Each line chart shows the number of characteristic-based components that minimize the mean-squared error in the validation period.

Figure 1.4: Number of features for each PC portfolio under the different models. The number of features is identified by recursively applying lasso on the set of components and picking the penalty factor that minimizes the mean-squared-error in the validation period.



Results from Figure 1.4 confirm the existence of significant time variability in the required number of features across time and across PC portfolios. The time variation in the number of features by itself implies that the predictive ability of characteristics is not constant, something that is expected given the time variation in factor portfolio risk premia. Interestingly, at certain periods the number of features falls down to zero, implying that at times characteristics provide no predictive information at all and the PC return forecasts shrink down to an intercept term. Conversely, a high number of features implies that a lot of the variation in the characteristics is useful in predicting PC portfolio returns. Such peaks and troughs in the number of features are observed at different points in time for the different PC portfolios, which implies that the importance of characteristics is also unstable across the main sources of variation and that each source should be approached independently in terms of model specification. Finally, with regards to the different methods used for the RHS, it is evident that PCA uses on average more features and has higher variability in the number of features across time compared to PLS. PCA components mix return-relevant and irrelevant information, making the selection of the optimal number of features more sensitive to the validation sample and as a result less stable. PLS condenses the characteristic information into fewer PCs than PCA and is more stable over time, although there is still significant time variability in the number of components being used.

1.4.3 Investment performance

In this section, we assess the performance of each model in terms of economic rather than statistical contribution and examine how return forecasts can be translated into factor timing strategies. We construct three different strategies and assess their performance using a monthly holding period and standard portfolio evaluation measures. The first strategy is a simple long-short strategy (LSS), or an LS portfolio of factor portfolios. Factor portfolios are grouped into equally-weighted deciles based on their return forecasts and a long-short strategy is constructed that goes long the top 10% and short the bottom 10% of the anomalies. Such a strategy focuses on the extremes of the conditional returns distribution and neglects factor portfolios that lie in the middle. Hence, LSS will work well as long as the models can identify anomalies with very high or very low expected returns at each period, even if they are indecisive about anomalies with conditional returns close to zero.

The second investment strategy is similar to the time-series factor momentum (TSFM) strategy by Gupta and Kelly (2019). TSFM scales factor portfolio returns $R_{t+1,\cdot}$ according to return forecasts $\hat{R}_{t+1,\cdot}$. The scaling vector $s_{t,n}$ is obtained by dividing return forecasts by individual factor in-sample monthly volatility and capping them at ± 2, as shown here:

$$s_{t,n} = \min\left(\max\left(\frac{1}{\sigma_{t,n}}\hat{R}_{t+1,n}, -2\right), 2\right). \tag{1.9}$$

The strategy goes long in factors with positive scores and short in factors with negative scores. The scores are rescaled to form unit dollar weights for the long and the short leg.[22] Multiplying next-period factor portfolio returns by their respective weights reveals the return of the strategy:

$$\mathrm{TSFM}_{t+1} = \frac{\sum_n 1_{\{s_{t,n}>0\}} R_{t+1,n} \times s_{t,n}}{\sum_n 1_{\{s_{t,n}>0\}} s_{t,n}} - \frac{\sum_n 1_{\{s_{t,n}\leq 0\}} R_{t+1,n} \times s_{t,n}}{\sum_n 1_{\{s_{t,n}\leq 0\}} s_{t,n}}. \tag{1.10}$$

The main difference between LSS and TSFM is that, while both are technically long-short, TSFM invests in the whole universe of factor portfolios and not in factor portfolios with extreme return forecasts only. Furthermore, the number of factor portfolios in each leg, as well as the relative weights, can differ for TSFM while remaining constant under LSS. More concretely, the sign of the return forecast determines whether the anomaly will be bought or sold, while the magnitude of the forecast determines the relative weight. Hence, under TSFM the long and the short legs can have a disproportional number of constituents, and in extreme cases, the strategy can converge to long or short only.

The last strategy, also in Gupta and Kelly (2019), is the cross-sectional version of TSFM (CSFM). The main difference between CSFM and TSFM is that the cross-sectional median is subtracted from the return forecasts before scaling with volatility. This strategy takes positions in factor portfolios that have outperformed or underperformed relative to their peers. For example, if return forecasts are positive for all factor portfolios, then

[22]Specifically, positive scores are divided by the sum of the positive scores, and the negative scores are divided by the sum of negative scores.

TSFM will take a long position in all of them, while CSFM will go long only in those with above-median return forecasts and short the rest. Hence, even if the models cannot identify the sign correctly, this strategy will still be profitable if forecasts are consistent in relative terms, similarly to LSS:

$$s_{t,n} = \min\left(\max\left(\frac{1}{\sigma_{t,n}}\hat{R}_{t+1,n} - \text{median}(\hat{R}_{t+1,\cdot}), -2\right), 2\right). \tag{1.11}$$

Table 1.4 presents the portfolio evaluation measures for the various models under the three strategies. First, we find that LSS delivers the highest average return among the three strategies across almost all the models, while CSFM and TSFM tend to have higher Sharpe ratios. Turning to panel A, results confirm the superiority of PLS over PCA for the RHS in the single-factor case, as also presented in Table 1.3. Evidently, models based on a single factor that concentrates the variation among multiple characteristics are unable to predict the cross-sectional dispersion of factor portfolio returns, implying again that a lot of variation in the characteristics is irrelevant in asset return prediction. As a result, strategies based on PCA and RPPCA deliver returns indistinguishable from zero, with returns for PCA even becoming negative. Conversely, when PLS is used for the RHS, all strategies deliver positive and significant returns, reflecting the ability of the method to concentrate return-relevant variation into a single predictor.

Panel C displays the results for the benchmark models. In line with prior literature (e.g., Gupta and Kelly 2019), factor momentum using a one-month formation period achieves the highest return among the benchmark models for the LSS strategy, while the 12-month signal delivers higher returns for TFSM and CFSM. Using the book-to-market ratio, issuer-repurchaser spread, or investor sentiment as predictors results in strategies with moderate return performance and Sharpe ratios. The historical sample average strategy delivers low average returns, albeit statistically significant. Such a strategy produces conservative return forecasts and, as a result, takes more static positions compared to the

Table 1.4: Portfolio evaluation measures for long-short (LSS), time-series (TSFM), and cross-sectional (CSFM) strategies under the different models for the sample period January 1990 to December 2019. Panel A displays results using a single latent factor to predict PC portfolio returns. Panel B shows the results where the optimal number of factors is selected by applying lasso on the whole set of latent factors. Panel C displays results for the benchmark models. Average return: average monthly return; Standard deviation: monthly standard deviation; Sharpe ratio: monthly Sharpe ratio; *t*-statistic: *t*-statistic on H_0: Average return = 0; Hit-rate: percentage of the total number of occasions that the strategy resulted in positive returns; Max drawdown: maximum cumulative loss. The best-performing model for each metric under each strategy is highlighted in bold.

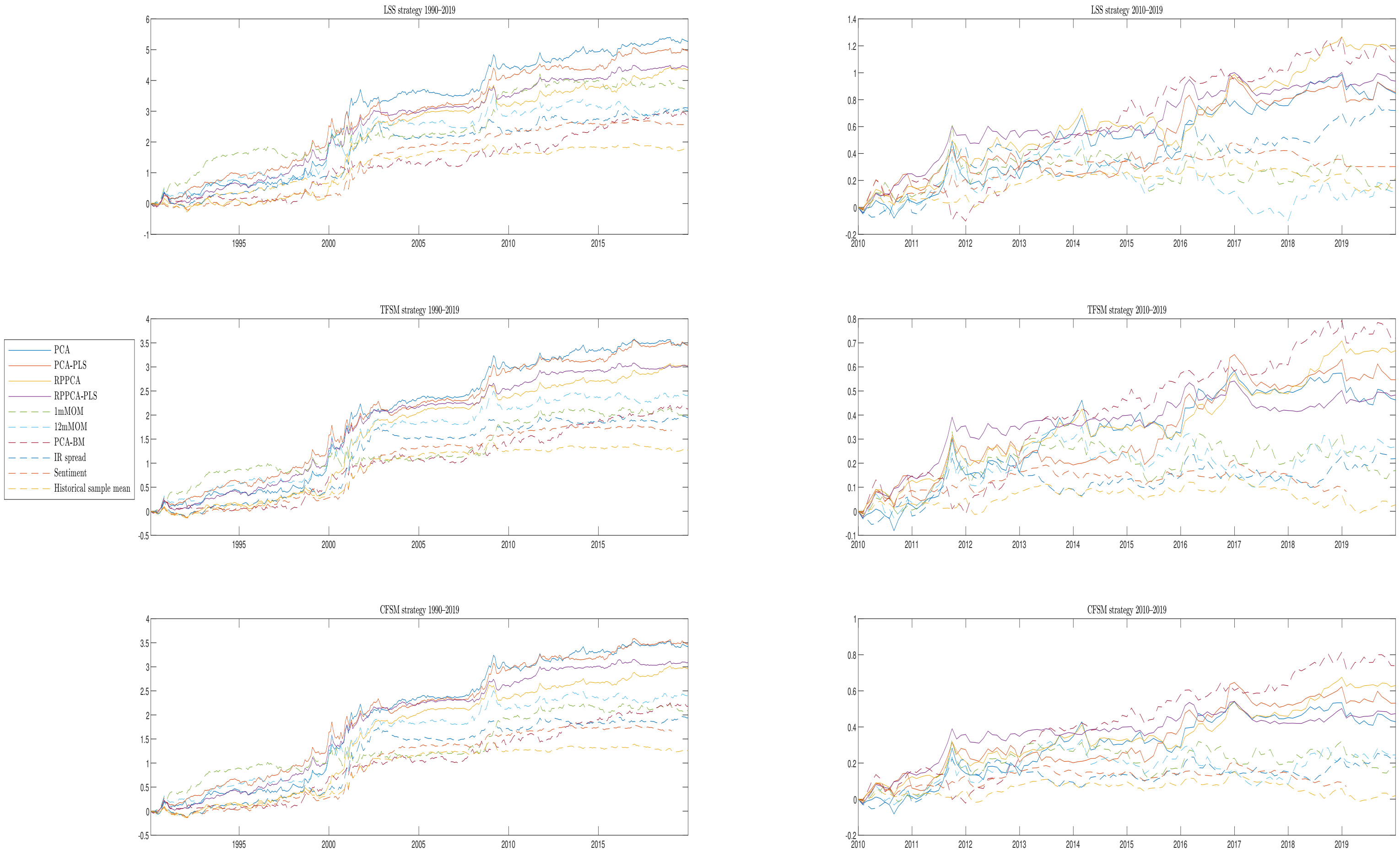
	Average return (%)			Standard deviation (%)			Sharpe ratio			*t*-statistic			Hit-rate (%)			Max drawdown (%)		
	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM
A. Single factor																		
PCA	-0.10	-0.03	-0.01	7.23	4.14	4.09	-0.01	-0.01	-0.00	-0.26	-0.13	-0.02	52.37	52.92	52.65	80.99	47.64	43.94
PCA-PLS	1.16	0.74	0.76	8.57	5.27	5.25	0.13	0.14	0.14	2.55	2.65	2.75	57.66	59.05	59.61	41.09	33.83	33.01
RPPCA	0.20	0.20	0.21	5.64	3.11	3.15	0.03	0.06	0.07	0.66	1.21	1.29	54.32	54.32	56.55	46.54	27.73	26.68
RPPCA-PLS	1.12	0.73	0.74	8.28	4.81	4.81	0.13	0.15	0.15	2.55	2.87	2.93	55.15	58.22	60.17	37.46	30.31	29.79
B. Time-varying number of factors using lasso																		
PCA	**1.47**	**0.97**	0.95	8.16	5.01	4.96	**0.18**	0.19	0.19	**3.40**	3.65	3.64	55.71	56.82	55.99	16.76	13.93	**12.16**
PCA-PLS	1.38	0.96	**0.97**	8.22	4.99	4.98	0.17	0.19	0.19	3.18	3.66	3.68	**61.00**	**62.67**	**61.56**	**15.00**	**13.19**	12.70
RPPCA	1.21	0.84	0.83	7.06	4.01	4.00	0.17	**0.21**	0.21	3.26	**3.99**	3.93	57.66	60.72	59.89	38.11	22.24	22.25
RPPCA-PLS	1.23	0.84	0.86	6.89	4.04	4.09	0.18	0.21	**0.21**	3.39	3.96	**3.98**	60.72	61.84	61.00	26.77	17.26	17.16
C. Benchmark models																		
1mMOM	1.06	0.56	0.58	8.81	4.95	4.96	0.12	0.11	0.12	2.28	2.13	2.22	57.10	56.82	57.10	18.45	16.94	17.32
12mMOM	0.84	0.67	0.67	8.69	5.08	5.12	0.10	0.13	0.13	1.84	2.51	2.48	55.43	55.99	56.82	25.96	17.89	18.28
PCA-BM	0.79	0.59	0.61	6.16	3.79	3.79	0.13	0.16	0.16	2.43	2.96	3.03	57.10	57.66	57.66	31.65	21.59	20.61
IR spread	0.87	0.54	0.54	6.49	3.92	3.92	0.13	0.14	0.14	2.49	2.63	2.61	52.87	52.65	52.92	34.74	25.51	25.71
Sentiment	0.74	0.47	0.46	5.29	3.13	3.22	0.14	0.15	0.14	2.60	2.82	2.72	56.32	53.48	52.92	21.80	20.15	19.80
Historical sample mean	0.48	0.35	0.35	**3.47**	**2.59**	**2.61**	0.14	0.14	0.13	2.64	2.59	2.56	58.22	57.94	58.22	19.19	19.22	19.03

rest of the models. Comparing results across panels, characteristic-based models that employ lasso outperform all benchmark models under all three strategies, demonstrating the benefits of conditioning factor portfolio returns on observable characteristics under a regularized framework.

In order to compare the performance of the various models across time, Figure 1.5 presents the cumulative return performance of the factor timing portfolios under the three investment strategies. For conciseness, we only display the performance for the characteristic-based models employing lasso together with the benchmark models. Graphs to the left show the cumulative performance over the whole out-of-sample period, and graphs to the right focus on the past 10 years. As it can be seen from the graphs, the one-month factor momentum outperforms the characteristic-based models in the early years of the out-of-sample period, up until the late 1990s. A spike in performance occurs for all strategies around 2000—that is, during the buildup of the dot-com bubble. Unlike the majority of the benchmarks, characteristic-based models do not plummet after the burst and continue to outperform thereafter. Furthermore, the performance of the characteristic-based models is relatively unaffected by the 2008 financial crisis, and a second spike in performance is observed as the economy enters the recovery phase in 2009. Hence, our strategies work well in periods of financial turmoil while still enjoying the upside potential of a bull market.

Finally, factor timing portfolios based on characteristics exhibit strong return performance in the post-2010 period. Looking at the graphs in the right panel of Figure 1.5, characteristic-based models display a positive trend in later years, while strategies based on issuer-repurchaser spread, investor sentiment, factor momentum, and in-sample average remain relatively stagnant. Out of all benchmark models, the historical sample mean remains the most stagnant, especially throughout the later years. Both approaches isolate the first five PCs of factor portfolio returns and use a characteristic-based measure to create forecasts.

Figure 1.5: Cumulative return performance of factor timing strategies. The figure displays the performance of LSS, TSFM, and CSFM for characteristic-based models using lasso and the benchmark models. Graphs to the left display the cumulative return performance over the whole sample period (January 1990 to December 2019), and graphs to the right display the cumulative performance over the past 10 years of the sample period (January 2010 to December 2019). All strategies begin with a zero dollar investment.



Notably, the book-to-market approach works equally well with the characteristic-based models in later years. As such, results highlight the importance of focusing on the main sources of variation and the ability of characteristics to explain the dynamics of factor portfolios. Characteristic-based models outperform the rest of the benchmarks under all three strategies, with the difference being more pronounced for the LSS strategy, as it focuses solely on the most prominent subset of factor portfolios. Overall, the profitability of the benchmark strategies erodes significantly in later years, suggesting that the informativeness of alternative predictors about future factor portfolio returns has faded.

It is also important to note that factor timing strategies based on observed characteristics yield positive returns in the most recent period, even though most factors have been found empirically to die out over time (Chordia, Subrahmanyam, and Tong 2014; Green, Hand, and Zhang 2017; McLean and Pontiff 2016). Corroborating this evidence, a comparison between Table 1.4 and Table A.4 reveals that characteristic-based factor timing strategies exhibit investment performance superior to that of unconditional factor portfolios. In that sense, our paper acknowledges the fact that unconditional risk premia lack robustness and shows that focusing on the predictability of conditional risk premia can help an investor expand her investment opportunity set. In a similar vein, Haddad, Kozak, and Santosh (2020) find that strong factor portfolio predictability implies a stochastic discount factor that is much more volatile than previously thought.

Lastly, a question that arises is what are the trading positions that our characteristic-based models take over time. The analysis presented in Section A.2.7 of the Appendix provides some interesting insights. First, even though prominent anomalies such as mom12m and retvol are heavily traded, the factor timing strategies rotate among multiple anomalies and do not focus on only a small subset with high unconditional returns. Second, several anomalies appear almost equally often in the long and the short legs. Finally, anomalies that have only a small impact on the PC portfolios are hardly considered by our factor timing strategies, which is expected given that their return forecasts are by construction tilted toward zero.

1.5 Alternative Approaches

In this section, we examine different estimation approaches. Our method uses a large collection of characteristics and combines different dimension reduction and regularization techniques to achieve robust out-of-sample predictability. As such, it is important to examine where the predictability stems from by evaluating the incremental effect of each contributor on the out-of-sample performance. Furthermore, it is important to assess whether the benchmark models can beat our characteristic-based models once dimension reduction and regularization techniques are also used in their cases.

Starting with the characteristics, the simplest approach is to forecast each anomaly using the time series of its own characteristic spread. Alternatively, one can forecast each anomaly individually using the whole collection of characteristics and can further employ a dimension reduction technique, such as PLS, or a regularization technique, such as lasso, for the RHS. Finally, one can create PC portfolios on the LHS without using any dimension reduction technique (but potentially using lasso) for the RHS. Moreover, the benchmark models can also be modified in various ways. For instance, factor momentum, issuer-repurchaser spread, and investor sentiment can be applied to the PC portfolios. For book-to-market ratio and issuer-repurchaser spread, the dataset can be expanded by using all the ratios and spreads simultaneously to predict each PC portfolio or individual anomaly. Finally, lasso can be applied to these richer information sets to account for overfitting. A detailed description of the models discussed in this section can be found in panel C of Table A.3.

Table 1.5 reports the total OOS R^2 and average cross-sectional correlation for the modified forecasting methods. Panel A shows that, in line with Haddad, Kozak, and Santosh (2020), predicting each anomaly by its own spread is not particularly successful, as it provides negative OOS R^2 and relatively low average cross-sectional correlation. When we incorporate the full set of characteristics for each anomaly, the OOS R^2 metric worsens possibly due to overfitting, but the average cross-sectional correlation improves

Table 1.5: Total OOS R^2 and average cross-sectional correlation of factor portfolio return forecasts based on various forecasting methods

	Total OOS R^2	Average cross-sectional correlation (%)
A. Modified characteristic-based models		
Anom-Own characteristic	-0.78	3.77
Anom-1 PLS	-2.51	2.94
Anom-All characteristics	-93.60	4.48
Anom-All characteristics lasso	-2.09	5.82
5 PCs-All characteristics	-53.93	6.19
5 PCs-All characteristics lasso	-1.14	6.26
B. Modified benchmark models		
5 PCs-1mMOM	-57.99	5.77
5 PCs-12mMOM	-4.72	3.73
5 PCs-5 BMs	-0.75	2.61
5 PCs-5 BMs lasso	-0.44	1.97
Anom-Own BM	0.20	4.29
Anom-All IR spreads	-31.73	7.11
Anom-All IR spreads lasso	-2.91	3.06
5 PCs-IR spread	-3.04	0.91
5 PCs-All IR spreads	-1.19	2.31
5 PCs-Sentiment	0.39	3.55
Sentiment-lasso	0.27	4.70

in two out of the three cases (the exception being the model that uses PLS on the RHS). When we further condense the information content of the anomalies into five PC portfolios, the average cross-sectional correlation increases even more. Nevertheless, the OOS R^2 remains negative and the cross-sectional correlation is still at the levels of 6%, clearly lower than the 8%–9% provided by our main models using dimension reduction also on the RHS (panel B of Table 1.3). Overall, the results of panel A corroborate the importance of using the full set of characteristics for factor timing purposes, while they further highlight the additional benefits that arise when incorporating dimension reduction and regularization techniques on both sides of the forecasting exercise.

Panel B shows the results for the alternative specifications of the benchmark models. Applying the momentum signal on the PCs instead of individual anomalies has an inconsistent effect on forecasting performance as it improves the OOS R^2 but reduces the cross-sectional correlation. Using the book-to-market ratios of all PC portfolios to predict each single one of them individually is not particularly fruitful, with both performance measures worsening compared to the main PCA-BM model. Predicting each anomaly by its own book-to-market ratio delivers a positive OOS R^2 and higher cross-sectional correlation, but it still falls behind the baseline BM model. In terms of the issuer-repurchaser spread, using the spreads of all portfolios instead of the spread of each individual portfolio, as in the baseline IR spread model, reduces the OOS R^2 but improves considerably the cross-sectional correlation.[23] Nevertheless, the use of lasso does not make any substantial contribution in this case, with the OOS R^2 still being on the negative side and the cross-sectional correlation decreasing. Finally, using the investor sentiment index to predict the PC portfolios or in combination with lasso to predict individual anomalies has little effect compared to the baseline Sentiment model. Overall, we find mixed results for the modified benchmark models, with the dimension reduction and regularization additions improving the models only occasionally. In any case, even the best modified models exhibit clearly worse performance than our main characteristic-based models.

Table 1.6 presents the portfolio evaluation results for the modified models. Starting with panel A, the investment performance of the modified characteristic-based models is broadly in line with the cross-sectional correlations from Table 1.5. In particular, all strategies exhibit good investment performance, while the average returns and Sharpe ratios tend to improve when, on top of using the whole set of portfolio characteristics, we further incorporate PCA and/or lasso in the forecasting exercise. Still, even the best-

[23]Using the collection of issuer-repurchaser spreads can be justified by considering the different factor portfolios as substitutes for the same investor. In that sense, time-varying characteristic mispricing can propagate from one factor to the rest due to changes in demand and supply for the different factors, making the issuer-repurchaser spreads of other factors important.

Table 1.6: Portfolio evaluation measures for long-short (LSS), time-series (TSFM), and cross-sectional (CSFM) strategies under the alternative specifications for the sample period January 1990 to December 2019. Average return: average monthly return; Standard deviation: monthly standard deviation; Sharpe ratio: monthly Sharpe ratio; t-statistic: t-statistic on H_0: Average return = 0; Hit-rate: percentage of the total number of occasions that the strategy resulted in positive returns; Max drawdown: maximum cumulative loss.

	Average return (%)			Standard deviation (%)			Sharpe ratio			t-statistic			Hit-rate (%)			Max drawdown (%)		
	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM
A. Modified characteristic-based models																		
Anom-Own characteristic	0.46	0.34	0.32	**4.04**	**2.48**	**2.53**	0.11	0.14	0.13	2.17	2.57	2.40	**60.72**	**60.72**	**61.00**	33.44	21.02	21.76
Anom-1 PLS	0.46	0.30	0.30	4.33	2.96	2.97	0.11	0.10	0.10	2.03	1.95	1.93	56.82	57.66	58.50	29.95	25.56	25.41
Anom-All characteristics	0.74	0.52	0.51	4.71	2.76	2.77	0.16	0.19	0.18	2.98	3.54	3.49	57.38	58.77	59.05	**11.65**	**8.74**	**9.16**
Anom-All characteristics lasso	1.02	**0.81**	**0.81**	5.54	4.05	4.06	**0.18**	**0.20**	**0.20**	**3.49**	**3.78**	**3.79**	58.77	59.05	57.38	19.40	16.62	16.50
5 PCs-All characteristics	1.02	0.67	0.65	7.98	4.65	4.72	0.13	0.14	0.14	2.42	2.72	2.60	54.04	54.60	54.60	33.36	25.60	26.61
5 PCs-All characteristics lasso	1.09	0.70	0.72	7.59	4.64	4.63	0.14	0.15	0.16	2.71	2.88	2.95	54.60	55.99	55.99	19.23	16.62	16.32
B. Modified benchmark models																		
5 PCs-1mMOM	1.13	0.64	0.65	9.24	5.52	5.54	0.12	0.12	0.12	2.32	2.20	2.22	57.38	55.71	54.60	25.74	19.89	19.82
5 PCs-12mMOM	0.76	0.58	0.58	9.40	5.90	5.85	0.08	0.10	0.10	1.53	1.88	1.89	52.65	53.76	53.20	36.46	28.19	27.61
5 PCs-5 BMs	0.28	0.22	0.25	7.37	4.86	4.81	0.04	0.05	0.05	0.73	0.86	0.98	55.99	52.37	55.15	47.72	37.71	36.62
5 PCs-5 BMs lasso	0.35	0.26	0.27	7.85	5.01	5.00	0.04	0.05	0.05	0.85	0.99	1.03	53.76	53.48	53.76	46.79	37.78	37.21
Anom-Own BM	0.70	0.48	0.48	4.37	2.60	2.69	0.16	0.18	0.18	3.05	3.50	3.37	57.10	56.27	55.99	28.28	21.53	21.61
Anom-All IR spreads	**1.24**	0.80	0.81	7.70	4.47	4.49	0.16	0.18	0.18	3.06	3.38	3.43	60.17	58.22	59.33	21.60	14.53	14.61
Anom-All IR spreads lasso	0.67	0.50	0.51	6.88	4.85	4.87	0.10	0.10	0.10	1.84	1.95	1.98	56.27	55.99	56.27	24.21	20.57	20.46
5 PCs-IR spread	0.07	-0.14	-0.09	6.28	4.52	4.22	0.01	-0.03	-0.02	0.22	-0.59	-0.39	48.19	47.08	47.08	77.69	99.97	87.40
5 PCs-All IR spreads	0.51	0.29	0.28	7.99	4.89	4.82	0.06	0.06	0.06	1.20	1.12	1.11	52.65	52.65	52.65	57.43	33.41	32.55
5 PCs-Sentiment	0.56	0.35	0.33	6.86	3.89	3.97	0.08	0.09	0.08	1.54	1.70	1.59	54.87	53.76	52.37	57.10	27.64	29.55
Sentiment-lasso	0.67	0.46	0.46	5.37	3.22	3.25	0.12	0.14	0.14	2.35	2.72	2.69	55.15	56.27	56.55	17.90	19.22	19.13

performing modified models fall behind the main ones presented in panel B of Table 1.4. Therefore, it is confirmed again that using multiple portfolio characteristics is indispensable for forming a successful factor timing strategy, but the dimension reduction and regularization techniques provide additional benefits. Turning to panel B, the PCA-based momentum models, the model that uses all issuer-repurchaser spreads, the model that uses each portfolio's book-to-market ratio, and the model that employs lasso together with market sentiment appear to be the strongest ones. This is unsurprising given that these models also deliver the highest cross-sectional correlations in Table 1.5. Importantly, even these alternative benchmark models exhibit weaker investment performance than our preferred characteristic-based models in panel B of Table 1.4. Overall, the alternative information sets have lower factor timing ability compared to the set of portfolio characteristics even if they are enhanced by employing dimension reduction or lasso.

1.6 Conclusion

We investigate the predictability of factor portfolios from their own portfolio characteristics, going over and above existing methods for predicting factor portfolio returns and examining the possibility that factor portfolios are predictable by characteristics other than their own. Our approach offers a natural continuation to the stock return predictability problem, and our findings shed light on the evolution of the underlying return drivers over time. Under our empirical framework, a large collection of stock characteristics is used to initially construct the LS portfolios and subsequently predict their performance. A key aspect of our methodology is the reduction of the dimensions of the predictability problem, which we achieve by independently shrinking the number of predictors and forecasting targets. Our approach provides a new framework for dealing with panel data, allowing each source of variation to be approximated by models of different complexity. By using a flexible model specification that combines lasso with dimension reduction techniques, we allow the number of predictors to vary across PC portfolios and over time. We find this approach to be especially fruitful, as it considerably improves results over a static single latent factor model.

In terms of factor portfolio predictability, we observe significant benefits from timing factor portfolio returns using observed characteristics. These benefits go over and above existing methods documented in the literature, highlighting the importance of considering the information in the characteristics in a collective way. Specifically, the dominant PC portfolios are highly predictable by the information contained in their characteristics, and this predictability can be easily extended to individual anomalies. In that sense, dimension reduction techniques not only accommodate the computational tractability of the estimation problem, but also improve forecasting and investment performance by enabling us to focus on the sources of variation that are most predictable. The performance of our factor timing strategies is superior to that of any individual anomaly and persistent over the later years of the sample period, demonstrating the benefits of timing over static factor investing. Hence, in the context of anomaly return prediction it is important to (i) account for the information contained in multiple characteristics, (ii) focus on the main sources of variation in factor portfolio returns since those are the most predictable, and (iii) apply some kind of time-varying regularization on the set of predictors to account for the time variability in characteristic informativeness. Overall, our findings have important implications for the use of machine learning methods in asset pricing applications and help justify the importance of observable characteristics in explaining the dynamics of factor portfolios.

Chapter 2

Power Sorting

This paper is a joint work with Anastasios Kagkadis, Harald Lohre, Ingmar Nolte, and Sandra Nolte. It was awarded first place in the Chicago Quantitative Alliance 30th Annual Academic Competition. We would like to thank Vikas Agarwal, Pedro Barroso, David Blitz, Amit Goyal, Foort Hamelink, Matthias Hanauer, Clint Howard, Illia Kovalenko, Gianluca De Nard, Peter Pope, Konark Saxena, Paul Schneider, Laurens Swinkels, Ruslan Tuneshev, Steven Young, as well as the participants at the CQA 2023 Fall Conference, the Inquire UK Autumn 2023 Webinar, the 2024 FMA Europe Conference, the 2024 EFMA Conference, the Robeco research seminar and the seminars at Lancaster University and University of St Andrews for their valuable comments and suggestions. The views and opinions expressed in this paper are solely those of the authors and may not necessarily align with those of Robeco or Ultramarin.

2.1 Introduction

When associating observable firm characteristics with equity returns, the classic approach in empirical asset pricing is to construct characteristic-sorted portfolios, commonly referred to as factor portfolios or simply, factors. Such factor portfolios have been widely used by academics to identify asset pricing anomalies and construct asset pricing models. They are also used by investment practitioners who look for systematic exposure to rewarded factors, provided these are investable. The conventional procedure for constructing factor portfolios involves ranking the stock universe by a characteristic, creating quantile portfolios, and analyzing the long-short portfolio of the two extreme quantile portfolios. Despite its popularity and intuitive appeal, this conventional approach has its limitations. First, it lacks an objective criterion for choosing the number of quantile portfolios, with that number usually remaining invariant for the characteristic at hand. Usually, ten portfolios are considered, even though there is little motivation behind such choice apart from ensuring a decent characteristic spread. Second, the conventional method cannot address variation in characteristics within quantile portfolios, as these are usually either equal or value-weighted.[1] In that respect, it also cannot account for potential non-linearities in the characteristic-return relations.[2] Third, the conventional weighting scheme is symmetric, implicitly assuming equal pricing ability of the characteristic on the long and the short side, while disregarding stocks in the middle. In that sense, it fails to explore the existence of monotonic patterns between returns and economic variables that are implied by finance theories (Patton and Timmermann 2010).

To illustrate the limitations of conventional factor portfolio construction, Figure 2.1 plots the return for selected factors across the full spectrum of the respective characteristic, using 100 quantile bins. The main insight from Figure 2.1 is that characteristics can relate to average returns in non-trivial ways and decile sorting provides a simplistic

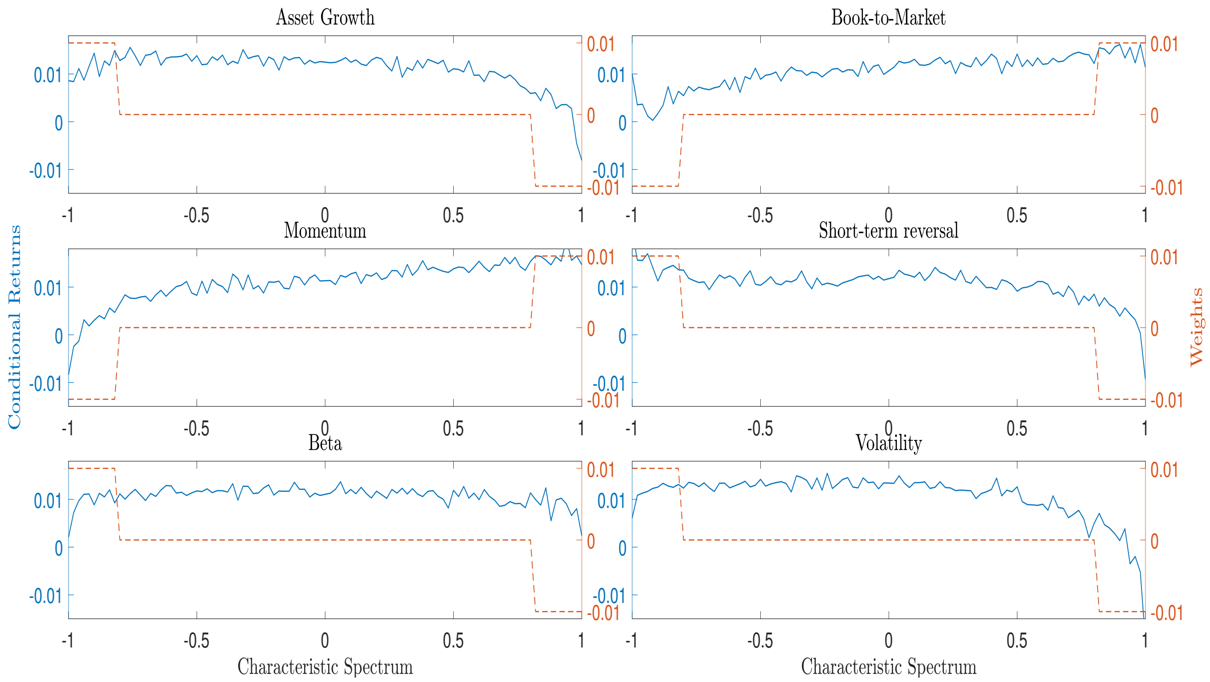
[1]Such fixed weighting schemes introduce other factor exposures and can thus have a confounding effect on factor return inference (Swade et al. 2023).

[2]Note that numerous leading finance theories predict that expected returns are highly non-linear functions of the underlying characteristics or state variables (e.g., Campbell and Shiller, 1988; Campbell and Cochrane, 1999; Bansal and Yaron, 2004; He and Krishnamurthy, 2013).

perspective to a more complex set of patterns between the two. One such pattern is the inverted "smile" shape, where both stocks with very high and very low characteristics underperform (e.g., beta), resulting in insignificant return differences across the two legs. In this case, investing in the corner decile portfolios delivers an insignificant long-short spread, implicitly declaring the characteristic as an unimportant return predictor. Another common pattern is the inverted "smirk" shape, where stocks on the short side underperform, but stocks on the long side display no significant outperformance (e.g., asset growth and volatility). In that case, the factor portfolio spread is primarily driven by the short side. Lastly, average returns in the tails of the characteristics can drift in opposite directions. That is, they might drift in the intended direction as implied by the overall relationship, leading to an amplified effect in the extreme quantile portfolios (e.g., short-term reversal), but they might also turn in the opposite direction, reducing the return spread (e.g., book-to-market). Regardless of the underlying pattern, Figure 2.1 suggests that the extreme quantiles shall be treated differently and stocks in between the two extremes are also worth of consideration in the construction of factors. Nonetheless, any potential weighting scheme should be economically sound and theoretically motivated to ensure that the resulting portfolios retain the underlying factor structure and avoid overfitting and data mining concerns. Put differently, allowing the weight vector to vary freely without imposing any structure or economic prior could lead to overfitted factor portfolios that are based on return patterns alone and therefore unable to capture the underlying economic driver.

In this paper, we develop a data-oriented power sorting procedure to directly model factor portfolio weights as a function of firm characteristics. This procedure extends to conventional long-short factor portfolios by allocating some weight to all assets, while still allowing to tilt more towards stocks with extreme characteristics if deemed appropriate. Unlike conventional sorting, power sorting does not require manual selection of quantile breakpoints and seeks to exploit variation in characteristics across the full characteristic spectrum rather than overlaying fixed-weighting schemes that could mask the factor's nature. Importantly, power sorting can capture asymmetries and non-linearities from

Figure 2.1: Conditional monthly returns and conventional equal-weighted decile-sorted factor portfolio weights for six characteristics. Characteristics are standardized in the $[-1, 1]$ range. The conditional returns are estimated by ranking stocks based on their underlying characteristic in the previous period and splitting them into 100 equal-weighted quantile portfolios. The blue line shows the average monthly return across portfolio groups. The dashed orange line shows the weight function for the factor portfolio that invests in the corner decile portfolios based on the underlying characteristic. The sample includes all common shares on NYSE, AMEX, and NASDAQ exchanges and covers the period from January 1980 to December 2021.



characteristics to returns, allowing for tailored treatment on the long and the short side and a deeper understanding of the behavior of the two complementary drivers of factor premia.

The power sorting procedure is based on the assumption of monotonicity between characteristic and return and is flexible enough to extract optimal performance from the underlying characteristic, while still creating portfolios that are theoretically guided and economically meaningful. Specifically, the cross-sectional weight vector for any given factor is obtained by expressing portfolio weights as a power series of the underlying characteristic rank. This formulation presents a tightly parameterized problem that accommodates a variety of monotonic weighting schemes based on just two parameters, one for the long and one for the short leg of the factor portfolio. The two parameters determine the concentration of the power factor portfolio on stocks with extreme characteristic scores and can be estimated based on any arbitrary objective function, such as maximizing the Sharpe ratio of the factor portfolio. Importantly, unlike conventional optimization routines, our approach achieves this without explicitly requiring the use of a variance-covariance matrix, thus avoiding associated estimation challenges. Higher parameter values lead to portfolios that are more concentrated in stocks with characteristic extremes, while lower values lead to a more diversified factor exposure by spreading portfolio weights more evenly across stocks. Additionally, differences in the two parameters allow for capturing asymmetries and non-linearities in the weight function; for instance, one may construct factors that take an aggressive stance on one leg and a more passive stance on the other.

