Market concentration, institutional demand, and risk factors in asset pricing

Nguyen, N.M.T.

Publication date
2025
Document Version
Final published version

This dissertation contributes to the literature on empirical asset pricing. Each chapter is motivated by a post-2000 development: the increasing concentration of stock markets, the growing role of institutional investors, and the widespread adoption of factor-based investing. Together, the chapters examine how stock market concentration, institutional demand, and factor overlap link to frictions that distort asset pricing and generate risk premia.

Ngoc Minh Thu Nguyen holds a BSc in Econometrics and Operations Research from Maastricht University and an MPhil in Economics from the Tinbergen Institute. During her research master's, she developed a strong interest in the interaction between financial markets and the real economy, particularly within the field of asset pricing. Building on these interests, she pursued a PhD in Finance at the University of Amsterdam, under the supervision of Aleksandar Andonov and Esther Eiling. She is currently a Lecturer in Finance at Rotterdam School of Management, Erasmus University.

Market Concentration, Institutional Demand, and Risk Factors in Asset Pricing

Ngoc Minh Thu Nguyen

878

tinbergen institute
GRADUATE PROGRAM

Market Concentration, Institutional Demand, and Risk Factors in Asset Pricing

Ngoc Minh Thu Nguyen

Universiteit van Amsterdam

Market Concentration, Institutional Demand, and Risk Factors in Asset Pricing

Ngoc Minh Thu Nguyen

ISBN: 978 90 361 0820 1

Cover design: *Crasborn Graphic Designers bno, Valkenburg a.d. Geul*

This book is no. **878** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found here.

Market Concentration, Institutional Demand, and Risk Factors in Asset Pricing

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. P.P.C.C. Verbeek

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Aula der Universiteit

op vrijdag 7 november 2025, te 11.00 uur

door Ngoc Minh Thu Nguyen

geboren te Ho Chi Minh

Promotiecommissie:

Promotor:	prof. dr. A. Andonov	Universiteit van Amsterdam
	dr. E. Eiling	Universiteit van Amsterdam
Overige leden:	prof. dr. A.W.A. Boot	Universiteit van Amsterdam
	prof. dr. R.J.A. Laeven	Universiteit van Amsterdam
	dr. S.R. Rottke	Universiteit van Amsterdam
	dr. E. Pazaj	Universiteit van Amsterdam
	prof. dr. F.C.J.M. de Jong	Tilburg University
	dr. E. Smajlbegovic	Erasmus Universiteit Rotterdam
	prof. dr. E.C. Perotti	Universiteit van Amsterdam

Faculteit Economie en Bedrijfskunde

Acknowledgements

My dream leading to this day began in May 2014, exactly 11 years ago. One Thursday morning, my dad read in the newspaper that Auckland International College in New Zealand was organizing scholarship interviews in Vietnam. The next day, he drove me on his scooter to the interview event. Weeks later, I received the scholarship. My mom then drove me on her scooter to complete application forms, visa paperwork, and other administrative tasks. Deciding to go abroad was easy for me because I have always been passionate about learning and exploring the world, but it was not easy for my family to let me leave home at the age of sixteen. Looking back, I am grateful that my family supported my decisions wholeheartedly from the beginning.

During high school in New Zealand, I took Economics for the first time, thanks to Ms. Kitty Keeman, whose inspiring teaching sparked my passion for economics. I still vividly remember her first lesson explaining the demand and supply of a good. Her Dutch background then inspired me to study in the Netherlands and complete my Bachelor's degree in Econometrics and Operations Research at Maastricht University. During my bachelor's years, multiple finance-related internships and my thesis in financial econometrics led me to specialize in finance. During my exchange at UCLA, I attended Prof. Rosa Matzkin's Advanced Econometrics course. Her excellence and dedication to research greatly inspired me to pursue a PhD.

At Tinbergen Institute, my passion narrowed to asset pricing, particularly after meeting my supervisors, Aleksandar Andonov and Esther Eiling. Their Empirical Asset Pricing course was pivotal in shaping my research direction. Pursuing my PhD in asset pricing has been a deeply rewarding experience. Even today, I wake up excited to

learn more in this field. I believe that asset pricing will remain my lifelong passion.

This journey would not have been possible without the invaluable support and inspiration from many people. First, I am deeply grateful to my supervisors, Aleksandar Andonov and Esther Eiling, for their continuous support and guidance. Aleks was always available whenever I needed advice. For example, we communicated on every platform: email, WhatsApp, and, during the early pandemic days, Zoom. I had many "wow" moments during our meetings when Aleks suggested smart explanations that often shifted the direction of my thinking. I have learned not only how to do research but also how to balance teaching, supervision, and academic responsibilities effectively from him. Aleks always stays up to date with the latest developments in the field and has generously shared valuable opportunities with me, such as the Demand System Asset Pricing workshop, which inspired my second paper, and the LTI@UniTo 2024 Asset Pricing Conference, where I was honored with the Best PhD Paper Award.

Esther consistently challenged my thinking, deepening my research by asking many critical questions. Even when I felt thoroughly prepared, I often left her office with ten more tasks on my to-do list. Our brainstorming sessions on her whiteboard significantly shaped my job market paper. Esther always carefully wrote down the feedback during my brownbag seminars and shared it with me afterwards. To this date, I still have separate folders of her notes on paper and digitally. I greatly admire how she thoughtfully considered my papers even outside our meetings, which continuously helped me polish my papers.

At the same time, Aleks and Esther encouraged me to work independently. This gave me the freedom to work on my favorite topics and to learn how to identify opportunities on my own, such as when I won the John A. Doukas PhD Best Paper Award at the EFMA Annual Meeting 2025. Beyond research, I deeply respect Aleks and Esther for their ability to maintain a successful academic career while prioritizing family life.

My sincere appreciation also extends to my dissertation committee for dedicating their valuable time to evaluate my work. Special thanks to Arnoud Boot for his support as the department head and for providing a reference letter for my job market applica-

tions. I am grateful to Roger Laeven, whose Asset Pricing Theory course set my passion for the field. Throughout my PhD years, I benefited tremendously from discussing my work with Simon Rottke, who provided candid advice to become a young researcher. Elisa Pazaj has been a role model to me, exemplifying success both professionally and personally as a female researcher.

I am deeply thankful to Enrico Perotti for attending all my brownbag seminars and offering valuable feedback, particularly on the intuitions. I truly enjoyed our PhD and reading group gatherings at Enrico's home. Frank de Jong's lunchtime chats and discussions were always a great time. Esad Smajlbegovic provided valuable guidance from my master's thesis through my PhD defense and offered extensive support throughout my job market journey.

I warmly thank the other faculty and staff members in the Finance Group. Spyros Terovitis generously offered mock interviews and made teaching Financial Economics and Quantitative Methods particularly efficient and enjoyable. Florian Peters provided thoughtful paper feedback and engaging dinner conversations. As the department head, Jeroen Ligterink fostered a supportive and welcoming departmental environment. Special thanks go to Martijn Dröes, Tolga Caskurlu, Torsten Jochem, Tomislav Ladika, Paul Smeets, Vladimir Vladimirov, Yenan Wang, Stefan Arping, Evgenia Zhivotova, Theresa Spickers, Philippe Versijp, Jens Martin, Pepijn Trietsch, and Sabine Ferdik for their invaluable support in teaching and job-market preparations. I also deeply appreciate the warm interactions and generous support from the faculty and staff at Amsterdam Law School, especially Fatjon Kaja, Edoardo Martino, Hossein Nabilou, Alessio Pacces, and Lorena Zevedei.

Outside the University of Amsterdam, I am particularly grateful to my co-author, Dongryeol Lee, whom I met through an online workshop by Ralph Koijen and Moto Yogo. Our collaboration, spanning two years without physical meetings, has greatly enhanced both my productivity and my approach to research. I also thank Christian Schlag, Winston Wei Dou, and Chi-Yang Tsou, whose insightful feedback considerably improved my research and shaped my professional attitude during the job market.

Many wonderful friendships have deeply enriched my PhD experience. Special thanks to my "Mermaid" group – Eva Mynott, Yasmine van der Straten, and Xinyi Wang – for our unforgettable trips to EFA Annual Meetings. My best friend Minh Phuong Tran, thank you for always cheering me up and being there whenever I needed you. Many thanks to my friends from the PhD Council and my office: Robin Aarts, Samara Cooper, Lisa Timm, and Danjun Xu, as well as friends and staff at Tinbergen Institute, especially Antonia Kurz, Aishameriane Schmidt, Christina Månsson, Ester van den Bragt, Arianne de Jong, and Eric Bartelsman. The Institute has been instrumental in preparing me both intellectually and mentally for this PhD journey.

Finally, my deepest gratitude goes to my family. To my dear boyfriend, Huu Tung Nguyen, meeting you during my first year of PhD brought immense joy to my life. Your support has helped me maintain a healthy work-life balance. To my beloved parents, Hue Ngoc Lu and Thanh Minh Nguyen, your unconditional trust and encouragement empowered me to explore and push beyond my limits. To my dear sister, Ngoc Minh Tam Nguyen, thank you for your support, especially for taking care of our Mom and Dad while I pursued my dreams abroad. My brother-in-law, Dung Nam Nguyen, and my cousins, Xuan Cat Tien Nguyen and Thien Bao Nguyen, thank you for always cheering me up and warmly welcoming me home in Vietnam.

Reflecting on my PhD journey fills me with happiness and gratitude. Thank you all from the bottom of my heart for accompanying and enriching my path.

Cám ơn Ba và Má!

Thanks, Dad and Mom!

List of Authors

Market Concentration, Capital Misallocation, and Asset Pricing

This chapter is single-authored.

Intermediary Asset Pricing Through the Lens of a Demand System

This chapter is co-authored with Dongryeol Lee.

I contributed the data sets, data construction, and the majority of the writing. Dongryeol Lee contributed to the methodology, code implementation, and provided feedback on writing. We collaborated on the research idea, programming, interpretation of results, and final revisions.

Overlapping Factors

This chapter is single-authored.

Financial Support

I am grateful to the Amsterdam University Fund for the financial support provided throughout my PhD program. This support has enabled me to pursue my research, attend conferences, and prepare for a successful academic career.

Contents

Chapter 1

Introduction

This dissertation explores the asset pricing implications of several changes in financial markets over recent decades. As the U.S. stock market has evolved, it has become increasingly dominated by a small set of large firms, so-called "superstars," while the number of publicly listed firms has steadily declined. Simultaneously, institutional investors, such as broker-dealers and hedge funds, now play a central role in market making, and factor-based investing has become an important strategy in asset management. These developments raise several important questions for asset pricing: How do dominant firms influence capital allocation and risk premia? What role does intermediary financial health play in shaping demand and price dynamics? And how should we interpret the overlap among factor-based investment strategies?

The goal of this dissertation is to address these questions by examining how market concentration, institutional demand, and risk factors contribute to asset pricing. Each chapter in this dissertation focuses on one of these topics. The common theme is that these market trends, whether related to firm dominance, financial intermediation, or investment strategies, link to frictions that distort asset prices and generate risk premia.

Chapter 2 investigates the implications of superstar firms on asset pricing. The rise of superstar firms, which dominate their markets in size and market power, has generated much attention in macroeconomics (e.g., Hsieh and Klenow, 2009; Gutierrez and Philippon, 2017; Autor, Dorn, Katz, Patterson, and Van Reenen, 2020). These firms

are associated with lower labor shares (Autor, Dorn, Katz, Patterson, and Van Reenen, 2020), lower aggregate output (Schlingemann and Stulz, 2022), lower aggregate productivity (Gutierrez and Philippon, 2019), and less efficient capital allocation (Bae, Bailey, and Kang, 2021). However, little is known about how their dominance affects the pricing of risks in financial markets. This chapter addresses this gap by looking at the consequences of superstar firms on efficient capital allocation and proposing capital misallocation between superstar and non-superstar firms as a priced macroeconomic risk factor. With quarterly U.S. firm-level data from 1975 to 2023, I use a common proxy for capital misallocation, i.e., the cross-sectional dispersion in marginal product of capital (MPK), and decompose it into misallocation within superstar firms, within non-superstar firms, and between the two groups – termed "MPK spread."

The intuition is that, in an efficient market, production capital flows to the most productive firms. If superstar firms disproportionately absorb capital due to their size or market power, the result is misallocation that dampens long-run economic growth. I show that changes in the MPK spread, which capture capital misallocation between superstar and non-superstar firms, are robustly priced in the cross-section of stock returns. Specifically, stocks with negative exposure to changes in the MPK spread outperform stocks with positive exposure by 4.8% annually. Moreover, I show that the MPK spread predicts long-run macroeconomic indicators such as consumption growth, industrial production, employment, and innovation, particularly among non-superstar firms.

This chapter contributes to the literature in three ways. First, it links a macroeconomic consequence on efficient capital allocation caused by superstar firms to asset pricing. Second, it introduces the MPK spread as a key state variable consistent with the Intertemporal Capital Asset Pricing Model (ICAPM). Third, it emphasizes that the only component of aggregate capital misallocation with predictive power for future economic growth and the pricing power in the cross-section of stock returns is capital misallocation between superstar and non-superstar firms.

Chapter 3 turns to the role of financial intermediaries, specifically broker-dealers

and hedge funds, in shaping asset prices through leverage constraints. The asset pricing literature has recognized the importance of intermediary constraints such as capital and leverage in risk pricing (e.g., Brunnermeier and Pedersen, 2009; He and Krishnamurthy, 2013; Adrian, Etula, and Muir, 2014; He, Kelly, and Manela, 2017). However, to the best of our knowledge, limited empirical evidence has quantified how constraints on intermediaries affect asset demand and asset prices. In this chapter, co-authored with Dongryeol Lee, we apply the demand system approach from Koijen and Yogo (2019) using institutional ownership data to estimate how changes in broker-dealer leverage influence the demand elasticity of different types of institutional investors. By estimating the price impacts due to each type of investors, we develop a framework that can explain how seemingly small shifts in leverage can amplify effects on the cross-sectional variation of asset returns.

The main intuition is that intermediaries play a key role in absorbing risk in financial markets. When their leverage is constrained, for instance, during economic downturns, they reduce their demand for risky assets, reducing asset prices even in the absence of changes to fundamentals. We find that a one-standard-deviation reduction in broker-dealer leverage leads to an average price decline of 1.6%, with stronger effects among stocks disproportionately held by hedge funds. The chapter contributes to the literature in two ways. First, it introduces a demand system that quantifies intermediary-induced price pressures. Second, it highlights the importance of broker-dealer balance sheets in understanding time-varying risk premia through the hedge fund channel.

Next, Chapter 4 examines the phenomenon of factor overlap in asset pricing models. Factor investing, in which investors construct portfolios based on stock characteristics such as value, momentum, and idiosyncratic volatility, has become widespread among institutional investors (e.g., Fama and French, 1993a; Asness, Moskowitz, and Pedersen, 2013; Hou, Xue, and Zhang, 2014). While each factor is constructed based on a unique characteristic, many stocks are simultaneously included in multiple factor portfolios. This chapter proposes that such overlap is not random but economically meaningful: Stocks with high factor overlap are exposed to multiple priced risk sources

and therefore exhibit stronger performance.

To test this hypothesis, I create a stock-level measure of factor overlap and sort stocks into decile portfolios based on their degree of overlap. I find that stocks with a high degree of overlap earn significantly higher returns than those with no overlap – up to 3% annually – and have stronger pricing power in the Fama-MacBeth two-stage regressions across different robust tests. The intuition behind this result is straightforward: Overlapping stocks are jointly exposed to various risk factors and thus provide greater compensation for bearing systematic risk. However, surprisingly, pure factor stocks that are included only in a single factor display no abnormal returns, although they, on average, have a higher characteristic spread than overlapping factor stocks. This chapter contributes to the literature by (1) developing a new empirical measure of factor overlap and (2) demonstrating that such overlap is a source of return predictability.

Together, the three chapters examine how several trends in financial markets, such as rising market concentration, financial intermediation, and factor-based investing, shape asset pricing. They demonstrate that capital misallocation and intermediary constraints introduce distinct sources of priced risk that are not easily captured by traditional models. On the other hand, the commonality across seemingly unrelated factors, through overlapping stocks, is a key driver of factor returns, challenging the conventional method in asset pricing of constructing factors independently.

The remaining structure of this dissertation is as follows. Chapter 2 focuses on market concentration, capital misallocation, and asset pricing. Chapter 3 studies asset pricing through the lens of a demand system shaped by broker-dealer leverage. Chapter 4 turns to overlapping factors and their implications for risk premia.

Chapter 2

Market Concentration, Capital Misallocation, and Asset Pricing[1]

Abstract

Superstar firms, which dominate stock markets through their large size and high markups, can distort efficient capital allocation. This paper empirically studies the asset pricing implications of superstar firms through the channel of capital misallocation, measured as the cross-sectional dispersion in the marginal product of capital (MPK). I decompose this measure into misallocation (1) among superstars, (2) among other firms, and (3) between these two groups. I find that only changes in the third component, termed the "MPK spread," are negatively priced in the cross-section of stock returns. Stocks with negative exposure to these changes outperform stocks with positive exposure by 4.1% per year. In the long run, a higher MPK spread predicts lower economic growth and aggregate stock returns, while in the short run, it predicts lower innovation growth. Consistent with the ICAPM framework, capital misallocation between superstar and non-superstar firms is a key state variable, and its changes capture a macroeconomic risk factor.

Keywords: Superstar firms, capital misallocation, economic growth, cross-sectional asset pricing.

JEL codes: E22, G11, G12, G14.

[1] I am grateful to my supervisors, Aleksandar Andonov and Esther Eiling, for their continuous support. I thank Dennis Bonam, Winston Wei Dou, Nan Li, Alberto Manconi, Agnieszka Markiewicz, Elisa Pazaj, Florian Peters, Simon Rottke, Majeed Simaan, Esad Smajlbegovic, Di Tian, Chi-Yang Tsou, Fabio Trojani, participants at the ICMA Doctoral Finance Symposium 2024, CREDIT 2024, 7th Asset Pricing Conference by LTI@UniTO 2024, Nederlandse Economendag 2024, AFA/ASSA Annual Meeting 2025, EAYE Annual Meeting 2025, EFMA Annual Meeting 2025, and the Finance Group Brownbag Seminar at the University of Amsterdam for helpful comments and discussions.

2.1 Introduction

In recent decades, the US stock markets have become increasingly concentrated. Not only did the number of listed firms decline by 50% between 1996 and 2012 (Doidge, Karolyi, and Stulz, 2017), but also a few numbers of highly successful firms, so-called "superstar" firms, have dominated the market (Kwon, Ma, and Zimmermann, 2024). These developments have led to several macroeconomic and corporate consequences, such as the decline in aggregate labor share (e.g., Autor, Dorn, Katz, Patterson, and Van Reenen, 2020) and contribution to GDP (e.g., Schlingemann and Stulz, 2022). However, little is known about the implications for asset pricing.

This paper shows the asset pricing consequences of superstar firms through the channel of capital misallocation. Bae, Bailey, and Kang (2021) find that stock market concentration is associated with capital misallocation as it impedes competition and innovation. As higher capital misallocation hinders economic growth (David, Hopenhayn, and Venkateswaran, 2016; Dou, Ji, Tian, and Wang, 2023), capital misallocation could be a candidate state variable that predicts changes in investment opportunities in the Intertemporal CAPM (ICAPM) framework.

My paper finds that capital misallocation between superstar and non-superstar firms is a key state variable and its changes capture a macroeconomic risk factor. Using the quarterly data of US-listed firms from Compustat from 1975:Q1 to 2023:Q4, I measure capital misallocation as the cross-sectional dispersion in the firm-level marginal product of capital (MPK) (David and Venkateswaran, 2019) and David, Schmid, and Zeke (2022). I decompose this measure into three components: (1) misallocation among superstar firms, (1) misallocation among non-superstar firms, and (3) misallocation between these two groups referred to as "MPK spread." I construct changes in the aggregate misallocation and each component and show three key results.

First, only changes in the MPK spread, indicating capital misallocation between superstar and non-superstar firms, are a priced risk factor. In different cross-sections of stock returns, changes in the MPK spread carry a negative price of risk, implying

that stocks with negative exposure to the changes earn a higher expected return. These stocks depreciate when changes in the MPK spread rise, making them more risky. In contrast, stocks with positive exposure to these changes earn a lower expected return. Their returns tend to increase when changes in the MPK spread rise, making them a hedge. The long-short portfolio sorted on individual stock exposure to changes in the MPK spread, referred to as the MPK spread-mimicking portfolio, earns an expected return of -4.1% per year.

Second, in contrast, changes in the aggregate capital misallocation and the other components do not carry significant prices of risk. A spanning test analysis confirms the dominance of the MPK spread over other components. Specifically, only the MPK spread-mimicking portfolio spans the portfolio exposed to changes in aggregate misallocation. Whereas, the portfolios exposed to changes in aggregate misallocation and other components do not span the MPK spread-mimicking portfolio.

Third, only the MPK spread is a key state variable as it predicts future economic growth. In the long term, a higher MPK spread predicts lower consumption growth, industrial production growth, and employment growth. In the short term, a higher MPK spread predicts lower aggregate innovation growth and innovation growth of non-superstar firms. In contrast, aggregate capital misallocation as well as the other components do not yield any predictive power. Therefore, only capital misallocation between superstar and non-superstar firms is a candidate state variable.

Consistent with the ICAPM, the MPK spread satisfies three restrictions proposed by Maio and Santa-Clara (2012). First, the MPK spread as a state variable negatively forecasts changes in investment opportunities, proxied by the aggregate stock returns. Second, changes in the MPK spread as a factor earn a negative price of risk in cross-sectional tests, consistent with the sign of forecast. Third, in the multi-factor model that includes market and changes in the MPK spread as the factors, the market price of risk estimated from the cross-sectional tests is economically plausible as an estimate of the coefficient of relative risk aversion (RRA) of the representative investor.

These findings imply that a higher spread in the mean productive use of produc-

tion capital between superstar and non-superstar firms discourages long-run economic growth. Thus, changes in capital misallocation between superstar and non-superstar firms represent negative news to investors whose marginal utility depends on consumption growth risks. In the ICAPM framework, these changes capture a macroeconomic risk factor, highlighting superstar firms' role in shaping the risk premium associated with capital misallocation.