Our primary objective is to establish a framework for factor portfolio construction that accommodates characteristic-specific treatment of the various characteristics with clear interpretability of the underlying model parameters. Regarding the characteristics, several studies have consistently emphasized the asymmetric impact of the long and the short side on factor portfolio performance (Ang et al., 2006; Stambaugh, Yu, and Yuan, 2012; Blitz, Baltussen, and van Vliet, 2020; Leung et al., 2021). Furthermore, many characteristics documented in the literature have been found to yield insignificant performance, when the portfolio construction method is taken as given (Hou, Xue, and Zhang, 2015; Green, Hand, and Zhang, 2017; Hou, Xue, and Zhang, 2020). It is worth noting, however, that slight modifications in factor construction can lead to significantly different conclusions about factor significance (Jensen, Kelly, and Pedersen, 2023; Soebhag, van Vliet, and Verwijmeren, 2023). Consequently, the conventional portfolio construction technique cannot efficiently extract the underlying risk premium for the vast majority of characteristics, and can yield misleading conclusions about their economic and statistical significance.

For example, Hou, Xue, and Zhang (2020) find that factor portfolios based on market friction proxies exhibit insignificant performance under a conventional long-short quantile approach. This finding resonates with an inverted smirk pattern where the relationship between market friction proxies and next-period returns is highly asymmetric and factor performance is driven by the short side that contains the most illiquid stocks. Similar conclusions can be drawn for inverted smile patterns observed in many accounting variables. Power sorting proves particularly effective in modeling such patterns and producing weighting schemes that can exploit variation in the short leg while maintaining a more diversified stance in the long leg. Furthermore, power sorting can enhance the performance of already successful monotonic factors by leveraging the variation on both sides.

With regard to the model parameters, several degrees of freedom are involved in the construction of factor portfolios. For example, increasing the number of quantile portfolios — from terciles to quintiles, deciles, or beyond — produces portfolios that are concentrated in stocks with extreme characteristics. Additionally, researchers can affect the weighting scheme through other construction choices, such as value- or equal-weighting stocks in the selected quantile. Both schemes can introduce unwanted factor exposures that may unduly confound the targeted characteristic. Equal-weighting amplifies the effect of small stocks, while value-weighting results in portfolios that are heavily skewed towards very large stocks, thereby masking factor behavior via size effects. To address this issue, researchers can use NYSE breakpoints and winsorize market capitalizations. Such choices are often framed as data pre-processing steps and their implicit effect on portfolio performance is usually overlooked.

Ultimately, our approach constitutes a sample-efficient solution for deriving portfolio weights in an objectively optimal way, thereby alleviating p-hacking concerns related to subjective portfolio construction choices. Additionally, by explicitly parameterizing weight concentration in the tails, our framework enables clear interpretability of the underlying model parameters, thus bridging the gap between ad-hoc portfolio sorts and black-box machine learning methods. Finally, one distinctive feature of our method is the introduction of a hyper-parameter that controls for the impact of size in the construction of factor portfolios. This parameter is determined in a robust and transparent manner based on specific criteria, such as the maximum weight assigned to any individual stock. As a result, power factor portfolios are sufficiently diversified, easily interpretable, and practically relevant, establishing a data-driven and discretion-free framework for constructing factor portfolios.

Our results demonstrate that power sorting outperforms conventional sorting in terms of various portfolio metrics, using a set of 85 well-established characteristics in an out-of-sample period from March 1980 to December 2021. For many factor portfolios, the outperformance arises from adopting a more aggressive stance on the short leg and a more conservative stance on the long leg. This in turn implies that the characteristic signal is strong for underperforming stocks but it tends to be weaker for outperforming stocks. In the case of equal-weighted portfolios, the average factor portfolio Sharpe ratio increases by 57%, while for value-weighted portfolios, it doubles. Importantly, the observed performance enhancement is highly statistically significant and cannot be attributed to increased turnover or tail risk considerations. Furthermore, these economic gains also carry important asset pricing implications, as they lead to the resurrection of many documented factors that were previously deemed insignificant. Specifically, the factor significance rate rises from 40% to 75.3% for equal-weighted portfolios, and from 18% to 55.3% for value-weighted portfolios, even when employing a strict t-stat threshold of three (Harvey, Liu, and Zhu 2016). Finally, we show that power sorting outperforms other sophisticated portfolio construction techniques that account for non-linearities such as the parametric portfolio policy of Brandt, Santa-Clara, and Valkanov (2009) and the efficient sorting approach of Ledoit, Wolf, and Zhao (2019).

Despite the univariate nature of the power sorting methodology, we provide evidence of its effectiveness in a multi-factor context too. Adopting an asset pricing perspective, we demonstrate that the incorporation of power factors into existing asset pricing models consistently improves their pricing ability as evidenced by a significantly higher model squared Sharpe ratio (Barillas et al. 2020). From an investment perspective, we highlight the empirical relevance of power sorting in combining individual factors into multi-factor portfolios. Our approach explicitly considers the asymmetric pricing abilities of different characteristics when combining signals, resulting in multi-factor portfolios with improved investment performance. The performance enhancement achieved through power sorting is substantial compared to single-characteristic strategies or equal weighted multi-factor approaches, particularly after accounting for size effects.

The remainder of the paper is structured as follows: Section 2.2 introduces the power sorting procedure and relates it to the conventional procedure and prior literature on characteristic-based portfolio construction. Section 2.3 explores power factor portfolio construction for a large set of characteristics and examines their out-of-sample performance on an individual and aggregate factor level. Section 2.4 compares power sorting to alternative methods proposed in the literature for factor construction and performs a variety of robustness tests to corroborate the validity of power-sorted factor portfolios. Section 2.5 concludes.

2.2 Power Sorting Methodology

The goal is to construct portfolios with exposure to some characteristic but in a way that one can best exploit its relationship to future returns. We begin by explaining the conventional portfolio construction technique, followed by the power sorting approach. The conventional sorting procedure is to rank the cross-section of stock returns according to a characteristic. The cross-sectional vector of characteristics, observable at the beginning of month t is denoted by $x_t := (x_{t,1}, \dots x_{t,N_t})'$, where N_t is the number of stocks available at time t. The vector of stock returns at the beginning of month $t+1$ is denoted by $r_{t+1} := (r_{t+1}, \dots, r_{t+1,N_t})'$. Finally, let $\{(1), (2), \dots, (N_t)\}$ be a permutation of $\{1, 2, \dots, N_t\}$ that results in ordered factor scores (from smallest to largest):[3]

$$x_{t,(1)} \leq x_{t,(2)} \leq \dots \leq x_{t,(N_t)}. \tag{2.1}$$

[3]This is assuming a positive relation between the characteristic and returns; otherwise the characteristic is inverted.

The essence of factor investing is the estimation of a weight vector $w_t := \left(w_{t,(1)}, \ldots, w_{t,(N_t)}\right)'$ for r_{t+1} based on x_t. A typical long-short portfolio satisfies,

$$\sum_{w_{t,(n)}<0} \left|w_{t,(n)}\right| = \sum_{w_{t,(n)}>0} \left|w_{t,(n)}\right| = 1, \tag{2.2}$$

so that we have a unit dollar short leg and a unit dollar long leg.

2.2.1 Conventional long-short quantile factor portfolios

The standard procedure is based on partitioning the characteristic space into equal-sized quantile bins. Let B be the number of quantile groups considered; for example, B equals 5 for quintiles, or B equals 10 for deciles. Let d be the largest integer that is smaller than or equal to N/B (the number of stocks in each quantile group). The k-th quantile of x_t, $q_{t,k} = 1, \ldots, B-1$, is equal to $x_{t,(k \cdot d)}$. The weighting vector under a standard long-short decile portfolio scheme (i.e., B equals 10) is denoted as w_t^{LS}. The standard long-short portfolio allocates equal weight to stocks belonging to the two corner portfolios and disregards the rest,

$$\begin{aligned} w_{t,(1)}^{\mathrm{LS}} &= \ldots = w_{t,(d)}^{\mathrm{LS}} = -1/d, \\ w_{t,(d+1)}^{\mathrm{LS}} &= \ldots = w_{t,(\mathrm{N_t}-d)}^{\mathrm{LS}} = 0, \\ w_{t,(\mathrm{N_t}-d+1)}^{\mathrm{LS}} &= \ldots = w_{t,(\mathrm{N_t})}^{\mathrm{LS}} = 1/d. \end{aligned} \tag{2.3}$$

The resulting portfolio return at time $t+1$ is denoted by $r_{t+1}^{\mathrm{LS}} = r_{t+1}' w_t^{\mathrm{LS}}$. Value-weighted versions of those portfolios can be constructed by weighting stocks within each group based on their market capitalization:

$$w_{t,(n)}^{\mathrm{LS}} = \begin{cases} -mcap_{t,(n)} / \sum\limits_{i=1}^{d} mcap_{t,(i)}, & \text{for } n \leq d \\ 0, & \text{for } d < n \leq N_t - d \\ mcap_{t,(n)} / \sum\limits_{i=N_{t-d+1}}^{N_t} mcap_{t,(i)} & \text{for } n > N_t - d, \end{cases} \tag{2.4}$$

where $mcap_{t,(n)}$ is the market capitalization of stock n at time t. Specifically, we construct capped value-weighted versions of the factor portfolios, following Jensen, Kelly, and Pedersen (2023b), such that we assign weights to stocks based on their market capitalization winsorized at the 80^{th} percentile of the NYSE.

Regardless of the underlying weighting scheme, this approach has some important implications. First, the process is dependent on the specific choice of quantile breakpoints (e.g., terciles, quantiles or deciles). In essence, B is a hyper-parameter that dictates the concentration of the long-short factor portfolio. Although deciles are commonly used, it is ultimately a choice parameter that can significantly affect inferences about the significance of factor premia (Soebhag, van Vliet, and Verwijmeren 2023). High values for B can potentially improve return performance but lead to undiversified portfolios that are less practical as they over-concentrate in a small number of stocks. Second, employing equal- and value-weighted weighting schemes in portfolio construction introduces ad-hoc variation that may obscure the underlying return signal. Third, the approach places equal emphasis on the long and the short leg, while it disregards the information about mid-rank stocks. This attribute renders the method inadequate to effectively capture non-linearities and asymmetries in the underlying characteristic-return relationship.

2.2.2 Power sorting

Pure power-sorted portfolios

We propose power sorting that uses the underlying characteristic rank to determine factor portfolio weights, but the weighting vector is directly derived for the whole cross-section of stocks without requiring any grouping. In each period, we cross-sectionally rank all stock characteristics and map them onto the interval $[-1, 1]$ centered around the median rank. As shown below, the standardized characteristic rank vector $\tilde{s}_{t,(n)}$ is obtained as:

$$s_{t,(n)} = \left\lfloor \text{rank}(x_{t,(n)}) - \frac{N_t + 1}{2} \right\rceil \tag{2.5}$$

$$\tilde{s}_{t,(n)} = \begin{cases} -\frac{s_{t,(n)}}{s_{t,(1)}} & \text{for } s_{t,(n)} < 0 \\ 0 & \text{for } s_{t,(n)} = 0 \\ \frac{s_{t,(n)}}{s_{t,(N_t)}} & \text{for } s_{t,(n)} > 0, \end{cases} \tag{2.6}$$

where $\text{rank}(\cdot)$ is the rank function and $\lfloor \cdot \rceil$ is the function rounding to the nearest integer.

One advantage of using characteristic ranks rather than raw scores to derive the weighting vector is that the former is unaffected by the distribution of the characteristics. Next, we translate scores into weights by normalizing them based on the respective sums of scores as outlined below in equation (2.7). Stocks with below median characteristic rank are assigned negative weights and stocks with above median characteristic rank are assigned positive weights. Specifically, positive scores are divided by the sum of all positive scores and negative scores are divided by the sum of all negative scores, ensuring a unit dollar investment for the long and the short side. Non-linearities and asymmetries in the weight function are incorporated by introducing two parameters, one for the long (p) and one for the short side (q). These two parameters are exponents that are applied to positive and negative characteristic ranks before transforming them into portfolio weights. For exposition purposes, we assume p and q to be constant across time, while in our empirical investigation, we demonstrate how time-variability in p and q can impact the shape of the factor portfolio weight function over time. Hence, we express positive and negative scores as two independent power series and their scaling factors as their power sums. The resulting weighting vector for the power sorting portfolio is given by:

$$w_t^{PS}(\tilde{s}_{t,(n)}; p, q) = w_{t,(n)}^{PS} = \begin{cases} -\frac{|\tilde{s}_{t,(n)}|^q}{\sum\limits_{\tilde{s}_{t,(n)}<0} |\tilde{s}_{t,(n)}|^q} & \text{for } \tilde{s}_{t,(n)} < 0 \\ 0 & \text{for } \tilde{s}_{t,(n)} = 0 \\ \frac{\tilde{s}_{t,(n)}^p}{\sum\limits_{\tilde{s}_{t,(n)}>0} \tilde{s}_{t,(n)}^p} & \text{for } \tilde{s}_{t,(n)} > 0. \end{cases} \tag{2.7}$$

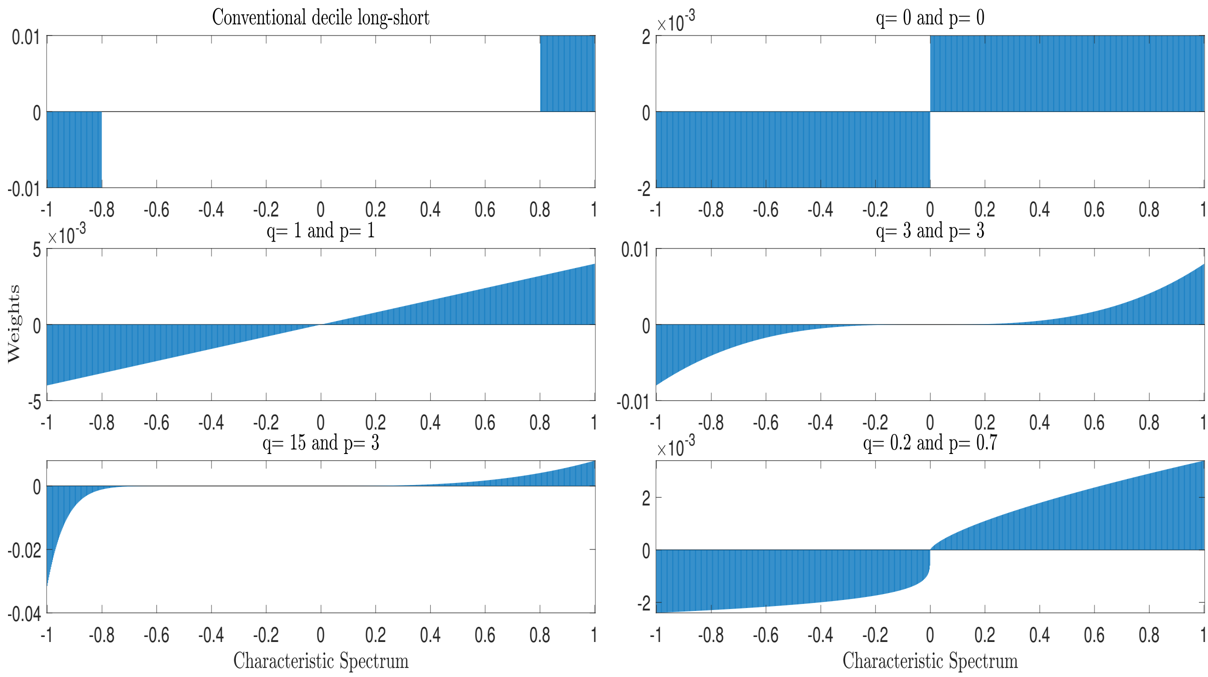
The two hyper-parameters p and q govern the concentration of the power sorting portfolio weights. Higher parameter values lead to portfolios that are more concentrated in the extreme ranks, as stocks with characteristic rank closer to the median shrink towards zero faster due to the function's exponential nature. Given that $\lim_{p \wedge q \to \infty} w_t^{PS} = (-1, 0, \ldots, 0, 1)$, all capital is allocated to the two stocks with the most extreme characteristics. This formulation provides a natural way to capture weight concentration in the tails and offers ample flexibility in modeling the underlying weighting function.

To illustrate, Figure 2.2 presents the resulting weighting function for various combinations of p and q, alongside the conventional long-short weighting scheme. When $p \wedge q = 0$, the function evenly distributes weights between stocks above and below the median, resembling a conventional long-short portfolio with two groups and reflecting a passive factor approach. When $p \wedge q = 1$, the function aligns with a linear rank weighting scheme, where absolute weights increase linearly for stocks with characteristic ranks further from the median.

For values between 0 and 1 in $p \wedge q$, absolute weights increase at a marginally decreasing rate around the median, while for $p \wedge q > 1$ the weights increase at a marginally increasing rate, over-weighting the extreme ranks. Notice that when $p \wedge q < 0$, portfolio weights concentrate towards the centre, resulting in lack of monotonicity. Therefore $p \wedge q = 0$ constitutes a natural lower bound for the parameter space in the context of factor portfolio construction.

In general, high values for p and q correspond to an aggressive factor stance, where stocks with the most extreme characteristic ranks are expected to contribute most to the factor premium. Importantly, differences in p and q introduce asymmetries in the weighting scheme, allowing one leg to be more concentrated/less diversified than the other.

Figure 2.2: The opportunity set of power sorting. The top-left chart displays a conventional equal-weighted decile-sorted long-short weighting scheme for a characteristic positively related to returns. The remaining charts display stock weights for different values of p and q under the power sorting scheme.



Value-weighted power factor portfolios

We next discuss the construction of value-weighted versions of the power factor portfolios. The rationale to use market capitalization for weighting stocks within each factor portfolio is to reflect the relative size of companies. Put differently, market capitalization weighting is likely to give sector and industry exposures similar to the overall market. However, under a conventional approach, such value-weighted factor portfolios tend to overweight mega-cap stocks, resulting in less diversified portfolios that cannot robustly capture the underlying factor premium. To this end, scholars have put forward ways to control the effect of market capitalization on portfolio composition. A well-known example is the Fama and French (1993) construction methodology, which gives half the weight to small stocks and the other half to big stocks. Jensen, Kelly, and Pedersen (2023b) winsorize market capitalizations at the NYSE 80^{th} percentile before calculating factor portfolio weights, which avoids excessive weights on mega-cap stocks while still emphasizing large stocks. These approaches, although masked as data pre-processing steps, allow for different degrees of freedom in the estimation of value-weighted portfolios and can have a significant impact on portfolio outcomes (Soebhag, van Vliet, and Verwijmeren 2023). Furthermore, the effect of such modifications on portfolio composition is usually unassessed. For this reason, we directly incorporate and parameterize the effect of size on the estimation of portfolio weights by computing the capitalization-adjusted versions of the power portfolio as:

$$w_{t,(n)}^{PS,cap} = \begin{cases} -\frac{|\tilde{s}_{t,(n)}|^{q} \cdot mcap_{t,(n)}^{h}}{\sum\limits_{\tilde{s}_{t,(n)}<0} |\tilde{s}_{t,(n)}|^{q} \cdot mcap_{t,(n)}^{h}} & \text{for } \tilde{s}_{t,(n)} < 0 \\ 0 & \text{for } \tilde{s}_{t,n} = 0 \\ \frac{\tilde{s}_{t,(n)}^{p} \cdot mcap_{t,(n)}^{h}}{\sum\limits_{\tilde{s}_{t,(n)}>0} \tilde{s}_{t,(n)}^{p} \cdot mcap_{t,(n)}^{h}} & \text{for } \tilde{s}_{t,(n)} > 0, \end{cases} \tag{2.8}$$

where parameter $h \in [0, 1]$ controls the concentration in mega-cap stocks. A value of $h = 1$ corresponds to the uncapped value-weighted versions, while a value of 0 corresponds to the pure characteristic-weighted power portfolios. Values between 0 and 1 regulate the effect of size in the estimation of weights and are crucial to avoiding corner allocations in mega-cap stocks. The reason is that the vector of ordered market capitalizations behaves as a power series with high exponential growth, as it is dominated by a handful of stocks of exponentially larger size than their peers. Hence, presuming no shrinkage on market caps ($h = 1$) means that the weighting vector of the value-weighted versions is the product of two power curves. This can lead to extreme concentrations in mega-cap stocks in cases where mega-cap stocks have extreme characteristic ranks and factor concentration ($p \vee q$) is high. As such, it is key to moderate the market capitalization component to avoid extreme mega-cap stock allocations.

Our approach allows for an efficient formulation of the weighting function and does not require any data pre-processing/manipulation step to avoid overconcentration, such as winsorization, NYSE breakpoints, grouping, or similar. The value of h can be either calibrated based on the desired maximum portfolio weight or prespecified as a constant value. To mitigate data mining concerns, we opt for a constant value of $h = 0.5$ for all power portfolios, which is equivalent to taking the square root of market capitalization. As a benchmark, we estimate the value-weighted long-short portfolios using winsorized market caps at the 80^{th} NYSE breakpoints, as in Jensen, Kelly, and Pedersen (2023b).

Managing weight concentration over time

The presented power sorting framework can naturally be extended to deal with extreme corner case allocations and account for time variation in p and q (i.e., consider p_t and q_t). First, the maximum weight for each leg in each period is always allocated to the stock with the maximum absolute standardized characteristic rank. To illustrate, the maximum weight of the long leg portfolio is given by:

$$w_{max,t}^{PS} = w_{N_t}^{PS} = \frac{1}{F(\tilde{s}_{t,(N_t)}, p_t)}, \tag{2.9}$$

where $F(\tilde{s}_{t,(N_t)}, p_t) = \sum_{\tilde{s}_{t,(n)}>0} \tilde{s}_{t,(n)}^{p_t}$ is a power sum that can be efficiently computed using Faulhaber's formula (Knuth 1993). Hence, $w_{N_t}^{PS}$ is decreasing in the number of available assets and increasing in the value of the power p_t, meaning that the effect of p_t on $w_{N_t}^{PS}$ is conditional on the number of available assets and therefore characteristic- and time-dependent. Equation (2.9) highlights that finding a single optimal combination of p and q over time would lead to inconsistent weight distributions due to the variation in the size of the equity cross-section. In other words, imposing a single optimal power exponent would yield a variety of maximum weights over time. This inconsistency poses challenges when comparing different characteristics and determining the parameter values that maximize in-sample performance.[4] As the number of available assets is known at time t, the maximum weight can be constrained by setting an upper threshold to the maximum power. The threshold is calculated by solving:

$$\frac{1}{F(\tilde{s}_{t,(N_t)}, p_t)} - w^{ceil} = 0, \tag{2.10}$$

[4] It should be noted that the optimal power exponents might not be directly compared across characteristics since they correspond to different availability of characteristics data.

where w^{ceil} is the targeted maximum weight. Depending on characteristic availability, the value of p_t solving equation (2.10) will vary. We opt for a maximum portfolio weight of 2% when estimating the upper threshold for p_t and q_t (labeled p_t^{max} and q_t^{max}, $t = 1, \ldots, T$) to ensure healthy portfolio diversification. In Section 2.4.2, we check for robustness of results with respect to the choice of maximum portfolio weight.

To ensure consistency in maximum weight concentration over time while still optimizing with respect to a single set of parameters, we define the concentration ratios for the two sides as $\tilde{p_t} = p_t/p_t^{max} \in [0, 1]$ and $\tilde{q_t} = q_t/q_t^{max} \in [0, 1]$, respectively. The concentration ratios correspond to the densities associated with the truncated distributions of p_t and q_t and are essentially standardized metrics that allow for a clear and intuitive interpretation. Specifically, a concentration ratio equal to one for either leg indicates that power factor portfolio performance is optimized when the weights are concentrated in the tail(s), with the maximum weight being no larger than w^{ceil}. Conversely, a value of zero implies that factor performance is optimized when a diversified stance is taken, equal-weighting stocks away from the median. This standardization allows for uniformity in the behavior of the weight distribution across time and characteristics and, hence, in the calibration of a single set of parameters. These optimal densities can then be mapped out to every period based on p_t^{max} and q_t^{max}, allowing for the optimal p_t and q_t to be time-varying.

2.2.3 Power sorting and related literature

Our paper contributes to the literature on characteristic-based portfolio choice for asset pricing and investment applications. Conventional characteristic sorting has been a workhorse in empirical asset pricing due to its simplicity and intuitive interpretation. Early empirical contributors of portfolio characteristic sorts include Basu (1977) and Banz (1981), while the approach was popularised by Fama and French (1992) and Jegadeesh and Titman (1993). Despite its popularity, prior literature has identified some practical and theoretical limitations of the conventional portfolio construction. Jacobs and Levy (1993) raise various practical concerns that underline long-short strategies, while Patton and Timmermann (2010) highlight the inability of long-short strategies to test for monotonicity between characteristics and returns. In contrast to standard portfolio sorts, power sorting imposes monotonicity in the characteristic-return relationship and leverages variation across the characteristic spectrum to derive factor portfolio weights. Therefore, it promises to align more closely with economic theory.

Alternative approaches to portfolio construction, like Frazzini and Pedersen (2014) or Koijen et al. (2018), utilize rank portfolios. Rank portfolios assign progressively higher weights to stocks as they deviate further from the characteristic median in a linear manner. This method proves effective for characteristics that demonstrate a monotonic relationship with returns, particularly when the effect is more pronounced for extreme values.[5] Notably, power sorting encompasses rank portfolios, allowing for a linear weighting function based on the characteristic rank when the underlying relationship is linear. However, power sorting goes beyond linear, enabling the incorporation of non-linear and asymmetric weights, thus offering greater flexibility in portfolio construction. A comprehensive comparison between rank portfolios and power portfolios is provided in Section 2.4.1.

Our study is also related to a strand of the literature that models portfolio weights as a function of underlying firm characteristics and employs optimization-based approaches for portfolio construction. Notable examples include Brandt (1999), Aït-Sahalia and Brandt (2001), Brandt and Santa-Clara (2006), and Brandt, Santa-Clara, and Valkanov (2009).[6] The main difference between power sorting and the parametric portfolio policy of Brandt, Santa-Clara, and Valkanov (2009) is that the former utilizes characteristic ranks and obtains non-linear weights by the usage of powers, while the latter derives non-

[5] Novy-Marx and Velikov (2022) propose a rank- and capitalization-weighted scheme to account for market capitalizations in rank portfolios.

[6] Building upon the parametric portfolio policy framework of Brandt, Santa-Clara, and Valkanov (2009), Ammann, Coqueret, and Schade (2016) introduce leverage constraints, DeMiguel et al. (2020) incorporate transaction costs, and Simon, Weibels, and Zimmermann (2023) integrate feed-forward neural networks to capture non-linear and interaction characteristic effects. Hjalmarsson and Manchev (2012) demonstrate that, within a mean-variance framework, the use of firm characteristics enables the reduction of the asset space to a set of characteristic-based portfolios. In an alternative approach, McGee and Olmo (2022) use non-parametric kernel methods to estimate the conditional moments of stock returns based on stock characteristics in a cross-sectional setting. These estimated moments are then used within a mean-variance objective function for portfolio construction.

linear weights by scaling linearly the higher moments of the characteristic distribution. In that respect, the success of Brandt, Santa-Clara, and Valkanov's (2009) method relies on whether the characteristic-return relation in the tails follows the shape of the underlying characteristic's distribution, while in ours this is not the case. Section 2.4.1 compares the performance of the two methods and discusses thoroughly their conceptual differences.

More broadly, the limitations of conventional portfolio sorting have prompted researchers to seek improvements in the construction of characteristic-based portfolios. Cattaneo et al. (2020) approach portfolio sorting as a non-parametric estimator, where the number of portfolios serves as a hyper-parameter, while Zhang, Wu, and Chen (2022) propose a listwise learn-to-rank loss function that sequentially selects pairs of stocks for the long and the short leg. Closer to our study, Ledoit, Wolf, and Zhao (2019) utilize the DCC-NL estimator developed by Engle, Ledoit, and Wolf (2019) to estimate "efficient" factor portfolios. These portfolios aim to minimize variance while maintaining the overall factor exposure of traditional long-short portfolios. To assess the impact of parameter shrinkage resulting from power sorting, we compare our method with the efficient sorting methodology of Ledoit, Wolf, and Zhao (2019) in Section 2.4.1.

Finally, our work relates to recent studies that construct characteristic-driven portfolios but with different objectives compared to ours. For instance, Fama and French (2020) utilize the cross-sectional regression approach of Fama and MacBeth (1973) to construct factors based on standardized characteristics.[7] Their findings reveal that these cross-sectional factors are more effective at explaining average returns compared to the original Fama-French-type factors. In a different context, Kim, Korajczyk, and Neuhierl (2021) introduce portfolios that aim to exploit mispricing information in the characteristics while hedging out systematic variation related to those characteristics. Similarly, Daniel et al. (2020) construct "characteristic efficient portfolios" by hedging away variation associated with unpriced risk using a hedge portfolio.

[7]The regression slopes correspond to the returns of the zero investment factor portfolios with unit exposure to their characteristic and zero exposure to all the other characteristics.

2.3 Optimal Power Sorting Portfolios

2.3.1 Characteristics and power thresholds

We replicate a large set of 85 characteristics that have been considered by Green, Hand, and Zhang (2017). The characteristics are calculated using data from the Center of Research on Securities (CRSP), Compustat, and the Institutional Brokers' Estimate System (I/B/E/S), covering the period from January 1980 to December 2021. The stock universe includes common stocks listed on NYSE, AMEX, and NASDAQ that have a record of month-end market capitalization on CRSP and a non-missing and non-negative common value of equity on Compustat. Additional information about the characteristics, including origination and characteristic description, can be found in Section A.1 of the Appendix.

For every month, stock returns for month *t+1* are matched against their respective characteristics in month t. For accounting data, we allow at least six months to pass from the firms' fiscal year-end before they become available and at least four months to pass for quarterly data. To mitigate the effect of microcaps, we remove stocks with a market capitalization below the 10^{th} percentile at the portfolio formation period.

For constructing conventional benchmark factor portfolios, we first group stocks into equal-weighted deciles based on their characteristic scores in the previous month and then go long and short in the two extreme deciles, depending on the prevailing characteristic-return relationship. For value-weighted results, we use a "capped value-weighting" scheme following Jensen, Kelly, and Pedersen (2023b). Factors are categorized into six groups based on economic rationale, (Hou, Xue, and Zhang 2015), namely: Momentum, Value, Investment, Profitability, Intangibles, and Trading Frictions.

A subtle but important detail is setting the maximum threshold for the hyper-parameters p_t and q_t. This threshold depends on the targeted maximum portfolio weight and the number of available assets. To ensure that the power portfolios are sufficiently diversified, we set the maximum weight to 2% and solve equation (2.9) for the values of p_t^{max} and q_t^{max}. Figure A.11 in the Appendix displays the time-variation in the maximum powers for the long and the short leg of the different characteristics. Evidently, those thresholds vary significantly across characteristics and time, further stressing the importance of using a standardized measure for optimization purposes and for conducting comparisons.

2.3.2 Estimation procedure

The construction of the power sorting portfolio for a given characteristic x_t requires an estimate of the powers $p_t \in [0, p_t^{max}]$ and $q_t \in [0, q_t^{max}]$ for each period t. To this end, we solve for the respective concentration ratios $\tilde{p}_t$ and $\tilde{q}_t$ that maximize the power sorting factor portfolio Sharpe ratio in the in-sample period and estimate the powers for the most recent cross-section by multiplying the ratios with the maximum power thresholds for the most recent period p_t^{max} and q_t^{max}. To mitigate data-mining concerns regarding the selection of the estimation window, we adopt an expanding window approach and consider the longest out-of-sample period possible. In particular, the out-of-sample period covers March 1980 to December 2021, while different estimation windows are explored in Section 2.4.2. To illustrate, assuming a Sharpe ratio maximization objective and based on an underlying rank-standardized characteristic $\tilde{s}_t := \tilde{s}_{t,1}, \ldots \tilde{s}_{t,N_t}$, the estimation problem at each investment date can be formulated as follows:

$$\{\hat{\tilde{p}}_t, \hat{\tilde{q}}_t\} = \underset{\tilde{p}_t \wedge \tilde{q}_t \in [0,1]}{\arg\max} \frac{\overline{r_t}^{PS}}{\sqrt{var(r_t^{PS})}}, \tag{2.11}$$

$$\overline{r_t}^{PS} = \frac{1}{t-1} \sum_{i=1}^{t-1} \sum_{j=1}^{N_i} r_{i+1,j} \cdot w_i^{PS}(\tilde{s}_{i,(j)}; p_i^{max} \cdot \tilde{p}_t, q_i^{max} \cdot \tilde{q}_t), \tag{2.12}$$

$$var(r_t^{PS}) = \frac{1}{t-2} \sum_{i=1}^{t-1} \left(\sum_{j=1}^{N_i} \left(r_{i+1,j} \cdot w_i^{PS}(\tilde{s}_{i,(j)}; p_i^{max} \cdot \tilde{p}_t, q_i^{max} \cdot \tilde{q}_t) \right) - \overline{r}^{PS} \right)^2, \tag{2.13}$$

which is a constrained optimization problem that can be solved numerically.[8] Notice that under this formulation there is no need to estimate the variance-covariance matrix (VCV) for individual stocks. Each combination of $\hat{\tilde{p}}_t$ and $\hat{\tilde{q}}_t$ practically corresponds to a set of cross-sectional weight vectors, and hence to a power portfolio return time-series for which the first and the second moments are computed directly. The out-of-sample power sorting portfolio return at time $t+1$ is then estimated as:

$$r_{t+1}^{PS} = r_{t+1}' \times w_t^{PS}(\tilde{s}_t; p_t^{max} \cdot \hat{\tilde{p}}_t, q_t^{max} \cdot \hat{\tilde{q}}_t). \tag{2.14}$$

Value-weighted results for the power versions of each factor are estimated as in equation (2.8), using the same maximum powers as in the pure characteristic weighted versions and a value of $h = 0.5$.

2.3.3 Power-sorted portfolios and concentration ratios

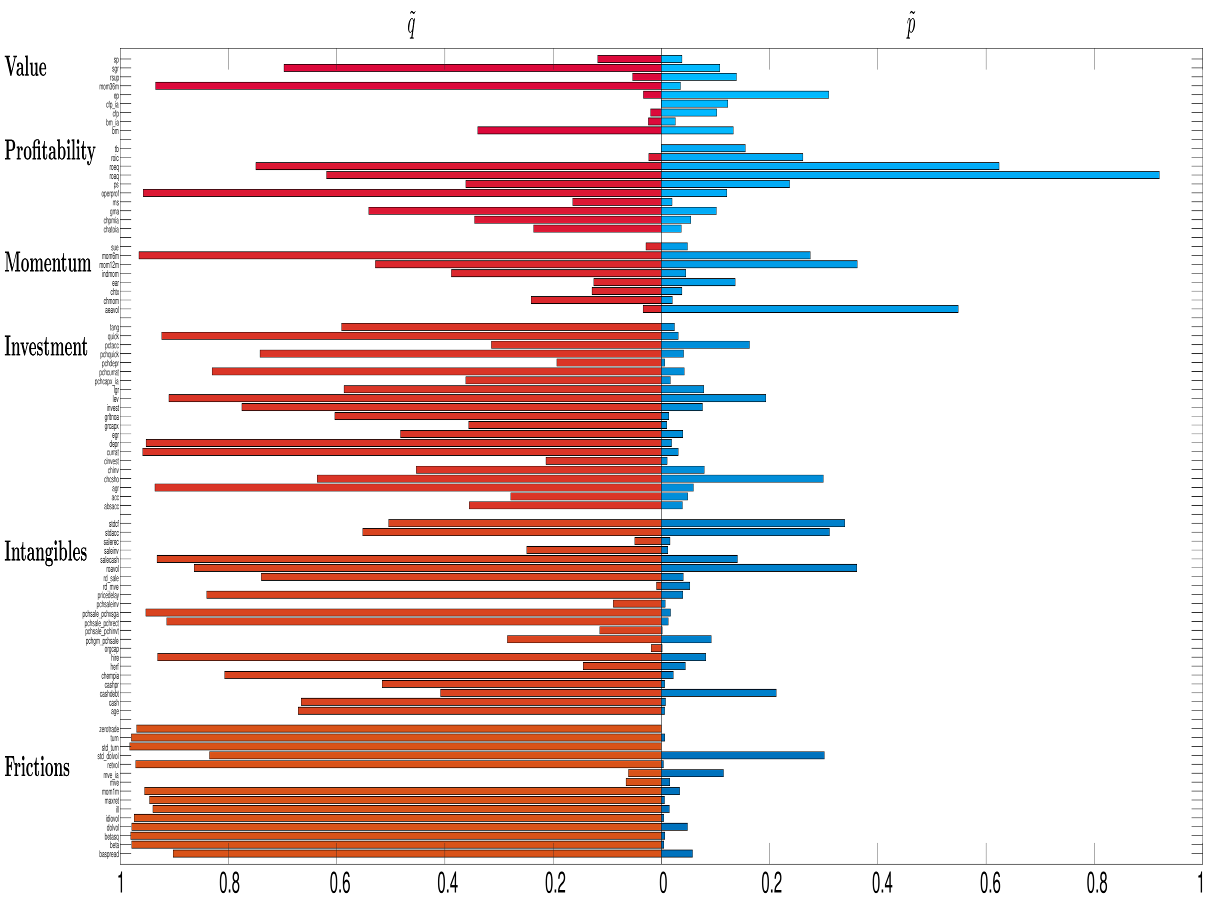
First, we examine the underlying form of the weight function for various characteristics. Power-sorted portfolios assign a portfolio weight to every stock that is uniquely determined by the $\tilde{p}_t$ and $\tilde{q}_t$ parameters. The use of an expanding estimation window implies that the out-of-sample parameters should gradually stabilize and converge to the optimal in-sample parameters as the sample expands. To foster intuition with respect to the underlying weight function, we present the average concentration ratios of each factor in Figure 2.3. Blue-shaded bars represent the average concentration ratio for the long side and red-shaded bars represent the average concentration ratio for the short side.

Figure 2.3 clearly illustrates that the optimal degree of concentration is highly asymmetric and skewed towards the short side for the majority of characteristics. That is the factor portfolio Sharpe ratio is maximized by adopting an aggressive stance on the short side and a more conservative stance on the long side. This finding indicates that stocks at the lower end of the conditional return distribution tend to perform very poorly, while

[8] In Section A.3.2 of the Appendix we report results under a return-spread maximization objective. These results are similar to the ones presented in the main paper.

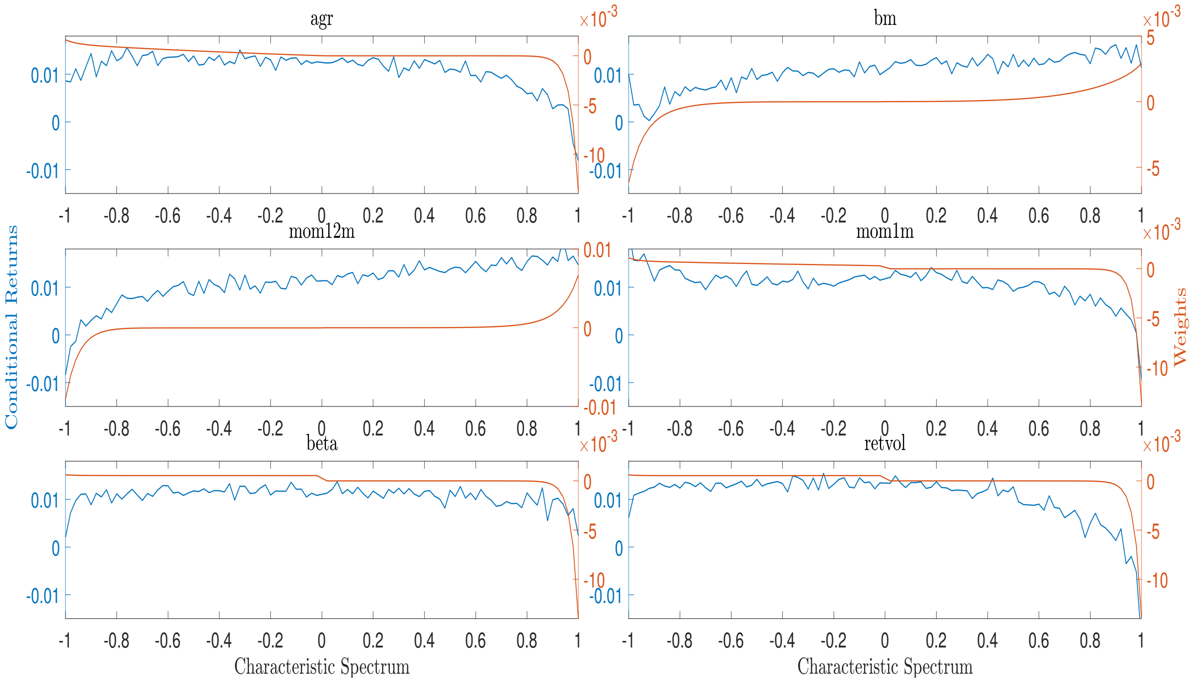
stocks' outperformance at the extreme upper end is less extreme. Nonetheless, lower values for $\tilde{p}_t$ compared to $\tilde{q}_t$ do not imply that the long leg is an insignificant contributor to factor portfolio performance. In fact, as we show later in the analysis, the long leg of the power sorting portfolios delivers positive and significant returns. Nevertheless, this asymmetry suggests that conditional returns in the long tail either remain relatively flat or that stocks in the corner of the long leg tend to underperform. Consequently, a lower concentration ratio in the long leg helps to avoid overinvesting in corner stocks that are likely to underperform when compared to their peers.

Figure 2.3: Average estimated concentration ratios $\hat{\tilde{p}}$ and $\hat{\tilde{q}}$ of each factor. Blue-shaded bars show the average concentration ratio for the long side and red-shaded bars show the average concentration ratio for the short side. Factors are sorted into six groups based on their economic rationale. The sample period is from March 1980 to December 2021.



To visualize the implications of the estimated concentration ratios in terms of portfolio weights, Figure 2.4 depicts the average weight function resulting from the selected values for $\hat{\tilde{p}}_t$ and $\hat{\tilde{q}}_t$ for the six factors presented in Figure 2.1. When compared to the conventional weighting scheme, power sorting is able to capture the underlying return patterns more accurately. In particular, our approach proves highly effective in dealing with characteristics that exhibit inverted "smirk" or inverted "smile" shapes, such as asset growth, return volatility, and beta. In such cases, power portfolios combine a high value for $\tilde{q}_t$ and a low value for $\tilde{p}_t$, thus producing inverted smirk weight schemes that increase exposure on stocks in the extremes of the short side, and reduce exposures on stocks in the extremes of the long side. Furthermore, for some characteristics, like momentum, the effect is amplified in the extremes and the algorithm opts for high values of $\tilde{p}_t$ and $\tilde{q}_t$, resonating with an aggressive stance in both the long and the short side to exploit variation in both tails.

Figure 2.4: Conditional monthly returns and power factor portfolio weight function for six characteristic-based factor portfolios. Characteristics are standardized in the $[-1, 1]$ range. Conditional returns are estimated by ranking stocks based on their underlying characteristic in the previous period and splitting them into 100 equal-weighted quantile portfolios. The blue line shows the average monthly return across portfolio groups. The orange line shows the average weight function for the factor portfolio as implied by the selected values for $\tilde{p}_t$ and $\tilde{q}_t$ across periods. The sample includes all common shares on NYSE, AMEX, and NASDAQ exchanges and covers the period from March 1980 to December 2021.



2.3.4 Power sorting versus conventional decile sorting

Table 2.1 compares the portfolio performance of power sorting against that of conventional decile sorting, presenting average portfolio statistics across factors for both equal-weighted and value-weighted cases. The results demonstrate the superiority of power sorting over the conventional approach across all portfolio metrics. Specifically, power sorted portfolios consistently exhibit significantly higher average returns and Sharpe ratios, with an average t-statistic above three for both value and equally weighted factors. Notably, for the value-weighted case, power sorting leads to an average Sharpe ratio that is twice as high as that achieved through the conventional approach (0.52 versus 0.26). Importantly, these results are not likely to be driven by increased trading costs or tail risk, as power factors exhibit on average a lower turnover and maximum drawdown compared to the conventional long-short portfolios. Finally, the resulting portfolios are more diversified, encompassing a higher number of effective names on both the long and short sides. The asymmetrically higher number of effective names for the long leg of the average power portfolio corroborates the patterns depicted in Figure 2.3, reflecting higher values of $\tilde{q}$ and a more aggressive stance for the short side.

Table 2.1: Average performance measures across factors for power and conventional long-short portfolios. Return: Average monthly return, Stand. dev.: Monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: Return=0, Hit rate: Percentage frequency of positive returns, Turnover: Average monthly turnover (bounded by 200%), # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample period is from March 1980 to December 2021.

	Equal-weighted		Value-weighted	
	Power	*Conventional*	*Power*	*Conventional*
Return (%)	0.77	0.51	0.62	0.32
Standard deviation (%)	4.04	4.21	4.17	4.39
Sharpe ratio	0.72	0.46	0.52	0.26
t-stat	4.64	2.96	3.40	1.71
Maximum drawdown (%)	−45.79	−55.35	−48.96	−59.22
Hit rate (%)	61.88	57.39	58.27	53.44
Turnover (%)	37.57	39.39	33.09	35.44
# of effective names long	1315.41	369.24	460.01	107.00
# of effective names short	535.60	370.33	229.79	98.42

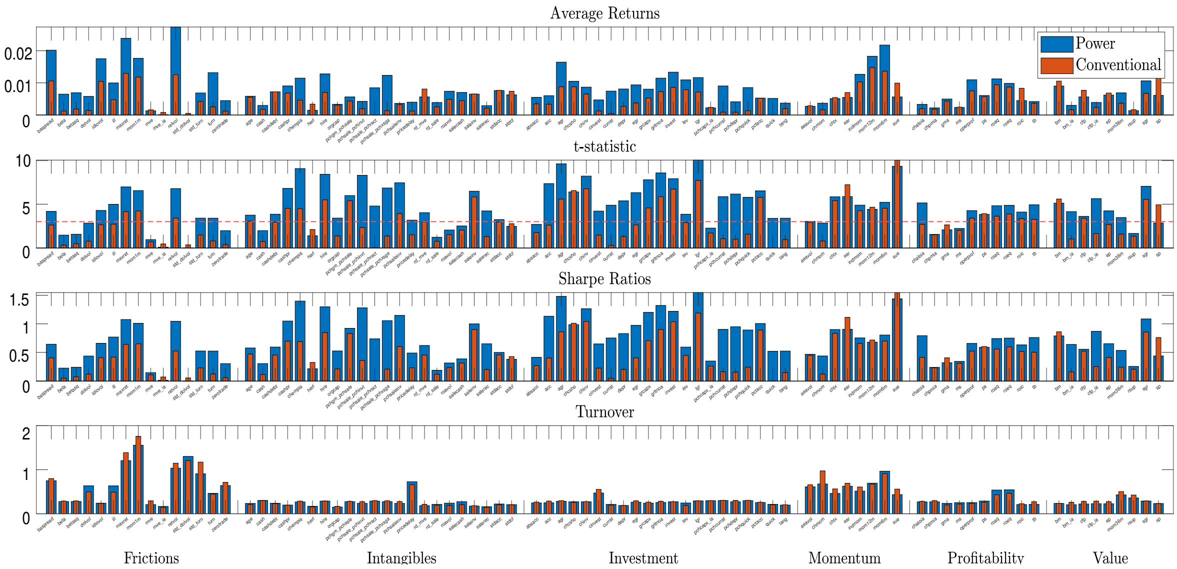
Figure 2.5 presents selected out-of-sample portfolio evaluation measures for individual factors, depicting the power-sorted versions in blue and the conventional long-short decile versions in orange. Panel A compares the pure power-sorted portfolios with equal-weighted decile benchmarks, while Panel B compares the capitalization-adjusted power versions with the capped value-weighted versions of the conventional long-short approach. Power sorting consistently leads to substantial gains in average returns and Sharpe ratios across the majority of factors, and these improvements cannot be attributed to increased turnover. Specifically, 75.3% of power versions have higher average returns, and 86% have higher Sharpe ratio. For value-weighted results, the respective numbers are 85.9% for returns and 96.5% for Sharpe ratios.

In addition, power sorting achieves a significantly higher significance rate for the average returns of factor portfolios, as indicated by a t-statistic above three (75.3% versus 40% for equal-weighted portfolios and 55.3% versus 18% for value-weighted portfolios). Hence, several factors deemed insignificant under the conventional weighting scheme become significant when the power weighting scheme is applied, even when using the stricter t-value threshold of three, as advocated by Harvey, Liu, and Zhu (2016). These results raise questions about the ability of decile sorting to efficiently extract the underlying signal from many characteristics, potentially leading to false rejections of factors.

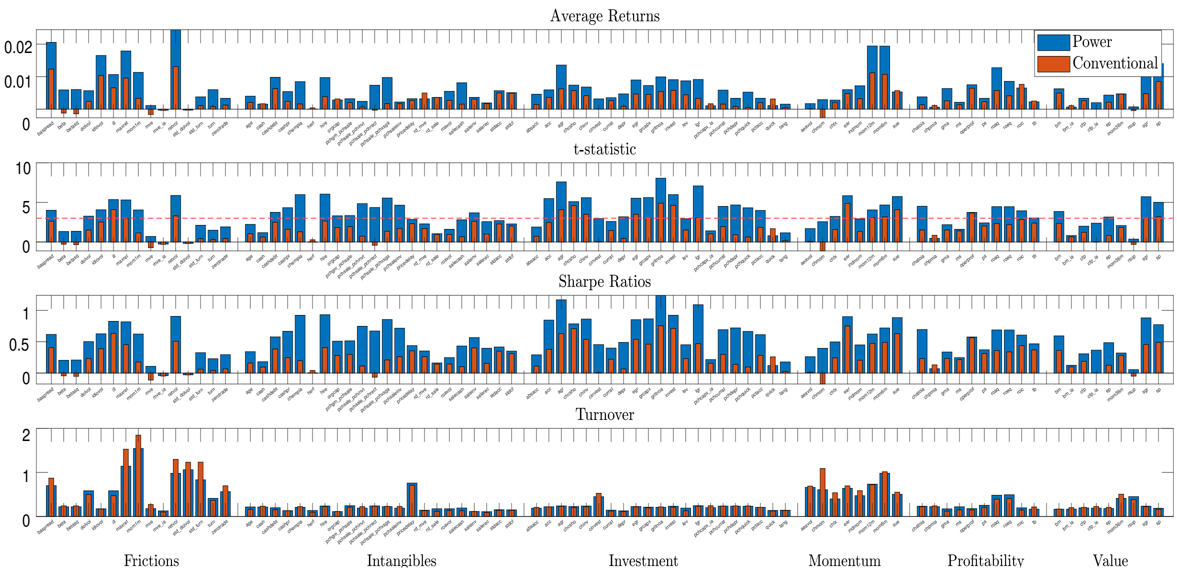
It is worth noting that results are fairly consistent across the pure power-sorted and value-adjusted versions of the power portfolios, while the benchmark results considerably deteriorate under value-weighting. In fact, in some cases adjusting for market capitalization leads to value-weighted portfolios with negative average returns under the conventional method. Conversely, returns remain positive under a power sorting approach. Hence, the incorporation and parameterization of the size effect into the factor weighting procedure preserves the underlying factor behavior and controls for any confounding effects that might otherwise arise in naive value-weighted decile sorts.

Figure 2.5: Portfolio evaluation measures for conventional long-sort and power versions. Panel A displays equal-weighted results and Panel B displays value-weighted results. (1) Average monthly return, (2) t-statistics on average monthly return, (3) Annualized Sharpe ratio, (4) Monthly turnover. The optimal powers are selected using an in-sample expanding window starting from January 1980 to December 2021. Factors are sorted into six groups based on their economic rationale.

A. Equal-weighted portfolios



B. Value-weighted portfolios



Finally, investment gains from power sorting are evenly distributed across factors, yet notable significance is observed for factors associated with Frictions, Investment, and Intangibles. These factor themes are recognized for their asymmetric nature (i.e., Ang et al. (2006), Cooper, Gulen, and Schill (2008)), confirming the effectiveness of power sorting in capitalizing on the specific patterns inherent to these factors.

2.3.5 Dissecting long and short factor portfolio legs

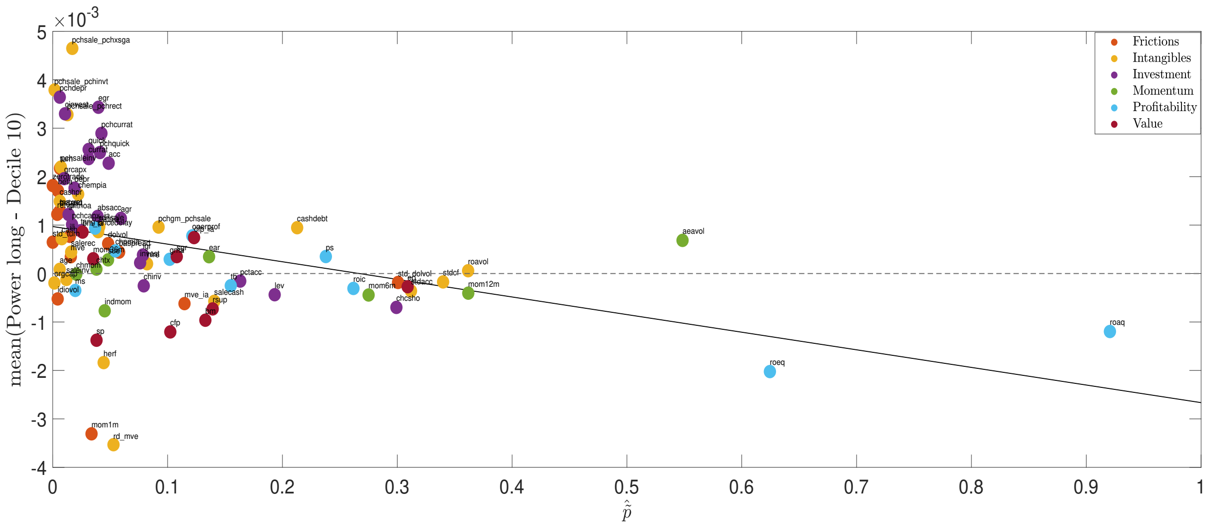
So far, we have established the presence of asymmetric and non-linear return effects of characteristics and demonstrated the effectiveness of power sorting in identifying and capitalizing on these patterns. To gain a deeper understanding, we further explore the implications of $\tilde{p}_t$ and $\tilde{q}_t$ on the long and short sides of each factor.

Figure 2.6 provides a visual depiction of the improvement in return performance for the long (Panel A) and short (Panel B) sides of the different factors, based on their corresponding values of $\tilde{p}_t$ and $\tilde{q}_t$. Each subfigure includes the line of best fit and the zero line. Data points positioned below the zero line indicate instances where the power leg underperforms its conventional decile leg. Panel A shows the increase in average return for the power long leg compared to decile ten, using the average optimal value of $\tilde{p}_t$. The relationship between the two is negative, suggesting that a more diversified approach that spreads weights across the long leg is preferable over concentrating solely on stocks in the extreme decile. Characteristics associated with Investment, Intangibles, and Market Friction proxies exhibit the most significant benefits from a low value of $\tilde{p}_t$. As already discussed, these variables demonstrate inverted smile and smirk patterns, indicating that conditional returns in the long tail either decrease or remain relatively flat. Consequently, a low $\tilde{p}_t$ reduces portfolio exposure to underperforming corner stocks in the long tail, enhancing diversification benefits and investment performance in the long leg.

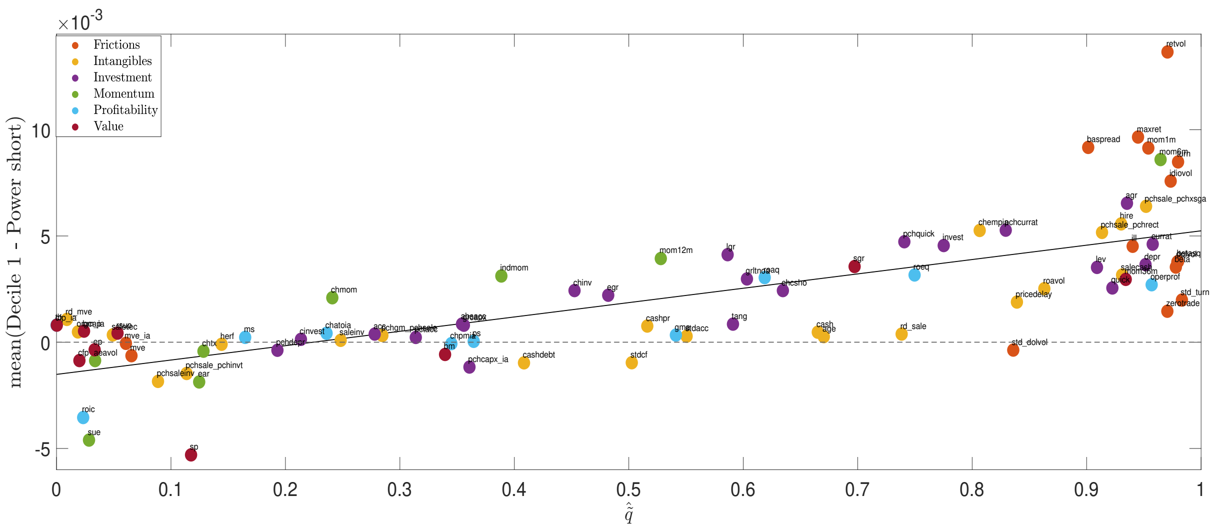
Panel B of Figure 2.6 shows the decrease in average return for the power short leg over decile one for the average optimal value of $\tilde{q}_t$. Higher values of $\tilde{q}_t$ are associated with lower average returns for the short side, increasing the long-short spread. Power sorting remains particularly effective for factors related to market frictions, such as maxret or retvol, as it capitalizes on the sharp decline in conditional returns on the short side. Still, intangibles and investment factors are now more spread across the line, implying heterogeneity in terms of optional concentration levels for the short leg. Intuitively, this result suggests that the different investment and intangible proxies agree on the long side but disagree on the short side.

Figure 2.6: Concentration ratios and excess returns. Panel A shows the increase in average returns for the long leg given the estimated value of $\tilde{p}_t$. Panel B shows the decrease in average return for the short leg given the estimated value of $\tilde{q}_t$. Each subfigure includes the line of best fit and the zero line. The sample period is from March 1980 to January 2021.

A. Long leg



B. Short leg



Note that stocks with the most extreme characteristics have a significant impact on the determination of optimal powers, as their weights increase exponentially. When using the median rank as the cutoff, the algorithm favors a low power in the long leg to avoid overinvestment in underperforming corner stocks relative to their peers. Due to the monotonic nature of the function, it has limited capacity to capture inflection points in the tails of the conditional return distribution. As a result, it adopts a passive approach by equally weighting stocks above the median to compensate. This pattern emerges also for the short leg of many value characteristics.

Overall, both sides of power portfolios outperform their conventional counterparts. The outperformance on the long side is driven by adopting a more balanced approach, spreading weights across a broader range of stocks. On the other hand, the outperformance on the short side is attributed to adopting a more aggressive stance, capitalizing on the specific patterns identified through power sorting.

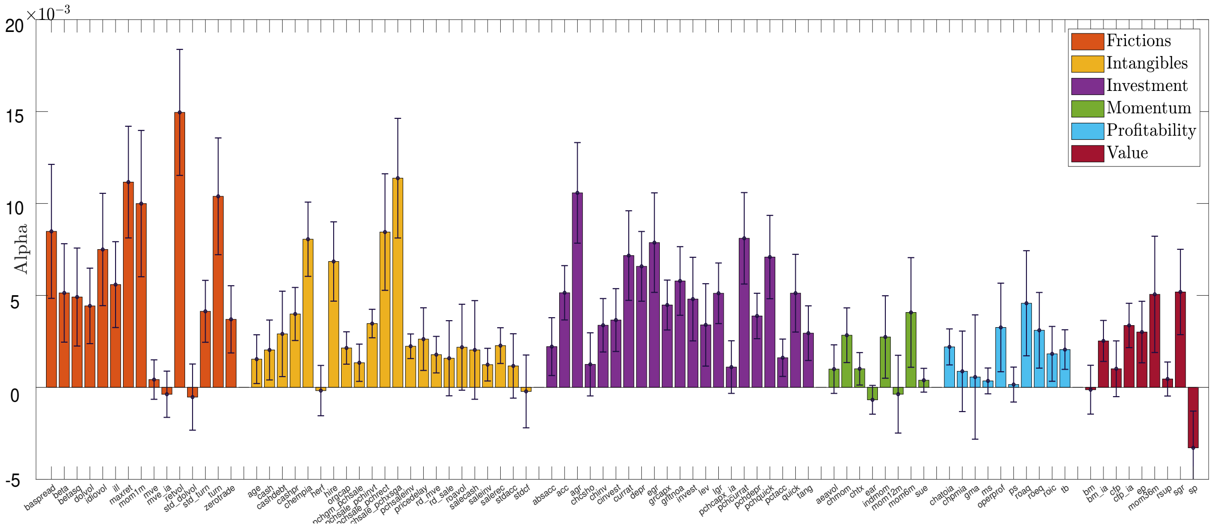
2.3.6 Spanning regressions

In Figure 2.7, we further report the monthly alphas from regressing power portfolio returns on those of their conventional long-short counterparts. The subfigures correspond to equal-weighted and value-weighted results and all estimates include their 95% confidence bounds. An interval that excludes (includes) zero indicates statistical significance (insignificance) at the 5% level.

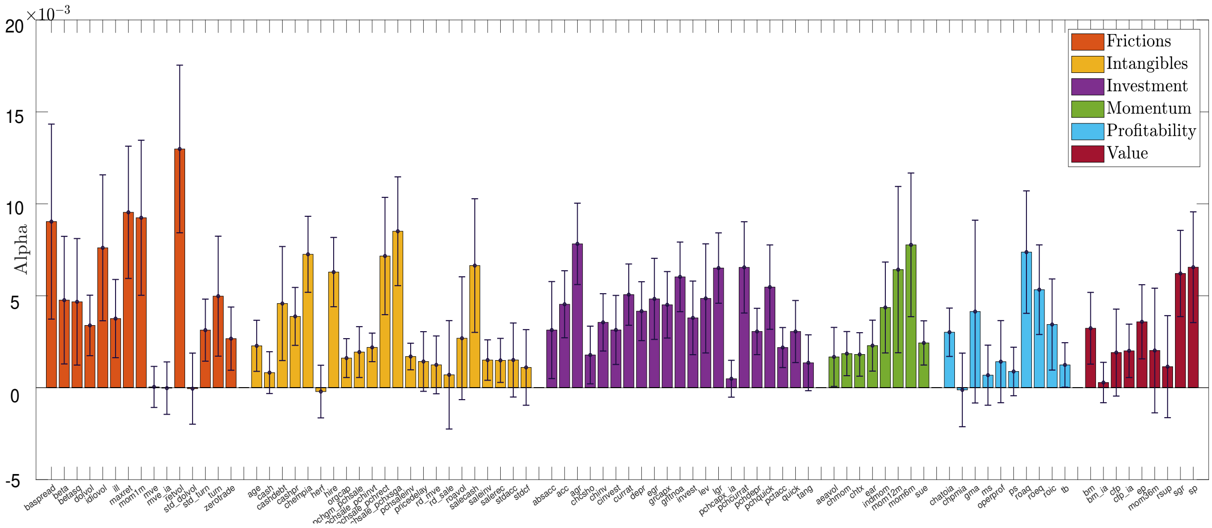
Out of the 85 alphas, 77 are positive and 62 are statistically significant for the equal-weighted case. In the value-weighted case, the corresponding numbers are 81 and 63, respectively. In fact, even factors that did not exhibit any significant improvement under power sorting in the equal-weighted case now exhibit alphas that are positive and significant. This improvement is particularly noticeable for momentum factors, which are infamous for experiencing a sharp decline in profitability with market capitalization (Hong, Lim, and Stein 2000). Hence, results underscore the importance of incorporating size considerations in the factor construction process to mitigate performance deterioration or factor dilution in the value-weighted case. Finally, several of the alphas are also economically significant, with the 18% annualized alpha for the past month's volatility (retvol) being particularly noteworthy.

Figure 2.7: Spanning regression alphas. Intercepts from univariate regressions of power portfolio returns on conventional long-short portfolio returns for the sample period March 1980 to December 2021. Panel A shows equal-weighted results and Panel B shows value-weighted results. Power portfolios are constructed using an expanding window and conventional long-short using equal-weighted decile sorting. Factors are sorted into six groups based on their economic rationale.

A. Equal-weighted portfolios



B. Value-weighted portfolios



2.3.7 Multi-factor portfolios

Power sorting can be utilized to construct multi-factor portfolios through either an averaging or a combination approach. The averaging approach (AVP) involves aggregating factor exposures into a single weight vector by averaging the weights of each stock across different power factor portfolios. Let $W_t^{PS} = (w_{t,1}^{PS}, w_{t,2}^{PS}, \ldots, w_{t,M}^{PS})$ be a $(N_t \times M)$ matrix of weights, where M is the number of available characteristics and $w_{t,m}^{PS}$ is the m^{th} column of W_t^{PS} based on characteristic $m = 1, \ldots, M$. Each $w_{t,m}^{PS}$ is estimated based on a set of values for $\hat{\tilde{p}}_t$ and $\hat{\tilde{q}}_t$, specific to the underlying characteristic. The average weight vector is then obtained as:

$$\overline{w}_t^P = \frac{1}{M} W_t^{PS} 1'_M, \tag{2.15}$$

where 1_M is a $(1 \times M)$ vector of ones. To ensure a unit sum for the long and short sides, the weights are re-standardized:

$$w_{t,(n)}^{AVP} = \begin{cases} -\dfrac{\overline{w}_{t,(n)}^{PS}}{\sum\limits_{\overline{w}_{t,(n)}^{PS}<0} \left|\overline{w}_{t,(n)}^{PS}\right|} & \text{for } \overline{w}_{t,(n)}^{PS} < 0 \\ 0 & \text{for } \overline{w}_{t,(n)}^{PS} = 0 \\ \dfrac{\overline{w}_{t,(n)}^{PS}}{\sum\limits_{\overline{w}_{t,(n)}^{PS}>0} \left|\overline{w}_{t,(n)}^{PS}\right|} & \text{for } \overline{w}_{t,(n)}^{PS} > 0. \end{cases} \tag{2.16}$$

This approach promises significant diversification benefits by mixing factor exposures and allows for the cut-off point for the long and the short side to deviate from the characteristic median rank. As a benchmark, we repeat the same procedure using equal-weighted decile weights (AVD).

In the second approach, we combine standardized characteristic ranks into an equal-weighted composite score, which serves as a signal for constructing a power-sorted multi-factor portfolio. This approach is called Power-sorted Multi-factor Equal-weight (PME) since each characteristic contributes equally to the combined signal. To illustrate, let $\tilde{S}_t = (\tilde{s}_{\cdot,1}, \ldots \tilde{s}_{\cdot,M})$ be an $(N_t \times M)$ matrix of standardized characteristic ranks for N_t

stocks at time t. The next step is to use the average standardized characteristic rank as the underlying signal to obtain the weight vector for the composite power portfolio:

$$w_t^{PME} = w_t^{PS}\left(\frac{1}{M}\tilde{S}_t 1_M'; \tilde{p}_t, \tilde{q}_t\right). \tag{2.17}$$

Again, $\tilde{p}_t$ and $\tilde{q}_t$ are estimated based on the Sharpe ratio maximization objective to derive w_t^{PME}. As a benchmark, we use the average characteristic rank in conventional decile sorting, which is referred as Decile Mutli-factor Equal-weight (DME).

In our third approach, we construct the power multi-factor portfolio by using the sum of weights across power-sorted factors as the underlying signal. This approach considers not only the underlying characteristic scores but also the values of the characteristic powers in determining the contribution of each characteristic to the composite score.[9] We name this approach Power-sorted Multi-factor Power portfolio (PMP), and its weights are derived as:

$$w_t^{PMP} = w_t^{PS}\left(\frac{1}{M}W_t^{PS} 1_M'; \tilde{p}_t, \tilde{q}_t\right). \tag{2.18}$$

Similarly to the previous case, the values of $\tilde{p}_t$ and $\tilde{q}_t$ are calibrated to maximize the portfolio Sharpe ratio. This approach assigns higher weights to characteristics that have a better ability to identify the extreme ends of the conditional return distribution. As a benchmark, we construct a decile-sorted multi-factor portfolio using the sum of weights from decile sorting as a ranking variable. We refer to this benchmark as Decile-sorted Multi-factor Decile-weighted (DMD).

Table 2.2 presents the out-of-sample performance of the three multi-factor strategies, comparing the utilization of power sorting to the conventional benchmark. One might anticipate a reduced opportunity set for power sorting in multi-factor portfolios due to the

[9] Consider as an example a hypothetical factor with $\tilde{q} = 1$ and $\tilde{p} = 0$. The concentration ratios indicate that a stock with a low characteristic should be allocated a highly negative weight, while a stock with a high characteristic should be allocated a moderately positive weight, reflecting the varying importance of the characteristic rank.

inclusion of multiple signals. However, our findings show that power sorting consistently outperforms the standard procedure across all construction schemes. This outperformance holds true for both equal-weighted and value-weighted portfolios, highlighting the robustness of our approach.

Table 2.2: Portfolio evaluation measures for multi-factor portfolios. AVP: Multi-factor portfolio based on the average portfolio weight from individual power portfolios. AVD: Multi-factor portfolio based on average portfolio weight from individual decile long-short portfolios. PME: Power portfolio on based the average characteristic rank. DME: Decile long-short portfolio based on the average characteristic rank. PMP: Power portfolio based on the rank implied by average power portfolio weights. DMD: Decile long-short portfolio based on the rank implied by the average decile long-short portfolio weights. The sample period covers March 1980 to December 2021. Panel A shows equal-weighted results and Panel B shows value-weighted results.

	AVP	AVD	PME	DME	PMP	DMD
A. Equal-weighted portfolios						
Return (%)	1.82	1.51	3.05	2.04	3.16	1.98
Standard deviation (%)	4.96	4.32	7.27	6.39	7.30	5.62
Sharpe Ratio	1.27	1.21	1.45	1.11	1.50	1.22
t-stat	8.28	7.89	9.44	7.19	9.72	7.92
Maximum drawdown (%)	−46.08	−41.80	−55.51	−56.59	−55.40	−53.10
Hit rate (%)	71.60	70.61	73.72	68.77	72.73	70.75
Turnover (%)	37.33	41.40	48.57	52.34	66.73	59.18
# of effective names long	1814.32	1425.83	1608.76	390.38	146.34	390.38
# of effective names short	600.71	935.10	101.38	390.38	141.27	390.38
B. Value-weighted portfolios						
Return (%)	1.36	0.84	2.80	1.72	2.50	1.32
Standard deviation (%)	4.57	3.54	7.40	6.42	7.07	5.39
Sharpe Ratio	1.03	0.82	1.31	0.93	1.23	0.85
t-stat	6.68	5.32	8.52	6.01	7.96	5.51
Maximum drawdown (%)	−41.21	−42.01	−61.54	−57.63	−56.61	−43.26
Hit rate (%)	67.26	63.31	69.37	64.03	68.77	61.46
Turnover (%)	34.00	42.89	44.64	47.84	52.56	51.64
# of effective names long	641.23	420.89	578.14	211.20	166.51	191.04
# of effective names short	510.43	433.92	79.87	67.56	89.25	69.45

Regarding specific strategies, AVP and AVD exhibit a similar risk-return profile in the equal-weighted case. However, the strength and robustness of the power sorting procedure become evident when market capitalization is incorporated into the construction of the multi-factor portfolio. In this case, power sorting experiences significantly lower performance deterioration compared to the conventional approach, while also maintaining lower turnover and drawdown risk. This discrepancy in value-weighted results for the two approaches further emphasizes the importance of effectively incorporating the size effect within the factor weighting procedure, demonstrating its positive impact on the risk-return profile of the multi-factor portfolio.

Moving on to the combination approaches, both PME and PMP display significant outperformance in terms of average returns and Sharpe ratios compared to their respective benchmarks. The notable performance advantage of PME over DME demonstrates the ability of power sorting in generating superior portfolios utilizing the same information source, emphasizing its effectiveness in extracting optimal performance from informative signals. Similarly, the superior performance of PMP over DMD highlights the advantages of combining power sorting with multiple characteristics and emphasizes the effectiveness of this approach in aggregating and integrating various characteristics into a composite signal.

Overall, across the different portfolio weighting methods, PMP stands out with the highest overall return performance, followed by PME. It is worth noting that the key factor driving the performance difference between PMP and PME lies in their long legs. PME takes a diversified approach in the long leg, as indicated by a high number of effective names, suggesting that the combined characteristic rank does not strongly differentiate returns in the long tail. On the other hand, PMP adopts an aggressive long stance, indicating that the combined power portfolio weights can effectively identify strong performers. This outcome highlights the significant effectiveness of power sorting, as it allows for characteristic-specific treatment of weights. By assigning more weight to characteristics with higher concentration ratios, PMP leverages power sorting to identify and capitalize on assets with robust performance potential.

2.3.8 Asset pricing tests

Finally, we examine the asset pricing implications of power sorting across existing asset pricing models using the squared Sharpe ratio test of Barillas et al. (2020). This test enables direct model comparison by quantifying the difference in squared Sharpe ratios between two models, eliminating the need for test assets. Our objective is to assess whether incorporating power-sorted factors into predetermined models enhances the squared Sharpe ratio and, consequently, the pricing ability of these models beyond what is achieved by conventional factors.

We consider three asset pricing models that can be constructed from our characteristic universe. The first model is the 5-factor model (FF5) introduced by Fama and French (2015), which extends the previous 3-factor model by adding profitability and investment factors. The second model, FF5M, follows the framework proposed by Fama and French (2018) augmented by the momentum factor. Our final model is the 4-factor model suggested by Hou, Xue, and Zhang (2015), which includes size, investment, profitability, and the market factor.

To ensure a meaningful comparison, we employ value-adjusted power-sorted factors (using $h = 0.5$) and compare them to the factors provided in the original studies, which we obtained from the authors' websites. It is worth noting that while the proposed models concentrate on similar economic drivers—namely, market, size, profitability, and investment—they diverge in their approaches to constructing the underlying variables. For instance, while both models utilize the percentage change in total assets (agr) as a proxy for investment, Fama and French (2015) emphasize operating profitability (operprof), whereas Hou, Xue, and Zhang (2015) examine return on equity (roeq) as a measure of profitability. Moreover, the original papers have adopted distinct methodologies for constructing these factors. For instance, Fama and French (2015) used independent 2×3 sorts based on size, although they acknowledge that this choice is quite arbitrary. The motivation behind the 2×3 sorting methodology is to capture the factor effect across different size groups, ensuring a balanced representation of small and large stocks. By implementing value-adjusted power sorting with a parameter value of $h = 0.5$, we effectively replicate this effect, as it guarantees the inclusion of smaller capitalization stocks in the factor, provided they possess a sufficient characteristic rank. On the other hand, Hou, Xue, and Zhang (2015) conducted a triple $2 \times 3 \times 3$ sort on their characteristics to achieve orthogonality among the predictors. This sorting method helps reduce the covariance among factors, thereby decreasing the variance component of the squared Sharpe ratio. Under a power-sorting framework, a similar effect could be achieved by fine-tuning the powers of the factors to minimize factor covariance or even directly maximize the model squared Sharpe ratio.

Table 2.3: Asset pricing models based power-sorted versus original factors. θ_P^2: Squared Sharpe ratio of factor model utilizing power-sorted factors. θ_O^2: Squared Sharpe ratio of factor model utilizing original factors. $\theta_P^2 - \theta_O^2$: Difference in squared Sharpe ratio. We conduct nonnested pairwise model comparisons with traded factors using sequential testing. We first reject the null-hypothesis that the difference between the market factor, which is the only overlapping factor, and a model that includes all the non-overlapping factors from both competing model versions is different from zero. We then test whether the squared Sharpe ratios of the nonnested models are different by computing the p-value as in Barillas et al. (2020).

	FF5	FF5M	HXZ
θ_P^2	0.236	0.281	0.238
θ_O^2	0.127	0.150	0.156
$\theta_P^2 - \theta_O^2$	0.097	0.114	0.085
p-value	0.006	0.002	0.058

Table 2.3 evidences that models incorporating power-sorted factors consistently outperform conventional models in terms of squared Sharpe ratio across all scenarios. These results are statistically significant at a 1% level for two out of three cases, with the q-theory model proposed by Hou, Xue, and Zhang (2015) exhibiting significance at the 10% level. These findings underscore the significant asset pricing implications of power sorting, demonstrating its capacity to enhance the performance of asset pricing models.

2.4 Benchmarking and Robustness

2.4.1 Alternative Benchmarks

Rank portfolios

It is natural to investigate how power sorting compares to alternative factor portfolio weighting schemes.[10] As shown in Figure 2.2, rank sorting constitutes a special case of power sorting. Hence, it is important to examine whether the incorporation of non-linearities and asymmetries through the use of powers adds value beyond the use of simple rank portfolios.

Table 2.4: Power sorting versus rank sorting. Return: Average monthly return, Standard deviation: Monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: Return=0, Hit rate: Percentage frequency of positive returns, Turnover: Average monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample period is from March 1980 to December 2021. In the value-weighted case we use a value $h = 0.5$ for both rank and power portfolios.