Following Autor, Dorn, Katz, Patterson, and Van Reenen (2020) and Cheng, Vyas, Wittenberg-Moerman, and Zhao (2024), I identify superstar firms as the top 5% firms in their industries based on market capitalization and market power. I estimate firm-level market power as markup, proxied by the ratio of sales to the cost of goods sold, using the method from De Loecker, Eeckhout, and Unger (2020)[2]. I measure firm-level MPK as the log ratio of output-to-capital. Capital includes both tangible capital (net property, plant, and equipment) and intangible capital estimated from Eisfeldt and Papanikolaou (2013) and Eisfeldt, Kim, and Papanikolaou (2020). Aggregate capital misallocation is then the dispersion in MPK across firms in each quarter. A time series of the changes in misallocation shows that periods of high capital misallocation coincide with recessions, consistent with the countercyclical pattern of capital misallocation documented in the literature (e.g., Eisfeldt and Rampini, 2006; Bloom, 2009; Bachmann and Bayer, 2014; Kehrig, 2015; David, Schmid, and Zeke, 2022).

To test the role of superstar firms, I examine the pricing of changes in aggregate misallocation, misallocation among superstar firms, misallocation among non-superstar firms, and the MPK spread. The cross-sectional Fama-MacBeth regressions show that changes in the MPK spread are significantly and negatively priced in the cross-section of 25 size$\times$book-to-market, 10 momentum, 25 size$\times$investment, 25 size$\times$operating profitability portfolios. This finding is robust to using Giglio and Xiu (2021)'s 202 portfolios. Whereas, the pricing of changes in aggregate misallocation and the other two components are not robust and consistent across different models and test portfolios.

Next, I create mimicking portfolios to examine individual stock exposure to changes

[2] Additionally, I also impose the condition that superstar firms have to increase their markup shares within their industries compared to the previous 12 quarters (3 years).

in aggregate misallocation and each component. For each stock, I regress the quarterly excess returns on the changes using a rolling window of 20 quarters. Then I sort stocks into quintiles based on their beta estimates in each quarter. The value-weighted portfolios show that the average excess returns decrease as the stock exposure to changes in the MPK spread increases. The long-short portfolio has an average annual excess return of -4.1%. The abnormal returns (alphas) of the long-short portfolio estimated from the CAPM, the Fama and French (1992b, 1993b)'s three-factor, and the Fama and French (2015a)'s five-factor models are significantly negative. The post-formation portfolios also show a negative beta in the lowest quintile and a positive beta in the highest quintile. These findings are robust when I construct an equally weighted portfolio.

Finally, to test whether the aggregate misallocation or its components is a state variable, I run the standard predictive regression of several future macroeconomic variables on the lagged aggregate misallocation and its components. I find that only the MPK spread predicts negative per capita real consumption growth (nondurables and services), industrial production growth, employment growth, and stock market returns in the next 4 to 12 quarters (1 to 3 years). The MPK spread also predicts negative aggregate innovation growth, especially innovation growth of non-superstar firms in the next 1 to 5 quarters. Thus, only capital misallocation between superstars and non-superstars shows both pricing power in the cross-section and predictability power in the time series.

The main asset pricing results are robust to various specifications. Alternative definitions of superstars based on market capitalization or sales show that the top 50 firms confirm the results in the cross-sectional asset pricing tests and predictability. Besides, the results are robust to using different industry classifications, a value-weighted measure for misallocation, separating misallocation by tangible and intangible capital, the pre-2000s and post-2000s subsamples, and using annual frequency.

The contribution of this paper is threefold. First, I link the macroeconomic consequences of superstars to asset pricing. Second, I propose capital misallocation between superstars and non-superstars as a state variable and its changes as a priced risk fac-

tor. Third, my findings highlight the role of superstars in shaping the price of risk associated with capital misallocation: Superstar firms, resulting in higher capital misallocation, could prevent innovation growth and deter economic growth, consistent with Bae, Bailey, and Kang (2021) and Kung and Schmid (2015). Investors are willing to pay a sizeable premium to eliminate such long-run uncertainty about economic growth.

Related literature. My paper contributes to three strands of literature. First, the literature has documented several macroeconomic consequences of superstar firms. For example, Autor, Dorn, Katz, Patterson, and Van Reenen (2020) argue that the rise of superstar firms leads to a decline in the aggregate labor share[3]. Schlingemann and Stulz (2022) argue that the market capitalization of listed firms has become less informative about firms' contribution to aggregate output[4]. Gutierrez and Philippon (2019) also find that the contribution of superstar firms to aggregate productivity growth has also decreased. My paper contributes to this literature by linking one consequence of superstar firms to asset pricing.

Particularly, there has been evidence that superstar firms deter efficient capital allocation. For instance, Bae, Bailey, and Kang (2021) find that a concentrated stock market is less likely to allocate capital to firms that could use capital more efficiently. Su (2022) also shows that capital misallocation increases when the economy includes superstar firms[5]. De Loecker, Eeckhout, and Unger (2020) and De Loecker and Eeckhout (2021) show that economic activity is more reallocated toward superstar firms. Neuhann and Sockin (2024) suggest that financial market concentration can dampen efficient capital

[3] Technological advances push sales toward the most successful firms in the service industry whose contribution to labor share is relatively low, leading to a decline in the aggregate labor share. Gutierrez and Philippon (2020) also find that the contribution of superstar firms to aggregate labor productivity has decreased.

[4] This is due to the decline in the manufacturing industry oriented in tangible capital and the shift towards the service industry which favors intangible capital.

[5] When firms face uncertainty in their product quality, superstar firms become riskier as they face more volatile fluctuations in their markups.

allocation[6]. Based on this evidence, my paper shows that changes in capital misallocation between superstar firms and non-superstar firms capture an asset pricing factor.

Second, several papers have shown that markups are an important source of capital misallocation. Peters (2020) finds that older firms improve their productivity away from competitors, raising the markups that they can optimally charge on their existing products, consistent with the superstar phenomenon. Capital misallocation rises as a result of a decline in creative destruction. David and Venkateswaran (2019) estimate that markup heterogeneity explains a large share of 14% in the capital misallocation in the US. Consistently, I find that rising markup yields a higher MPK level for the superstars[7].

Furthermore, capital misallocation is an important deterrent to economic growth (Eisfeldt and Rampini, 2006, 2008; Eisfeldt and Shi, 2018). My paper adds to the literature by looking at different components of an empirical measure of capital misallocation. I find that only the component that captures the misallocation between superstar firms and the rest of the firms negatively predicts proxies for economic growth. Other components do not have predictive power for economic growth.

Third, recent papers document asset pricing implications of superstar firms and market concentration. For example, Emery and Koëter (2024) show that rising stock market concentration increases the size premium. Cheng, Vyas, Wittenberg-Moerman, and Zhao (2024) show in product markets that firms with high exposure to superstars experience weaker financial performance and higher risk. Hou and Robinson (2006)

[6] Other papers also support the fact that the presence of superstars creates market inefficiency. For instance, Grullon, Larkin, and Michaely (2019) show firms in concentrated industries become more profitable by higher profit margins, rather than higher productivity. Thus, market concentration can yield profits to winners rather than enhancing the whole economy. Covarrubias, Gutierrez, and Philippon (2020) document a "bad concentration" in the market after the 2000s, since concentration increases the barrier to entry and reduces innovation. Gopinath, Kalemli-Özcan, Karabarbounis, and Villegas-Sanchez (2017) also document an increase in capital misallocation in South Europe and observe that firms with higher net worth, but not necessarily higher productivity, attract more capital, causing productivity losses.

[7] Baqaee and Farhi (2020) show that reallocation of market share toward high-markup firms improves the aggregate TFP growth in the U.S. The intuition is that firms with high markups were too small to begin with, so the reallocation of labor and capital toward these firms improves TFP growth over time. My paper studies superstar firms that are very large in terms of size and markup each quarter and considers capital misallocation in a static production function.

shows at the industry level that firms in more concentrated industries earn lower expected returns. My paper finds evidence for market-wide concentration and that stocks negatively (positively) exposed to changes in misallocation between superstar and non-superstar firms earn a higher (lower) expected return.

The mechanism of my paper is closely related to Dou, Ji, Tian, and Wang (2023) who show that changes in aggregate misallocation carry a negative price of risk. As capital misallocation prevents optimal R&D and innovation, capital misallocation captures news about long-run economic growth[8]. The source of friction that drives misallocation is financial constraints, while my paper studies size and market power as the main sources of friction.

Finally, my paper examines the effect of capital misallocation on risk premia. Whereas, David, Schmid, and Zeke (2022) study the role of risks in generating misallocation and find that firms' exposure to systematic risk is an important source of misallocation. Misallocation increases in times when risk premia are high and thus is countercyclical. My paper confirms the countercyclicality of the capital misallocation between superstar and non-superstar firms.

The remaining organization of the paper is as follows. Section 2.2 discusses in detail the data and variable construction and derives a decomposition based on superstar vs non-superstar portfolios. Section 2.3 shows main cross-sectional asset pricing results. Section 2.4 examines the predictability for future economic growth and tests the restrictions for ICAPM. Finally, Section 2.5 discusses different tests for robustness, and Section 2.6 concludes.

2.2 Data construction

In this section, I discuss the data and describe the method to identify superstar firms. I discuss the empirical measure of capital misallocation. Furthermore, I decompose the aggregate capital misallocation into three components: misallocation among super-

[8] Changes in misallocation are a risk factor as capital misallocation influences the investors' stochastic discount factor (SDF) through its effect on consumption growth.

star firms, misallocation among other firms, and misallocation between superstars and other firms.

2.2.1 Sample and variable construction

The sample contains CRSP common stocks (share codes 10 or 11) traded on NYSE, AMEX, or Nasdaq (exchange codes 1, 2, or 3). Following the standard literature, I exclude the financial sector (SIC 6,000 - 6,999), utilities (SIC 4,900 - 4,999), and public administration (SIC 9,000 - 9,999), since firms in these sectors have little capital. Then, I compute quarterly stock returns from the monthly stock file to merge with firm characteristics from Compustat Fundamentals Quarterly. I take stock returns as of the end of a quarter and accounting variables at the end of the previous quarter to ensure that accounting data are public on the trading date and that market participants have access to accounting variables. The results are also robust to without this timing convention.

I obtain firm market capitalization from Compustat (mkvaltq), which equals price times shares outstanding (prccq $\times$ cshoq) if the value is missing. The resulting market cap is closely identical to when I compute price times shares outstanding at quarter-end from CRSP. All nominal variables are adjusted for inflation using the GDP deflator (GDPDEF series from FRED at quarterly frequency). I also exclude micro-cap stocks whose prices are less than \$1 and observations with negative values for sale (Compustat saleq) and physical capital (ppentq). The final sample contains 590,154 unique firm-quarter observations from 1975:Q1 to 2023:Q4, as capital is only available for most firms after 1974:Q4.

The variable for intangible capital (intanq) is often missing in Compustat. Therefore, I estimate intangible capital using the perpetual inventory method from Eisfeldt and Papanikolaou (2013) and Eisfeldt, Kim, and Papanikolaou (2020). Specifically,

$$INT_{it} = (1-\delta)INT_{it-1} + SG\&A_{it} \tag{2.1}$$

where the initial value is $INT_{i0} = SG\&A_{i1}/(g+\delta), g = 0.1$, and $\delta = 0.2$. The estimation

uses 100% SG&A (Compustat xsgaq). The Appendix reports the robust result when using 30% of (SG&A minus R&D) plus 100% of R&D as in Peters and Taylor (2017).

Following David, Schmid, and Zeke (2022) and David, Hopenhayn, and Venkateswaran (2016), I compute the firm-level MPK as the log ratio of output to lagged capital[9]. I use sales as output and net property, plant, and equipment (ppentq) as tangible capital plus intangible capital estimated from Eisfeldt and Papanikolaou (2013). I exclude firm-quarter observations with a ratio of intangible to tangible capital exceeding 10, because unlikely it is more than 10 times the mean ratio as discussed in David, Schmid, and Zeke (2022). Besides, I winsorize the log MPK at 1% and 99% in each quarter to avoid outliers.

In the main analysis, the test portfolios for the Fama-MacBeth regressions include 25 size and book-to-market, 10 momentum, 25 size and investment, 25 size and operating profitability sorted portfolios. In the robustness section, I use the 202 portfolios used in Giglio and Xiu (2021): 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolios sorted by size and momentum. I retrieve all factor data at a quarterly frequency and annual frequency for a robust check from Kenneth French's data library[10].

In addition, I construct several macroeconomic variables. The per capita real consumption is the total real consumption of nondurable goods and services, divided by the total population, obtained from the US Bureau of Economic Analysis (BEA). The industrial production index measures the real output of all relevant establishments in the US, retrieved from the Federal Reserve Bank of St. Louis (FRED). The monthly aggregate US unemployment rate is from the Bureau of Labor Statistics (BLS), compounded to quarterly frequency.

Besides, following Bae, Bailey, and Kang (2021), I construct the aggregate innovation proxy as a natural logarithm of one plus the number of patents divided by the popu-

[9] Changes in capital misallocation capture the recessions better when we take the lagged capital.
[10] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

lation. The aggregate number of patents are from the All Technology (Utility Patents) Reports from the US Patent and Trademark Office (USPTO)[11]. Similarly, I compute the innovation proxy in each portfolio as the natural logarithm of one plus the total number of patents divided by the total market capitalization in each portfolio. I use firm-level patent data from Kogan, Papanikolaou, Seru, and Stoffman (2017).

2.2.2 Identifying superstar firms

The literature often characterizes superstars as a top number of highly successful firms in the economy. For example, Bae, Bailey, and Kang (2021) select the largest 5 or 10 firms in each country based on their market cap at the end of each year. Schlingemann and Stulz (2022) select the top 1, 3, 5, or 10 firms in the whole economy or 3 in each of Fama-French's 48 industries based on their market cap each year. Although there is no unified definition, in this paper, I identify superstars based on Autor, Dorn, Katz, Patterson, and Van Reenen (2020)'s discussion that superstars are firms *increasingly dominating* their industries, leading to rising concentration.

In each quarter, I select the top 5% firms in their SIC two-digit industries based on their market power. To proxy firms' market power, I estimate the firm-level markup ratio from De Loecker, Eeckhout, and Unger (2020), i.e. 0.85 times the ratio of sale to the cost of goods sold ($0.85 \times$ saleq/cogsq). De Loecker, Eeckhout, and Unger (2020) show that the output elasticity of variable input 0.85 is time-invariant and contributes little to the rise in markups, so I also fix this coefficient across all stocks. In some cases, the industries include less than 20 firms in a given quarter, so I follow Cheng, Vyas, Wittenberg-Moerman, and Zhao (2024) to assign the largest markup firm as the superstar in that industry.

As many papers identify mega-firms such as Apple and Amazon as superstars, I multiply this markup ratio by the firm-level market cap to take into account the size

[11] https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_at.htm

of firms ($0.85 \times$ saleq/cogsq $\times$ market cap)[12]. Finally, I also impose the condition that superstar firms have to increase their markup shares within their industries compared to the previous 12 quarters (3 years) to follow that intuition from Autor, Dorn, Katz, Patterson, and Van Reenen (2020) that superstars *increasingly dominate* the market[13].

2.2.3 Capital misallocation

In the absence of any friction, MPK should equalize across firms. The MPK dispersion can help to measure capital misallocation because it is proportional to aggregate TFP loss (Midrigan and Xu, 2014). Capital misallocation is defined as the firm-level log MPK dispersion. Following David, Schmid, and Zeke (2022) and David, Hopenhayn, and Venkateswaran (2016), I compute the misallocation as the cross-sectional variance across all firm MPKs within each quarter t as

$$\sigma^2_{mpk,t} = \frac{1}{N-1}\sum_{i=1}^{N}(mpk_{it} - \mu_{mpk,t})^2. \tag{2.2}$$

where $\sigma^2_{mpk,t}$ and $\mu_{mpk,t} = \frac{1}{N}\sum_{i=1}^{N} mpk_{it}$ denote the variance and the mean MPK of the economy.

The intuition of the capital misallocation is as follows. To minimize the total factor productivity (TFP) loss, the economy needs to allocate capital efficiently. The TFP loss is at the minimum if firm-level MPKs equalize, i.e. $\sigma^2_{mpk,t} = 0$, the allocation that maximizes the static production of all firms in the economy. If MPK varies substantially across firms, it implies that the economy forgoes the opportunity to increase the aggregate output by reallocating capital from low MPK to high MPK firms. A lower MPK among firms in the cross-section implies misallocated capital that high MPK firms could have more efficiently utilized.

Finally, capital misallocation, i.e. MPK dispersion, is the cross-sectional variance of

[12] Cheng, Vyas, Wittenberg-Moerman, and Zhao (2024) multiply the market ratio by the sales. Robust results show that my main findings are robust to the choice of either sales or market cap to identify superstars.

[13] Figure A1 shows the rise in market concentration among superstar firms and the rise of aggregate capital misallocation.

firm-level MPK in each quarter. Furthermore, I propose changes in capital misallocation as the log changes of capital misallocation with respect to the 4-quarter lag. I take annual log changes in remove seasonal fluctuations of sales that influence the measure

$$\Delta\sigma^2_{mpk,t} = \sigma^2_{mpk,t} - \sigma^2_{mpk,t-4}. \tag{2.3}$$

Figure 2.1 shows the MPK distribution over time. Panel A shows that the MPK distribution is right-skewed and the distribution becomes more heavily tailed over time. Panel B plots the percentiles of the MPK distribution in each quarter. Superstars are firms in the further right tail of MPK distribution, so the skewness of MPK distribution captures MPK of superstars. Positive changes in the skewness of MPK indicate increases in MPK of superstars relative to other firms, which raises the MPK spread.

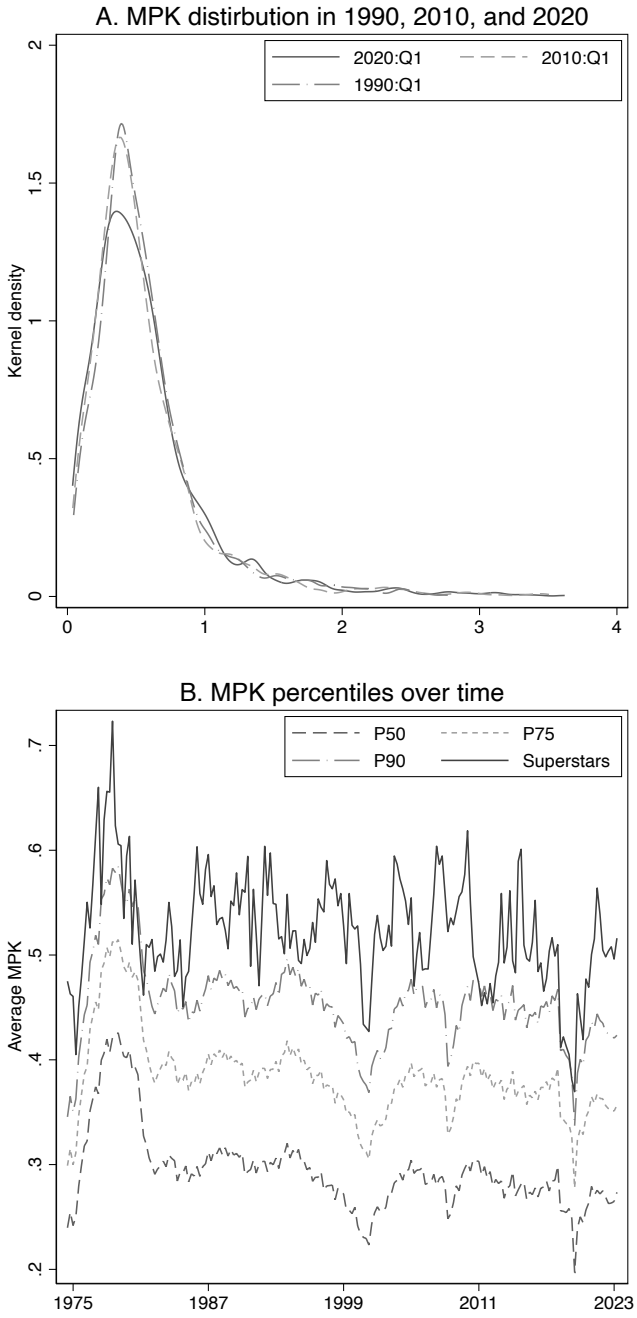
To examine which component of misallocation is important for asset pricing and which subset of firms drives the misallocation, I decompose the capital misallocation into contributions by each subset of firms. In the case of two groups (superstar firms and non-superstar firms), the aggregate MPK dispersion can be decomposed into

$$\underbrace{\sigma^2_{mpk}}_{\text{Total misallocation } (\text{Misall}_{\text{total}})} = \underbrace{\frac{N_0 - 1}{N-1}\sigma^2_{mpk,0}}_{\text{Misallocation within non-superstars } (\text{Misall}_{\text{rest}})} + \underbrace{\frac{N_* - 1}{N-1}\sigma^2_{mpk,*}}_{\text{Misallocation within superstars } (\text{Misall}_{\text{top}})} + \underbrace{\frac{N_0 N_*}{N(N-1)}(\mu_{mpk,0} - \mu_{mpk,*})^2}_{\text{MPK spread}} \tag{2.4}$$

Section A1 in the Appendix shows the derivation. Intuitively, I decompose the capital misallocation into three components: the misallocation among superstars, the misallocation among non-superstars, and the misallocation due to the (squared) difference in the *mean* MPK between the two portfolios, referred to as "MPK spread." Finally, we can decompose the changes in misallocation into

$$\Delta\text{Misall}_{\text{total}} = \Delta\text{Misall}_{\text{rest}} + \Delta\text{Misall}_{\text{top}} + \Delta\text{MPK spread} \tag{2.5}$$

Figure 2.1: **MPK distribution over time**



Description. This figure plots MPK distribution across all firms in each year. Panel A plots the distribution density in 1990, 2010, and 2020. Panel B plots different moments of the distribution of average MPK from 1975:Q1 to 2023:Q4, compared to the average MPK of superstars. Firm-level MPK is the output-to-capital ratio. I use sales as output and net property, plant, and equipment (ppentq) as physical capital plus intangible capital estimated from Eisfeldt and Papanikolaou (2013).