	Equal-weighted		Value-weighted	
	Power	*Rank*	*Power*	*Rank*
Return (%)	0.77	0.32	0.62	0.22
Standard deviation (%)	4.04	2.78	4.17	2.68
Sharpe ratio	0.72	0.45	0.52	0.30
t-stat	4.64	2.91	3.40	1.97
Maximum drawdown (%)	−45.79	−41.20	−48.96	−41.80
Hit rate (%)	61.88	57.35	58.27	54.57
Turnover (%)	37.57	28.47	33.09	23.28
# of effective names long	1,315.41	1,329.93	460.01	543.98
# of effective names short	535.57	1,331.95	229.79	622.76

[10] In this section, we compare the performance of alternative approaches at a univariate level, while Section A.3.3 of the Appendix presents results for multi-factor strategies applied to the alternative benchmarks.

Table 2.4 presents the average portfolio results for power portfolios and rank portfolios, encompassing both value- and equal-weighted cases. On average, power portfolios deliver a considerably higher annualized Sharpe ratio by providing more than double the return without doubling the risk. In contrast, rank portfolios generally demonstrate lower risk and turnover given broadly diversified positions. Specifically, their weight function corresponds to the rank in a linear manner, resulting in minor weight adjustments. Conversely, power portfolios can adopt more concentrated positions and vary their level of concentration over time, thus introducing an additional layer of turnover. This effect is more noticeable on the short side, where power portfolios tend to exhibit higher concentration, resulting also in a lower number of effective names.

Parametric Portfolio Policy

Next, we compare power sorting against the Parametric Portfolio Policy (PPP) of Brandt, Santa-Clara, and Valkanov (2009) for single characteristics in isolation. Under a PPP framework, the cross-sectional portfolio weight vector is defined as a linear function of the underlying firm characteristic:

$$w_t^{PPP}(\bar{w}_t, \hat{x}_t; \theta) = \bar{w}_t + \frac{1}{N_t}\theta\hat{x}_t, \tag{2.19}$$

where $\bar{w}_t$ denotes benchmark portfolio weights, θ the coefficient to be estimated, $1/N_t$ is a normalization term, and $\hat{x}_t$ is the cross-sectional characteristic vector at date t, standardized cross-sectionally to have zero mean and unit standard deviation. In our case, the benchmark portfolio weight is determined by equally weighting stocks above and below the mean characteristic value. This approach allows θ to have a similar interpretability to that of p and q, where a value of zero reflects a passive stance and higher values shift the weight distribution towards stocks with more extreme characteristics. As in the case of power sorting, we maximize the in-sample Sharpe ratio and impose a cap on the estimate-constrained version, wherein θ is limited to not exceed a maximum weight of

2%.[11] Finally, to ensure unit dollar portfolio legs we normalize positive and negative weights to an absolute unit sum. Hence, as $\theta \to \infty$, the portfolio weight vector converges to $\hat{x}_t$ normalized such that positive (negative) scores sum up to 1 (-1).

Similar to power sorting, PPP directly derives portfolio weights from firm characteristics and can be optimized for any arbitrary objective function, such as maximizing investors' utility or portfolio Sharpe ratio. Despite their similarities, there are some apparent differences between the two approaches. First, power sorting utilizes characteristic ranks and introduces non-linearities through the use of powers, while PPP employs a linear specification and standardized characteristics to derive portfolio weights. Despite PPP's linear approach, the resulting portfolio weights are, in fact, non-linear, with the non-linearities being driven by the higher moments of the cross-sectional characteristic distribution. Even though recent research suggests that the higher moments of the characteristics could potentially be informative about expected returns (e.g., Kagkadis et al. (2024b)), it is crucial to recognize that these higher moments are heavily contingent upon the underlying feature construction scheme. For instance, when examining momentum, Figure 2.4 emphasizes the joint importance of the long and the short side of the momentum factor. However, since cumulative returns cannot be lower than –100%, the momentum distribution exhibits positive skewness by construction. Consequently, PPP is biased into assigning more extreme weights to the long side. In contrast, power sorting, by utilizing ranks and introducing powers, allows deliberate manipulation of the higher moments of the weight distribution in a way most pertinent to returns.

In addition, a notable distinction between the two approaches arises with regards to the stability of the weight distribution over time. Under power sorting, the shape of the weight distribution remains approximately constant as long as p and q do not change significantly, while under PPP, it can exhibit significant variation, even with a stable θ, due to strong time-variation in the skewness and kurtosis of the underlying characteristic's

[11] Note that in rare cases the characteristic distribution is so skewed towards the one side that the respective benchmark portfolio ends up containing fewer than 50 stocks and hence the constraint is not achievable. Equally, for platykurtic characteristic distributions the constraint is redundant.

distribution. In that sense, the comparative analysis of power sorting and PPP sheds light on the extent to which rapid deviations of characteristics' higher moments from their long-term values are informative for future returns.

Table 2.5: Power sorting versus Parametric Portfolio Policy. Return: Average monthly return, Stand. dev.: Monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: Return=0, Hit rate: Percentage frequency of positive returns, Turnover: Average monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample period is from March 1980 to December 2021. PPP-constrained applies a maximum weight constrain of 2%.

	Power	PPP-constrained	PPP-unconstrained
Return (%)	0.77	0.43	0.46
Stand. dev. (%)	4.04	3.18	3.36
Sharpe ratio	0.72	0.49	0.50
t-stat	4.64	3.21	3.23
Max drawdown (%)	−45.79	−44.02	−45.10
Hit-rate (%)	61.88	58.12	58.23
Turnover (%)	37.57	34.21	34.75
# of eff. names long	1315.41	1398.32	1360.83
# of eff. names short	535.57	884.34	843.51

Table 2.5 presents the average portfolio results for power sorting and PPP. To ensure comparability with power sorting, PPP is implemented by maximizing Sharpe ratio as in Section 2.3.2. First, we observe that imposing a maximum theta constraint on PPP portfolios has minimal impact on their performance, albeit marginally decreasing their average risk profile. Second, the results show that power portfolios provide average returns that are markedly higher than the ones offered by the PPP portfolios (0.77% per month for power sorting vs 0.43% and 0.46% per month for PPP). These higher average returns are also translated into higher Sharpe ratios. In essence, power sorting is able to capitalize better the information content in the tails of the various characteristic distributions with only a slight increase in turnover compared to PPP.

Overall, these results underscore the importance of independently addressing and parameterizing the two legs of factor portfolios. Furthermore, leveraging higher moments within the cross-sectional characteristic distribution to influence factor portfolio weights does not appear particularly fruitful. Instead, employing a constant approach via power transformations effectively moderates the weight vector and improves investment performance.

Efficient sorting portfolios

Lastly, we evaluate the performance of power sorting in comparison to the "efficient sorting" approach proposed by Ledoit, Wolf, and Zhao (2019). The term "efficient" refers to minimum variance-optimized factor portfolios that preserve the characteristic spread of the original long-short decile portfolio. Specifically, the weight vector w_t^{EF} at each point in time is estimated as:

$$\min_{w_t^{EF}} w_t^{EF\prime} \hat{H}_t w_t^{EF} \tag{2.20}$$

$$\text{subject to} \quad x_t' w_t^{EF} = x_t' w_t^{LS} \text{ and} \tag{2.21}$$

$$\sum_{w_{t,i}^{EF}<0} \left|w_{t,i}^{EF}\right| = \sum_{w_{t,i}^{EF}>0} \left|w_{t,i}^{EF}\right| = 1 \tag{2.22}$$

where $\hat{H}_t$ is an estimator of the (conditional) VCV. The resulting portfolio is supposed to have the same exposure to the underlying characteristics as the original long-short portfolio because of (2.21), but smaller variance, and therefore a higher Sharpe ratio because of (2.20).[12] For the estimation of H_t, we employ the Quadratic-Inverse Shrinkage estimator proposed by Ledoit and Wolf (2022). Specifically, at each investment date, we estimate $\hat{H}_t$ for stocks with available return history over the most recent five years (i.e., 1,260 days), which considerably reduces the viable investment universe in the comparison.[13] Finally, we winsorize the cross-sectional characteristic vector m_t at each period t, following the

[12]Note that we additionally incorporate a maximum weight constraint of 2% to align with the power sorting framework and to prevent the minimum-variance optimizer from generating excessively large and imbalanced positions for the long and short side.

[13]The effect of this constraint on the sample size is illustrated in Figure A.12 in the Appendix.

methodology outlined in Ledoit, Wolf, and Zhao (2019).

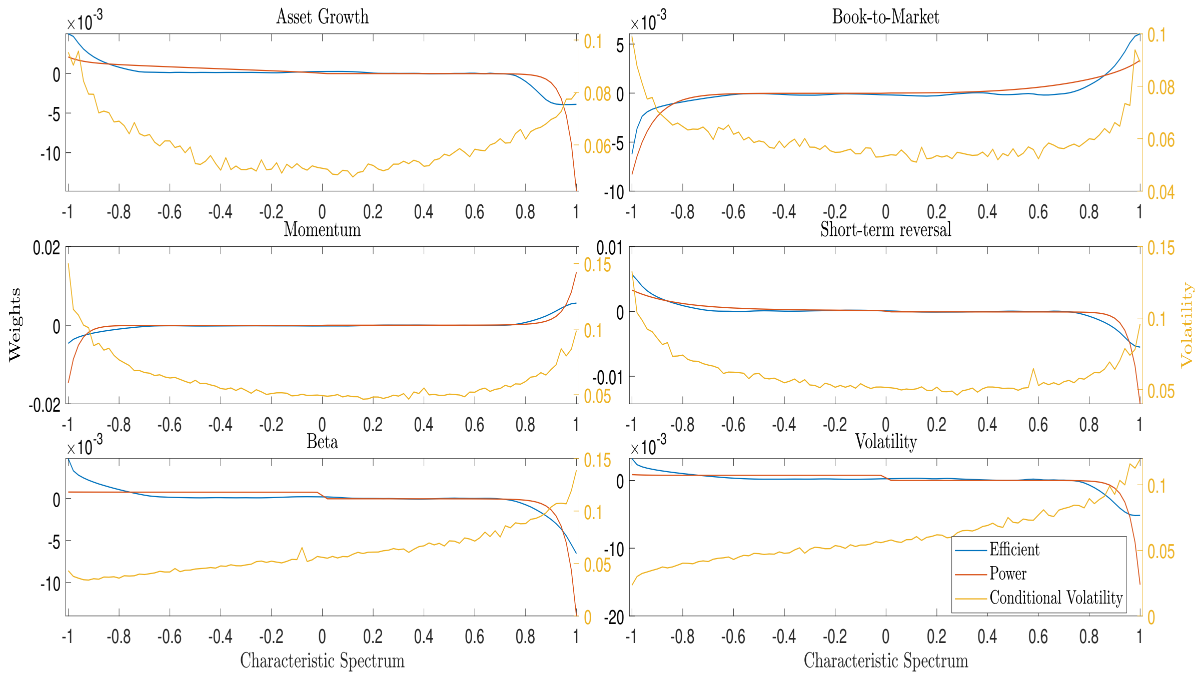
A virtue of power sorting is that there is no need for computing a VCV at an individual stock level. We though investigate whether enriching the power sorting procedure by $\hat{H}_t$ is beneficial, thus making efficient sorting and power sorting more comparable. In particular, we modify equation (2.13) as:

$$var(r^{PS}) = w_t^{PS\prime} \hat{H}_t w_t^{PS}. \tag{2.23}$$

Figure 2.8 illustrates the average weight function for selected factors under power sorting and efficient sorting, alongside the conditional volatility across quantile groups. Evidently, the two approaches differ in terms of portfolio construction, reflecting their distinct underlying objective functions. Specifically, power sorting aims to exploit variations in conditional returns to maximize the factor portfolio Sharpe ratio, while efficient sorting focuses on minimizing variance while maintaining the same characteristic spread.

To gain insights into how these different objectives translate into portfolio decisions, consider the volatility factor as an illustrative example. Recall from Figure 2.1 that stocks with the highest volatility exhibit relatively lower average returns, while those with the lowest volatility do not demonstrate significant outperformance. Consequently, the power-sorted portfolio adopts an aggressive stance on the short side and a more diversified one on the long side to capitalize on this pattern. In contrast, the efficient sorting portfolio aims to minimize variance by reducing exposure to stocks with the highest volatility on the short side and increasing exposure to those with the lowest volatility on the long side. Similar conclusions can also be drawn for beta, while in other cases efficient sorting tends to take a more passive stance, particularly on the short side. Only in the case of book-to-market ratio, efficient sorting adopts a more aggressive stance than power sorting, even though this behavior does not align with conditional returns. This result lies in the fact that as we move towards the extremes (high and low book-to-market ratios), the covariance between the long and short positions increases, leading to a reduction in the overall long-short variance. Hence, efficient sorting falls short of fully capturing the relationship between characteristics and returns beyond what is implied by covariance alone. On the other hand, power sorting integrates characteristic, return, and variance information, directly targeting the Sharpe ratio, while preserving the factor structure through the imposition of monotonicity.

Figure 2.8: Average weight function for efficient sorting and power sorting. Characteristics are standardized in the $[-1, 1]$ range. The conditional volatilities (orange lines) are estimated by ranking stocks based on their underlying characteristic in the previous period and splitting them into 100 equal-weighted quantile portfolios. The blue lines represent the weight function under efficient sorting, while the dashed blue lines depict the weight function under power sorting. The sample includes all common shares on NYSE, AMEX, and NASDAQ exchanges and covers the period from January 1980 to December 2021.



Next, Table 2.6 presents the average portfolio evaluation measures using the viable sample (that only includes stocks that have five years of return history at a given point in time) for the four approaches: the conventional approach, the efficient sorting approach, the original power sorting approach, and the modified power sorting approach utilizing the stock-level VCV, which we label as Power-VCV. Consistent with the findings of Ledoit, Wolf, and Zhao (2019), efficient sorting consistently reduces factor portfolio variance (2.90% for efficient sorting vs 3.96% for conventional sorting). However, this reduction in variance comes at a slight cost of lower average returns. This result aligns with the notion that characteristics are intertwined with the covariance structure of returns (Kelly, Pruitt, and Su 2019). Consequently, when attempting to limit the underlying factor portfolio variance, there is an unavoidable trade-off with the underlying risk premia. Additionally, we note that the number of effective names for the efficient sorting approach implies a more symmetric stance that focuses on both extremes, even though the characteristic-return relationship is often asymmetric.

Table 2.6: Power sorting versus efficient sorting. Return: Average monthly return, Stand. dev.: Monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: Return=0, Hit rate: Percentage frequency of positive returns, Turnover: Average monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample includes stocks with an available return history of five years at each investment date through the sample period from March 1980 to December 2021.

	Power	Power-VCV	Efficient	Conventional
Return (%)	0.50	0.44	0.31	0.36
Standard deviation (%)	3.73	3.46	2.90	3.96
Sharpe ratio	0.52	0.50	0.37	0.34
t-stat	3.35	3.24	2.43	2.19
Maximum drawdown (%)	−45.50	−43.63	−44.71	−55.62
Hit rate (%)	58.89	58.87	56.07	55.78
Turnover (%)	35.72	35.96	40.83	38.46
# of effective names long	940.97	1,031.10	529.46	270.23
# of effective names short	434.85	478.56	493.05	264.10

On the other hand, power sorting effectively captures the inherent asymmetries in many characteristics, leading to a significant increase in average factor portfolio returns, along with a slight decrease in portfolio variance compared to the conventional method. As a result, the average Sharpe ratios and t-statistics show notable enhancements. Specifically, our findings demonstrate a 53% increase in the average t-stat through power sorting, compared to an 11% increase with efficient sorting. Importantly, this result remains consistent regardless of whether the variance is estimated directly from the power portfolio time-series or using a VCV approach. Finally, note that while power sorting portfolios may exhibit higher volatility than efficient sorting portfolios, this increased volatility does not translate into higher drawdown risk, with average turnover being also lower.

2.4.2 Robustness Tests

Next, we analyze whether the presented results generalize to different sub-periods and methodological alternations. We show that power sorting generates performance that is robust to the choice of maximum weight thresholds, different size adjustment levels, and different sub-periods.

Lookback window

In the base case we employ an expanding window for estimating the optimal concentration ratios. Here, we explore the out-of-sample power using different rolling windows ranging from 12 months up to 10 years. With the expanding window, the estimated concentration ratios converge toward the values that were most effective through the whole sample, while a rolling window is more adaptive. Shorter windows adapt more dynamically to recent information, potentially introducing higher variation in the concentration ratios and resulting in more pro-cyclical strategies.

Table 2.7 presents the average portfolio evaluation measures across power portfolios for different lookback windows. To ensure consistency regarding the length of the evaluation period, results are assessed for the out-of-sample period from January 1990 onward. Our findings reveal that both short and long windows yield similar return performances, with the expanding window showing a slight advantage on average. Generally, longer lookback windows achieve comparable investment performance while maintaining significantly lower turnover, making them more desirable from a practical standpoint. Overall, results remain consistent across different formation periods, with power portfolios consistently outperforming the conventional benchmark, and the results not being driven by higher turnover or tail risk.

Table 2.7: Robustness with respect to different lookback windows. Return: average monthly return, Standard deviation: monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: return=0, Hit rate: percentage frequency of positive returns, Turnover: monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample period is from January 1990 to December 2021. Panel A shows equal-weighted and Panel B shows value-weighted results.

Lookback window	12 Months	36 Months	60 Months	120 Months	Expanding Window	Conventional
A. Equal-weighted portfolios						
Return (%)	0.65	0.60	0.62	0.63	0.65	0.43
Standard deviation (%)	4.50	4.53	4.49	4.41	4.30	4.55
Sharpe ratio	0.54	0.50	0.53	0.55	0.58	0.38
t-stat	3.04	2.86	3.00	3.11	3.31	2.14
Maximum drawdown (%)	−46.14	−49.75	−49.44	−48.93	−45.70	−54.04
Hit rate (%)	58.61	58.81	59.35	59.79	60.05	55.99
Turnover (%)	50.16	42.54	39.97	37.84	36.39	39.03
# of effective names long	1,133.67	1,201.18	1,261.32	1,311.78	1,369.61	374.18
# of effective names short	709.94	594.08	564.05	523.48	518.53	377.85
B. Value-weighted portfolios						
Return (%)	0.48	0.45	0.47	0.48	0.53	0.26
Standard deviation (%)	4.46	4.51	4.52	4.50	4.37	4.66
Sharpe ratio	0.39	0.36	0.38	0.39	0.44	0.21
t-stat	2.20	2.05	2.17	2.24	2.53	1.16
Maximum drawdown (%)	−47.60	−51.77	−51.28	−50.95	−48.18	−59.06
Hit rate (%)	55.77	55.74	56.34	56.45	56.82	52.65
Turnover (%)	48.23	39.27	36.24	33.81	32.61	35.44
# of effective names long	412.18	432.73	448.25	465.18	454.03	109.18
# of effective names short	306.44	260.62	239.16	216.16	187.25	100.92

Maximum weights

In the main analysis, we opted for a maximum portfolio weight of 2% to ensure that the power portfolios are properly diversified. Table 2.8 shows how average results change for alternative choices of maximum portfolio weight, ranging from 0.5% to 10%. Higher maximum weights lead to higher values for p_t^{max} and q_t^{max} and hence to weight distributions that are potentially more concentrated in the tails, delivering higher returns at the expense of higher risk and turnover.

Table 2.8: Robustness with respect to the maximum stock weight. Panel A shows equal-weighted results and Panel B shows value-weighted results. Return: average monthly return, Standard deviation: monthly standard deviation Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: return=0, Hit rate: percentage frequency of positive returns Turnover: monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample covers the period from March 1980 to December 2021. Panel A shows equal-weighted results and Panel B shows value-weighted results.

w^{ceil}	0.5%	1%	2%	3%	4%	5%	10%
A. Equal-weighted portfolios							
Return (%)	0.53	0.67	0.77	0.83	0.85	0.86	0.87
Standard deviation (%)	3.26	3.72	4.04	4.21	4.33	4.40	4.49
Sharpe ratio	0.65	0.69	0.72	0.73	0.72	0.72	0.71
t-stat	4.19	4.50	4.64	4.72	4.71	4.70	4.64
Maximum drawdown (%)	−41.07	−44.28	−45.79	−46.10	−47.34	−47.28	−48.11
Hit rate (%)	60.68	61.65	61.88	62.09	62.21	62.11	62.18
Turnover (%)	32.49	35.55	37.57	38.62	39.29	39.69	40.01
# of effective names long	1413.22	1342.05	1315.41	1298.95	1290.90	1288.57	1287.28
# of effective names short	816.46	612.07	535.57	508.47	493.30	487.75	491.10
B. Value-weighted portfolios							
Return (%)	0.39	0.52	0.62	0.67	0.70	0.71	0.72
Standard deviation (%)	3.34	3.81	4.17	4.35	4.48	4.57	4.67
Sharpe ratio	0.44	0.49	0.52	0.53	0.54	0.54	0.53
t-stat	2.84	3.20	3.40	3.47	3.49	3.49	3.46
Maximum drawdown (%)	−44.83	−47.47	−48.96	−49.62	−50.27	−50.84	−52.02
Hit rate (%)	56.78	57.74	58.27	58.34	58.34	58.53	58.60
Turnover (%)	31.12	34.08	36.02	37.24	38.01	38.41	38.70
# of effective names long	1306.86	1197.04	1144.98	1128.32	1119.88	1121.61	1117.87
# of effective names short	371.61	275.21	229.79	214.08	202.74	197.20	196.74

Increasing the upper weight threshold up to 10% can result in higher average return gains for portfolio performance, with the Sharpe ratio remaining practically unchanged. In the equal-weighted case, the maximum Sharpe ratio is achieved at a 3% weight threshold (0.73). For value-weighted data, the maximum Sharpe ratio is achieved with a weight concentration of 5% (0.54), suggesting that a higher power threshold is required to extract optimal performance after accounting for market capitalization. Importantly, power sorting does not appear to excessively increase concentration in both legs, even when higher maximum weight thresholds are allowed. This can be attributed, in part, to the objective of maximizing the power portfolio Sharpe ratio, which helps maintain a balance between concentration and diversification.

Conversely, the enforcement of high diversification via a low weight threshold may moderate the effectiveness of power sorting in exploiting return-relevant characteristic variation (i.e., setting $w^{ceil} = 0.5\%$). Nevertheless, values below 1% can lead to maximum power thresholds below one for the different factors ($p_t^{max} \wedge q_t^{max} < 1$), rendering them insufficient upper bounds for examining concentration in the tails.

Concentration in mega-cap stocks

In the base case, we employed $h = 0.5$ to address extreme concentration in mega-cap stocks when evaluating value-weighted results. Here, we examine the implications of different values of h on the performance of value-weighted power portfolios. Additionally, we consider different variations of the conventional decile sorts to assess the sensitivity of the conventional approach with regard to the treatment of size effects. First, we compute "pure" value-weighted portfolios without winsorizing market-caps of individual stocks, effectively setting $h = 1$ in the power sorting framework. Second, since $h = 0.5$ is equivalent to using the square root, we analyze the effect of employing the square root of the market cap within the conventional decile sorting approach.

Table 2.9 displays the average portfolio evaluation results for values of h ranging from 0 (equal-weighted) to 1 (pure value-weighted), along with the different benchmark variations. Lower values yield better portfolio performance as they minimize the effect of size on portfolio composition. Nonetheless, even when there is no size adjustment for the value-weighted power portfolios ($h = 1$), power sorting outperforms capped and unadjusted value-weighted decile sorting. Moreover, the performance of $h = 0.5$ significantly surpasses that of the square root approach in the conventional framework, reflecting the ability of the method to effectively incorporate size effects in portfolio construction.

Table 2.9: Robustness with respect to concentration on mega-cap stocks. Conv. VW (Capped): Stocks are weighted by their market cap winsorized at the NYSE 80^{th} percentile. Conv. VW (No Adj.): Stocks are weighted by their market cap without any adjustment. Conv. VW (Square root): Stocks are weighted by the square root of their market cap. Return: average monthly return, Standard deviation: monthly standard deviation Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: return=0, Hit rate: percentage frequency of positive returns, Turnover: monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample period covers March 1980 to December 2021.

	Power Sorting					Conventional VW		
	$h = 0$	$h = 0.25$	$h = 0.5$	$h = 0.75$	$h = 1$	*Capped*	*No Adj.*	*Square root*
Return (%)	0.77	0.71	0.62	0.52	0.43	0.32	0.29	0.39
Standard deviation (%)	4.04	4.09	4.17	4.41	4.74	4.39	4.78	4.20
Sharpe Ratio	0.72	0.64	0.52	0.39	0.29	0.26	0.20	0.33
t-stat	4.64	4.15	3.40	2.54	1.85	1.71	1.31	2.19
Maximum drawdown (%)	-45.79	-47.10	-48.96	-52.91	-58.50	-59.22	-65.05	56.81
Hit rate (%)	61.88	60.53	58.27	55.77	54.20	53.44	52.82	55.19
Turnover (%)	37.57	35.05	33.09	32.54	33.11	35.44	35.85	34.71
# of effective names long	1,315.41	975.46	460.01	180.73	73.60	107.00	37.09	165.09
# of effective names short	535.57	409.18	229.79	116.87	67.47	98.42	37.21	158.65

Sub-period analysis

Finally, in Table 2.10 we conduct a decade-by-decade analysis which shows the average portfolio evaluation measures for power portfolios and the conventional approach for the four sub-periods. The magnitude of the difference between power and conventional sorting covaries with the efficacy of factor investing as a whole, corroborating that results are driven by extracting optimal performance from the underlying factors rather than introducing other effects on the portfolio construction procedure. Confirming our full-sample analysis, the added value of power sorting is consistently positive within the chosen sub-period and is not driven by specific time periods. However, it is important to highlight that the performance of factor investing as a whole exhibits a noticeable decline in later years, as observed in previous studies (McLean and Pontiff 2016).

Table 2.10: Robustness across different sub-periods. Return: average monthly return, Standard deviation: monthly standard deviation Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: return=0, Hit rate: percentage frequency of positive returns, Turnover: monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample period covers March 1980 to December 2021. Panel A shows equal-weighted results and Panel B shows value-weighted results.

	1980/1989	1990/1999	2000/2010	2011/2021
A. Equal-weighted portfolios				
A.1 Power sorting				
Return (%)	1.19	0.90	0.68	0.43
Standard deviation (%)	2.81	3.53	5.64	3.39
Sharpe ratio	1.40	0.97	0.50	0.44
t-stat	4.42	3.07	1.58	1.52
Maximum drawdown (%)	−16.01	−25.14	−33.87	−31.97
Hit rate (%)	67.81	63.75	60.21	57.01
Turnover (%)	41.49	38.50	38.52	32.92
# of effective names long	1138.13	1595.78	1461.20	1115.79
# of effective names short	590.50	610.45	548.69	421.27
A.2 Conventional				
Return (%)	0.83	0.60	0.43	0.24
Standard deviation (%)	2.76	3.68	5.94	3.70
Sharpe ratio	1.00	0.66	0.31	0.23
t-stat	3.16	2.09	0.99	0.79
Maximum drawdown (%)	−20.68	−30.80	−42.56	−38.83
Hit rate (%)	62.33	59.25	55.86	53.13
Turnover (%)	40.59	40.67	40.11	36.81
# of effective names long	352.98	450.04	405.37	288.44
# of effective names short	345.68	455.04	411.05	289.38
B. Value-weighted portfolios				
B.1 Power sorting				
Return (%)	0.95	0.77	0.56	0.29
Standard deviation (%)	2.95	3.84	5.79	3.38
Sharpe ratio	1.02	0.72	0.37	0.29
t-stat	3.21	2.27	1.18	1.03
Maximum drawdown (%)	−19.49	−28.47	−36.20	−31.21
Hit rate (%)	62.71	60.50	56.27	54.60
Turnover (%)	38.33	33.06	33.58	28.54
# of effective names long	448.13	541.06	463.29	402.63
# of effective names short	288.98	257.16	213.02	174.07
B.2 Conventional				
Return (%)	0.46	0.33	0.43	0.13
Standard deviation (%)	3.28	3.98	6.10	3.59
Sharpe ratio	0.48	0.29	0.28	0.11
t-stat	1.51	0.92	0.87	0.40
Maximum drawdown (%)	−28.29	−37.71	−42.26	−36.64
Hit rate (%)	55.60	54.86	53.33	50.63
Turnover (%)	37.33	35.34	35.47	34.00
# of effective names long	100.05	115.65	121.79	93.90
# of effective names short	90.42	109.48	108.19	88.25

2.5 Conclusion

We propose power sorting as a framework for constructing equity factors to improve upon conventional quantile sorting. Our method hinges on the assumption of a monotonic relationship between factor characteristics and returns. It is geared at creating refined versions of the factors while facilitating the construction of economically meaningful and sufficiently diversified portfolios. We deem the power sorting procedure as an effective compromise between conventional portfolio sorts and machine learning methods. While the former easily fails to account for characteristic-specific information, the latter is usually criticized for its lack of interpretability and its black box character. By striking a balance between interpretability and computational efficiency, our framework offers practical advantages. Under our modeling procedure, concentration ratios directly translate to weight concentration levels, allowing for a simple and intuitive interpretation of the model parameters.

We present several important empirical findings. First, we document the existence of asymmetric and non-linear patterns between characteristics and returns. Such patterns contradict the notion that the return signal is always amplified at the extremes and motivate separate treatment of the long and the short side of factor portfolios. As a consequence, off-the-shelf procedures may struggle to harvest the underlying factor premiums, which, in turn, can lead to false rejections of individual characteristics. The limitations of the conventional approach become more evident when dealing with value-weighted portfolios, as it fails to adequately account for confounding size effects. Unlike standard approaches, our method is designed to extract optimal performance from the vast majority of characteristics by allowing the weight function to be characteristic-specific and effectively incorporating size-effects in the construction of factors.

Building on these insights, we investigate the performance gains resulting from power sorting compared to the conventional quantile approach. Power sorting can generate average returns and Sharpe ratios that are up to double those achieved through conventional quantile sorting. These gains are both economically and statistically significant, survive size-adjustments, and are not driven by increased turnover or tail risk. Furthermore, the benefits persist when considering alternative optimization-based portfolio formulation approaches, suggesting that the use of exponential functions to model factor portfolio weights introduces structure to the weight vector that is beneficial in terms of out-of-sample performance.

The outperformance of power-sorted factor portfolios primarily stems from taking an aggressive stance on the short leg and adopting a more diversified one on the long leg. Hence, our results demonstrate that various characteristics are effective in identifying underperforming stocks, although they may provide mixed signals for outperforming stocks. Nonetheless, power sorting boosts performance in both the long and short leg of the various factor portfolios.

Lastly, the benefits of power sorting extend to a multi-factor level. For instance, by adopting power-sorted factors in existing asset pricing models, we can enhance the squared Sharpe ratio of the underlying model, thus increasing its ability to capture the cross-section of stock returns. In the context of multi-factor strategies, power sorting implicitly accounts for the informativeness of characteristics across the characteristic spectrum, yielding multi-factor portfolios with improved risk-return properties compared to simple equal-weighted schemes and individual factors.

Chapter 3

Multi-Factor Power Sorting

This paper is a joint work with Anastasios Kagkadis, Harald Lohre, Ingmar Nolte, and Sandra Nolte. We would like to thank Robert Korajczyk, Carsten Rother, and the participants at the 4th Frontiers of Factor Investing Conference for their valuable comments and suggestions. The views and opinions expressed in this paper are solely those of the authors and may not necessarily reflect those of Robeco or Ultramarin. Send correspondence to Nikolaos Vasilas, n.vasilas@lancaster.ac.uk.

3.1 Introduction

The asset pricing literature posits that firm characteristics convey information about future stock returns. When a firm characteristic is a strong cross-sectional return predictor, a long-short portfolio based on that characteristic should exhibit robust return performance over time, providing exposure to the associated premium. While such portfolios' robust performance suggests the importance of the underlying characteristic, it is insufficient to declare a characteristic a strong return predictor, especially given the increasing number of contenders in the factor zoo. A key challenge in empirical asset pricing is assessing whether a new candidate factor brings additional information to the existing set of factors, ultimately identifying and constructing a set of factors that can jointly explain asset returns (Cochrane 2011; Green, Hand, and Zhang 2017; Freyberger, Neuhierl, and Weber 2020).[1]

In this paper, we extend the power sorting methodology of Kagkadis et al. (2024a) to a multivariate level by incorporating the dependence structure of firm characteristics in the construction of the power sorting factor portfolios. Rather than maximizing the individual Sharpe ratios of each factor, we jointly estimate the set of weight functions for the factor portfolios to maximize their squared Sharpe ratio (Barillas et al. 2020). This approach enables leveraging potential interactions among factors within the portfolio construction process, thereby providing valuable insights into the interplay of different characteristics. In our analysis, we observe a shift from the short side to the long side when many characteristics are considered jointly. These findings complement Kagkadis et al. (2024a), who show that the short leg is the most significant contributor to univariate power sorting factor portfolio performance. This suggests that on the short side, different characteristics tend to capture rather similar effects, while on the long side, they provide more complementary information. Consequently, our methodology holds significant implications, making it relevant from both practical and academic perspectives.

[1] In this context, the underlying assumption is that asset returns are explained by a reduced form factor model or as a (linear) function of a small number of factors.

The emergence of various characteristics on the long side under multivariate power sorting supports the view that information on these characteristics is particularly useful from a practical perspective. While characteristics' information can be challenging to exploit on the short side due to shorting constraints, it can also guide investment decisions for long-only investors. More importantly, multivariate power sorting leads to significant improvements in risk-adjusted performance compared to portfolios that merely combine univariate power sorting portfolios. Specifically, for the out-of-sample period from January 2000 to December 2021, multivariate power sorting delivers a Sharpe ratio of 1.12 versus 0.76 for the univariate approach. Additionally, the tail risk of multivariate power sorting is half that of univariate power sorting, suggesting that joint optimization benefits factor diversification and reduces the risk associated with heavily loading on specific factor themes.

From an academic perspective, multivariate power sorting provides a powerful, structured, and intuitive framework for constructing characteristic-based factor models. We thus construct fully specified dynamic characteristic-based models for the cross-section of expected stock returns that optimize the pricing performance of any given characteristic set while maintaining a high level of interpretability and explainability in the estimation process. In doing so, we aim to reconcile observed and latent asset pricing models, which have emerged in the empirical asset pricing literature over recent decades. While observed models are easier to interpret and link to economic theories, they often rely on a fixed set of factors and portfolio construction schemes. On the other hand, latent models, while empirically strong, can be opaque and non-intuitive, posing challenges to their wide adoption for empirical asset pricing applications.

Our approach bridges this gap by recursively adding power sorting factors to the Capital Asset Pricing Model (CAPM; Sharpe 1964), re-estimating the power parameters at each step until the marginal contribution of the added factor to the squared Sharpe ratio is statistically insignificant. Several studies, such as Fama and French (2018) and Swade et al. (2024), employ the squared Sharpe ratio as a metric to assess and compare asset pricing models. However, none consider directly optimizing this metric during the model construction process. To the best of our knowledge, this is a unique advantage of multivariate power sorting. Our recursive approach allows us to identify the highest marginal contributor in terms of pricing ability, aiming for a parsimonious yet statistically robust factor specification. Intuitively, a given candidate power sorting factor should exhibit a high Sharpe ratio for the underlying characteristic to be a strong cross-sectional predictor. In the case of multiple factors and a linear specification, the candidate factor should also exhibit low covariance with the rest of the factor universe, implying that it provides additional information. By maximizing the squared Sharpe ratio, we effectively obtain power sorting portfolios that jointly span the tangency portfolio for the full investment universe, maximizing the attainable Sharpe ratio. This approach allows us to dynamically build factor models that implicitly incorporate interaction effects in their construction process, providing valuable insights into the dynamics of factors. More importantly, the underlying portfolios remain investable and economically relevant, allowing us to interpret the economic forces captured by asset pricing factors end-to-end.

Our procedure uncovers five power sorting factors on top of the market portfolio that are important in explaining the dynamics of stock returns in the US stock market. The selected factors in order of contribution are: asset growth (Cooper, Gulen, and Schill 2008), volatility (Ang et al. 2006), earnings announcement return (Kishore et al. 2008), unexpected quarterly earnings (Rendleman, Jones, and Latane 1982), and market capitalization (Banz 1981). Thus, our results align with the prior literature in terms of identifying characteristics that are known to provide independent information (Green, Hand, and Zhang 2017) and are part of leading characteristic-based models (i.e., Fama and French 1992, 1993, 2015, 2018 ; Hou, Xue, and Zhang 2015; and Hou et al. 2021)

Finally, we compare the performance of our power sorting 6-factor model (PS6) with existing leading academic factor models put forward by the literature to explain stock returns. PS6 attains a significantly higher out-of-sample squared Sharpe ratio compared to other models, declaring it the unanimous winner in the absence of test assets. To solidify our inferences regarding the asset pricing ability of PS6, we conduct extensive tests on a broad range of test assets. Specifically, we examine how inferences vary based on different portfolio construction choices, utilizing various sorting algorithms and weighting schemes. In addition, we address concerns about the construction or selection of test assets, and evaluate the models' ability to price the 199 factors from Chen and Zimmermann (2022) with complete return data in the out-of-sample period. Although PS6 generally delivers smaller absolute alphas against different sets of test assets, we find that the selection of test assets is crucial, as different models perform better under different construction schemes. For instance, a model using value-weighted decile-sorting portfolios is more successful at pricing the same type of portfolios compared to a model using double-sorting portfolios based on size or any other sorting scheme. As different academic models utilize various methods for constructing factors, this subtle but important detail is often overlooked, resulting in an unfair advantage for the model that matches the construction process of the test assets. Thus, our paper also implicitly contributes to the ongoing debate between a right-hand-side and left-hand-side approach for evaluating competing factor models.

The remainder of the paper is structured as follows: Section 3.2 provides a literature synopsis on characteristic-based portfolio construction and factor models. Section 3.3 introduces the multivariate power sorting procedure and relates it to the univariate approach. Section 3.4 demonstrates the implications of the multivariate procedure in terms of multi-factor portfolios. Section 3.5 introduces an iterative factor selection procedure for the construction of PS6 and compares the proposed models to others in the literature. Section 3.6 concludes.

3.2 Literature Overview

Our paper contributes to the growing literature on factor models for equity returns. Given the breadth of this literature, we focus on a brief overview in this subsection. Henceforth, let $t = 1, \dots, T$ represent time periods, $n = 1, \dots, N$ denote the number of stocks, and $k = 1, \dots, K$ indicate the number of factors. In these models, the stochastic discount factor (SDF) M_t is represented as a function of a small number of K factor portfolio returns,

$$M_t = 1 - (f_t - \mathbb{E}[f_t])\, b, \tag{3.1}$$

where f_t is a $1 \times K$ vector of factor returns at time t and b is a $K \times 1$ vector of SDF weights corresponding to the conditional mean variance efficient portfolio. The SDF M_t satisfies $\mathbb{E}[r_{n,t} M_t] = 0$ for any stock return in excess of the risk-free rate $r_{n,t}$. This setup maps to a factor model for excess returns of the form,

$$r_{n,t} = \beta_{n,k} f_{k,t} + \epsilon_{n,t} \quad \text{for} \quad n = 1, \ldots, N, \quad t = 1, \ldots, T, \quad k = 1, \ldots, K, \tag{3.2}$$

where $\beta_{n,k}$ denotes the factor exposure (or loadings) of stock n to factor k, and $\epsilon_{n,t}$ captures its idiosyncratic errors.[2] Although the above linear specification is inherently simple, it offers flexibility in modeling asset returns. The factors $f_{k,t}$ can be either observed or latent, constructed via characteristic-based sorts (e.g., Fama and French 2015), or statistical criteria (e.g., Kozak, Nagel, and Santosh 2018). Similarly, the betas can be conventional OLS regression slopes or dependent on other economic and asset-specific variables (e.g., Ferson and Harvey 1991; Kelly, Pruitt, and Su 2019).

Starting with the strand focusing on observed factors, studies in this field typically select a small number of characteristics and form equal or value-weighted portfolios as factor proxies. The most prominent example is Fama and French (1993), who, building on the CAPM, add size and value factors to the market portfolio to form a three-factor model (FF3). Motivated by the dividend discount model, Fama and French (2015) add profitability and investment factors to the three-factor model (FF5). In Fama and French (2018), the momentum factor from Jegadeesh and Titman (1993) is considered, forming a six-factor model (FF6). Another notable example is Hou, Xue, and Zhang (2015). Taking inspiration from Q-theory, they propose a four-factor model comprising size, investment, profitability, and the market factor. This model was later extended with expected growth

[2]For the model to be empirically valid, these portfolios should explain the cross-section of expected returns, as well as the co-movement of stock returns.

by Hou et al. (2021). Other recent factor models based on observable portfolios include the mispricing model by Stambaugh and Yuan (2017), the behavioral factors model by Daniel, Hirshleifer, and Sun (2020), and the revised six-factor model by Barillas et al. (2020).

Studies on latent factor models are inspired by Ross's (1976) Arbitrage Pricing Theory (APT), which assumes linear relationships between risk premia and factor betas. Successful models under this framework should exhibit a strong factor structure with low idiosyncratic volatility. Kozak, Nagel, and Santosh (2018) emphasize the absence of near-arbitrage opportunities, using principal components of anomaly returns as pricing factors. Lettau and Pelger (2020a) augment PCA with a cross-sectional pricing error to better capture average asset returns. Similarly, Kelly, Pruitt, and Su (2019) propose Instrumental Principal Components (IPCA), allowing time-varying factor loadings based on firm characteristics. Finally, Gu, Kelly, and Xiu (2021) address the limitation of PCA's assumed linearity using autoencoders.

Our work is more closely aligned with recent studies that construct characteristic-based portfolios with varying objectives. For example, Fama and French (2020) utilize the cross-sectional regression method of Fama and MacBeth (1973) to create factors based on standardized characteristics, demonstrating that these cross-sectional factors better explain average returns compared to the original Fama-French factors. Similarly, Kim, Korajczyk, and Neuhierl (2021) develop portfolios that leverage mispricing information in characteristics while hedging systematic risk. Likewise, Daniel et al. (2020) create "characteristic efficient portfolios" by using a hedge portfolio to eliminate unpriced risk variation.

Our study is also related to a strand of literature that parameterizes portfolio weights as a function of underlying firm characteristics for solving optimal portfolio formation problems. One of the earliest and most prominent examples is the Parametric Portfolio Policy (PPP) of Brandt, Santa-Clara, and Valkanov (2009). Under a PPP framework, the cross-sectional portfolio weight vector is defined as a linear function of the underlying asset characteristics, and the parameters are estimated by solving a simple utility optimization problem. Recent non-linear extensions of this framework that employ flexible neural network specifications include Cong et al. (2021) and Simon, Weibels, and Zimmermann (2023).

Finally, our paper relates to the strand of literature examining the multitude of firm characteristics. Green, Hand, and Zhang (2017) use Fama-MacBeth regressions to identify 12 characteristics as independent determinants of average monthly returns, although they observe a decrease in this number in later periods. Freyberger, Neuhierl, and Weber (2020) propose a non-parametric approach based on adaptive group lasso and find 13 characteristics that provide independent information. Using IPCA, Kelly, Pruitt, and Su (2019) identify 8 characteristics as significant. Meanwhile, DeMiguel et al. (2020) investigate the impact of transaction costs by extending the parametric portfolio policy (PPP) of Brandt, Santa-Clara, and Valkanov (2009), showing that incorporating transaction costs increases the number of important characteristics from 6 to 15. Lastly, Swade et al. (2024) identify 15 factors as relevant for spanning the entire factor zoo from an alpha perspective.

3.3 Methodology

3.3.1 Univariate Power Sorting

We begin by providing an overview of the original power sorting procedure. In this framework, factor portfolio weights are expressed as a power series of the underlying characteristic rank. For ease of presentation, assume a balanced panel of stocks. Let $x_{t,k} := (x_{t,k,1}, \dots x_{t,k,N})'$ denote the cross-sectional vector of characteristic scores for $k = 1, \dots, K$ characteristics and N stocks at time t and $r_{t+1} := (r_{t+1,1}, \dots, r_{t+1,N})'$ the vector of stock returns from month t to month $t+1$. Equally, let $s_{t,k} = (s_{t,k,1}, \dots, s_{t,k,N})'$ denote the vector of characteristic ranks of $x_{t,k}$ mapped onto the interval $[-1, 1]$ and centered around the median rank and $m_t = (m_{t,1}, \dots, m_{t,N})$ be the vector of market capitalizations for the N stocks at month t. The cross-sectional weight vector $w_{t,k} = (w_{t,k,1}, \dots, w'_{t,k,N})$ for the power sorting portfolio is given by:

$$w_{t,k,n} = w_{t,k}(s_{t,k,n}, m_{t,n}; p_k, q_k, h) = \begin{cases} -\frac{|s_{t,k,n}|^{q_k} \cdot m_{t,n}^h}{\sum\limits_{s_{t,k,n}<0} |s_{t,k,n}|^{q_k} \cdot m_{t,n}^h} & \text{for } s_{t,k,n} < 0 \\ 0 & \text{for } s_{t,k,n} = 0 \\ \frac{s_{t,k,n}^{p_k} \cdot m_{t,n}^h}{\sum\limits_{s_{t,k,n}>0} s_{t,k,n}^{p_k} \cdot m_{t,n}^h} & \text{for } s_{t,k,n} > 0. \end{cases} \tag{3.3}$$

The behaviour of the weight vector is dictated by two exponent parameters p_k and q_k that are applied to positive and negative characteristic ranks before transforming them into portfolio weights. Those exponents can be tuned across multiple cross-sections to maximize any arbitrary objective function, such as maximizing the average return and Sharpe ratio of the power sorting portfolio. Note that the incorporation of market capitalization in the estimation of power portfolio weights ensures that the underlying portfolios are "capitalization-adjusted", thus avoiding potential extreme allocations in microcaps. In that context, $h \in [0, 1]$ acts as a shrinkage parameter, regulating the concentration in mega-cap stocks.[3] Once the optimal powers are estimated, the respective portfolio weights are obtained for each cross-section by plugging the latest vector of characteristics ranks into equation (3.3). The return of the power sorting portfolio related to characteristic k from t to $t+1$ is then given by $r_{t+1,k}^{PS} = w'_{t,k} \cdot r_{t+1}$.

3.3.2 Multivariate Power Sorting

Under a multivariate framework the objective is no longer to maximize risk-adjusted performance of individual factors, but to maximize a joint performance metric based on the underlying factor set. Therefore, we consider the squared Sharpe ratio metric of Barillas et al. (2020) as a way to maximize the pricing power of a candidate factor model. Hence, for a set of K factors let $p = (p_1, \dots, p_K)'$ and $q = (q_1, \dots, q_K)'$ be the vectors of

[3] The vector of ordered market capitalizations behaves as a power series with high exponential growth, dominated by a few exponentially large stocks. Therefore, not applying any shrinkage ($h = 1$) results in the value-adjusted weighting vector being the product of two power curves, leading to potentially extreme concentrations in mega-cap stocks.

power parameters. Equally, let $R^{PS} = (r_1^{PS}, \dots, r_K^{PS}) \in \mathbb{R}^{T\times K}$ denote the matrix of power sorting portfolio returns based on p, q. The objective is to identify the set of parameter vectors p and q that jointly maximize the squared Sharpe ratio of the underlying set of factors. Specifically, the optimal parameters are estimated by solving the following optimization problem:

$$\{\hat{p}, \hat{q}\} = \arg\max \quad \bar{R}^{PS}\hat{\Omega}^{-1}\bar{R}^{PS\prime} \tag{3.4}$$

$$\text{Subject to:} \quad 0 \leq p_k \leq p_k^{\max}, \quad 0 \leq q_k \leq q_k^{\max} \quad \text{for } k = 1, \dots, K, \tag{3.5}$$

where $\bar{R}^{PS}$ is the vector of average power sorting factor portfolio returns and $\hat{\Omega}$ is the sample covariance matrix of the factors. The power parameters are bounded from below at zero and from above at q_k^{max} and p_k^{max}, based on a given maximum single stock weight.[4] Note that by restricting the parameters to be non-negative we impose an economically motivated prior that the underlying factor weighting scheme remains monotonic, thus avoiding spurious results related to taking offsetting positions in correlated factors.

3.4 Power Sorting Portfolios

3.4.1 Data

Our empirical analysis is based on a comprehensive set of 85 characteristics as outlined by Green, Hand, and Zhang (2017). These characteristics have been computed using data sourced from the Center for Research in Securities (CRSP), Compustat, and the Institutional Brokers' Estimate System (I/B/E/S), spanning the timeframe from January 1980 to December 2021. Our stock universe encompasses common stocks listed on NYSE, AMEX, and NASDAQ, meeting specific criteria such as having month-end market capitalization records on CRSP and possessing non-missing and non-negative common equity

[4]Due to its monotonic nature, it is easy to constrain the search space for the powers to avoid extreme weight allocations by solving $\frac{1}{\sum_n s_{t,k,n}^{p_k}} - w^{ceil} = 0$ to obtain p_k^{max}, where w^{ceil} is the maximum desired weight. The same logic applies to the short leg for obtaining q_k^{max}.

values on Compustat. For further details regarding the characteristics, including their origin and descriptions, please refer to Section A.1 of the Appendix.

To mitigate the effect of microcaps, we remove stocks with a market capitalization below the 10^{th} percentile at the portfolio formation period. Furthermore, we set the value of h at 0.5 for value-adjusted power sorting. This efficiently achieves a balanced mix of large and small cap stocks, ensuring the incorporation of smaller capitalization stocks into the factor, provided they meet a certain characteristic rank threshold. As a result, power sorting factor portfolios are sufficiently diversified, easily interpretable, and practically relevant, establishing a data-driven and discretion-free framework for constructing factor portfolios.

3.4.2 Univariate versus Multivariate Power Sorting

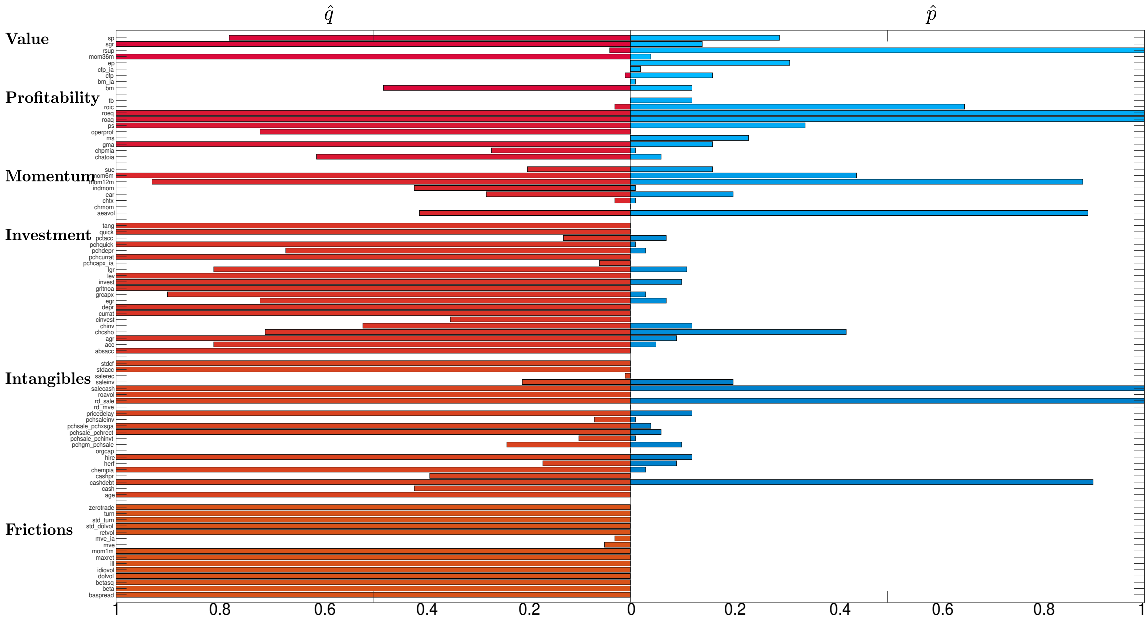
First, we demonstrate how a multivariate approach can lead to more nuanced and reliable conclusions compared to traditional univariate inference methods. To this end, we explore how the optimal powers of a given power sorting portfolio evolve when tuned within the framework of joint versus individual objectives. For illustration purposes we use the whole sample period to estimate the optimal powers that maximize individual factor portfolio Sharpe ratios versus the squared Sharpe ratio of all 85 factors.

Figure 3.1 presents the optimal powers under the two objectives, standardized by the upper bounds to allow for easy comparison across factors.[5] Blue-shaded bars represent the average standardized powers for the long side and red-shaded bars represent the average standardized powers for the short side. In the univariate scenario in Panel A, maximizing the Sharpe ratio of individual power sorting portfolios leans towards an aggressive stance on the short side, juxtaposed with a more conservative stance on the long side. This reflects in disproportionately higher powers allocated to the short side compared to the long side, suggesting that the various characteristics effectively identify underperforming stocks but may yield mixed signals for outperforming ones.

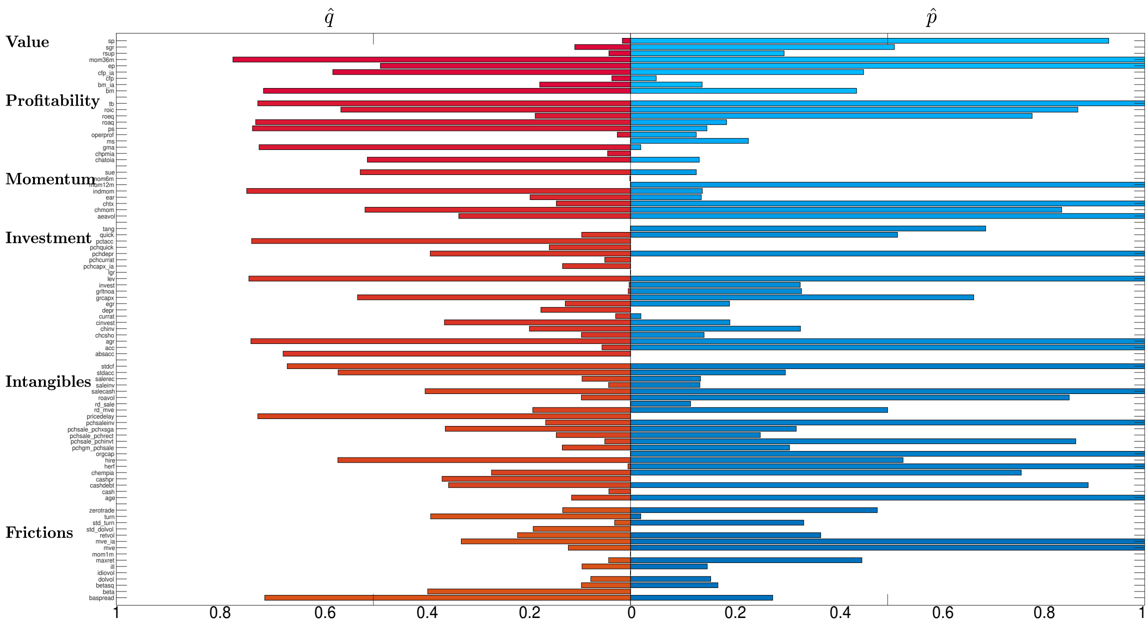
[5]Note that the upper bound constraint is heavily influenced by the size of the underlying cross-section and can vary for characteristics with different data availabilities.

Figure 3.1: Average standardized powers of each factor for univariate versus multivariate optimization. Panel A displays the optimal standardized powers that maximize the univariate Sharpe ratio and Panel B displays the optimal standardized powers that maximize squared Sharpe ratio for all factors. Blue-shaded bars show the standardized powers for the long side and red-shaded bars show the standardized powers for the short side. Factors are sorted into six groups based on their economic rationale. The sample period is from March 1980 to December 2021.

A. Powers optimized individually



B. Powers optimized jointly



Conversely, the long side becomes more significant than the short side when characteristics are considered jointly (Panel B). This is evident as the powers for many individual factors reach their maximum values, even for anomalies typically driven by the short side, such as volatility. Intuitively, joint optimization avoids increasing concentration on the short factor legs to control for their high covariance. This insight reveals that the long side of factors provides quite complementary information, whereas the short side rather tends to converge on similar signals across different factors. This finding aligns with Blitz, Baltussen, and van Vliet (2020), who show that individual factors exhibit nearly zero correlation on their long sides but positive correlations on their short sides.

3.4.3 Multi-factor portfolios

We proceed by examining the economic implications of multi-factor power sorting. For this empirical exercise, we use the first 20 years of the sample period to construct the corresponding multi-factor strategies and test their performance from January 2000 onward. Our approach builds on the strategies proposed by Kagkadis et al. (2024a), where univariately optimized power portfolios are combined into multi-factor portfolios through averaging. This method can be viewed as a forecast combination technique, where optimal weights are determined individually for each factor and then pooled together to form the multi-factor strategy, effectively guarding against overfitting and model instability (Rapach, Strauss, and Zhou 2010).

Arguably, estimation error in the sample moments, such as factor expected returns and their covariance structure, can propagate to the estimation of the squared Sharpe ratio, indicating potential drawbacks of using a joint optimization approach. However, one potential limitation of the univariate optimization and forecast combination approach is that characteristics with high powers might dominate the multi-factor strategy, especially if they are correlated. As such, our comparative analysis aims to shed light on the debate about joint optimization versus the combination approach in terms of out-of-sample performance.

The first strategy we consider is the weight averaging approach (AVP) which involves aggregating factor portfolio weights by averaging the weights of each stock across different power sorting factor portfolios into a single weight vector. Let $W_t = (w_{t,1}, w_{t,2}, \ldots, w_{t,K})$ be the $(N \times K)$ matrix of weights for K characteristics at time t. The average weight vector $\overline{w}_t \in \mathbb{R}^N$ at time t is then obtained as:

$$\overline{w}_t = \frac{1}{K} W_t S'_K, \tag{3.6}$$

where S_K is a $(1 \times K)$ vector reflecting the economic sign of the underlying power sorting portfolio (-1 or 1), determined in-sample. In other words, we invert the weight sign for characteristics for which the power portfolio has a negative average return in the in-sample period. This is crucial for the multi-factor case, as the maximization of the squared Sharpe ratio does not distinguish if the squared average return is positive or negative. Finally, to ensure a unit sum for the long and short sides, the weights are re-standardized to:

$$w_{t,n}^{AVP} = \begin{cases} -\frac{\overline{w}_{t,n}}{\sum\limits_{\overline{w}_{t,n}<0} |\overline{w}_{t,n}|} & \text{for } \overline{w}_{t,n} < 0 \\ 0 & \text{for } \overline{w}_{t,n} = 0 \\ \frac{\overline{w}_{t,n}}{\sum\limits_{\overline{w}_{t,n}>0} |\overline{w}_{t,n}|} & \text{for } \overline{w}_{t,n} > 0. \end{cases} \tag{3.7}$$

The second multi-factor strategy uses the average weight vector $\bar{w}_t$ as the underlying ranking vector to form a power sorting multi-factor portfolio (PMP). This approach takes into account both the underlying characteristic scores and the values of the characteristic powers to determine each characteristic's contribution to the composite score. Hence, groups of characteristics that exhibit strong concentration on either side can significantly impact the multi-factor strategy unless considered in a multivariate context. Note that both AVP and PMP can be applied under univariate and multivariate frameworks, depending on how W_t is determined. This means using the powers that maximize the individual Sharpe ratios of the various factors or those that maximize the squared Sharpe ratio of the set comprising all factors. Therefore, for each strategy, we consider two ways of building multi-factor portfolios: either by employing univariate (UNI) or multivariate (MULTI) power sorting for deriving power portfolio weights.[6]

Table 3.1: Multi-factor portfolios. The table shows different variants of multi-factor power sorting strategies. AVP-UNI: Multi-factor portfolio based on the average portfolio weight from univariate power sorting. AVP-MULTI: Multi-factor portfolio based on the average portfolio weight from multivariate power sorting. PMP-UNI: Power portfolio based on the rank implied by portfolio weights from univariate power sorting. PMP-MULTI: Power portfolio based on the rank implied by portfolio weights from multivariate power sorting. Return: Average monthly return, Standard deviation: Monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, VaR: Value-at-risk at a level of 95%, CVaR: conditional value-at risk at a level of 95%, t-stat: t-statistic on H_0: Return=0, Hit rate: Percentage frequency of positive returns, Turnover: Average monthly turnover (bounded by 200%), # of effective names long: Number of effective names for the long leg, # of effective names short: Number of effective names for the short leg. The sample period covers January 2000 to December 2021.

	AVP-UNI	AVP-MULTI	PMP-UNI	PMP-MULTI
Return (%)	0.89	0.62	0.99	0.82
Standard deviation (%)	5.33	3.47	4.51	2.55
Sharpe ratio	0.58	0.62	0.76	1.12
VaR 95%	−7.35	−4.52	−5.29	−2.69
CVaR 95%	−13.00	−9.02	−8.46	−4.10
t-stat	2.74	2.93	3.59	5.25
Maximum drawdown (%)	−39.48	−25.82	−30.42	−18.89
Hit rate (%)	60.67	60.30	58.65	65.04
Turnover (%)	30.82	38.12	28.16	35.91
# of effective names long	554.06	433.22	720.73	691.04
# of effective names short	437.38	633.01	42.60	50.11

Table 3.1 presents the portfolio evaluation measures for the different multi-factor portfolios in the out-of-sample period, comparing the utilization of univariate power sorting to the multivariate approach. AVP results in similar performance across the two approaches as it mixes factor exposures without an explicit objective. Univariate power sorting achieves higher average return at the expense of higher volatility and drawdown risk. As a result, the multivariate approach delivers higher Sharpe ratio (0.62 versus 0.58) although the difference is economically insignificant. However, once the multi-factor strat-

[6] Hereafter, the terms "multivariate power sorting" and "multi-factor power sorting" are used interchangeably.

egy is constructed with an explicit objective, the benefits of multivariate power sorting become evident. Shifting our focus to PMP, the difference in risk-adjusted performance between multivariate and univariate power sorting becomes way more pronounced with multivariate power sorting delivering annualized Sharpe ratio of 1.12 versus 0.76 for the univariate power sorting. More importantly, the various tail risk measures for PMP-MULTI are about half of those for PMP-UNI. Specifically, PMP-MULTI exhibits an expected shortfall (at a CVar level of 95%) of -4.10% with maximum drawdown of -18.89% compared to -8.46% and -30.42% for PMP-UNI. This result stems from not overloading on factor themes with high univariate powers and highlights the importance of accounting for the covariance structure of different factors from a risk management perspective. Finally, it is worth noting that the more balanced factor stance that arises from multivariate power sorting naturally leads to less constant bets and thus slightly higher turnover compared to the univariate approach.

3.5 Power Sorting-driven Factor Model

In this section, we examine the implications of multivariate power sorting for constructing characteristic-based factor models. We begin by explaining our iterative factor selection approach, followed by the empirical application and the selected factors comprising the model. Finally, we evaluate the performance of the proposed model and benchmark it against two well-known factor models: the Fama-French 5-factor model augmented with momentum (Fama and French 2018) and the Q-theory-based model augmented with expected growth (Hou et al. 2021).

3.5.1 Iterative Factor Selection

The objective is to identify a parsimonious factor representation that explains asset returns. Specifically, our aim is to discover a set of power sorting factor portfolios capable of spanning the tangency portfolio. Similar to the previous exercise, we use the first 20 years of the sample period to construct our factor model and test its performance in the second half. Beginning with the market portfolio, we iteratively incorporate new power sorting factor portfolios from a pool of candidate predictors into an expanding factor model, estimating the squared Sharpe ratio at each step. The power sorting portfolio that maximally increases the squared Sharpe ratio is added to the factor model. We repeat this process with the augmented factor model until the incremental increase in squared Sharpe ratio between nested and augmented models becomes insignificant.

The innovative aspect of our approach lies in the recursive estimation of the powers at each iteration, allowing us to capture the interaction effects between the existing factor set and the new candidate predictor. In other words, the recursive power tuning ensures that the power sorting factors are constructed such that the increase in the squared Sharpe ratio (or the marginal contribution of the candidate predictor in the factor model) is maximized. Practically, this translates to the candidate power sorting factor portfolio achieving a high alpha with low idiosyncratic variance in the spanning regression on the existing factor set (Fama and French 2018). Finally, it is worth noting that our procedure constitutes a "greedy" approach in pursuit of sparsity, as we consistently evaluate the marginal contribution of adding one additional factor at each step, excluding the possibility of subsequent factors significantly altering the squared Sharpe ratio. Formally, the selection strategy can be stated as follows:

Step 1. Let l be the number of iterations and τ the number of in-sample periods. Set $l := 0$ and begin with a single factor CAPM for the period January 1980 to December 2000.

$$r_{n,t} = \alpha_n + \beta_n r_t^m + \epsilon_{n,t} \quad n = 1, \ldots, N, \quad t = 1, \ldots, \tau, \tag{3.8}$$

where $r^m = (r_1^m, \ldots, r_\tau^m)$ denotes the vector of market excess return in the in-sample period.

Step 2. Test $K - l$ different augmented factor models by adding each factor to the existing set and re-estimating the parameters to maximize the model squared Sharpe ratio. Specifically, let $f_k := (r^m, r_k^{PS})$ for $k = 1, \ldots, K - l$ be a $(T \times (l + 2))$ matrix of factor portfolio returns including the market and the k^{th} power sorting factor portfolio. For each candidate factor we estimate

$$Sh_k^2 = \max_{\hat{p}_k, \hat{q}_k} \bar{f}_k \hat{\Omega}_k^{-1} \bar{f}_k' \qquad \text{for } k = 1, \ldots, K - l, \tag{3.9}$$

where $\bar{f}_k$ is the vector of average returns and $\hat{\Omega} = (f_k - \bar{f}_k)'(f_k - \bar{f}_k)/(\tau - 1)$ is the in-sample covariance matrix of the factor set including the market and the k^{th} factor.

Step 3. Sort the candidate factor models by their squared Sharpe ratio and select the model with the highest value.

Step 4. Test whether the difference in squared Sharpe ratio between the selected model and the nested model from the previous iteration is statistically significant by computing the p-value as in Barillas et al. (2020).

Step 5. Stop if the p-value is above the desired significance level (i.e., 10%). Otherwise, update the existing set with the selected factor model, increment l by 1, and proceed to Step 2.

3.5.2 Selected Factors

In this section, we report the results of the iterative power sorting factor selection. Table 3.2 displays the selected factor at each iteration, alongside the underlying factor model's squared Sharpe ratio, the difference in squared Sharpe ratio from the previous iteration, and the associated p-value. Figure A.13 in the Appendix also provides individual squared Sharpe ratio results for all characteristics at each iteration. Naturally, the inclusion of an additional factor increases the squared Sharpe ratio of the given model, although the marginal contribution for later iterations becomes smaller, leading to less statistically significant increases. Depending on the chosen significance level, the iterative algorithm will terminate after four to six iterations. For example, at a 10% significance level, the improvement in the squared Sharpe ratio ceases to be statistically significant after the inclusion of the sixth factor. When the significance level is reduced to 5% or 1%, the number of factors is reduced to five and four, respectively. Therefore, our statistical inferences regarding the required number of factors in asset pricing models align with the general academic consensus (Barillas et al. 2020; Fama and French 2015; Hou et al. 2021; Stambaugh and Yuan 2017).

The first iteration reveals that asset growth (*agr*; Cooper, Gulen, and Schill 2008) is the most important individual contributor. Such investment factors are a key component in many well-known characteristic models, such as the five-factor model by Fama and French (2015) and the q-models by Hou, Xue, and Zhang (2015) and Hou et al. (2021), both of which provide strong, though different, theoretical foundations. Notably, the investment factor category has the highest marginal importance as a whole. As shown in the first subfigure of Figure A.13, the second (growth in long-term net operating assets,

Table 3.2: Iterative Factor Selection. The table displays the factor that led to the highest increase in the squared Sharpe ratio between the nested and the augmented models at each iteration. The first column shows the squared Sharpe ratio of the model incorporating the additional factor. The second column shows the difference in squared Sharpe ratio between the nested and the augmented models, and the third column shows the associated p-value as per Barillas et al. (2020). The sample period for the estimation of the squared Sharpe ratio is from March 1980 to December 2000.

	SR2	SR2dif	p-value
Market	0.109		
Asset Growth	0.637	0.528	0.000
Volatility	1.026	0.389	0.000
Earnings Announcement Return	1.512	0.486	0.001
Unexpected Quarterly Earnings	1.750	0.237	0.041
Market Capitalization	2.004	0.254	0.079
Max Daily Return Last Month	2.162	0.157	0.208

grltnow), third (annual percentage change in total liabilities, lgr), and fifth best contributor (annual change in PPE scaled by total assets, invest) also belong to the investment factor category, corroborating the selection of asset growth in the first iteration. To foster intuition on how the factor is constructed, Figure 3.2 illustrates the standardized powers for each factor in the factor model at each iteration. With reference to Figure 3.2, the predictability of asset growth stems from the short leg, which includes high asset growth stocks. This is indicated by the high standardized power for the short side and the standardized power close to zero for the long side.

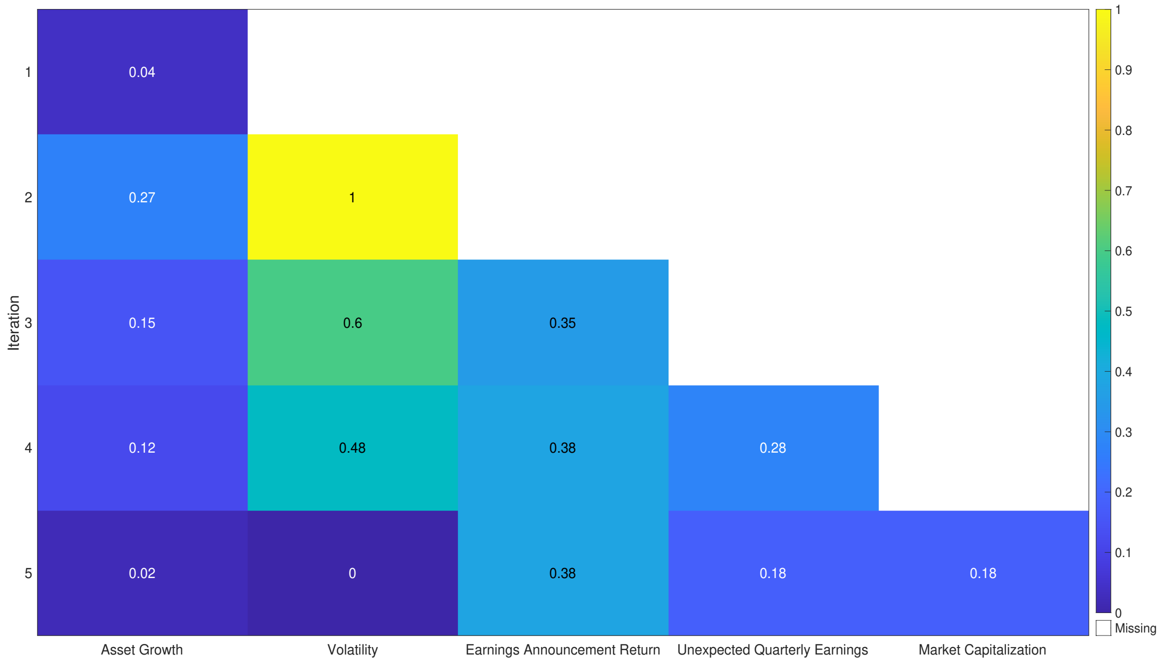
In the second iteration, past-month volatility (*retvol*) from Ang et al. (2006) emerges as the most significant contributor. Although the volatility factor is commonly identified as an important characteristic in prior empirical studies (DeMiguel et al. 2020; Green, Hand, and Zhang 2017) and integrated into risk models used by quantitative asset managers, such as Axioma and BARRA, it is not included in academic factor models given its defiance of a risk-based explanation, challenging its theoretical justification. Specifically, the low-volatility anomaly, which shows that high-beta stocks often underperform relative to low-beta stocks, is at odds with the CAPM's prediction that higher beta should be associated with higher returns. Referring to Figure 3.2, it is also worth noting that for two short-side-dominant factors, such as asset growth and past-month volatility, which are both driven by the short leg including high asset growth/volatility stocks, not both will fully engage on the short side. Specifically, in the case of volatility, both the short and long sides are important, while the significance of high asset growth diminishes compared to the previous iteration as the standardized power for the short leg drops from 0.96 to 0.52. This result highlights the importance of recursive estimation of the powers at each iteration in capturing the dynamics between the different factors in the model. Additionally, looking at the second subfigure of Figure A.13, volatility slightly surpasses momentum in terms of increasing the squared Sharpe ratio. The literature has implicitly or explicitly documented the non-trivial interplay between volatility and momentum without providing a clear direction (e.g., Grinblatt and Moskowitz 2004; Daniel and Moskowitz 2016; Kim, Tse, and Wald 2016; Fan et al. 2022). Consequently, linear models without interaction effects may favor one feature over the other, suggesting that the selection of volatility may displace traditional momentum in the model.

Nonetheless, in the third and fourth iterations, alternative factors from the momentum category, specifically earnings announcement return (*ear*; Kishore et al. 2008) and unexpected quarterly earnings (*sue*; Rendleman, Jones, and Latane 1982), are identified. Earnings announcement return refers to the excess return of a firm around the earnings announcement date compared to a benchmark portfolio of stocks with similar risk profile. Unexpected quarterly earnings measure the difference between realized earnings and investors' expected earnings, derived either from a time series model of earnings or analyst forecasts, and standardized by a measure of earnings uncertainty. They both fall under the wider definition of momentum as they both lead to significant drift in future returns similar to conventional momentum. Novy-Marx (2015a) argues that past stock performance is subsumed by earnings surprise measures, and that price momentum is merely a weak expression of earnings momentum, reflecting the tendency for stocks with strong recent earnings announcements to outperform those with weak earnings announcements. Furthermore, earnings surprise measures include a profitability component and covary with certain profitability variations, such as return-on-equity (Novy-Marx 2015b). This explains the similar contribution of *sue* and *roe* to the model's squared Sharpe ratio in the fourth iteration. Finally, as highlighted by Kishore et al. (2008), earnings announcement returns and standardized unexpected earnings strategies are largely independent, suggesting that investors underreact to earnings and other information in earnings announcements, which helps explain the inclusion of both characteristics in our factor model. Similar to *retvol*, *ear* and *sue* are often used by practitioners but rarely make it into academic models. Figure 3.2 illustrates that both *ear* and *sue* provide valuable information on their long sides, while the short side of *ear* remains inactive across all iterations.

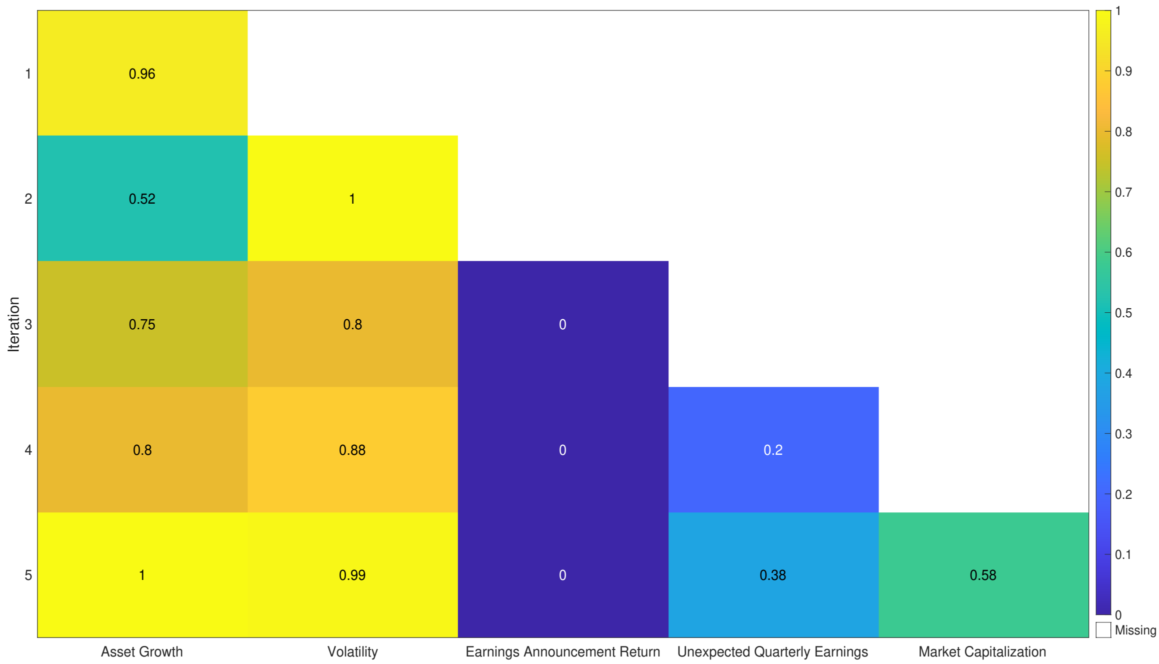
The fifth iteration identifies size as measured by market capitalization (*mve*; Banz 1981) as the final feature that makes a statistically significant contribution in terms of squared Sharpe ratio. The size factor has remained a foundational part of many academic factor models, as it is known to add explanatory power to the cross-section of stock returns (Fama and French 1993, 2015, 2018; Stambaugh and Yuan 2017). It is also important to note that the marginal benefits of size may be due to its conditional efficacy for other

Figure 3.2: Standardized powers for each factor in the power sorting factor model per iteration. Panel A displays the standardized powers for the long legs and Panel B displays the standardized powers for the short legs. The sample period for the estimation of the powers is March 1980 to December 2000.