Figure 2.2 plots the time series of the aggregate misallocation and its components. Periods of high misallocation tend to coincide with the NBER recessions, implying the countercyclicality of capital misallocation. Consistent with David, Schmid, and Zeke (2022), MPK dispersion rises during economic downturns. The MPK spread, i.e. the misallocation between superstars and non-superstars, particularly displays clear cyclical patterns.

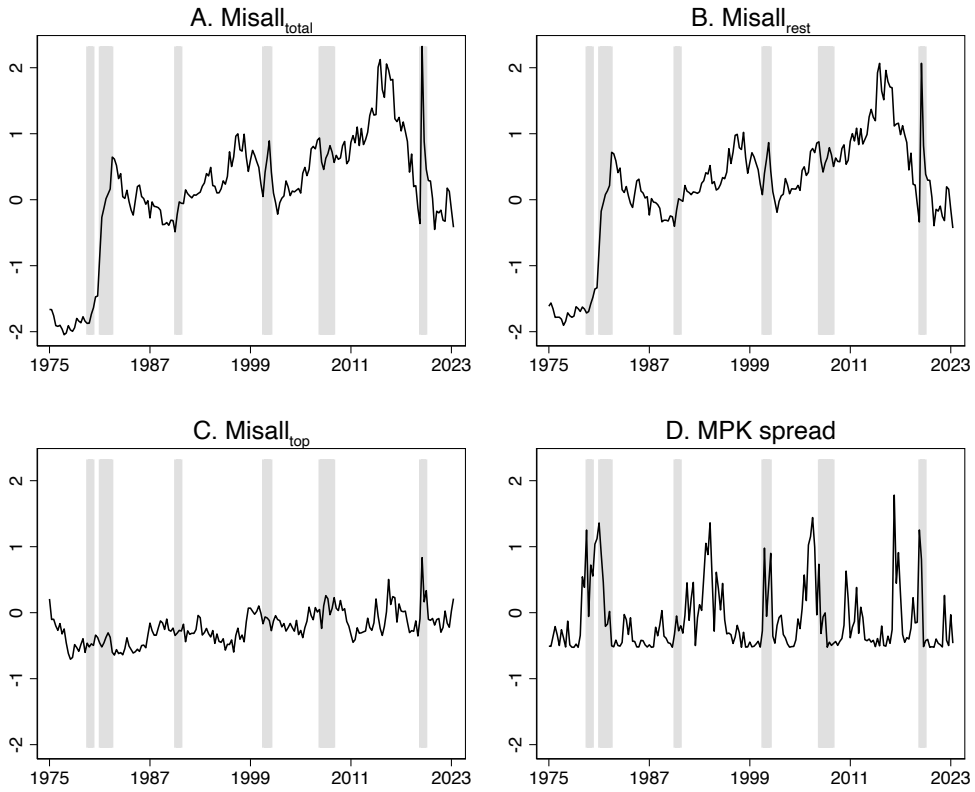
In an efficient market, capital flows towards its most productive uses, i.e. firms with the highest MPK would attract more investment. Intuitively, when superstar firms dominate this capital inflow, it may not always be due to their superior productivity. If these firms' higher MPKs result from market distortions, such as barriers to entry for competitors, then the higher MPK reflects an inefficiency rather than a pure productivity advantage. This misallocation can lead to slower economic growth by preventing capital from reaching potentially innovative but smaller competitors.

2.3 Asset pricing results

Bae, Bailey, and Kang (2021) show that stock market concentration is associated with capital misallocation as it impedes competition and innovation. As higher capital misallocation indicates slower economic growth (David, Hopenhayn, and Venkateswaran, 2016; Dou, Ji, Tian, and Wang, 2023), capital misallocation could be a candidate state variable that predicts changes in investment opportunities under the Intertemporal CAPM (ICAPM) framework.

In this section, I show the cross-sectional asset pricing results of changes in capital misallocation and its components. I first run the Fama-MacBeth regressions and then form the factor-mimicking portfolios sorted on individual stock exposure to the changes.

Figure 2.2: **Capital misallocation against NBER recessions**



Description. This figure plots the aggregate misallocation, misallocation to non-superstars, misallocation to superstars, and the MPK spread. The sample is from 1975:Q1 to 2023:Q4. In each quarter, capital misallocation $\sigma^2_{mpk,t}$ is the cross-sectional dispersion of MPK across firms. The aggregate misallocation is be decomposed into:

$$\underbrace{\sigma^2_{mpk}}_{\text{Total misallocation (Misall}_{\text{total}})} = \underbrace{\frac{N_0-1}{N-1}\sigma^2_{mpk,0}}_{\text{Misallocation within non-superstars (Misall}_{\text{rest}})} + \underbrace{\frac{N_*-1}{N-1}\sigma^2_{mpk,*}}_{\text{Misallocation within superstars (Misall}_{\text{top}})} + \underbrace{\frac{N_0 N_*}{N(N-1)}(\mu_{mpk,0}-\mu_{mpk,*})^2}_{\text{MPK spread}}$$

2.3.1 Cross-sectional analysis

In this section, I examine whether the changes in aggregate misallocation and each component are priced in the cross-section of expected stock returns. Following the standard Fama-MacBeth two-stage procedure, for each test portfolio i, I first estimate the factor loadings (β_{ik}) using time-series regressions of portfolio excess returns on the risk factor(s)

$$R^e_{it} = \alpha_i + \sum_k \beta_{ik} f_{kt} + u_{it}, \tag{2.6}$$

where k is the number of risk factors. Second, for each time t, I estimate the price of risk for each factor ($\lambda_{k,t}$) using cross-sectional regressions

$$R^e_{it} = \lambda_{0,t} + \sum_k \lambda_{kt} \hat{\beta}_{ik} + \varepsilon_{i,t}. \tag{2.7}$$

The risk premium for each factor k is then the time-series average of the price of risk $\lambda_k = \frac{1}{T}\sum_{t=1}^{T} \hat{\lambda}_{kt}$. As a convenience to compare the magnitudes of the asset pricing results, I standardize the changes in aggregate misallocation and its components. Table 2.1 reports results for the second-stage regressions. Panel A shows that results in the cross-section of 25 size $\times$ book-to-market and 10 momentum portfolios, the CAPM cannot explain the difference in average returns across portfolios since the price of risk is statistically insignificant in Model (1). The adjusted R-squared is close to zero. In contrast, when Model (2) includes the market factor and changes in aggregate misallocation, the changes in aggregate misallocation have a negative price of risk, although not significant.

When the model includes changes in each component as a separate factor, Model (5), (8), and (9) show that changes in the MPK spread carry a negative price of risk. The price of risk is significantly priced at the 5% level across all models. Adding this factor to the CAPM increases the cross-section R-squared by 65%. In contrast, adding other changes does not improve the fit any further. The findings are consistent in Panel B

Table 2.1: **Cross-sectional asset pricing tests**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Pricing 25 size×book-to-market and 10 momentum portfolios									
Constant	12.090	10.792	10.895	11.405	14.109	13.051	12.544	14.467	13.972
	(3.67)***	(3.25)***	(3.28)***	(3.47)***	(4.26)***	(4.15)***	(4.01)***	(4.42)***	(4.43)***
	[3.67]***	[2.99]***	[3.05]***	[3.14]***	[2.52]**	[2.90]***	[3.07]***	[2.62]***	[2.56]**
MKT	-0.257	-0.057	-0.056	-0.318	-1.191	-0.771	-0.608	-1.253	-1.094
	(-0.25)	(-0.05)	(-0.05)	(-0.31)	(-1.17)	(-0.78)	(-0.62)	(-1.23)	(-1.10)
	[-0.22]	[-0.04]	[-0.04]	[-0.25]	[-0.65]	[-0.50]	[-0.43]	[-0.69]	[-0.60]
$\Delta\text{Misall}_{\text{total}}$		-0.435				0.501	0.342	0.374	
		(-0.99)				(1.36)	(0.97)	(0.89)	
		[-0.90]				[0.94]	[0.73]	[0.52]	
$\Delta\text{Misall}_{\text{rest}}$			-0.410			0.590			-0.034
			(-0.87)			(1.56)			(-0.10)
			[-0.80]			[1.08]			[-0.06]
$\Delta\text{Misall}_{\text{top}}$				-0.353			-0.494		0.205
				(-1.40)			(-1.99)**		(0.91)
				[-1.25]			[-1.50]		[0.52]
ΔMPK spread					-1.077			-1.066	-1.036
					(-3.54)***			(-3.88)***	(-3.71)***
					[-2.09]**			[-2.29]**	[-2.13]**
R^2	0.012	0.064	0.050	0.131	0.668	0.200	0.148	0.669	0.688
Adj. R^2	-0.018	0.006	-0.009	0.077	0.647	0.122	0.066	0.637	0.646
RMSE	2.701	2.669	2.689	2.572	1.590	2.508	2.588	1.613	1.592

Fama-Macbeth t-statistics in parentheses
Shanken t-statistics in square brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 size × book-to-market, 10 momentum, 25 size×investment, and 25 size×operating profitability portfolios. The sample runs from 1975:Q1 to 2023:Q4. Returns and risk premia are reported in percent per year (quarterly percentages multiplied by four).

Table 2.1: **Cross-sectional asset pricing tests - continued**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel B: Pricing 25 size×book-to-market, 25 size×investment, 25 size×operating profitability portfolios									
Constant	8.642	8.574	8.539	9.785	9.186	9.676	9.904	8.998	10.594
	(2.82)***	(2.88)***	(2.86)***	(3.17)***	(3.01)***	(3.13)***	(3.19)***	(3.02)***	(3.38)***
	[2.45]**	[2.51]**	[2.50]**	[2.64]***	[2.44]**	[2.51]**	[2.49]**	[2.44]**	[2.42]**
MKT	0.166	0.183	0.192	-0.120	0.014	-0.090	-0.149	0.059	-0.344
	(0.17)	(0.19)	(0.20)	(-0.12)	(0.01)	(-0.09)	(-0.15)	(0.06)	(-0.35)
	[0.13]	[0.14]	[0.15]	[-0.09]	[0.01]	[-0.07]	[-0.11]	[0.04]	[-0.23]
SMB	0.668	0.667	0.667	0.722	0.661	0.739	0.749	0.657	0.747
	(1.75)*	(1.74)*	(1.74)*	(1.90)*	(1.73)*	(1.95)*	(1.98)**	(1.71)*	(1.97)**
	[1.16]	[1.16]	[1.16]	[1.22]	[1.10]	[1.23]	[1.23]	[1.09]	[1.16]
HML	1.054	1.051	1.049	1.064	0.958	1.053	1.037	0.936	0.897
	(2.23)**	(2.23)**	(2.22)**	(2.25)**	(2.04)**	(2.23)**	(2.20)**	(1.99)**	(1.91)*
	[1.49]	[1.49]	[1.48]	[1.46]	[1.30]	[1.42]	[1.37]	[1.27]	[1.13]
UMD	3.057	3.006	2.964	2.940	3.390	2.395	2.335	3.239	2.522
	(2.90)***	(3.02)***	(2.98)***	(2.75)***	(3.18)***	(2.35)**	(2.30)**	(3.20)***	(2.47)**
	[2.30]**	[2.38]**	[2.35]**	[2.11]**	[2.39]**	[1.73]*	[1.66]*	[2.38]**	[1.65]
ΔMisall$_{\text{total}}$		-0.051				-0.157	-0.174	-0.055	
		(-0.32)				(-0.96)	(-1.05)	(-0.34)	
		[-0.26]				[-0.72]	[-0.78]	[-0.26]	
ΔMisall$_{\text{rest}}$			-0.069			-0.216			-0.261
			(-0.42)			(-1.25)			(-1.48)
			[-0.34]			[-0.95]			[-1.01]
ΔMisall$_{\text{top}}$				0.260			0.405		0.423
				(1.48)			(2.05)**		(2.12)**
				[1.19]			[1.57]		[1.49]
ΔMPK spread					-0.443			-0.462	-0.426
					(-2.89)***			(-3.03)***	(-2.85)***
					[-2.21]**			[-2.31]**	[-1.94]*
R^2	0.719	0.720	0.720	0.737	0.738	0.744	0.751	0.740	0.780
Adj. R^2	0.703	0.699	0.700	0.718	0.718	0.722	0.729	0.717	0.757
RMSE	1.273	1.281	1.280	1.242	1.240	1.232	1.216	1.242	1.152