A. Standardized powers for the long leg



B. Standardized powers for the short leg



factors, as size can interact with other characteristics (e.g., Blitz and Hanauer 2020; Chen, Pelger, and Zhu 2024). Such interactions are hard to capture in a linear setting without being explicitly modeled, which is partly why other risk factors are usually neutral with respect to size. In our setting, we find that the incorporation of size allows the short-side dominant nature of asset growth and volatility to surface, as the standardized power for the short leg reaches the upper bounds, and the power for the long leg goes to zero.

From the sixth iteration, the maximum daily return of the previous month (*maxret*; Bali, Cakici, and Whitelaw 2011) is selected as the seventh factor, although its marginal contribution in terms of squared Sharpe ratio is highly insignificant. Furthermore, Figure A.13 demonstrates that the contribution of *maxret* is very similar to the contributions of other factors across all economic themes, making it difficult to justify the incorporation of this individual characteristic as a single best contributor. Therefore, we stop our procedure at the fifth iteration, resulting in a six-factor model of power sorting factors, henceforth labeled as PS6.

3.5.3 Comparative Analysis

Next, we compare the PS6 factor model to classic academic factor models using both left-hand-side (LHS) and right-hand-side (RHS) approaches. Table 3.3 presents this comparison, contrasting our model with the five-factor q-model of Hou et al. (2021) (Q5) and the six-factor model of Fama and French (2018) (FF6).

Specifically, Panel A presents the squared Sharpe ratios of the various models for the out-of-sample period. It also displays the difference in squared Sharpe ratios between PS6 and the competing models, along with the p-value related to the null hypothesis that the difference is statistically insignificant. Among the three models, PS6 exhibits the highest out-of-sample squared Sharpe ratio of 0.376, which is more than twice that of Q5 and 2.7 times that of FF6. The difference is statistically significant, barely crossing the 5% threshold for Q5, while being highly significant for FF6.

Although the out-of-sample Sharpe ratio achieved is lower than that achieved in-sample, largely due to the decay in anomaly performance over later periods (McLean and Pontiff 2016; Chen, Lopez-Lira, and Zimmermann 2022), our iterative power sorting procedure successfully selects and constructs factors that meet the underlying objectives,

Table 3.3: Overall Performance of Factor Models.Panel A displays the squared Sharpe ratio for PS6, Q5, and FF6 models, together with the squared Sharpe ratio test of Barillas et al. 2020. Panels B to F display the average absolute alphas and the number of significant alphas under a t-stat threshold of 1.96 and 3 for different universes of test assets.(January 2000 - December 2021)

	PS6	Q5	FF6
A. Difference in squared Sharpe ratio			
SR2	0.376	0.180	0.139
SR2dif		0.196	0.238
p-value		0.053	0.003
B. 85 Equal-weighted decile sorting portfolios			
$Avg(\lvert\alpha\rvert)$	0.228	0.223	0.240
# $\lvert t\rvert > 1.96$	12	10	21
# $\lvert t\rvert > 3$	2	3	9
C. 85 Value-weighted decile sorting portfolios			
$Avg(\lvert\alpha\rvert)$	0.228	0.174	0.188
# $\lvert t\rvert > 1.96$	9	4	11
# $\lvert t\rvert > 3$	2	0	1
D. 85 Equal-weighted power sorting portfolios			
$Avg(\lvert\alpha\rvert)$	0.445	0.533	0.502
# $\lvert t\rvert > 1.96$	33	43	43
# $\lvert t\rvert > 3$	9	14	16
E. 85 Value-weighted power sorting portfolios			
$Avg(\lvert\alpha\rvert)$	0.303	0.388	0.362
# $\lvert t\rvert > 1.96$	12	20	20
# $\lvert t\rvert > 3$	2	4	4
F. 199 Factors from Chen and Zimmermann (2022)			
$Avg(\lvert\alpha\rvert)$	0.317	0.366	0.372
# $\lvert t\rvert > 1.96$	57	71	83
# $\lvert t\rvert > 3$	31	26	46

delivering a significantly higher squared Sharpe ratio compared to competing models. Therefore, it emerges as the best performing model in the absence of test assets. However, relying solely on the squared Sharpe ratio may not suffice to declare a model as superior, emphasizing the necessity of using a diverse set of testing assets to draw reliable conclusions.

For this reason, Panels B to F display the average absolute alpha obtained by regressing each factor on each of the three models, along with the count of statistically significant alphas using conventional thresholds of 1.96 and a more conservative threshold of 3.00 to address data-mining concerns (Harvey, Liu, and Zhu 2016). Specifically, in Panels B to E, we construct portfolios using combinations of value adjustment or no value adjustment with power sorting or decile sorting, resulting in four different variations: power sorting without value adjustment, power sorting with value adjustment, decile sorting without value adjustment, and decile sorting with value adjustment.[7] These portfolios are used as test assets. Panel F serves as an external validation sample, comprising 199 factors from Chen and Zimmermann (2022) for which we have complete time-series information for the period January 2000 to December 2021.

Across the three models, FF6 performs the worst, showing higher average absolute alphas and leaving more anomalies unexplained across different groups of test assets. Q5 performs well, closely competing with PS6. Specifically, our results indicate that Q5 prices decile-sorting portfolios more effectively than PS6, aligning better with factors under the conventional definition. Naturally, PS6 excels in explaining power sorting portfolios, capturing inherent non-linearities in various characteristics more effectively.

When examining the factor dataset from Chen and Zimmermann (2022), which offers a fair comparison without potential biases introduced in our factor construction process, results are mixed. PS6 exhibits smaller average absolute alphas overall compared to

[7]For power sorting, value adjustment is implemented with $h = 0.5$, while no value adjustment is implemented with $h = 0.5$. For decile sorting, when no value adjustment is applied, deciles are equally weighted; when value adjustment is used, we employ a "capped value-weighting" scheme following Jensen, Kelly, and Pedersen (2023a).

Q5 and leaves fewer anomalies unexplained under the conventional threshold, but more under the stricter threshold. Hence, although based on the RHS approach PS6 is the clear winner, conclusions about relative performance based on the LHS approach between PS6 and Q5 are inconclusive and hinge on the selection of test assets. Overall, our findings affirm the empirical observations of Hou et al. (2021) regarding the sensitivity of relative performance inferences to the choice of testing assets.

3.6 Conclusion

In this paper, we extend the power sorting methodology to a multivariate framework by maximizing the squared Sharpe ratio of the underlying set of factors instead of focusing solely on individual power portfolio Sharpe ratios. Our contributions are twofold.

First, we demonstrate that when considering multiple characteristics jointly, the focus shifts from the short side to the long side of the power sorting portfolios. This insight underscores the relevance of long-side factors, making characteristics information particularly useful for long-only investors. More importantly, the tangible benefits of the multivariate approach become evident in constructing multi-factor strategies. We show that risk-adjusted performance significantly improves and risk is almost halved when power parameters are estimated jointly rather than individually. These risk benefits arise from more diversified factor exposures, which mitigate the risk of experiencing a crash due to loading heavily on a specific factor theme.

Second, by integrating power sorting with an iterative factor selection approach, we propose a structured framework for constructing factor models given a set of characteristics. Our approach thus strikes a balance between observed and latent factor models, as the factor portfolios remain tradable but dynamically change in structure based on the underlying characteristic set. Using our procedure, we identify a six-factor model that outperforms classic asset pricing models in terms of squared Sharpe ratio. Asset pricing tests further support the validity of our model, although relative performance inferences become test-asset specific.

Concluding remarks

In this dissertation, we have explored innovative approaches to factor investing by addressing key challenges in empirical asset pricing literature. Our research has underscored the significance of firm characteristics in predicting stock returns and introduced novel methodologies to enhance the effectiveness of factor portfolios as both a research and investment tool.

First, we examined the relevance of firm characteristics for factor timing by extending stock return predictability to the portfolio level and proposing a new framework for managing panel data. This approach has notable implications for applying machine learning methods in asset pricing and highlights the value of observable characteristics in explaining the dynamics of factor portfolios.

Second, we developed the power sorting procedure—a data-oriented method that addresses the limitations of conventional quantile approaches. This method represents an effective compromise between traditional portfolio sorts and advanced machine learning techniques, leading to refined factor versions and the revival of previously dismissed characteristics. Importantly, the method offers practical advantages by remaining simple and easily reproducible for fellow researchers.

Third, we extended the power sorting methodology to a multivariate setting, revealing how considering various characteristics jointly can uncover interaction effects and shift importance towards the long side of factors. The construction of a six-factor model that spans the tangent portfolio illustrates the potential of the multivariate approach in identifying the most crucial characteristics and understanding their interactions.

Overall, the findings of this dissertation deepen our understanding of factor investing by refining existing methods and providing new perspectives on portfolio construction. The introduction of power sorting and the exploration of multivariate extensions offer researchers and practitioners advanced tools to enhance factor portfolio strategies. These methodologies not only improve statistical and economic performance but also provide a more nuanced understanding of the drivers behind factor premia.

We hope that this work will support and inspire future research in factor investing, contributing to the ongoing development of more effective and insightful asset pricing tools and methodologies.

Appendix A

Appendix

A.1 List of characteristics

Table A.1: Listing of firm characteristics used in the study, including the source and the exact definition.

Acronym	Author(s)	Journal	Definition	Group
absacc	Bandyopadhyay, Huang, & Wirjanto	2010, WP	Absolute value of acc.	Investment
acc	Sloan	1996, TAR	Annual income before extraordinary items (ib) minus operating cash flows (oancf) divided by average total assets (at); if oancf is missing then set to change in act – change in che – change in lct + change in dlc + change in txp–dp.	Investment
aeavol	Lerman, Livnat, and Mendenhall	2008, WP	Average daily trading volume (vol) for 3 days around earnings announcement minus average daily volume for 1-month ending 2 weeks before earnings announcement divided by 1-month average daily volume. Earnings announcement day from Compustat quarterly (rdq)	Momentum
age	Jiang, Lee, & Zhang	2005, RAS	Number of years since first Compustat coverage.	Intangibles
agr	Cooper, Gulen & Schill	2008, JF	Annual percentage change in total assets (at).	Investment
baspread	Amihud & Mendelson	1989, JF	Monthly average of daily bid-ask spread divided by average of daily spread.	Frictions
beta	Fama & MacBeth	1973, JPE	Estimated market beta from weekly returns and equal weighted market returns for 3 years ending month t-1 with at least 52 weeks of returns.	Frictions
betasq	Fama & MacBeth	1973, JPE	Market beta squared.	Frictions
bm	Rosenberg, Reid, & Lanstein	1985, JPM	Book value of equity (ceq) divided by fiscal year end market capitalization.	Value
bm_ia	Asness, Porter & Stevens	2000, WP	Industry adjusted book-to-market ratio.	Value
cash	Palazzo	2012, JFE	Cash and cash equivalents divided by average total assets	Intangibles
cashdebt	Ou & Penman	1989, JAE	Earnings before depreciation and extraordinary items (ib+dp) divided by avg. total liabilities (lt).	Intangibles
cashpr	Chandrashekar & Rao	2009, WP	Fiscal year end market capitalization plus long-term debt (dltt) minus total assets (at) divided by cash and equivalents (che).	Intangibles

Acronym	Author(s)	Journal	Definition	Group
cfp	Desai, Rajgopal & Venkatachalam	2004, TAR	Operating cash flows divided by fiscal year end market capitalization.	Value
cfp_ia	Asness, Porter & Stevens	2000, WP	Industry adjusted cfp.	Value
chatoia	Soliman	2008, TAR	2-digit SIC fiscal year mean adjusted change in sales (sale) divided by average total assets (at).	Profitability
chcsho	Pontiff & Woodgate	2008, JF	Annual percentage change in shares outstanding (csho).	Investment
chempia	Asness, Porter & Stevens	1994, WP	Industry-adjusted change in number of employees.	Intangibles
chinv	Thomas & Zhang	2002, RAS	Change in inventory (inv) scaled by average total assets (at).	Investment
chmom	Gettleman & Marks	2006, WP	Cumulative returns from months t-6 to t-1 minus months t-12 to t-7.	Momentum
chpmia	Soliman	2008, TAR	2-digit SIC fiscal year mean adjusted change in income before extraordinary items (ib) divided by sales (sale).	Profitability
chtx	Thomas and Zhang	2011, JAR	Percent change in total taxes (txtq) from quarter t-4 to t	Momentum
cinvest	Titman, Wei, and Xie	2004, JFQA	Change over one quarter in net PP&E (ppentq) divided by sales (saleq) - average of this variable for prior 3 quarters; if saleq = 0, then scale by 0.01	Investment
currat	Ou & Penman	1989, JAE	Current assets / current liabilities.	Investment
depr	Holthausen & Larcker	1992, JAE	Depreciation over PPE.	Investment
dolvol	Chordia, Subrahmanyam, & Anshuman	2001, JFE	Natural log of trading volume times price per share from month t-2.	Frictions
ear	Kishore et al.	2008, WP	Sum of daily returns in three days around earnings announcement. Earnings announcement from Compustat quarterly file (rdq)	Momentum
egr	Richardson, Sloan, Soliman & Tuna	2005, JAE	Annual percentage change in book value of equity (ceq).	Investment
ep	Basu	1977, JF	Annual income before extraordinary items (ib) divided by end of fiscal year market capitalization.	Value

Acronym	Author(s)	Journal	Definition	Group
gma	Novy-Marx	2013, JFE	Revenues (revt) minus cost of goods sold (cogs) divided by lagged total assets (at).	Profitability
grcapx	Anderson & Garcia-Feijoo	2006, JF	Percentage change in capital expenditures from year t-2 to year t.	Investment
grltnoa	Fairfield, Whisenant & Yohn	2003, TAR	Growth in long term net operating assets.	Investment
herf	Hou & Robinson	2006, JF	2-digit SIC fiscal year sales concentration (sum of squared percentage of sales in industry for each company).	Intangibles
hire	Bazdresch, Belo & Lin	2014, JPE	Percentage change in number of employees (emp).	Intangibles
idiovol	Ali, Hwang, & Trombley	2003, JFE	Standard deviation of residuals of weekly returns on weekly equal weighted market returns for 3 years prior to month end.	Frictions
ill	Amihud	2002, JFM	Average of daily (absolute return/dollar volume).	Frictions
indmom	Moskowitz & Grinblatt	1999, JF	Equal weighted average industry 12-month returns.	Momentum
invest	Chen & Zhang	2010, JF	Annual change in gross property, plant, and equipment (ppegt) + annual change in inventories (invt) all scaled by lagged total assets (at).	Investment
lev	Bhandari	1988, JF	Total liabilities (lt) divided by fiscal year end market capitalization.	Investment
lgr	Richardson, Sloan, Soliman & Tuna	2005, JAE	Annual percentage change in total liabilities (lt).	Investment
maxret	Bali, Cakici & Whitelaw	2011, JFE	Maximum daily return from returns during calendar month t-1.	Frictions
mom12m	Jegadeesh	1990, JF	11-month cumulative returns ending one month before month end.	Momentum
mom1m	Jegadeesh & Titman	1993, JF	1-month cumulative return.	Frictions
mom36m	Jegadeesh & Titman	1993, JF	Cumulative returns from months t-36 to t-13.	Value
mom6m	Jegadeesh & Titman	1993, JF	5-month cumulative returns ending one month before month end.	Momentum
ms	Mohanram	2005, RAS	Sum of 8 indicator variables for fundamental performance	Profitability

Acronym	Author(s)	Journal	Definition	Group
mve	Banz	1981, JFE	Natural log of market capitalization at end of month t-1.	Frictions
mve_ia	Asness, Porter, & Stevens	2000, WP	2-digit SIC industry-adjusted fiscal year end market capitalization.	Frictions
orgcap	Eisfeldt and Papanikolaou	2013, JF	Capitalized SG&A expenses	Intangibles
pchcapx_ia	Abarbanell & Bushee	1998, TAR	2-digit SIC fiscal year mean adjusted percentage change in capital expenditures (capx).	Investment
pchcurrat	Ou & Penman	1989, JAE	Percentage change in currat.	Investment
pchdepr	Holthausen & Larcker	1992, JAE	Percentage change in depreciation.	Investment
pchgm_pchsale	Abarbanell & Bushee	1998, TAR	Percentage change in gross margin (sale-cogs) minus percentage change in sales (sale).	Intangibles
pchquick	Ou & Penman	1989, JAE	Percentage change in quick.	Investment
pchsale_pchinvt	Abarbanell & Bushee	1998, TAR	Annual percentage change in sales (sale) minus annual percentage change in inventory (invt).	Intangibles
pchsale_pchrect	Abarbanell & Bushee	1998, TAR	Annual percentage change in sales (sale) minus annual percentage change in receivables (rect).	Intangibles
pchsale_pchxsga	Abarbanell & Bushee	1998, TAR	Annual percentage change in sales (sale) minus annual percentage change in SG&A (xsga).	Intangibles
pchsaleinv	Ou & Penman	1989, JAE	Percentage change in sales-to-inventory ratio.	Intangibles
pctacc	Hafzalla, Lundholm & Van Winkle	2011, TAR	Same as acc except that the numerator is divided by the absolute value of ib; if ib = 0 then ib set to 0.01 for denominator.	Investment
pricedelay	Hou & Moskowitz	2005, RFS	The proportion of variation in weekly returns for 36 months ending in month t explained by 4 lags of weekly market returns incremental to contemporaneous market return.	Intangibles
ps	Piotroski	2000, JAR	Sum of 9 indicator variables to form fundamental health score.	Profitability
quick	Ou & Penman	1989, JAE	(current assets – inventory) / current liabilities.	Investment
rd_mve	Guo, Lev & Shi	2006, JBFA	R&D expense divided by end of fiscal year market capitalization.	Intangibles

Acronym	Author(s)	Journal	Definition	Group
rd_sale	Guo, Lev & Shi	2006, JBFA	R&D expense divided by sales (xrd/sale).	Intangibles
retvol	Ang et al.	2006, JF	Standard deviation of daily returns from month t-1.	Frictions
roaq	Balakrishnan, Bartov, and Faurel	2010, JAE	Income before extraordinary items (ibq) divided by one quarter lagged total assets (atq)	Profitability
roavol	Francis et al.	2004, TAR	Standard deviation for 16 quarters of income before extraordinary items (ibq) divided by average total assets (atq)	Intangibles
roeq	Hou, Xue, and Zhang	2015 RFS	Earnings before extraordinary items divided by lagged common shareholders equity	Profitability
roic	Brown & Rowe	2007, WP	Annual earnings before interest and taxes (ebit) minus non-operating income (nopi) divided by non-cash enterprise value (ceq+lt–che).	Profitability
rsup	Kama	2009, JBFA	Sales from quarter t minus sales from quarter t-4 (saleq) divided by fiscal-quarter-end market capitalization (cshoq * prccq)	Value
salecash	Ou& Penman	1989, JAE	Annual sales divided by cash and cash equivalents.	Intangibles
saleinv	Ou& Penman	1989, JAE	Annual sales divided by total inventory.	Intangibles
salerec	Ou& Penman	1989, JAE	Annual sales divided by accounts receivable.	Intangibles
sgr	Lakonishok, Shleifer & Vishny	1994, JF	Annual percentage change in sales (sale).	Value
sp	Barbee, Mukherji, & Raines	1996, FAJ	Annual revenue (sale) divided by fiscal year end market capitalization.	Value
std_dolvol	Chordia, Subrahmanyam, & Anshuman	2001, JFE	Monthly standard deviation of daily dollar trading volume.	Frictions
std_turn	Chordia, Subrahmanyam, & Anshuman	2001, JFE	Monthly standard deviation of daily share turnover.	Frictions
stdacc	Bandyopadhyay, Huang, and Wirjanto	2010, WP	Standard deviation for 16 quarters of accruals (acc measured with quarterly Compustat) scaled by sales; if saleq = 0, then scale by 0.01	Intangibles

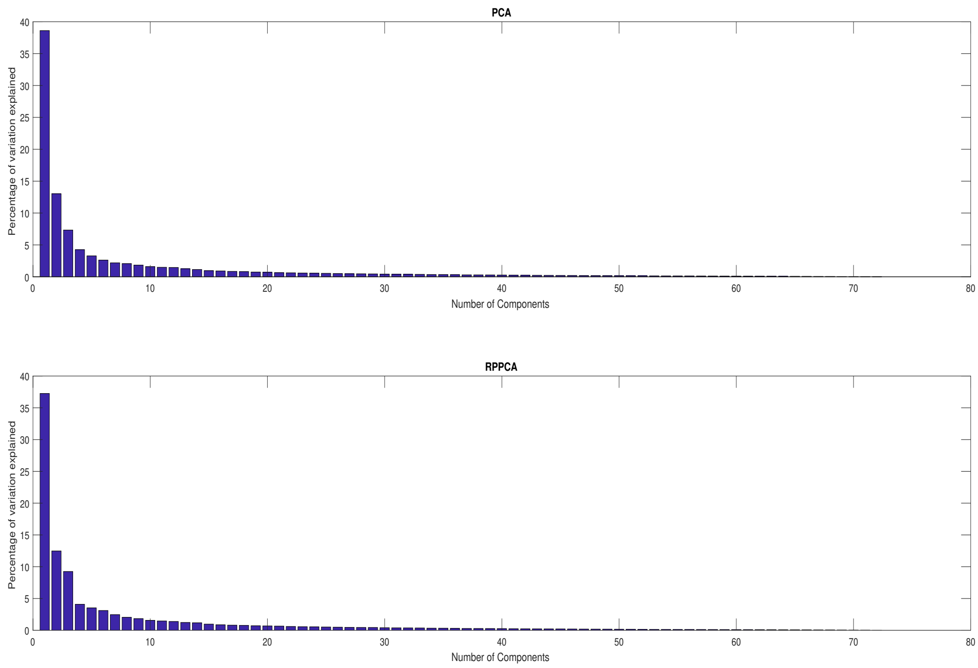
Acronym	Author(s)	Journal	Definition	Group
stdcf	Huang	2009, JEF	Standard deviation for 16 quarters of cash flows divided by sales (saleq); if saleq = 0, then scale by 0.01. Cash flows defined as ibq minus quarterly accrual	Intangibles
sue	Rendleman, Jones, and Latane	1982, JFE	Unexpected quarterly earnings divided by fiscal-quarter-end market cap. Unexpected earnings is I/B/E/S actual earnings minus median forecasted earnings if available, else it is the seasonally differenced quarterly earnings before extraordinary	Momentum
tang	Almeida & Campello	2007, RFS	Cash holdings + 0.715 $\times$ receivables + 0.547 $\times$ inventory + 0.535 $\times$ PPE/total assets.	Investment
tb	Lev & Nissim	2004, TAR	Tax income, calculated from current tax expense divided by maximum federal tax rate, divided by income before extraordinary items.	Profitability
turn	Datar, Naik, & Radcliffe	1998, JFM	Average monthly trading volume for most recent 3 months scaled by number of shares outstanding in current month.	Frictions
zerotrade	Liu	2006, JFE	Turnover weighted number of zero trading days for most recent 1 month.	Frictions

A.2 Supplementary Material to Chapter 1

A.2.1 Factor variation explained by the PC portfolios

One of the key elements of our predictive approach is condensing the information of the factor portfolios using either conventional PCA or the RPPCA of Lettau and Pelger 2020a. While the estimation of the PC portfolios is straightforward, the number of PC portfolios to be retained remains an empirical question. Figure A.1 shows that irrespective of whether we use PCA or RPPCA the variation explained by the principal components is very similarly. In particular, the first component captures around 37% of the total variation, the second component captures around 13%, the third component captures around 8%, while the fourth and fifth components capture around 4%. After the fifth component the decrease in explained variation is quite gradual but the components contribute very little to the total variation. Given the aforementioned pattern and in order to be consistent with Haddad, Kozak, and Santosh 2020, we focus our analysis on five PC portfolios.

Figure A.1: Percentage of the variation explained by each PC of factor portfolio returns under PCA and RPPCA for the sample period January 1970 to December 2019.



A.2.2 Details of the dimension reduction and regularization techniques

For exposition purposes we omit the time dimension when explaining the different statistical methods. In the methodology section of the main manuscript the time subscript t indicates the recursive estimation of the different objects.

Principal Component Analysis (PCA)

The first and most popular dimension reduction method is PCA. The method produces linear combinations of the original data (PCs) while best preserving the covariance structure among the variables. Each PC successively contains as much new information about the observed variables and dimension reduction can be accommodated by focusing on the first few (dominant) PCs, while omitting the rest which are usually noise-related. Let Σ, be the $(N \times N)$ variance-covariance matrix of the $(T \times N)$ factor portfolio return matrix R. Consider the eigendecomposition of Σ:

$$\Sigma = W \Lambda W' = \sum_{i=1}^{N} \lambda_i w_i w_i', \tag{A.1}$$

where W is a $(N \times N)$ matrix whose i^{th} column w_i is the eigenvector of Σ and Λ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues in decreasing order. The i^{th} eigenvector w_i, solves:

$$\begin{aligned} w_1 &= \underset{\|w_1\|=1}{\operatorname{argmax}} \left\{ w_1' \Sigma w_1 \right\}, \\ w_2 &= \underset{\|w_2\|=1}{\operatorname{argmax}} \left\{ w_2' \Sigma w_2 \right\} \text{ s.t. } w_1' \Sigma w_2 = 0, \\ &\vdots \\ w_N &= \underset{\|w_N\|=1}{\operatorname{argmax}} \left\{ w_N' \Sigma w_N \right\} \quad \text{s.t.} \quad w_M' \Sigma w_N = 0 \quad \forall \quad M < N. \end{aligned} \tag{A.2}$$

Practically, the solution in Equation (A.2) is obtained via a singular value decomposition

(SVD) R. The $(T \times N)$ matrix of PCs is then obtained by multiplying the matrix of factor portfolio returns with the eigenvectors, $Z = RW$. Notice that since W is an orthogonal matrix, this is equivalent to regressing the factor portfolio returns on the eigenvectors. PCA is also used to regularize the characteristics of each PC portfolio, H_i. This logic is identical to Principal Component Regression (PCR) where the predictors are transformed to their PCs and the coefficients of low variance PCs are set to zero.

Risk Premium PCA (RPPCA)

In general, PCA extracts factors that best explain time-series variation in the data. The variance-covariance matrix of factor portfolio returns can also be written as $\Sigma = \frac{1}{T}R'R - \bar{R}\bar{R}'$, where $\bar{R}$ is an $(N \times 1)$ vector of average portfolio returns. Since average returns are subtracted, PCA utilizes information from the second moment while it neglects information from the first moment of the data. Some factors may have weak explanatory power in terms of variance if they only affect a small proportion of assets, but may still be important in an asset pricing context. In this case, conventional PCA is unable to detect the true factors (Onatski 2012). Under an Arbitrage Pricing Theory framework, exposure to systemic risk factors should be able to explain the cross-section of expected asset returns (Ross 1976). As such, latent factors should be able to simultaneously capture time-series variation and explain the cross-section of average returns.

Lettau and Pelger 2020a propose a new estimator by augmenting PCA with a penalty term to account for pricing errors in average returns. RPPCA is a generalization of PCA, regularized by a cross-sectional pricing error and can be implemented by simple eigenvalue decomposition of the variance-covariance matrix of asset returns after a simple transformation:

$$\frac{1}{T}R'R + \gamma\bar{R}\bar{R}'. \tag{A.3}$$

Essentially, the method applies PCA to the variance-covariance matrix with over-weighted means. The resulting PCs jointly minimize the unexplained variation and the cross-sectional pricing error. The choice of the tuning parameter γ determines the relative weight of the cross-sectional pricing error compared to the time-series error. For conventional PCA $\gamma = -1$, while $\gamma = 0$ is equivalent to applying PCA to the second moment matrix. Values of $\gamma > -1$ can lead to the detection of weak factors with high Sharpe ratios. We opt for a constant value of $\gamma = 10$, as it provides a balance between explaining time-series variation and detecting weak factors.[1] The use of RPPCA should help us focus on factor portfolios with high average returns as by definition those will have a higher weight on dominant components.

Again, we apply SVD on $\frac{1}{T}R'R + 10\bar{R}\bar{R}'$ and retain the first five eigenvectors to calculate the PC portfolios $Z_i \in \mathbb{R}^{(T\times 1)}$, $i = 1, \ldots, 5$. Since the purpose of RPPCA is to detect weak factors within asset returns and given that characteristics are standardized due to their difference in scale, it would be insensible to apply it on $H_i \in \mathbb{R}^{(T\times M)}$, $i = 1, \ldots, 5$. Instead, we apply SDV on each $\frac{1}{T}H_i'H_i - \bar{H}_i\bar{H}_i'$, which converges back to conventional PCA.

Partial Least Squares (PLS) One of the limitations of PCA is that it focuses on condensing the covariation within the predictors. However, some of the characteristics may have no predictive power, meaning that PCA-based PCs can contain information that is ultimately useless in the forecasting exercise. In contrast, PLS constructs linear combinations of the characteristics based on their relationship with future returns by directly exploiting the covariance between the two. The method can be used to rotate H_i into linear combinations that best explain Z_i while still being orthogonal to each other.

[1] A value of $\gamma = 10$ is also consistent with what the authors identify as optimal in their empirical exercise.

The vector of weights for the i^{th} PC is estimated recursively by solving:

$$q_1^i = \underset{\|q_1^i\|=1}{\operatorname{argmax}} \left\{ q_1^{i\prime} H_i' Z_i Z_i' H_i q_1^i \right\},$$

$$q_2^i = \underset{\|q_i^2\|=1}{\operatorname{argmax}} \left\{ q_i^{2\prime} H_i' Z_i Z_i' H_i q_i^2 \right\} \text{ s.t. } q_i^{1\prime} H_i' Z_i Z_i' H_i q_i^2 = 0,$$

$$\vdots$$

$$q_N^i = \underset{\|q_i^N\|=1}{\operatorname{argmax}} \left\{ q_i^{N\prime} H_i' Z_i Z_i' H_i q_i^N \right\} \quad \text{s.t.} \quad q_i^{M\prime} H_i' Z_i Z_i' H_i q_i^N = 0 \quad \forall \quad M < N. \tag{A.4}$$

Equation $(A.4)$ highlights the distinction between PLS and PCA. Specifically, by making a comparison between Equation $(A.2)$ and Equation $(A.4)$, it is clear that PCA finds linear combinations that maximize the variance of H_i while PLS finds combinations of weights that maximize the squared covariance between Z_i and H_i, or the product of the variance of the predictors with the squared correlation with the forecasting target. In other words, PLS diverges from the solution that best describes H_i in order to find components that can better predict future returns. Practically, Equation $(A.4)$ can be efficiently solved using the SIMPLS algorithm by De Jong 1993. Again, we calculate the PLS components $X_i = H_i Q_i$ and either retain the first component or apply LASSO on X_i to predict each $\hat{Z}_{t+1,i}$.

LASSO

Another important aspect of our estimation procedure is the use of LASSO to account for overfitting and control for model complexity. LASSO imposes sparsity by selecting a subset of features and setting the remaining coefficients to zero. This is achieved by slightly modifying the OLS objective function to incorporate a penalty for the sum of the absolute value of the coefficients. For instance, at each time t we estimate the $\beta \in \mathbb{R}^{1\times M+1}$ of PC portfolio $Z_i \in \mathbb{R}^{T\times 1}$ on a set of characteristic PCs $X_i \in \mathbb{R}^{T\times M}$ as:

$$\min_{\beta_0,\beta} \sum_{t=1}^{T} (Z_i - \beta_i^0 - \sum_{m=1}^{M} \beta_i^m X_i^m)^2 + \delta \sum_{m=1}^{M} |\beta_i^m|, \tag{A.5}$$

where δ is a hyperparameter that determines the degree of regularization such as:

$$\sum_{m=1}^{M} |\beta_i^m| \leq \delta. \tag{A.6}$$

High values for δ result in solutions that set many coefficients exactly equal to zero, delivering parsimonious models. Using the coordinate decent algorithm by Friedman, Hastie, and Tibshirani 2010, we fit many values of δ simultaneously and pick the one that minimizes the forecasting error in the validation period.

A.2.3 Sources of variation in the PC characteristics

The characteristics of long-short factor portfolios are initially calculated by value-weighting characteristics of stocks within each decile portfolio and then subtracting the value of the bottom decile from the top. At this stage no standardization is applied, meaning that the average characteristic across factor portfolios still preserves its time series trend and the cross-sectional variance for any given characteristic changes over time. The factor portfolio characteristics are then standardized cross-sectionally by subtracting each month the cross-sectional characteristic mean and dividing by the cross-sectional characteristic standard deviation. These standardized characteristics are then transformed into PC portfolio characteristics by being multiplied with $w_{t,i}$, allowing us to focus on the cross-sectional dispersion in the data.

Table A.2 presents a stylized hypothetical example with three characteristics and one PC portfolio for one time period. Each factor portfolio has its own characteristic and two other characteristics. As in our empirical exercise, the characteristics are cross-sectionally standardized to have zero cross-sectional mean and unit cross-sectional variance. In this example, the PC portfolio loads positively on momentum and value and negatively on reversal. Momentum and value characteristics are positively related to returns, while reversal is negatively related. As a result, the PC portfolio has high momentum and value scores and low reversal score, resulting in a high return for that month. Provided that characteristics jointly explain returns and that this relationship holds on average over time, PC portfolio returns will be high when PC momentum and value characteristics are high and PC reversal is low.

Table A.2: Stylized example of estimating PC portfolio characteristics and returns using three anomalies.

	Momentum Portfolio	Value Portfolio	Reversal Portfolio	PC Portfolio
PC Portfolio Weights	0.75	0.5	-0.25	
Momentum Characteristic	1	0	-1	1
Value Characteristic	0	1	-1	0.75
Reversal Characteristic	-1	0	1	-1
Return $t+1$	1.0%	0.8%	-0.2%	1.2%

Given our procedure, it is important to understand the sources of variation in the characteristics of each PC portfolio that ultimately lead to PC portfolio return predictability. Let us start with the diagonal elements in the characteristic matrix in Table A.2 (e.g., the momentum of the momentum portfolio). The diagonal elements will always have the highest scores across rows (1 in our hypothetical example), since factor portfolios have the highest score for their own characteristic by construction. Still, the diagonal elements would remain constant across time only if each characteristic's distribution across the portfolio cross-section remained identical in terms of skewness and kurtosis. However, we observe significant variability in those higher moments for all characteristics over time. In particular, Figures A.2 and A.3 display the cross-sectional skewness and kurtosis of the 72 characteristics across all the factor portfolios over the whole sample period. As it can be seen, there is significant variation in the skewness and kurtosis of the characteristics across time, which suggests that the diagonal elements do change affecting also the characteristic scores of the PC portfolio.

With regards to the off-diagonal elements of the characteristic matrix in Table A.2 (e.g., the momentum characteristic of the value portfolio), those can further change from month to month depending on the characteristics of stocks within each factor portfolio. Provided that different factor portfolios do not contain the exact same stocks, each characteristic can differ significantly across factor portfolios and can vary over time in non-standard ways. For example, it is possible that for a given month the correlation between stock momentum and value is high and hence the standardized momentum score of the two factor portfolios is similar, while in another month the correlation might be low and hence the momentum score of the two factor portfolios will be completely different. To shed light on the time-varying interrelations arising among characteristics, we display in Figures A.4 and A.5 the mean and standard deviation of the monthly cross-sectional correlation across all the different characteristic pairs. Apropos Figure A.4, average cross-sectional correlations across characteristics are fairly low (typically between –0.2 and 0.2), with the exception being market friction and volatility proxies. Figure A.5 further shows that there is significant time variability on those correlations (standard deviation values mostly range from 0.2 to 0.4), with again volatility proxies being the most profound exception, while, as expected, correlations also remain relatively constant for characteristic pairs with similar economic interpretation. Overall, our results suggest that time variation in correlations among stock characteristics is another reason for which the characteristics of the PC portfolios exhibit significant time-series variability.

The above analyses show that, despite the fact that we use the same weighting vector $w_{t,i}$ for all the months of a given in-sample period, the PC characteristics naturally vary across months *even within the same predictive iteration/in-sample period.* Obviously, an additional source of variation in the characteristics of the PCs arises from the recursive estimation of the weighting vector $w_{t,i}$ across predictive iterations. Because of the variability in the covariance structure of factor portfolio returns (and the change in their average returns in the case of RPPCA), the weighting vectors change across time, affecting the characteristic construction process, as the PC characteristics are calculated by multiplying factor portfolio characteristics with $w_{i,t}$. To visualise how much the weighting vectors vary, Figures A.6 and A.7 display heatmaps that correspond to the (absolute value of the) factor loadings of each PC on the different anomalies across time, for PCA and RP-PCA respectively. As expected, there is some time variability in the weights. However, we observe that in most of the cases the loadings (especially the most important ones) remain quite stable throughout the whole sample period.

Overall, the PC characteristics change across months (within the same predictive iteration) because the characteristic distributions across portfolios exhibit time-varying skewness and kurtosis and because the characteristic themselves have time-varying correlations. An additional but less important source of variation stems from the recursive estimation of PC portfolios and consequently weighting vectors.

Figure A.2: Monthly cross-sectional skewness of 72 characteristics across factor portfolios. The sample period is 01/1970-12/2019.

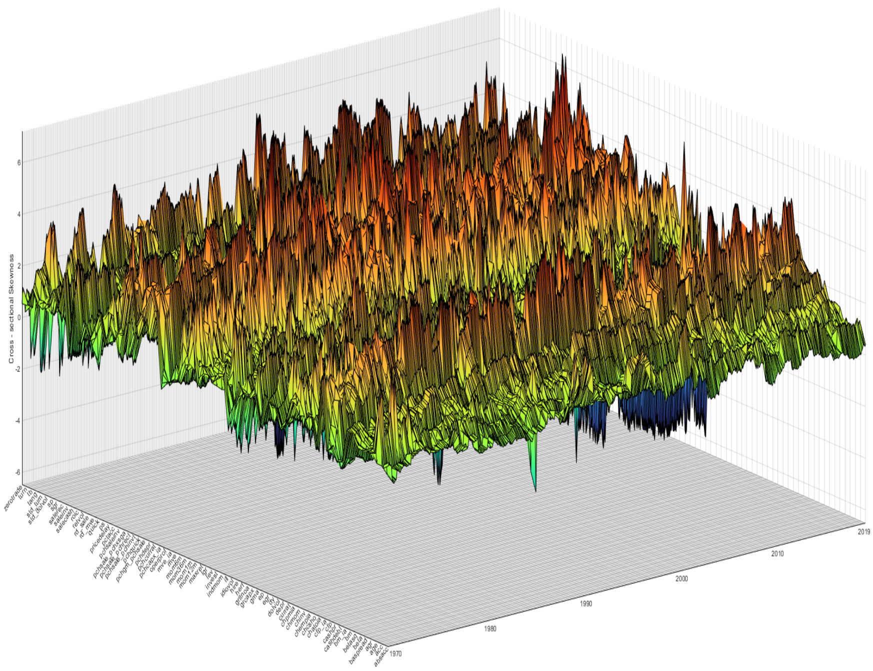


Figure A.3: Monthly cross-sectional kurtosis of 72 characteristics across factor portfolios. The sample period is 01/1970-12/2019.

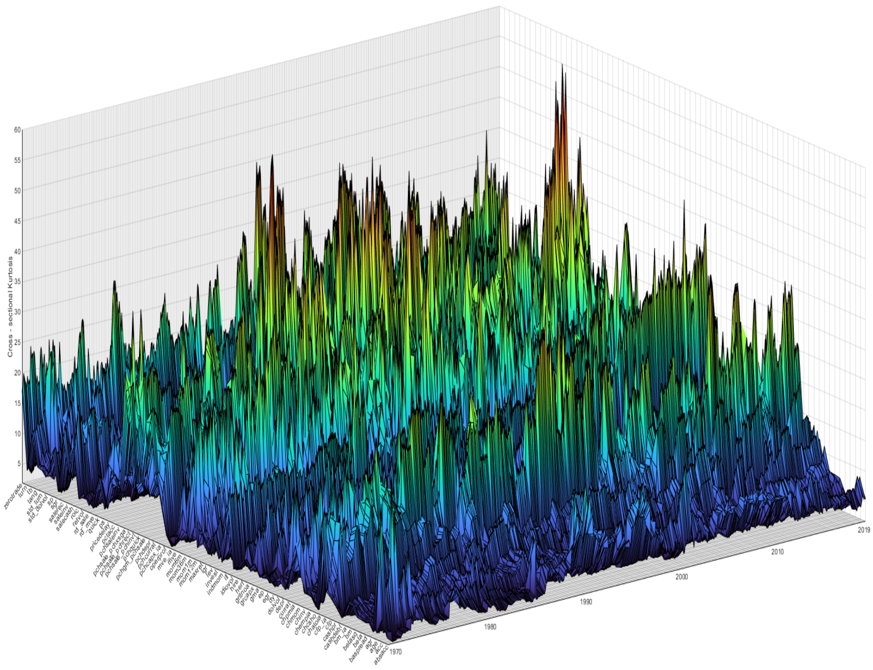


Figure A.4: Average monthly cross-sectional correlations across 72 characteristics. The sample period is 01/1970-12/2019.

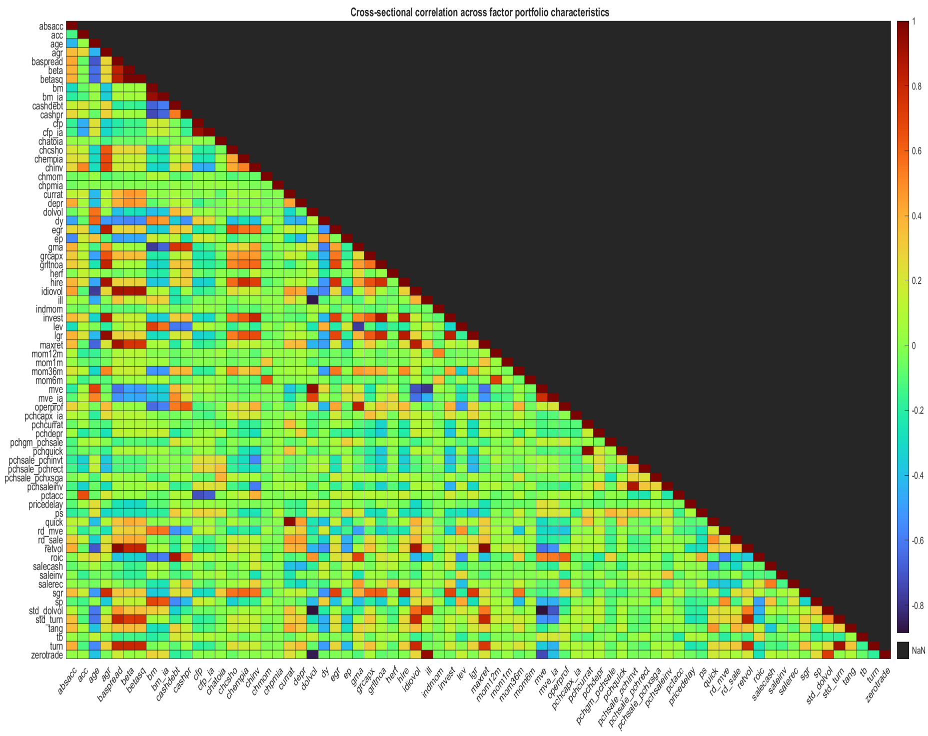


Figure A.5: Standard deviation of monthly cross-sectional correlations across 72 characteristics. The sample period is 01/1970-12/2019.

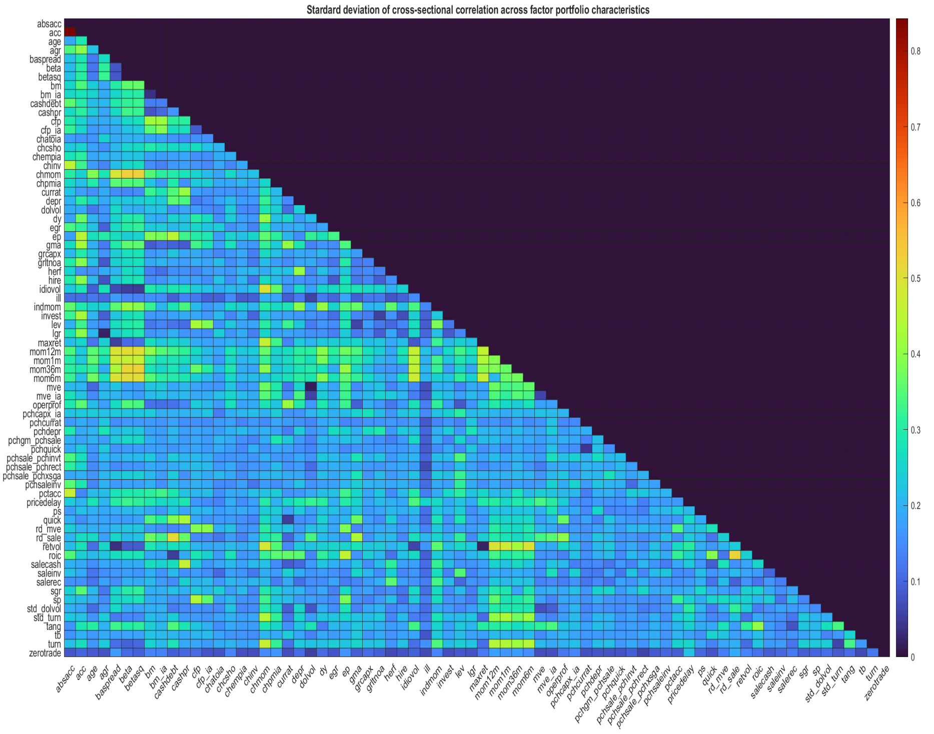


Figure A.6: Recursively estimated PC portfolio weights on 72 factor portfolios for the out-of-sample period 01/1990-12/2019. PC portfolios are constructed using PCA.

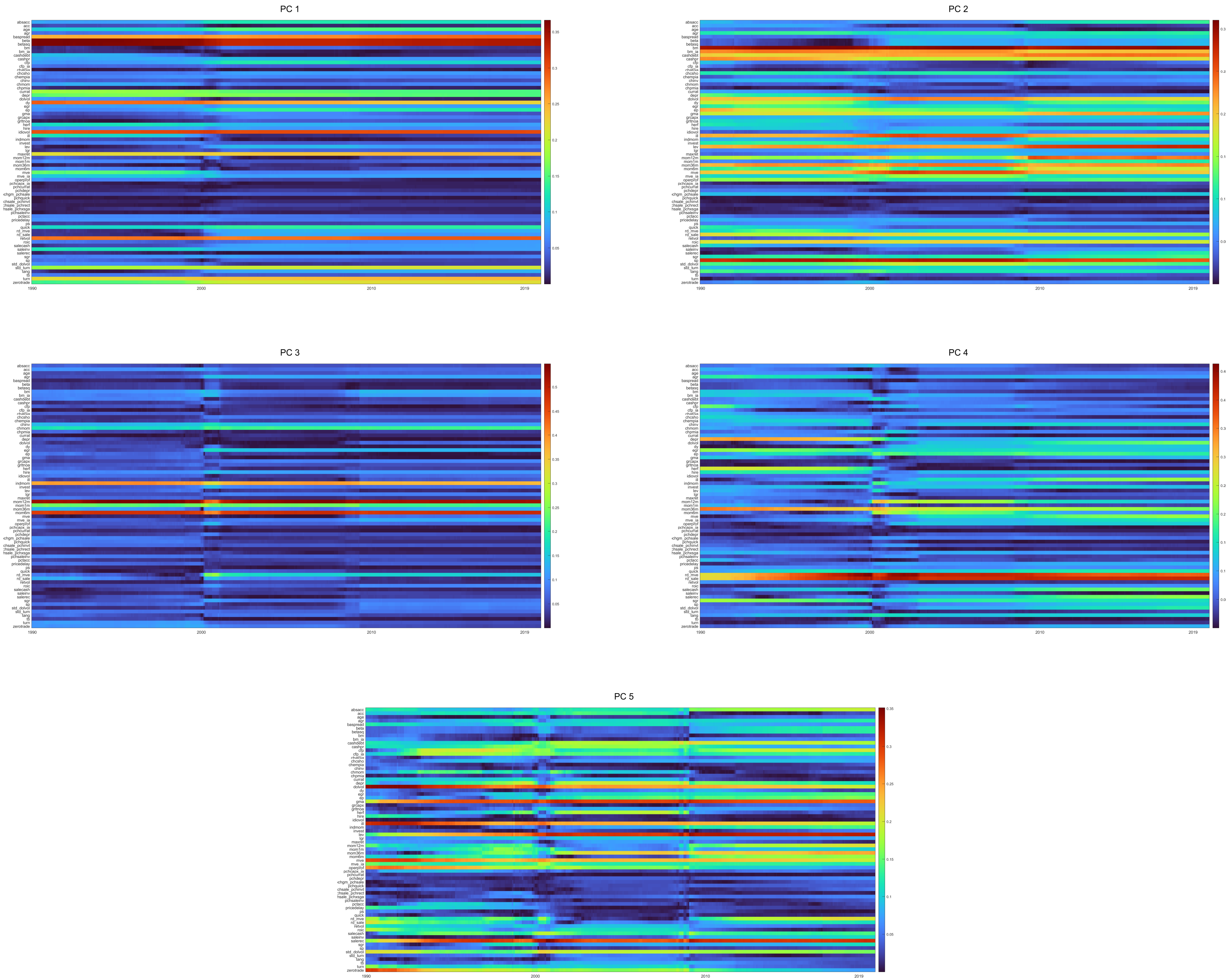
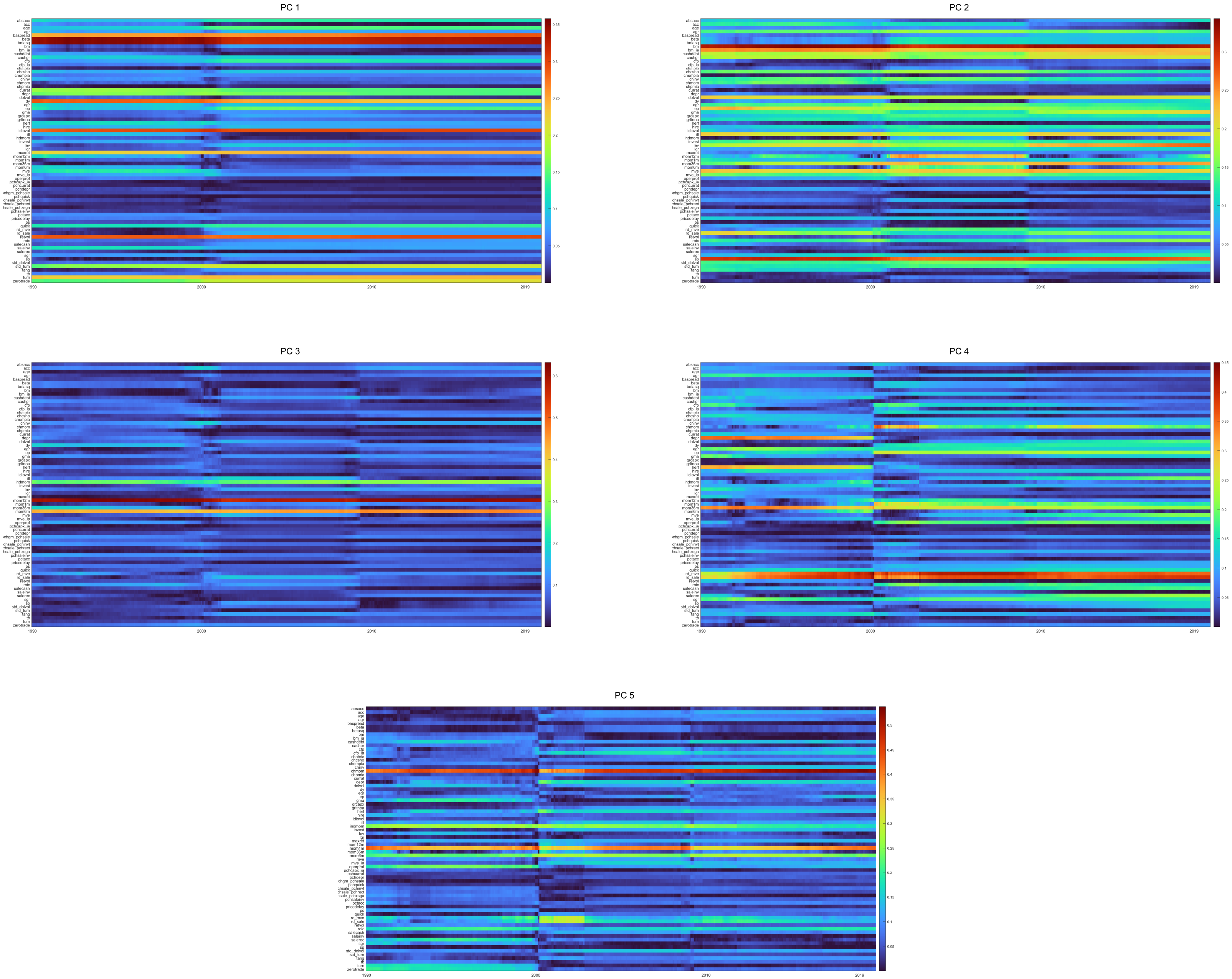


Figure A.7: Recursively estimated PC portfolio weights on 72 factor portfolios for the out-of-sample period 01/1990-12/2019. PC portfolios are constructed using RPPCA.



A.2.4 Details of the predictive models

The main analysis of the paper focuses on four models that incorporate PCA or RPPCA for condensing the variation in factor portfolio returns and PCA or PLS for condensing the variation in the portfolio characteristics. We further consider two cases, one where we retain only a single latent factor from the characteristics and another one where we retain all the characteristic factors but employ LASSO to select the important ones at each forecasting period. In addition, we select a series of alternative benchmark models that rely on predictors such as the past return of the factor portfolios, the issuer-repurchaser spread, market sentiment etc. Finally, we modify our main models in order to investigate what the most important pillars of our successful predictive method are, and also modify the benchmark models in order to examine whether adding additional statistical features to them delivers performance similar to the one provided by the portfolio characteristics. For convenience, Table A.3 presents the list with all the different models used in the study, together with a detailed description of how we handle the left-hand-side and the right-hand-side of the forecasting problem.

A.2.5 Factor statistics

In Table A.4, we report some summary statistics for the 72 factor portfolios that are based on our stock characteristics. Specifically, we report the average monthly return with the respective t-statistic, the monthly volatility, and the monthly Sharpe ratio of each portfolio. As discussed in the main paper, there is high cross-sectional variation in the performance of the portfolios with only 22 of them having an average return significantly different from zero.

Table A.3: Detailed description of all the predictive models used in the study. The table includes the acronym for each model and the description of the modelling approach for each side of the return predictability exercise.

Acronym	**Left-hand-side**	**Right-hand-side**
Panel A: Characteristic-based models (with and without LASSO)		
PCA	PCA	PCA
PCA-PLS	PCA	PLS
RPPCA	RPPCA	PCA
RPPCA-PLS	RPPCA	PLS
Panel B: Baseline Benchmark models		
1mMOM	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its past month return.
12mMOM	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its 12-month average past return.
PCA-BM	PCA	No dimension reduction. Each PC portfolio is predicted by its own book-to-market ratio.
IR Spread	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its own issuer-repurchaser spread.
Sentiment	No dimension reduction.	No dimension reduction. Each anomaly is predicted by the investors sentiment index.
Historical Sample Mean	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its in-sample average return.
Panel C: Alternative Specifications		
Anom-Own characteristic	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its own characteristic spread.
Anom-1 PLS	No dimension reduction.	PLS. One characteristic-based component is retained.
Anom-All characteristics	No dimension reduction.	No dimension reduction. Each anomaly is predicted by the collection of all its characteristics.
Anom-All characteristics LASSO	No dimension reduction.	LASSO is applied on the collection of characteristics.
5 PCs-All characteristics	PCA	No dimension reduction. Each PC portfolio is predicted by the collection of its characteristics.
5 PCs-All characteristics LASSO	PCA	LASSO is applied on the collection of characteristics.
5 PCs-1mMOM	PCA	No dimension reduction. Each PC portfolio is predicted by its past return.
5 PCs-12mMOM	PCA	No dimension reduction. Each PC portfolio is predicted by its 12-month average past return.
5 PCs-5 BMs	PCA	No dimension reduction. The 5 book-to-market ratios of all PC portfolios are used to predict each one individually.
5 PCs-5 BMs LASSO	PCA	LASSO is applied on the collection of the 5 book-to-market ratios.
Anom-Own BM	No dimension reduction.	No dimension reduction. Each anomaly is predicted its own book-to-market ratio.
Anom-All IR Spreads	No dimension reduction.	No dimension reduction. Each anomaly is predicted by the collection of all issuer-repurchaser spreads.
Anom-All IR Spreads LASSO	No dimension reduction	LASSO is applied on the collection of all issuer-repurchaser spreads.
5 PCs-IR Spread	PCA	No dimension reduction. Each PC portfolio is predicted by its own IR Spread.
5 PCs-All IR Spreads	PCA	No dimension reduction. Each PC portfolio is predicted by the collection of PC IR Spreads.
5 PCs-Sentiment	PCA	No dimension reduction. Each PC is predicted by the investor sentiment index.
Sentiment-LASSO	No dimension reduction.	LASSO is applied to the investor sentiment index.

Table A.4: Descriptive statistics of factor portfolios for the sample period January 1970 to December 2019. Average Return: Average monthly return, Standard Deviation: Monthly standard deviation, Sharpe Ratio: Monthly Sharpe ratio, t-statistic: t-statistic for the null hypothesis that the average monthly return is equal to zero.

	Average Return	**Standard Deviation**	**Sharpe Ratio**	**t-statistic**
absacc	−0.111	3.982	−0.028	−0.680
acc	−0.419	3.009	−0.139	−3.410
age	−0.031	3.896	−0.008	−0.196
agr	−0.399	3.143	−0.127	−3.110
baspread	−0.235	6.796	−0.035	−0.846
beta	−0.153	7.854	−0.020	−0.478
betasq	−0.146	7.851	−0.019	−0.456
bm	0.377	4.480	0.084	2.061
bm_ia	0.250	3.642	0.069	1.677
cashdebt	0.101	3.884	0.026	0.635
cashpr	−0.336	3.545	−0.095	−2.323
cfp	0.298	4.181	0.071	1.744
cfp_ia	0.328	2.889	0.114	2.781
chatoia	0.269	2.647	0.102	2.490
chcsho	−0.588	2.901	−0.203	−4.963
chempia	0.065	2.874	0.023	0.553
chinv	−0.537	2.955	−0.182	−4.448
chmom	−0.574	4.628	−0.124	−3.037
chpmia	0.028	3.094	0.009	0.220
currat	−0.052	3.975	−0.013	−0.322
depr	0.058	4.376	0.013	0.322
dolvol	−0.210	3.495	−0.060	−1.471
dy	−0.060	5.823	−0.010	−0.254
egr	−0.429	3.292	−0.130	−3.191
ep	0.613	4.691	0.131	3.198
gma	0.090	3.832	0.023	0.573
grcapx	−0.358	2.941	−0.122	−2.979
grltnoa	−0.270	3.010	−0.090	−2.199
herf	0.053	3.508	0.015	0.371
hire	−0.217	3.135	−0.069	−1.694
idiovol	−0.196	6.923	−0.028	−0.692
ill	0.051	3.688	0.014	0.340
indmom	0.175	4.837	0.036	0.887
invest	−0.395	3.003	−0.132	−3.223
lev	0.098	4.550	0.022	0.529
lgr	−0.189	2.638	−0.072	−1.758
maxret	−0.367	5.778	−0.063	−1.554
mom12m	1.080	6.492	0.166	4.071
mom1m	−0.353	5.089	−0.069	−1.696
mom36m	−0.204	4.891	−0.042	−1.019
mom6m	0.624	5.896	0.106	2.590
mve	−0.161	4.015	−0.040	−0.982

	Average Return	Standard Deviation	Sharpe Ratio	t-statistic
mve_ia	−0.131	3.212	−0.041	−0.997
operprof	0.248	3.001	0.083	2.021
pchcapx_ia	0.066	2.932	0.022	0.549
pchcurrat	−0.202	1.906	−0.106	−2.597
pchdepr	0.158	2.518	0.063	1.536
pchgm_pchsale	0.125	2.723	0.046	1.126
pchquick	−0.063	2.042	−0.031	−0.761
pchsale_pchinvt	0.256	2.513	0.102	2.489
pchsale_pchrect	0.004	2.399	0.001	0.036
pchsale_pchxsga	−0.078	3.005	−0.026	−0.632
pchsaleinv	0.224	2.457	0.091	2.230
pctacc	−0.175	3.127	−0.056	−1.366
pricedelay	−0.077	2.717	−0.028	−0.695
ps	0.250	2.337	0.107	2.614
quick	−0.099	3.777	−0.026	−0.641
rd_mve	0.243	4.748	0.051	1.253
rd_sale	−0.109	4.541	−0.024	−0.585
retvol	−0.406	6.782	−0.060	−1.464
roic	0.283	3.725	0.076	1.859
salecash	0.019	3.327	0.006	0.141
saleinv	0.146	3.040	0.048	1.175
salerec	0.252	3.486	0.072	1.770
sgr	−0.112	3.374	−0.033	−0.809
sp	0.377	4.183	0.090	2.208
std_dolvol	0.175	3.080	0.057	1.391
std_turn	0.090	5.163	0.017	0.425
tang	0.155	3.423	0.045	1.105
tb	0.195	2.852	0.068	1.676
turn	−0.073	5.779	−0.013	−0.308
zerotrade	−0.052	5.454	−0.010	−0.235

A.2.6 Interpreting the PC portfolios

One disadvantage of extracting latent common factors from factor portfolio returns is that the resulting PC portfolios do not have a straightforward economic interpretation. In order to tackle this problem, we regress recursively each PC portfolio return on each of the 72 anomalies and retain the R^2. Next, we estimate the average R^2 across months for each PC portfolio and each anomaly. As an example, we present in Figure A.8 the results of the PCs stemming from PCA. We observe that the first PC portfolio loads more on volatility variables such as beta, idiovol, maxret, retvol and std_turn. The second PC portfolio loads more on value variables such as bm, bm_ia and sp, as well as on profitability and leverage variables such as gma, cashdebt and lev. The third PC portfolio is clearly affected by the momentum characteristics, while the fourth PC portfolio is mainly driven by r&d variables. The fifth portfolio loads more on salerec and to a lesser extent on gma.

The above results stem from a time-series aggregation of the estimated R^2s. Another interesting question that arises is whether the relation between the PCs and the underlying characteristics is stable across time. To this end, we select the most important characteristic for each PC portfolio and plot the R^2 of all of them across time in Figure A.9. In particular, we plot the R^2 values of beta, bm, mom12m, rd_sale, and salerec. Each characteristic dominates a respective PC portfolio and it is evident from Figure A.9 that all the loadings are remarkably stable across time. Overall, we conclude that the PC portfolios extracted with statistical techniques have reasonable economic interpretation and the recursive estimation does not impact negatively this interpretation.

Figure A.8: Average R^2 across month from recursive regressions of each PC portfolio on each anomaly. The PC portfolios are constructed using PCA.

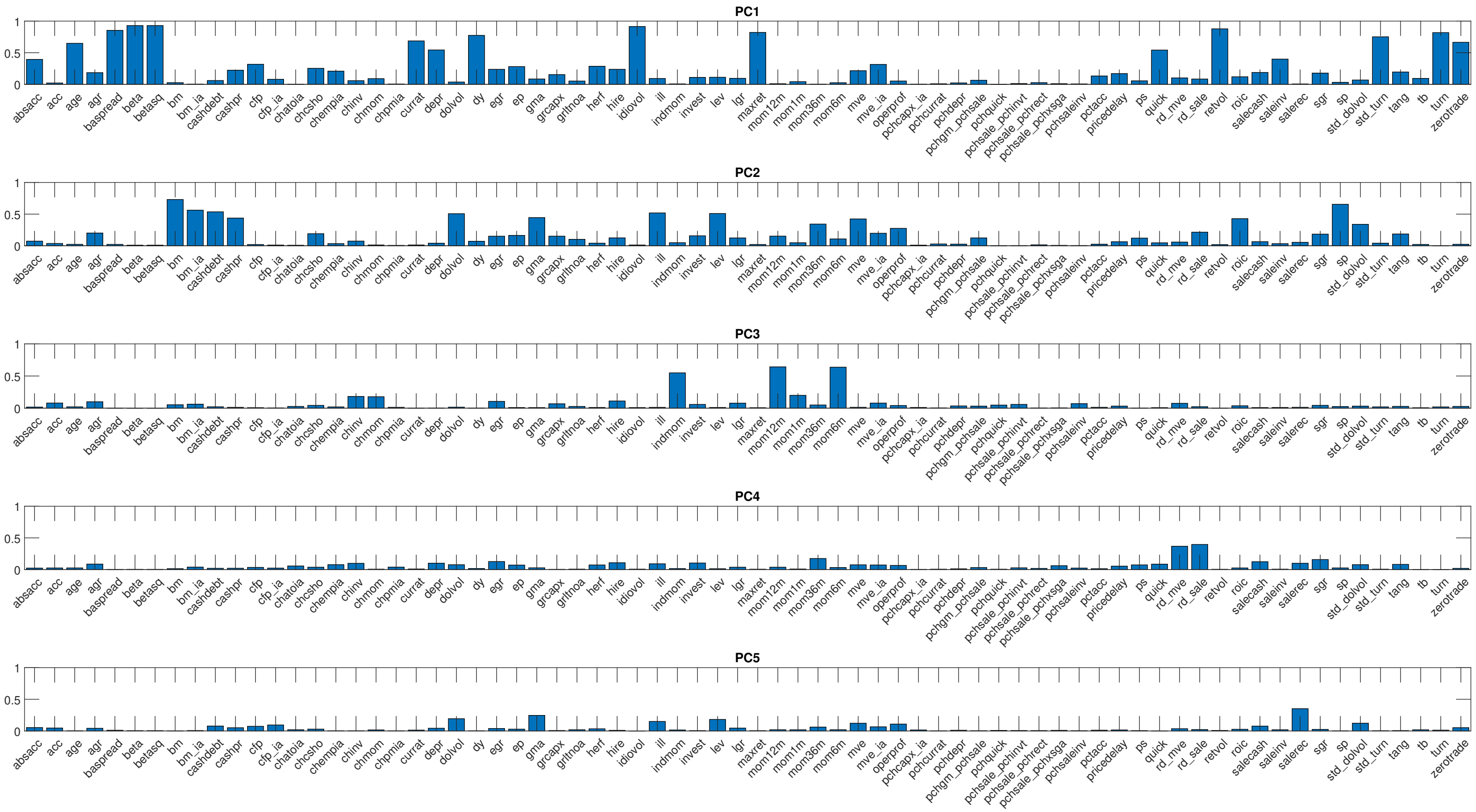
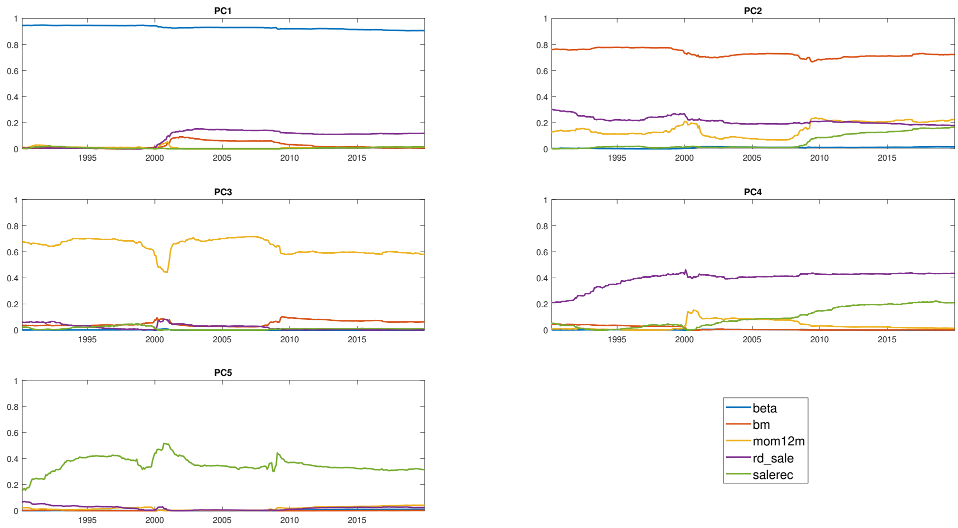


Figure A.9: R^2 across months from a recursive regression of each PC portfolio on each of the following anomalies: beta, bm, mom12m, rd_sale, and salerec. Each anomaly loads heavily on one PC portfolio and contributes little to the rest. The PC portfolios are constructed using PCA.

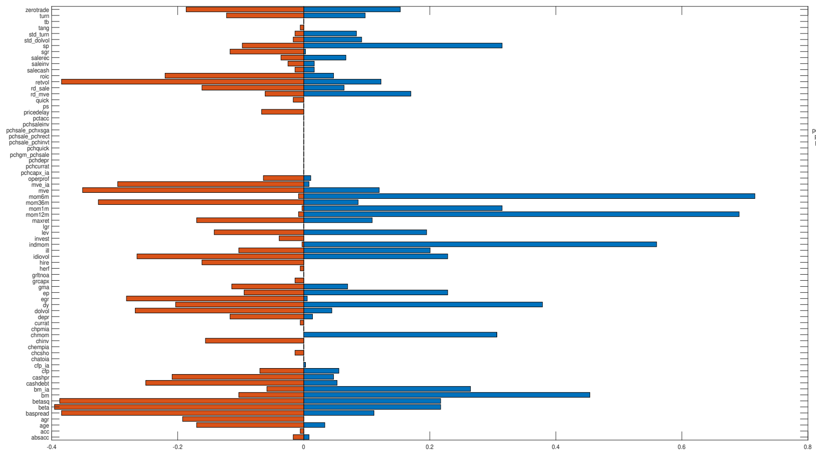


A.2.7 Factor timing strategy constituents

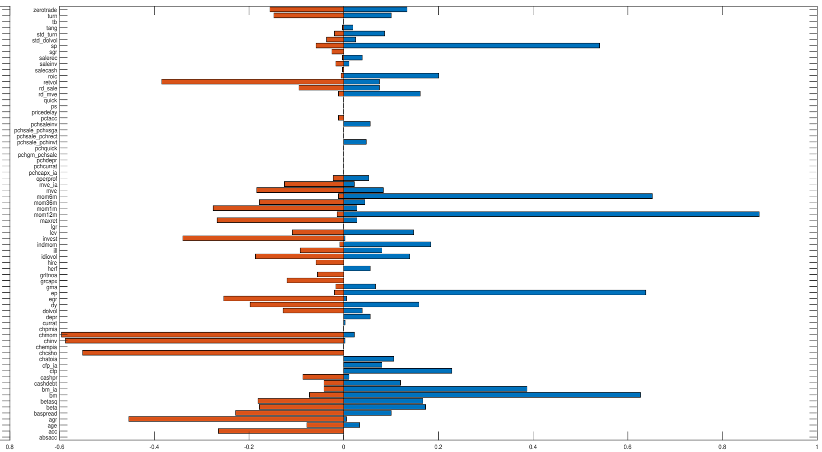
In this section we examine the trading positions of our factor timing portfolios. Specifically, using the forecasts from the LASSO-based models and the LSS, we investigate how often each anomaly is traded. The anomalies traded under LSS are also the ones with the highest absolute weights under TSFM and CSFM, so focusing on this case only is representative of the general investing approach. Figure A.10 displays the percentage trading frequency of each anomaly. Blue bars imply long and orange bars imply short positions. Clearly, the benefits of factor timing strategies arise from rotating among multiple anomalies and not by focusing on a handful of picks. Yet, all strategies tend to go long on anomalies with high average returns, such as mom6m and mom12m, and short anomalies with negative average returns, like beta, chmom and retvol. Furthermore, some anomalies appear almost equally often in the short and the long legs of our factor timing portfolio, implying higher volatility in conditional returns. For the models that use PCA, these anomalies are usually related to market frictions, which also have higher volatility and thus load more heavily on the first PCs. When RPPCA is used for the left-hand-side, the anomalies more regularly traded are those with higher absolute average returns. Finally, anomalies that do not load heavily on the dominant components, using either PCA or RPPCA, stay out of the investable universe as their small loadings compress their individual return forecasts close to zero. Consequently, the use of PCs for the left-hand-side has an impact on the factor timing portfolio formation.

Figure A.10: Percentage frequency of LSS constituents over the out-of-sample period. Blue bars suggest long and orange bars suggest short positions. (a) PCA (b) RPPCA (c) PCA-PLS (d) RPPCA-PLS.

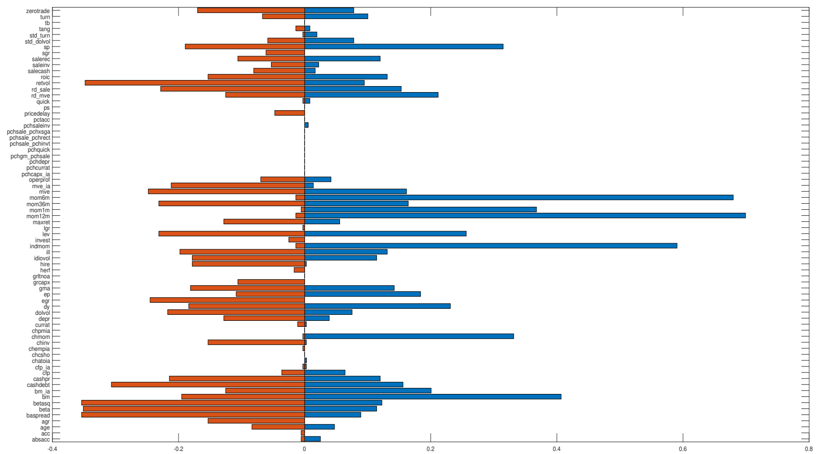
(a) PCA



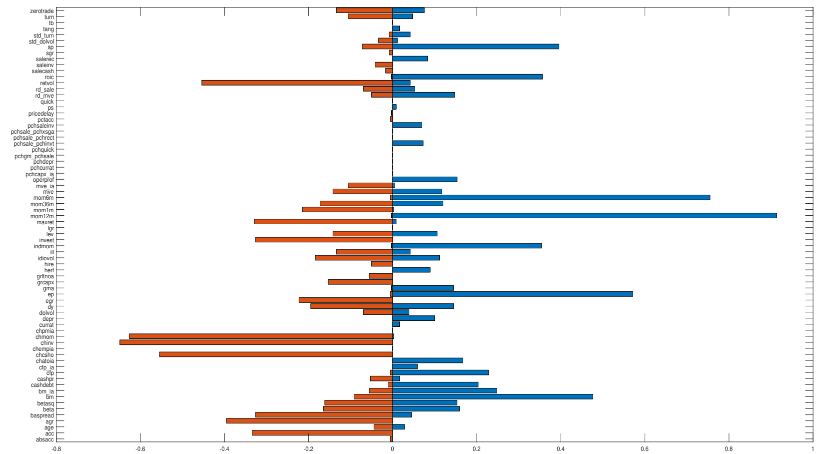
(b) RPPCA



(c) PCA-PLS



(d) RPPCA-PLS

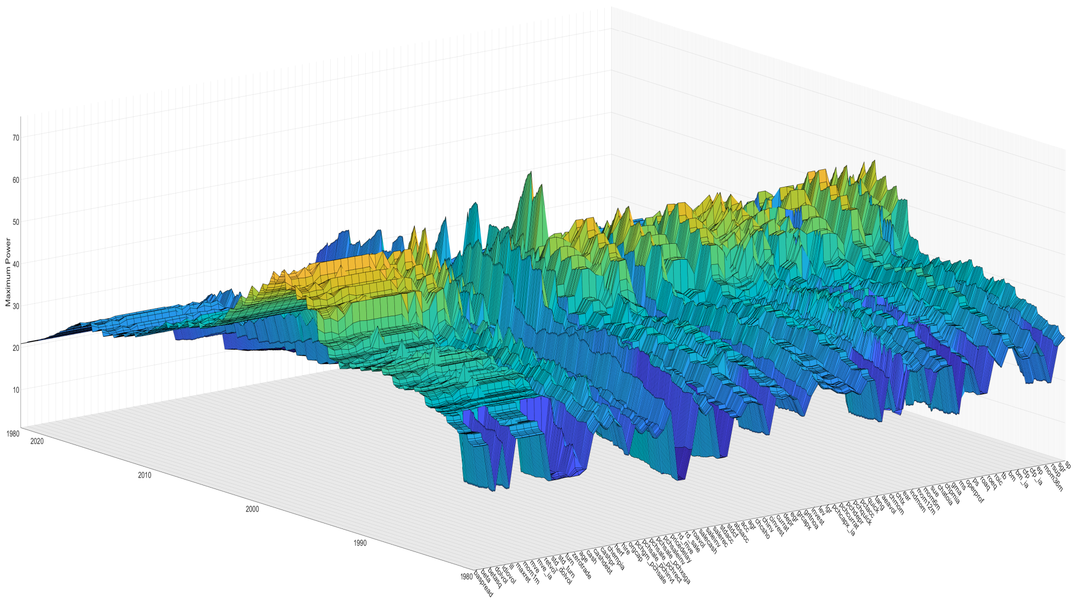


A.3 Supplementary Material to Chapter 2

A.3.1 Supplementary Figures

Figure A.11: Upper Thresholds for q_t^{max} (a) and p_t^{max} (b) for the 85 characteristics using a Maximum Weight Constraint of (w^{ceil}) 2%. The figures illustrate the time variability in the maximum threshold due to the varying number of the cross-sections across different characteristics. The maximum power thresholds vary with characteristics, reflecting the different characteristic variabilities within each characteristic due to the time-varying size of the cross-sections and between the long and short legs of the same characteristic due to the presence of ties in the underlying characteristic distribution. The sample period is from January 1980 to December 2021.