Fama-Macbeth t-statistics in parentheses

Shanken t-statistics in square brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

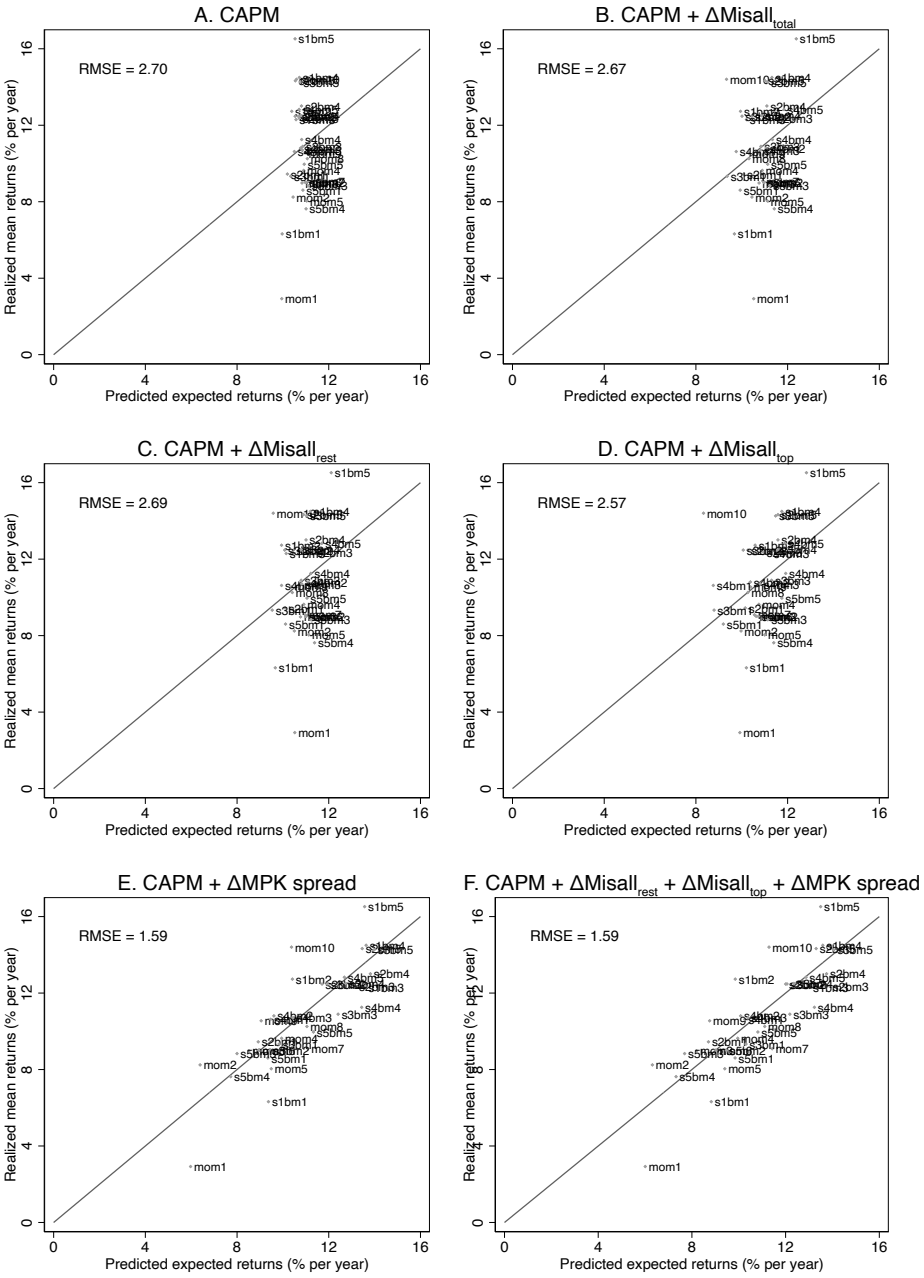
when the models include the Fama and French (1992b, 1993b) three factors and momentum factor and the cross-section includes 25 size $\times$ investment and 25 size $\times$ operating profitability portfolios. While the significance of the price of risk of other decomposed changes is not robust to different t-ratios, the significance of the price of risk of changes in the MPK spread remains consistent.

Intuitively, a positive change in the MPK spread raises the difference in the mean MPK between superstars and non-superstars. Superstars on average have higher MPK, while other firms on average have lower MPK, leading to an increase in the MPK dispersion. The economy as a result operates as a higher capital misallocation. Thus, an increase in capital misallocation represents negative news for investors whose marginal utility depends on future consumption growth. If the rise in the MPK spread lowers economic growth, changes in the MPK spread carry a negative price of risk. While other components do not yield significant results, these results imply that the price of risk of aggregate capital misallocation is mainly driven by the difference in the MPK level of superstars compared with other firms in the economy.

Figure 2.3 plots the realized versus predicted returns from the cross-section of 25 size $\times$ book-to-market and 10 momentum portfolios. Panel E shows that adding changes in the MPK spread to the CAPM reduces the pricing error (RMSE) to 1.59. Panel F reports the same pricing error when the model includes other changes. These results confirm an important role of superstars in asset pricing: The negative price of risk of capital changes in misallocation terms from the difference in the MPK level between the superstars and non-superstars.

The value premium, i.e. firms with high book-to-market have on average higher expected returns, is well documented in the literature. Parker and Julliard (2005b) discuss that value stocks have high average returns because they pay off poorly before and early in recessions, captured by ultimate consumption risk. Thus, the expected excess return of value-minus-growth stocks predicts consumption growth. Since higher capital misallocation also predicts a lower consumption growth rate (Dou, Ji, Tian, and Wang, 2023; David, Hopenhayn, and Venkateswaran, 2016), firms' book-to-market and

Figure 2.3: **Realized versus predicted mean returns**



Description. This figure shows the cross-sectional asset pricing tests from the CAPM. Test portfolios include 25 size $\times$ book-to-market portfolios and 10 momentum portfolios. Each panel plots the realized mean excess returns of the portfolios against the mean excess returns predicted by the CAPM with the changes in aggregate misallocation and its components. The sample runs from 1975:Q1 to 2023:Q4. Returns are in percent per year.

exposure to changes in misallocation must align in the sign direction.

Based on these mechanisms, I examine the factor loadings in the 25 portfolios formed on size $\times$ book-to-market ratio. I test separately the exposure to the changes in aggregate misallocation and changes in each component, controlling for the market risk. If the negative price of risk lines up with the portfolios, mechanically firms with higher book-to-market ratios are more negatively exposed to changes in the MPK spread. In other words, value stocks should have lower betas for changes in the MPK spread than growth stocks.

Table 2.2 reports the first-stage betas across the portfolios. In Panel A, the decreasing trend in the beta of the changes in aggregate misallocation, when portfolios are sorted from low to high book-to-market, reflects the mechanics for the negative price of risk. However, no betas are insignificant. In Panel B, the changes in misallocation in the non-superstar portfolio show no significant betas. Thus, these two factors are seemingly spurious factors. In Panel C, the changes in misallocation in the superstar portfolio, most betas are negative and large but weakly significant. Only changes in the MPK spread shown in Panel D display a clear decreasing trend across the book-to-market portfolios within each size. Most betas (24/25) are significant, consistent with the hypothesis that value stocks load more negative exposure to the changes than growth stocks.

2.3.2 Factor-mimicking portfolios

Next, I create portfolios sorted on individual stock exposure to changes in capital misallocation and each component. To estimate the firm-level exposure, for each stock, I regress the quarterly excess returns either on changes in aggregate misallocation or its components by a rolling window of 20 quarters (with a minimum of 12 quarters available).

$$R^e_{it} = \alpha_i + \beta_{it}\Delta z_t + \varepsilon_{it}, \quad t = t-20 \to t. \quad z \in \{\text{Misall}_{\text{total}}, \text{Misall}_{\text{rest}}, \text{Misall}_{\text{top}}, \text{MPK spread}\} \tag{2.8}$$

Table 2.2: **Exposure in the 25 size × book-to-market portfolios**

Panel A: Loading on $\Delta Misall_{total}$

	β					$t(\beta)$				
	Low	2	3	4	High	Low	2	3	4	High
Small	-0.220	-0.335	-0.666	-3.174	-5.842	-0.041	-0.076	-0.145	-0.652	-0.985
2	-0.647	-1.099	-3.184	-2.498	-2.959	-0.134	-0.250	-0.719	-0.534	-0.529
3	1.237	-0.140	-1.845	-2.488	-3.073	0.269	-0.035	-0.467	-0.537	-0.567
4	0.273	-2.662	-2.317	-3.399	-5.468	0.073	-0.735	-0.613	-0.715	-1.114
Big	0.507	-2.237	-3.021	-3.660	-3.272	0.174	-0.825	-1.021	-1.003	-0.694

Panel B: Loading on $\Delta Misall_{rest}$

	β					$t(\beta)$				
	Low	2	3	4	High	Low	2	3	4	High
Small	0.971	0.678	0.418	-2.144	-4.515	0.177	0.153	0.091	-0.444	-0.771
2	0.411	-0.240	-2.320	-1.524	-1.853	0.084	-0.055	-0.527	-0.329	-0.337
3	1.929	0.641	-0.960	-1.506	-2.159	0.417	0.163	-0.244	-0.329	-0.406
4	0.848	-1.959	-1.527	-2.358	-4.421	0.222	-0.539	-0.406	-0.503	-0.908
Big	0.954	-1.578	-2.483	-2.914	-2.423	0.325	-0.575	-0.840	-0.804	-0.518

Panel C: Loading on $\Delta Misall_{top}$

	β					$t(\beta)$				
	Low	2	3	4	High	Low	2	3	4	High
Small	-7.608	-6.667*	-7.897**	-7.387**	-8.365*	-1.599	-1.673	-2.153	-1.961	-1.845
2	-6.447	-6.946*	-7.139**	-6.877*	-7.681*	-1.538	-1.916	-1.967	-1.882	-1.850
3	-5.878	-6.773**	-5.761*	-6.445*	-7.388*	-1.447	-2.012	-1.724	-1.787	-1.795
4	-5.252	-4.536	-4.652	-6.766**	-6.383*	-1.519	-1.428	-1.474	-1.979	-1.774
Big	-4.117	-3.042	-1.537	-2.283	-5.030	-1.594	-1.317	-0.590	-0.861	-1.336

Panel D: Loading on ΔMPK spread

	β					$t(\beta)$				
	Low	2	3	4	High	Low	2	3	4	High
Small	-12.973**	-11.059**	-11.839**	-12.092**	-16.284**	-2.539	-2.449	-2.560	-2.338	-2.333
2	-11.703**	-9.488**	-10.191**	-11.290**	-12.863**	-2.442	-2.368	-2.358	-2.423	-2.070
3	-6.934*	-8.291**	-10.171***	-11.427**	-10.713*	-1.739	-2.274	-2.700	-2.542	-1.718
4	-5.955	-8.478**	-9.350**	-12.357***	-13.146**	-1.544	-2.453	-2.541	-2.717	-2.507
Big	-4.536*	-7.993***	-7.024**	-9.558**	-10.299**	-1.649	-2.811	-2.346	-2.412	-2.156

Description. This table reports the factor loadings on changes in aggregate misallocation and its components for each of the 25 size × book-to-market portfolios, i.e. the betas from the first stage of Fama-MacBeth regressions. Particularly, for each test portfolio i, I estimate the factor loadings using time-series regression of the excess returns against the risk factor(s): $R^e_{it} = c_i + \sum_k \beta_{ik} f_{kt} + u_{it}$. The sample runs from 1975:Q1 to 2023:Q4.

The stock's exposure to the changes in misallocation equals the misallocation-beta estimated from these regressions. Each quarter, I sort stocks into quintiles based on their misallocation-beta, lagging by one quarter. I hold and rebalance the portfolio every quarter.

Table 2.3 reports the average (expected) returns and the abnormal returns (alpha) when stocks are sorted into portfolios based on the exposure to the MPK spread. Portfolio returns are value-weighted by market cap. Across the portfolios, the average excess returns decrease as the exposure to changes in the MPK spread rises from Quintile 1 (11.3%) to Quintile 5 (7.2%). Importantly, stocks in Quintile 1 have a negative exposure to exposure to changes in the MPK spread. The expected return reduces when misallocation increases, i.e. a positive change in misallocation, making these stocks risky. Hence, they carry higher risk premia. Whereas, stocks in Quintile 5 have a low expected return. These stocks have a positive exposure to changes in the MPK spread. That is, their expected returns increase when misallocation rises, making them a hedge. Thus, these stocks carry lower risk premia. The long-short portfolio (Q5–Q1) has an average excess return of -4.1% per year with a t-statistic of -2.72, consistent with the negative price of risk of changes in the MPK spread.

The long-short portfolio has a negative and significant abnormal return (alpha) in all asset pricing models - the CAPM, Fama and French (1992b, 1993b) three-factor, and the Fama and French (2015a) five-factor models. Interestingly, the market beta is insignificant in the long-short portfolios across all models. The Appendix shows the same set of results for the exposure to the changes in aggregate misallocation and changes in the misallocation within each portfolio. Yet, the results are insignificant. These results strengthen the mechanism for the negative price of risk of the changes in the MPK spread.

Table A2 reports the results using the equally weighted returns. The average returns in each portfolio become higher but the long-short returns remain negative at -3.2% per year and significant at the 5% level. Across the CAPM, Fama and French (1992b, 1993b) three-factor, and the Fama and French (2015a) five-factor models, the alphas are

Table 2.3: **Exposure to changes in the MPK spread (value-weighted results)**

	Q1 (low)	Q2	Q3	Q4	Q5 (high)	Q5–Q1
Panel A: Expected return						
R^e	11.298***	9.468***	9.286***	7.052***	7.158***	-4.140***
	(6.34)	(5.74)	(5.71)	(4.69)	(5.72)	(-2.72)
Panel B: CAPM						
MKT	1.220***	0.983***	0.915***	0.988***	1.182***	-0.038
	(23.05)	(46.49)	(38.53)	(39.35)	(28.79)	(-0.63)
α_{CAPM}	0.594	0.843	1.258*	-1.617**	-3.213**	-3.807**
	(0.41)	(1.16)	(1.79)	(-2.08)	(-2.57)	(-2.12)
Panel C: FF3 + UMD						
MKT	1.149***	0.980***	0.927***	0.997***	1.094***	-0.055
	(20.50)	(36.01)	(34.89)	(34.31)	(24.06)	(-0.82)
SMB	0.174**	-0.033	-0.083**	-0.045	0.205***	0.031
	(2.25)	(-0.73)	(-2.03)	(-0.89)	(2.76)	(0.32)
HML	-0.191**	-0.016	0.002	-0.047	-0.181***	0.010
	(-2.46)	(-0.46)	(0.06)	(-1.25)	(-3.87)	(0.11)
UMD	0.005	-0.038	-0.043	0.034	-0.052	-0.056
	(0.09)	(-1.27)	(-1.48)	(1.09)	(-0.91)	(-0.87)
$\alpha_{FF3+UMD}$	1.329	1.253	1.647**	-1.673**	-2.056	-3.384*
	(0.83)	(1.51)	(2.27)	(-2.09)	(-1.59)	(-1.68)
Panel D: FF5						
MKT	1.131***	0.997***	0.961***	1.019***	1.086***	-0.045
	(21.48)	(35.80)	(38.03)	(37.63)	(23.46)	(-0.66)
SMB	0.132*	-0.022	-0.045	-0.029	0.191**	0.059
	(1.82)	(-0.49)	(-1.09)	(-0.60)	(2.53)	(0.62)
HML	-0.231**	-0.066	-0.064	-0.166***	-0.139*	0.093
	(-2.48)	(-1.50)	(-1.55)	(-4.02)	(-1.84)	(0.81)
RMW	-1.044**	0.003	0.618***	0.466**	-0.605**	0.439
	(-2.50)	(0.02)	(3.54)	(2.19)	(-2.11)	(0.84)
CMA	0.589	0.513*	0.506*	0.821***	-0.097	-0.686
	(1.17)	(1.71)	(1.97)	(3.35)	(-0.19)	(-1.07)
α_{FF5}	2.419	0.569	0.103	-2.455***	-1.694	-4.113**
	(1.64)	(0.71)	(0.15)	(-3.14)	(-1.38)	(-2.06)

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports value-weighted average excess returns and alphas in annual percentage for portfolios sorted on exposure to changes in the MPK spread. For each stock, I regress the quarterly excess returns either on changes in misallocation or each component by a rolling window of 20 quarters (with a minimum of 12 quarters available). Each quarter, I sort stocks into quintiles based on their misallocation-beta, lagging by one quarter. I hold and rebalance the portfolio every quarter. The sample runs from 1975:Q1 to 2023:Q4.

Table 2.4: **Post-formation portfolio exposure**

	Q1 (low)	Q2	Q3	Q4	Q5 (high)	Q5–Q1
Panel A: Value-weighted portfolios						
MKT	1.281***	1.029***	0.864***	0.954***	1.137***	-0.144**
	(24.58)	(43.55)	(37.05)	(40.69)	(27.94)	(-2.51)
ΔMPK spread	-0.666**	0.063	-0.195	0.148	0.207	0.873**
	(-2.19)	(0.35)	(-1.16)	(0.92)	(0.77)	(2.23)
α	-0.019	-0.046	0.082	0.191	-0.118	-0.099
	(-0.05)	(-0.27)	(0.50)	(1.00)	(-0.42)	(-0.20)
Panel B: Equally weighted portfolios						
MKT	1.461***	1.138***	1.111***	1.136***	1.311***	-0.150***
	(20.62)	(21.93)	(22.25)	(19.91)	(18.32)	(-3.51)
ΔMPK spread	-1.327***	-0.479	-0.441	-0.325	-0.701	0.626**
	(-2.74)	(-1.43)	(-1.44)	(-1.04)	(-1.54)	(2.16)
α	0.853	0.847**	0.727**	0.902**	0.828*	-0.025
	(1.59)	(2.15)	(2.13)	(2.41)	(1.69)	(-0.07)

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the betas from the regression of portfolio returns, whose portfolios are sorted on the stock exposure to changes in the MPK spread, against the market and changes in the MPK spread. The sample runs from 1975:Q1 to 2023:Q4.

significantly negative so the standard asset pricing models cannot explain the abnormal returns to the exposure to changes in the MPK spread.

Table A3 reports the value-weighted average characteristic of the stocks in each quintile portfolio. The results show that, on average, stocks negatively exposed to changes in the MPK spread (Quintile 1) tend to be smaller (low market cap), lower market power (markup), value firma, young (lower duration), but more innovative than stocks positively exposes to changes (Quintile 5). Thus, non-superstar firms tend to be negatively exposed to changes in the MPK spread. Whereas, superstar firms tend to be positively exposed to theses changes. Consistently, non-superstar firms tend to be risky, and superstar firms tend to provide a hedge during economic downturns against changes in the MPK spread.

After sorting stocks into the portfolios, we can examine in the post-formation period whether the beta of each portfolio lines up. Table 2.4 reports the coefficients from regressing value-weighted/equally weighted returns of each portfolio on the changes in MPK spread and the market factor. The post-formation portfolios show a negative

Table 2.5: **Portfolio spanning tests**

Panel A: Benchmark = Long-short returns to ΔMisall$_{\text{total}}$-mimicking portfolios			
Test portfolio = Returns to	ΔMisall$_{\text{rest}}$	ΔMisall$_{\text{top}}$	ΔMPK spread
Jensen alpha	-0.003	0.004	-0.044**
	(-0.49)	(0.37)	(-2.45)
Panel B: Benchmark = Long-short returns to ΔMPK spread-mimicking portfolios			
Test portfolio = Returns to	ΔMisall$_{\text{rest}}$	ΔMisall$_{\text{top}}$	ΔMisall$_{\text{total}}$
Jensen alpha	-0.001	0.004	0.005
	(-0.03)	(0.39)	(0.23)

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the spanning test alphas and the corresponding t-statistics. Panel A shows the time-series regressions of the returns to the changes in aggregate misallocation on returns to each component. Panel B shows the regressions of the returns to changes in the MPK spread on the returns to changes in misallocation in superstars, non-superstars, and all firms. The sample runs from 1975:Q1 to 2023:Q4.

beta in the lowest quintile and a positive beta in the highest quintile. Furthermore, the intercept is insignificant, implying that no returns remain unexplained in the portfolio. Thus, the exposure to MPK spread of the long-short portfolio returns is consistent.

2.3.3 Spanning tests

Next, I study the diversification benefits of each portfolio. This section tests whether the portfolio exposed to changes in the MPK spread adds to the mean-variance efficiency of other portfolios, and vice versa. I regress the returns of the test assets on the market and the returns of the benchmark assets. If the test assets exactly price the benchmark assets, then the intercept alphas should equal zero. This is known as the Jensen measure. Under the null hypothesis, the benchmark assets span the test assets. If the Jensen measure is significantly different from zero, then adding the test assets to the benchmark improves the mean-variance efficiency.

Table 2.5 shows the results. In Panel A, the benchmark assets are the returns to the long-short portfolio exposed to the changes in aggregate misallocation. The Jensen measure is significantly different from zero for the portfolio exposed to changes in the

MPK spread as the test asset. Therefore, adding the portfolio exposed to changes in the MPK spread to the portfolio exposed to the changes in aggregate misallocation improves the mean-variance efficiency.