(a) q_t^{max}



(b) p_t^{max}

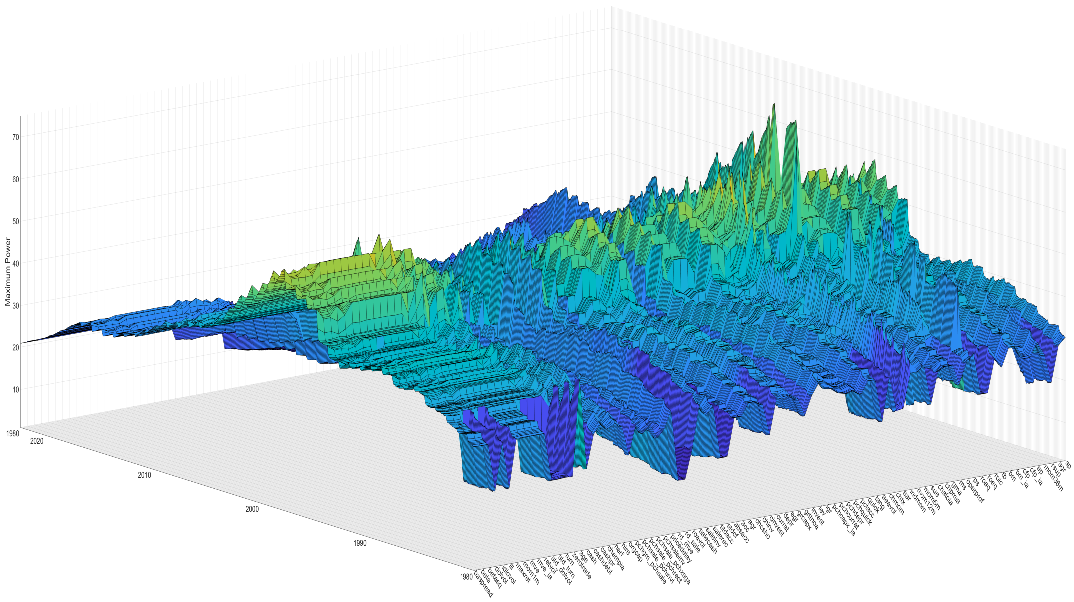
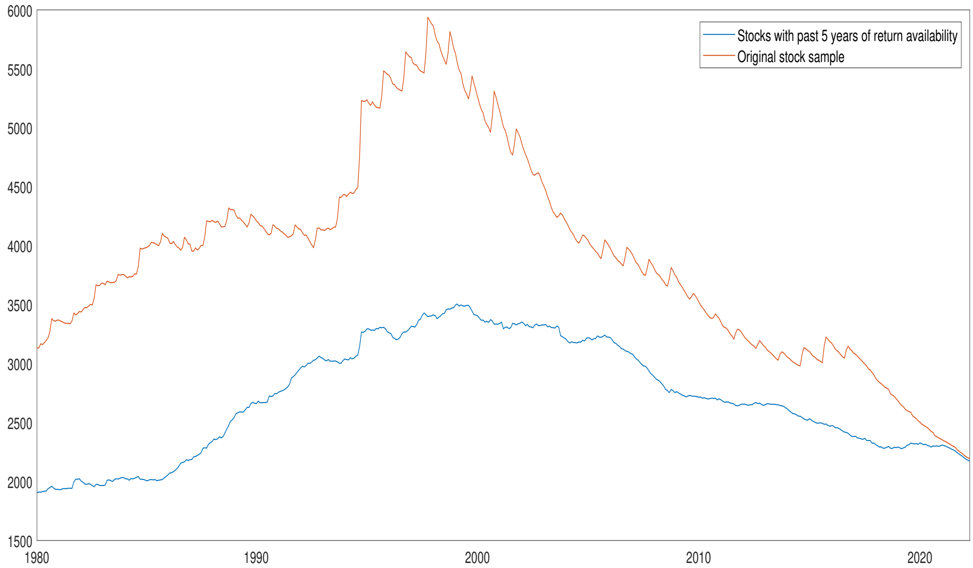


Figure A.12: Cross-sectional stock size for original sample and sub-sample with five years of daily return data. The chart illustrates the difference in the number of stocks in the sample when five years of past daily data are required. The disparity is particularly pronounced in earlier periods, notably during the build-up of the dot com bubble, but becomes less noticeable in later years.



A.3.2 Return-spread maximization objective

In the base case, power portfolios are constructed based on a Sharpe ratio maximization objective. To explore the sensitivity of the approach to the underlying objective, we here construct power portfolios under a return maximization objective. Table A.5 presents the average portfolio statistics for these power portfolios, alongside the decile-sorted benchmarks. The results show strong consistency with those in Table 2.1, confirming the significant outperformance of power sorting over the conventional benchmark.

As expected, the switch from Sharpe ratio to returns leads to power portfolios with a higher average return but a lower Sharpe ratio, in line with the new objective. This is achieved by slightly increasing concentration in the tails, as evident from a lower number of effective names for both the long and short sides.

Table A.5: Power portfolios under a return maximization objective. Return: Average monthly return, Standard deviation.: Monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: Return=0, Maximum drawdown: Maximum drawdown, Hit rate: Percentage frequency of positive returns, Turnover: Average monthly turnover (bounded by 200%), # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample period is from March 1980 to December 2021.

	Equal-weighted		Value-weighted	
	Power	Conventional	Power	Conventional
Return (%)	0.91	0.51	0.69	0.32
Standard deviation (%)	5.03	4.21	4.95	4.39
Sharpe ratio	0.65	0.46	0.48	0.26
t-stat	4.24	2.96	3.12	1.71
Maximum drawdown (%)	−56.22	−55.35	−57.05	−59.22
Hit rate (%)	61.21	57.39	57.83	53.44
Turnover (%)	39.39	39.93	35.05	35.44
# of effective names long	1168.99	369.24	425.66	107.00
# of effective names short	267.78	370.33	155.07	98.42

A.3.3 Multi-factor strategies for alternative benchmarks

We briefly examine how the results reported in Section 2.4.1 generalize for multi-factor strategies. Table A.6 reports the results for the three mutli-factor strategies, comparing rank portfolios to power portfolios. Evidently, multi-factor portfolios based on rank portfolios demonstrate very similar performance across the three strategies, indicating the method's inability to effectively combine signals in an objective-oriented fashion. In contrast, power portfolio-based strategies clearly showcase the method's capability to combine individual factors into multi-factor portfolios with maximum Sharpe ratio.

When comparing averaging strategies, AVP outperforms AVR in terms of average returns and Sharpe ratios, while also exhibiting less tail risk. In the case of the combination approaches, which are more aggressive in nature, PME and PMP deliver more than double the return compared to the rank-based strategies, leading to significantly higher Sharpe ratios and t-statistics. Overall, our results demonstrate that the strict enforcement of a linear weighting scheme hinders performance at both a univariate and a multivariate level. Furthermore, a simple rank-based approach leads to portfolios that inherit a passive stance, limiting the effective extraction of underlying signals or the combination of different signals.

Table A.6: Portfolio evaluation measures for multi-factor power and rank portfolios. AVP: Multi-factor portfolio based on the average portfolio weight from individual power portfolios. AVR: Mutli-factor portfolio based on the average portfolio weight from individual rank portfolios. PME: Power portfolio based on the average characteristic rank. RME: Rank portfolio based on the average characteristic rank. PMP: Power portfolio based on the rank implied by average power portfolio weights. RMR: Rank portfolio based on the rank implied by average rank portfolio weights. Panel A shows equal-weighted results and Panel B shows value-weighted results. The sample period is from March 1980 to December 2021.

	AVP	AVR	PME	RME	PMP	RMR
A. Equal-weighted portfolios						
Return (%)	1.82	1.45	3.05	1.27	3.16	1.26
Standard deviation (%)	4.96	4.60	7.27	4.29	7.30	4.16
Sharpe ratio	1.27	1.09	1.45	1.03	1.50	1.05
t-stat	8.28	7.09	9.44	6.66	9.72	6.82
Maximum drawdown (%)	−46.08	−41.81	−55.51	−39.17	−55.40	−38.11
Hit rate (%)	71.60	69.03	73.72	67.19	72.73	68.77
Turnover (%)	37.33	32.82	48.57	32.16	66.73	31.47
# of effective names long	1814.32	1392.70	1608.76	1463.92	146.34	1463.92
# of effective names short	600.71	1150.37	101.38	1463.92	141.27	1463.92
B. Value-weighted portfolios						
Return (%)	1.36	0.97	2.80	1.00	2.50	0.98
Standard deviation (%)	4.57	4.23	7.40	4.33	7.07	4.20
Sharpe ratio	1.03	0.80	1.31	0.80	1.23	0.81
t-stat	6.68	5.17	8.52	5.22	7.96	5.25
Maximum drawdown (%)	−41.21	−42.03	−61.54	−43.12	−56.61	−42.29
Hit rate (%)	67.26	63.31	69.37	62.45	68.77	62.45
Turnover (%)	34.00	29.57	44.64	27.04	52.56	26.30
# of effective names long	641.23	404.75	578.14	562.56	166.51	556.31
# of effective names short	510.43	1062.73	79.87	849.04	89.25	821.85

Next, our attention turns to PPP. Beginning with the averaging strategy, power sorting outperforms PPP, achieving higher returns with lower risk, while metrics for portfolio concentration and turnover remain largely comparable between the two methods.

In the case of PPPME, which employs average standardized characteristics instead of ranks, a lower risk-return trade-off is achieved compared to PME, albeit with reduced turnover and portfolio weight concentration. It's important to note that the use of average standardized characteristic scores, rather than ranks, implies an unequal contribution of characteristics to the composite score.

Table A.7: Portfolio evaluation measures for multi-factor power and PPP portfolios. AVP: Multi-factor portfolio based on the average portfolio weight from individual power portfolios. AVPPP: Mutli-factor portfolio based on the average portfolio weight from individual PPP portfolios. PME: Power portfolio based on the average characteristic rank. PPPME: PPP portfolio based on the average standardized characteristic score. PMP: Power portfolio based on the rank implied by average power portfolio weights. PMPP: PPP portfolio using as input the average PPP portfolio weights. The sample period is from March 1980 to December 2021.

	AVP	AVPPP	PME	PPPME	PMP	PMPPP
A. Equal-weighted portfolios						
Return (%)	1.82	1.59	3.05	1.71	3.16	1.16
Standard deviation (%)	4.96	5.03	7.27	5.23	7.30	4.00
Sharpe ratio	1.27	1.09	1.45	1.13	1.50	1.00
t-stat	8.28	7.10	9.44	7.36	9.72	6.48
Maximum drawdown (%)	−46.08	−49.86	−55.51	−50.68	−55.40	−39.96
Hit rate (%)	71.60	69.70	73.72	71.49	72.73	67.13
Turnover (%)	37.33	37.11	48.57	41.06	66.73	34.89
# of effective names long	1814.32	1817.63	1608.76	1349.67	146.34	2262.59
# of effective names short	600.71	599.62	101.38	731.17	141.27	1177.60

Finally, adopting the sum of weights as the underlying signal leads to further divergence in performance and a more passive stance for PMPPP. This passivity arises from extreme skewness in features related to market friction proxies, which are indicative of specific sets of illiquid stocks. These outliers in the short tail impede the strategy from attaining high values of theta to mitigate over-concentration in a handful of stocks on the short side. Unlike a ranking-based aggregation, the significant relative distance of these outliers can result in extreme weight allocations with only slight adjustments in the underlying parameter. This outcome underscores the limitations of employing a single parameter to capture variation in a long-short sense, as well as utilizing characteristic distributions with distinct properties, failing to effectively aggregate information across various characteristics and adequately capture the heterogeneity in behavior across tails.

Moving on, we see how efficient and power sorting generalize to a multi-factor level using the updated sample. In Panel A of Table A.8, we begin by analyzing the averaging strategy. Power sorting exhibits the highest average return, while efficient sorting exhibits the smallest volatility and highest Sharpe ratio. Nonetheless, the performance differences across all variations, including the conventional approach, are relatively small. This can be attributed to the averaging strategy, which blends exposures without adequately differentiating the strength of the underlying signal.

Panel B presents the results for the strategy that utilizes the average characteristic rank. In the case of efficient sorting, the characteristics are cross-sectionally standardized and added together, rather than using their ranks, to maintain consistency with the original framework. The optimization-based approaches clearly outperform the conventional approach when the underlying signal is informative, resulting in a significant 30% increase in the Sharpe ratio. While the power sorting approaches primarily increase average returns through asymmetric concentration in the tails, the efficient sorting approach achieves similar results by reducing variance. Moreover, the efficient sorting approach demonstrates the lowest drawdown, albeit at a cost of 20% higher turnover per month compared to power sorting.

Table A.8: Portfolio evaluation measures for multi-factor power and efficient portfolios. Return: Average monthly return, Standard deviation: Monthly standard deviation, Sharpe ratio: Annualized Sharpe ratio, t-stat: t-statistic on H_0: Return=0, Maximum drawdown: Maximum drawdown, Hit rate: Percentage frequency of positive returns, Turnover: Average monthly turnover bounded by 200%, # of effective names long: Number of effective names (i.e., sum of squared weights raised to -1) for the long leg, # of effective names short: Number of effective names (i.e., sum of squared weights raised to -1) for the short leg. The sample includes stocks with an available return history of five years at each investment date through the period from March 1980 to December 2021. Panel A shows the Weight averaging strategy, Panel B shows the average characteristic rank strategy, and Panel C shows the average weight rank strategy.

	Power	Power VCV	Efficient	Conventional
A. Weight Averaging Strategy				
Return (%)	1.38	1.28	1.19	1.14
Standard deviation (%)	4.58	4.67	3.52	3.81
Sharpe ratio	1.04	0.95	1.17	1.04
t-stat	6.78	6.15	7.59	6.75
Maximum drawdown (%)	−43.85	−45.96	−38.96	−41.82
Hit rate (%)	68.97	66.01	70.36	70.36
Turnover (%)	36.63	35.76	39.96	42.14
# of effective names long	1244.12	1242.10	995.91	990.89
# of effective names short	434.08	488.97	601.37	632.24
B. Average Characteristic Rank Strategy				
Return (%)	2.25	1.93	1.52	1.53
Standard deviation (%)	6.41	5.51	4.31	5.61
Sharpe ratio	1.22	1.22	1.22	0.95
t-stat	7.90	7.90	7.92	6.14
Maximum drawdown (%)	−55.28	−55.28	−50.02	−57.66
Hit rate (%)	70.16	69.96	71.34	68.18
Turnover (%)	47.26	46.71	66.76	52.68
# of effective names long	1051.80	1195.59	440.34	270.94
# of effective names short	103.19	294.13	271.34	270.94
C. Average Weight Rank Strategy				
Return (%)	2.39	2.29	1.23	1.55
Standard deviation (%)	6.99	7.37	3.97	5.03
Sharpe ratio	1.18	1.08	1.07	1.06
t-stat	7.68	6.99	6.96	6.91
Maximum drawdown (%)	−50.76	−59.18	−45.41	−51.68
Hit rate (%)	70.95	69.37	70.16	70.55
Turnover (%)	61.99	63.35	71.56	59.73
# of effective names long	99.17	98.74	443.71	270.94
# of effective names short	134.26	128.99	413.40	270.94

Finally, Panel C displays the findings for the strategy that utilizes the sum of weights as the underlying signal. Once again, power sorting emerges with the highest Sharpe ratio by maximizing average returns, despite exhibiting the highest variance. In contrast, efficient sorting demonstrates the lowest volatility, although it comes with the highest turnover. The conventional approach falls in the middle, achieving a Sharpe ratio comparable to that of the sophisticated approaches.

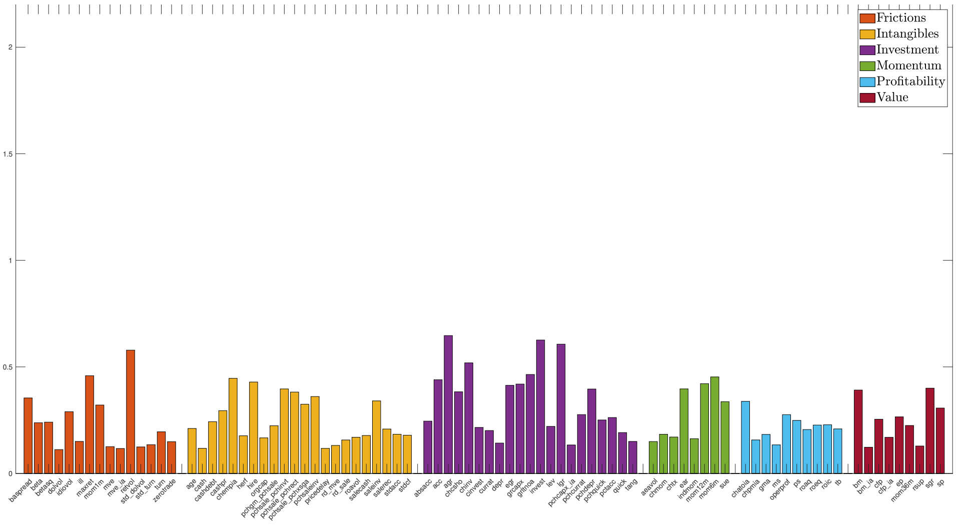
Notably, the sophisticated approaches inherently differ in their approach to factor portfolio construction. Specifically, when there is a strong underlying signal, as is the case with the sum of power portfolio weights, power sorting adopts an aggressive stance by increasing the concentration ratios to maximize performance. This results in a significant improvement in average returns, albeit at the expense of higher volatility. In contrast, efficient sorting does not distinguish between weak and strong signals, consistently striving to minimize variance, even if it slightly reduces the underlying premia. However, it is worth noting that a variance reduction objective can also be achieved through a power sorting framework.

A.4 Supplementary Material to Chapter 3

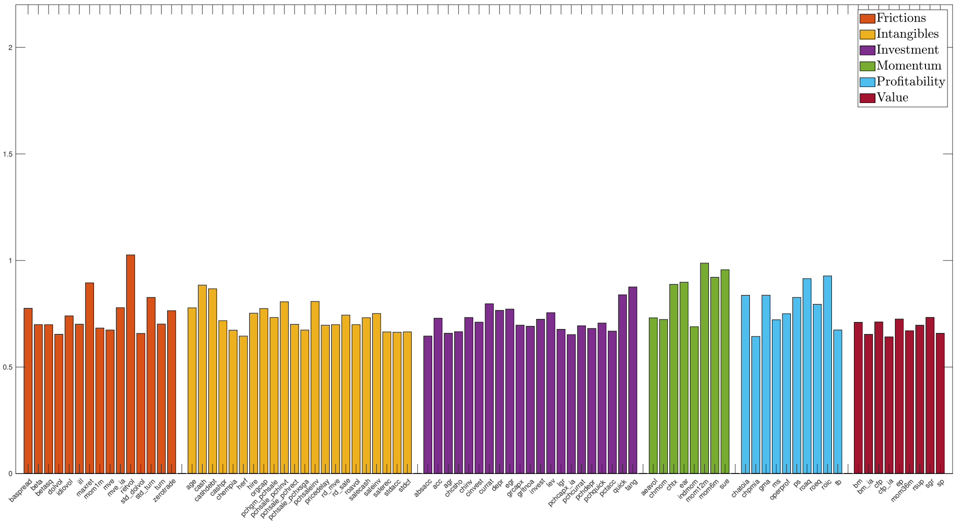
A.4.1 Squared Sharpe ratio from adding each factor in each iteration

Figure A.13: Squared Sharpe ratio from adding each factor to the factor model of the previous iteration

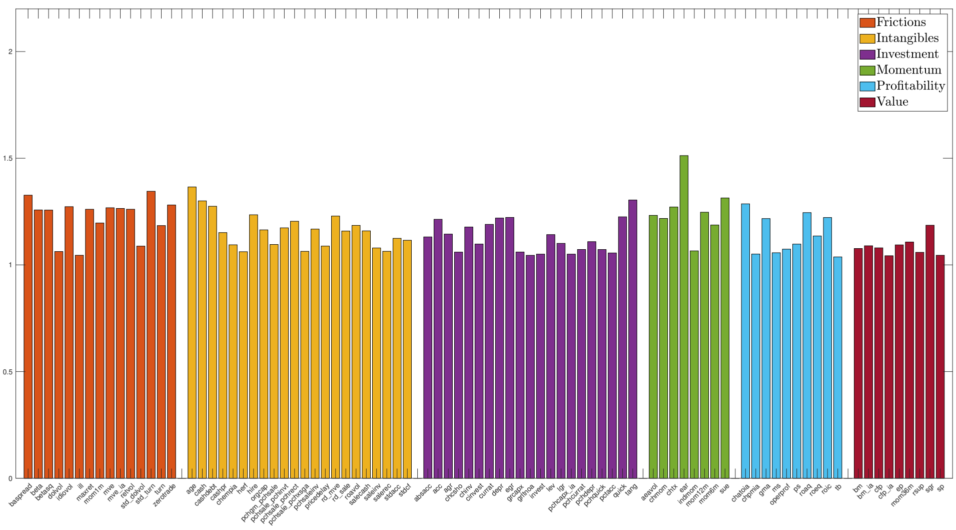
(a) Iteration 1



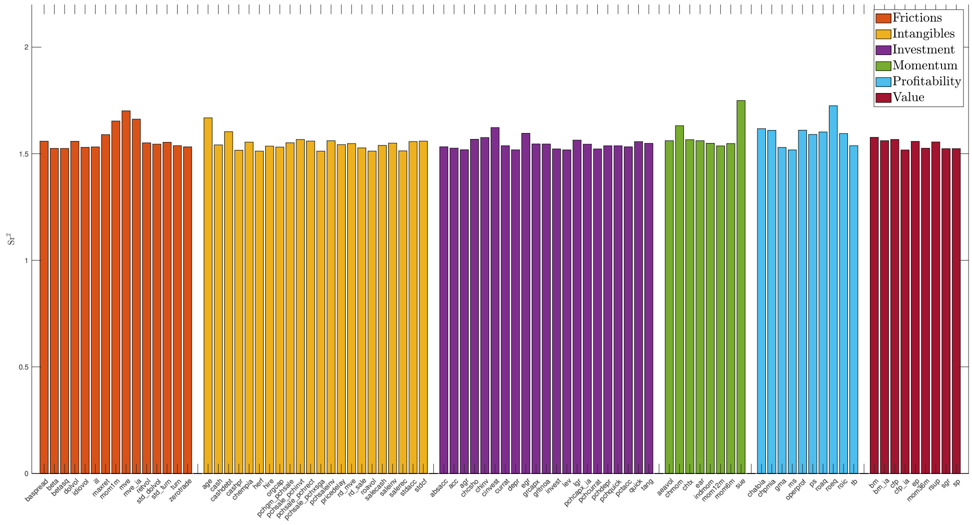
(b) Iteration 2



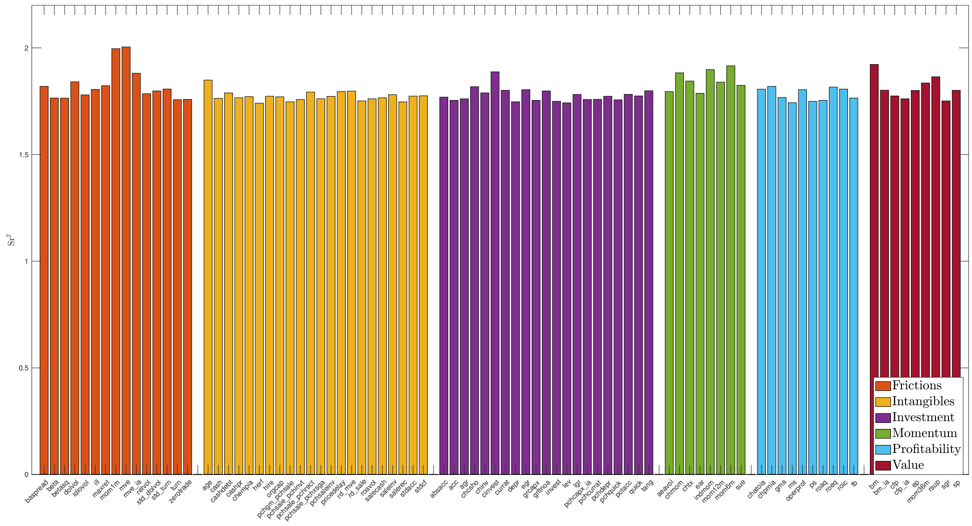
(c) Iteration 3



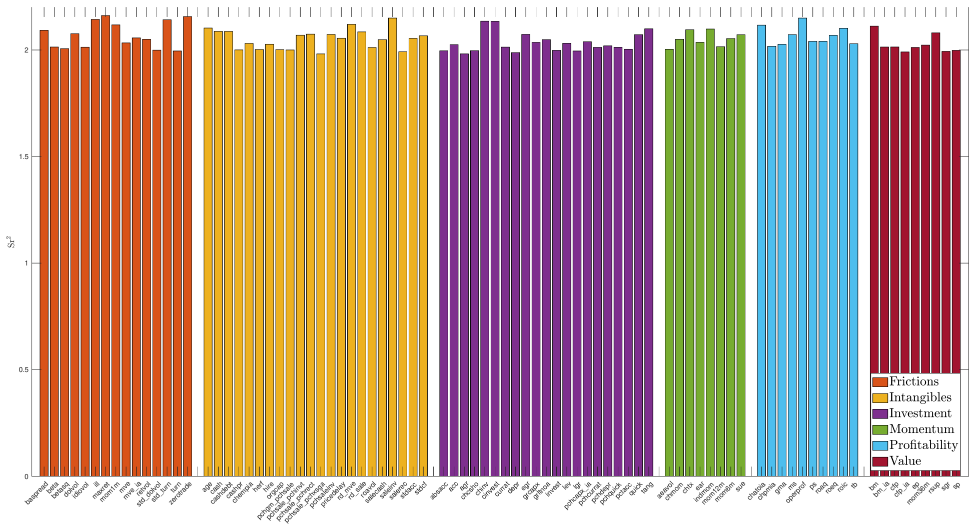
(d) Iteration 4



(e) Iteration 5



(f) Iteration 6



References

Aït-Sahalia, Y., and M. W. Brandt. 2001. Variable selection for portfolio choice. *Journal of Finance* 56(4):1297–1351.

Akbas, F., W. J. Armstrong, S. Sorescu, and A. Subrahmanyam. 2016. Capital market efficiency and arbitrage efficacy. *Journal of Financial and Quantitative Analysis* 51(2):387–413.

Ammann, M., G. Coqueret, and J.-P. Schade. 2016. Characteristics-based portfolio choice with leverage constraints. *Journal of Banking & Finance* 70:23–37.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61(1):259–299.

Arnott, R. D., M. Clements, V. Kalesnik, and J. T. Linnainmaa. 2021. Factor momentum. Available at SSRN: https://ssrn.com/abstract=3116974 or http://dx.doi.org/10.2139/ssrn.3116974.

Asness, C., S. Chandra, A. Ilmanen, and R. Israel. 2017. Contrarian factor timing is deceptively difficult. *Journal of Portfolio Management* 43(5):72–87.

Baba-Yara, F., M. Boons, and A. Tamoni. 2021. Value return predictability across asset classes and commonalities in risk premia. *Review of Finance* 25(2):449–84.

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

Baker, M., and J. Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61(4):1645–80.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99(2):427–446.

Bansal, R., and A. Yaron. 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59(4):1481–1509.

Banz, R. W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9(1):3–18.

Barillas, F., R. Kan, C. Robotti, and J. Shanken. 2020. Model comparison with Sharpe ratios. *Journal of Financial and Quantitative Analysis* 55(6):1840–1874.

Basu, S. 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance* 32(3):663–682.

Blitz, D., G. Baltussen, and P. van Vliet. 2020. When equity factors drop their shorts. *Financial Analysts Journal* 76(4):73–99.

Blitz, D., and M. X. Hanauer. 2020. Settling the size matter. *Available at SSRN 3686583*.

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 P. Santa-Clara. 2006. Dynamic portfolio selection by augmenting the asset space. *Journal of Finance* 61(5):2187–2217.

Brandt, M. W., P. Santa-Clara, and R. Valkanov. 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies* 22(9):3411–3447.

Campbell, J. Y., and J. H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107(2):205–251.

Campbell, J. Y., and R. J. Shiller. 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 S. B. Thompson. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21(4):1509–31.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and E. Schaumburg. 2020. Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics* 102(3):531–551.

Chen, A. Y., A. Lopez-Lira, and T. Zimmermann. 2022. Does peer-reviewed theory help predict the cross-section of stock returns? *arXiv preprint arXiv:2212.10317*.

Chen, A. Y., and T. Zimmermann. 2022. Open Source Cross-Sectional Asset Pricing. *Critical Finance Review* 27(2):207–264.

Chen, L., M. Pelger, and J. Zhu. 2024. Deep learning in asset pricing. *Management Science* 70(2):714–750.

Chordia, T., A. Subrahmanyam, and Q. Tong. 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics* 58(1):41–58.

Cochrane, J. H. 2011. Presidential address: Discount rates. *Journal of Finance* 66(4):1047–108.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang. 2021. AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI. *Available at SSRN 3554486*.

Connor, G., and R. A. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15(3):373–94.

———. 1988. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics* 21(2):255–89.

Cooper, M. J., H. Gulen, and M. J. Schill. 2008. Asset growth and the cross-section of stock returns. *Journal of Finance* 63(4):1609–1651.

Daniel, K., D. Hirshleifer, and L. Sun. 2020. Short-and long-horizon behavioral factors. *Review of Financial Studies* 33(4):1673–1736.

Daniel, K., and T. J. Moskowitz. 2016. Momentum crashes. *Journal of Financial Economics* 122(2):221–47.

Daniel, K., L. Mota, S. Rottke, and T. Santos. 2020. The cross-section of risk and returns. *Review of Financial Studies* 33(5):1927–1979.

De Jong, S. 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems* 18(3):251–263.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal. 2020. A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies* 33(5):2180–2222.

Dichtl, H., W. Drobetz, H. Lohre, C. Rother, and P. Vosskamp. 2019. Optimal timing and tilting of equity factors. *Financial Analysts Journal* 75(4):84–102.

Ehsani, S., and J. T. Linnainmaa. 2022. Factor momentum and the momentum factor. *Journal of Finance* 77(3):1877–1919.

Engle, R. F., O. Ledoit, and M. Wolf. 2019. Large dynamic covariance matrices. *Journal of Business & Economic Statistics* 37(2):363–375.

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

———. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1):3–56.

———. 2008. Dissecting anomalies. *Journal of Finance* 63(4):1653–78.

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

———. 2018. Choosing factors. *Journal of Financial Economics* 128(2):234–252.

———. 2020. Comparing cross-section and time-series factor models. *Review of Financial Studies* 33(5):1891–1926.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81(3):607–636.

Fan, J., Y. Liao, and W. Wang. 2016. Projected principal component analysis in factor models. *Annals of Statistics* 44(1):219.

Fan, M., F. Kearney, Y. Li, and J. Liu. 2022. Momentum and the Cross-section of Stock Volatility. *Journal of Economic Dynamics and Control* 144:104524.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the factor zoo: A test of new factors. *Journal of Finance* 75(3):1327–70.

Ferson, W. E., and C. R. Harvey. 1991. The variation of economic risk premiums. *Journal of political economy* 99(2):385–415.

Frazzini, A., and L. H. Pedersen. 2014. Betting against beta. *Journal of Financial Economics* 111(1):1–25.

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies* 33(5):2326–77.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1):1.

Giglio, S., and D. Xiu. 2021. Asset pricing with omitted factors. *Journal of Political Economy* 129(7):1974–90.

Green, J., J. R. Hand, and F. Zhang. 2017. The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies* 30(12):4389–4436.

Greenwood, R., and S. G. Hanson. 2012. Share issuance and factor timing. *Journal of Finance* 67(2):761–98.

Grinblatt, M., and T. J. Moskowitz. 2004. Predicting stock price movements from past returns: The role of consistency and tax-loss selling. *Journal of Financial Economics* 71(3):541–579.

Grundy, B. D., and J. S. Martin. 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. *Review of Financial Studies* 14(1):29–78.

Gu, S., B. Kelly, and D. Xiu. 2020. Empirical asset pricing via machine learning. *Review of Financial Studies* 33(5):2223–73.

———. 2021. Autoencoder asset pricing models. *Journal of Econometrics* 222(1):429–450.

Gupta, T., and B. Kelly. 2019. Factor momentum everywhere. *Journal of Portfolio Management* 45(3):13–36.

Haddad, V., S. Kozak, and S. Santosh. 2020. Factor timing. *Review of Financial Studies* 33(5):1980–2018.

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

He, Z., and A. Krishnamurthy. 2013. Intermediary asset pricing. *American Economic Review* 103(2):732–770.

Hjalmarsson, E., and P. Manchev. 2012. Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance* 36(5):1392–1401.

Hong, H., T. Lim, and J. C. Stein. 2000. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance* 55(1):265–295.

Hou, K., H. Mo, C. Xue, and L. Zhang. 2021. An augmented q-factor model with expected growth. *Review of Finance* 25(1):1–41.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28(3):650–705.

———. 2020. Replicating anomalies. *Review of Financial Studies* 33(5):2019–2133.

Huang, S. 2022. The momentum gap and return predictability. *Review of Financial Studies* 35(7):3303–36.

Jacobs, B. I., and K. N. Levy. 1993. Long/short equity investing. *Journal of Portfolio Management* 20(1):52.

Jacobs, H. 2015. What explains the dynamics of 100 anomalies? *Journal of Banking & Finance* 57:65–85.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48(1):65–91.

Jensen, T. I., B. Kelly, and L. H. Pedersen. 2023a. Is there a replication crisis in finance? *The Journal of Finance* 78(5):2465–2518.

Jensen, T. I., B. T. Kelly, and L. H. Pedersen. 2023b. Is there a replication crisis in finance? *Journal of Finance,* forthcoming.

Kagkadis, A., H. Lohre, I. Nolte, N. Sandra, and N. Vasilas. 2024a. Power Sorting. *Available at SSRN: https://ssrn.com/abstract=4552208*.

Kagkadis, A., I. Nolte, S. Nolte, and N. Vasilas. 2024b. Factor timing with portfolio characteristics. *The Review of Asset Pricing Studies* 14(1):84–118.

Kelly, B. T., S. Malamud, and K. Zhou. Forthcoming. The virtue of complexity in return prediction. *Journal of Finance*.

Kelly, B. T., and S. Pruitt. 2013. Market expectations in the cross-section of present values. *Journal of Finance* 68(5):1721–56.

Kelly, B. T., S. Pruitt, and Y. Su. 2019. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics* 134(3):501–524.

Keloharju, M., J. T. Linnainmaa, and P. Nyberg. 2016. Return seasonalities. *Journal of Finance* 71(4):1557–90.

Kim, A. Y., Y. Tse, and J. K. Wald. 2016. Time series momentum and volatility scaling. *Journal of Financial Markets* 30:103–124.

Kim, S., R. A. Korajczyk, and A. Neuhierl. 2021. Arbitrage portfolios. *Review of Financial Studies* 34(6):2813–2856.

Kishore, R., M. W. Brandt, P. Santa-Clara, and M. Venkatachalam. 2008. Earnings announcements are full of surprises. *Available at SSRN 909563*.

Knuth, D. E. 1993. Johann Faulhaber and sums of powers. *Mathematics of Computation* 61(203):2 77–294.

Koijen, R. S., T. J. Moskowitz, L. H. Pedersen, and E. B. Vrugt. 2018. Carry. *Journal of Financial Economics* 127(2):197–225.

Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73(3):1183–1223.

———. 2020. Shrinking the cross-section. *Journal of Financial Economics* 135(2):271–92.

Ledoit, O., and M. Wolf. 2022. Quadratic shrinkage for large covariance matrices. *Bernoulli* 28(3):1519–1547.

Ledoit, O., M. Wolf, and Z. Zhao. 2019. Efficient sorting: A more powerful test for cross-sectional anomalies. *Journal of Financial Econometrics* 17(4):645–686.

Lettau, M., and M. Pelger. 2020a. Estimating latent asset-pricing factors. *Journal of Econometrics* 218(1):1–31.

———. 2020b. Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies* 33(5):2274–2325.

Leung, E., H. Lohre, D. Mischlich, Y. Shea, and M. Stroh. 2021. The promises and pitfalls of machine learning for predicting stock returns. *Journal of Financial Data Science* 3(2):21–50.

McGee, R. J., and J. Olmo. 2022. Optimal characteristic portfolios. *Quantitative Finance* 22(10):1 853–1870.

McLean, R. D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71(1):5–32.

Nagel, S. 2021. *Machine learning in asset pricing.* Princeton, NJ: Princeton University Press.

Novy-Marx, R. 2015a. *Fundamentally, momentum is fundamental momentum.* Tech. rep. National Bureau of Economic Research.

———. 2015b. *How can a q-theoretic model price momentum?* Tech. rep. National Bureau of Economic Research.

Novy-Marx, R., and M. Velikov. 2022. Betting against betting against beta. *Journal of Financial Economics* 143(1):80–106.

Onatski, A. 2010. Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* 92(4):1004–16.

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

Patton, A. J., and A. Timmermann. 2010. Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts. *Journal of Financial Economics* 98(3):605–625.

Rapach, D. E., J. K. Strauss, and G. Zhou. 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies* 23(2):821–862.

Rendleman, R. J., C. P. Jones, and H. A. Latane. 1982. Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *Journal of Financial Economics* 10(3):269–287.

Ross, S. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13(3):341–360.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19(3):425–442.

Simon, F., S. Weibels, and T. Zimmermann. 2023. *Deep parametric portfolio policies.* Tech. rep. CFR Working Paper.

Soebhag, A., B. van Vliet, and P. Verwijmeren. 2023. Non-standard errors in asset pricing: Mind your sorts. *Available at SSRN.*

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104(2):288–302.

Stambaugh, R. F., and Y. Yuan. 2017. Mispricing factors. *Review of Financial Studies* 30(4):1270–1315.

Stephen, R. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13(3):341–360.

Swade, A., M. X. Hanauer, H. Lohre, and D. Blitz. 2024. Factor zoo (. zip). *Journal of Portfolio Management* 50(3):11–31.

Swade, A., S. Nolte, M. Shackleton, and H. Lohre. 2023. Why do equally weighted portfolios beat value-weighted ones? *Journal of Portfolio Management* 49(5):167–187.

Wold, S., A. Ruhe, H. Wold, and W. J. Dunn III. 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing* 5(3):735–43.

Zhang, X., L. Wu, and Z. Chen. 2022. Constructing long-short stock portfolio with a new listwise learn-to-rank algorithm. *Quantitative Finance* 22(2):321–331.