Panel B shows the same set of results except that the benchmark assets are the returns to the long-short portfolio exposed to changes in the MPK spread. The portfolio exposed to changes in the MPK spread prices all benchmark portfolios. Hence, Adding the portfolio exposed to changes in the MPK spread improves the mean-variance efficiency of the portfolio exposed to the changes in the aggregate misallocation and other portfolios.

2.4 Predicting economic growth

A candidate for ICAPM state variables must forecast the investment opportunities. In this section, I examine the predictability of the MPK spread for future economic growth. Particularly, I use the non-standardized variables to avoid forward-looking bias and run the standard predictive regressions.

Panel A of Table 2.6 reports the results of the following predictive regression:

$$\Delta CG_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k}, \quad z \in \{\text{Misall}_{\text{total}}, \text{Misall}_{\text{rest}}, \text{Misall}_{\text{top}}, \text{MPK spread}\} \quad (2.9)$$

where $CG_{t:t+k}$ is the per capita real consumption growth (nondurable and services) in k quarters. In each quarter, real consumption per capita is the total real consumption of nondurable goods and services, divided by the total population. The predictive variables z_t are the aggregate misallocation, misallocation among non-superstars, misallocation among superstars, and the MPK spread. The columns show results for $k = 1, 4, 8$ and 12 quarters. I calculate standard errors based on Hodrick (1992) and based on Newey and West (1987) with $k - 1$ lags.

From 1 to 8 quarters ahead, the MPK spread negatively predicts changes in per capita real consumption growth. The coefficient is significant at 1% level, with an adjusted R-squared of more than 6% across the quarters. The magnitude of the forecast

Table 2.6: **Predicting proxies for economic growth**

	$\text{Misall}_{\text{total}}$			$\text{Misall}_{\text{rest}}$			$\text{Misall}_{\text{top}}$			MPK spread		
	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE
Panel A: Per capita real consumption growth												
$k = 1$	-0.091	0.006	1.008	-0.086	0.005	1.008	-0.697	0.057	0.982	-0.271	0.039	0.991
	(-0.59)			(-0.56)			(-1.13)			(-3.39)***		
	[-0.59]			[-0.56]			[-1.12]			[-3.38]***		
$k = 4$	0.067	0.001	1.855	0.080	0.001	1.855	-0.550	0.010	1.847	-0.750	0.089	1.772
	(0.36)			(0.41)			(-1.16)			(-3.24)***		
	[0.32]			[0.37]			[-1.10]			[-3.00]***		
$k = 8$	0.035	0.000	2.508	0.053	0.000	2.508	-1.077	0.022	2.480	-0.756	0.050	2.444
	(0.09)			(0.13)			(-0.99)			(-2.59)***		
	[0.08]			[0.12]			[-0.96]			[-2.64]***		
$k = 12$	-0.166	0.002	3.058	-0.151	0.002	3.058	-1.832	0.043	2.995	-0.333	0.007	3.051
	(-0.32)			(-0.28)			(-1.14)			(-0.87)		
	[-0.32]			[-0.28]			[-1.13]			[-0.88]		
Panel B: Industrial production growth												
$k = 1$	-0.278	0.020	1.753	-0.272	0.017	1.755	-1.458	0.081	1.697	-0.282	0.014	1.759
	(-1.20)			(-1.17)			(-1.73)*			(-1.57)		
	[-1.20]			[-1.16]			[-1.72]*			[-1.56]		
$k = 4$	-0.524	0.012	4.288	-0.522	0.011	4.290	-1.964	0.025	4.260	-1.176	0.040	4.225
	(-1.03)			(-0.99)			(-1.39)			(-2.19)**		
	[-0.93]			[-0.89]			[-1.29]			[-2.13]**		
$k = 8$	-0.914	0.017	6.345	-0.923	0.016	6.348	-2.350	0.016	6.347	-2.423	0.079	6.139
	(-0.72)			(-0.70)			(-0.93)			(-3.38)***		
	[-0.69]			[-0.67]			[-0.91]			[-3.34]***		
$k = 12$	-1.363	0.026	7.559	-1.402	0.026	7.561	-2.054	0.009	7.628	-1.662	0.027	7.558
	(-0.78)			(-0.78)			(-0.67)			(-1.70)*		
	[-0.77]			[-0.76]			[-0.67]			[-1.71]*		
Panel C: Employment growth												
$k = 1$	-0.093	0.007	0.994	-0.089	0.006	0.995	-0.585	0.041	0.977	-0.153	0.013	0.991
	(-0.64)			(-0.62)			(-1.01)			(-2.30)**		
	[-0.64]			[-0.62]			[-1.00]			[-2.28]**		
$k = 4$	0.148	0.007	1.538	0.163	0.008	1.537	-0.429	0.009	1.537	-0.497	0.056	1.500
	(1.04)			(1.10)			(-1.01)			(-2.49)**		
	[0.94]			[1.00]			[-0.94]			[-2.38]**		
$k = 8$	0.294	0.016	2.114	0.317	0.017	2.112	-0.501	0.007	2.123	-0.887	0.096	2.026
	(0.83)			(0.87)			(-0.58)			(-4.04)***		
	[0.80]			[0.83]			[-0.56]			[-3.98]***		
$k = 12$	0.260	0.009	2.459	0.282	0.010	2.458	-0.560	0.006	2.463	-0.585	0.032	2.431
	(0.54)			(0.56)			(-0.50)			(-2.24)**		
	[0.53]			[0.55]			[-0.49]			[-2.28]**		

t-ratio of Hodrick (1992) with k-1 lags in parentheses. *t*-ratio of Newey-West (1987) with k-1 lags in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the results of the following predictive regression:

$$Q_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k},$$

where $Q \in \{CG, IP, E\}$. The macroeconomic data to construct this measure are obtained from the Bureau of Economic Analysis. The predictive variables z_t are aggregate misallocation, misallocation among non-superstars, misallocation among superstars, and the MPK spread. The sample runs from 1975:Q1 to 2022:Q4.

is also larger at a longer horizon. The aggregate misallocation, misallocation among non-superstars, and misallocation among superstars on the other hand are statistically and economically insignificant.

Panel B of Table 2.6 reports the results of the following predictive regression:

$$\Delta IP_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k} \tag{2.10}$$

where $IP_{t:t+k}$ is the industrial production growth in k quarters. From 8 to 12 quarters forward, the MPK spread also predicts negative changes in industrial production growth. Other variables in contrast yield no significant predictive results.

Panel C of Table 2.6 reports the results of the following predictive regression:

$$\Delta E_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k} \tag{2.11}$$

where $E_{t:t+k}$ is the log employment growth in k quarters. The MPK spread on the other hand predicts higher changes in employment growth. Other variables in contrast yield no significant predictive results.

Consistently, changes in the MPK spread also robustly predict negative economic growth. Table 2.11 reports the same set of regressions when changes in aggregate misallocation and the MPK spread act as predictors. Changes in the MPK spread negatively predict consumption growth, industrial growth, and employment growth. Although the changes in aggregate misallocation have some predictive power in some models, the positive sign is counter-intuitive since higher misallocation as in the literature predicts lower economic growth and hence consumption growth (David, Hopenhayn, and Venkateswaran, 2016).

Intuitively, the predictability results support the mechanism that superstar firms attract more capital disproportionally due to their market power rather than their productivity. If the driving factor is productivity, it would be considered good news for the economy, making it unclear why it would be associated with negative outcomes such as slower economic growth, and declining employment. Therefore, the results suggest

that the observed effects are not tied to productive efficiency but rather to the influence of market power, which could negatively impact these economic indicators.

Finally, I test the mechanism that the MPK spread dampens innovation activity in the short run. I test the predictability of the MPK spread for innovation growth using the following predictive regression:

$$\Delta I_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k} \tag{2.12}$$

where $I_{t:t+k}$ is the innovation growth in k quarters.

Table 2.7 reports the results for $k = 1, 2, 3, 4, 5$ and 6 quarters. I construct the innovation proxy I_t as the natural logarithm of one plus the total number of patent applications divided by the total firm market cap on the aggregate level, within superstars, and within non-superstars portfolios. Panel A shows that the MPK spread predicts negative changes in innovation growth. Both statistical significance and economic significance are large comparing to the aggregate capital misallocation.

Panel B uses changes in innovation growth among non-superstars as the dependent variable. In all quarters, the MPK spread significantly predicts negative changes in innovation growth among non-superstars. Both economic significance and statistical significance increase with the horizon from 1 to 5 quarters. Whereas, the coefficients for aggregate misallocation across all quarters are not significant.

Yet, when Panel C uses changes in innovation growth among superstars as the dependent variable, no coefficients are statistically and economically significant. These results imply a higher discrepancy in the mean productive use of capital between superstars and non-superstars discourages innovation activity in the economy, especially among non-superstars, leading to lower economic growth in the long run.

This implication is consistent with Bae, Bailey, and Kang (2021) who also find that higher stock market concentration is associated with lower innovation activity. It is also consistent with Kung and Schmid (2015) who show that innovation endogenously generates long-run fluctuation in economic growth. In contrast, the aggregate capital misallocation, the misallocation among superstars and among non-superstars do not

Table 2.7: **Predicting innovation growth**

	$\text{Misall}_{\text{total}}$			$\text{Misall}_{\text{rest}}$			$\text{Misall}_{\text{top}}$			MPK spread		
	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE
Panel A: Aggregate innovation growth												
$k=1$	-0.001	0.000	0.025	-0.002	0.000	0.025	0.488	0.005	0.025	-6.516	0.013	0.025
	(-0.06)			(-0.08)			(0.90)			(-2.23)**		
	[-0.06]			[-0.08]			[0.89]			[-2.22]**		
$k=2$	0.007	0.001	0.033	0.006	0.000	0.033	0.868	0.009	0.032	-7.089	0.009	0.032
	(0.22)			(0.19)			(1.14)			(-1.91)*		
	[0.20]			[0.17]			[1.01]			[-1.77]*		
$k=3$	0.019	0.004	0.036	0.018	0.003	0.036	1.210	0.015	0.036	-12.314	0.023	0.036
	(0.50)			(0.48)			(1.25)			(-2.46)**		
	[0.47]			[0.45]			[1.17]			[-2.37]**		
$k=4$	0.037	0.014	0.036	0.036	0.013	0.036	1.653	0.028	0.036	-10.875	0.018	0.036
	(0.82)			(0.80)			(1.38)			(-2.27)**		
	[0.79]			[0.77]			[1.32]			[-2.19]**		
$k=5$	0.047	0.020	0.038	0.046	0.019	0.038	2.133	0.041	0.037	-8.159	0.009	0.038
	(0.91)			(0.88)			(1.41)			(-1.10)		
	[0.87]			[0.85]			[1.36]			[-1.09]		
$k=6$	0.065	0.035	0.040	0.064	0.033	0.040	2.850	0.066	0.039	-4.687	0.003	0.040
	(1.11)			(1.08)			(1.58)			(-0.48)		
	[1.09]			[1.06]			[1.55]			[-0.48]		
Panel B: Innovation growth of non-superstar firms												
$k=1$	-0.002	0.000	0.030	-0.003	0.000	0.030	0.538	0.004	0.030	-9.215	0.019	0.029
	(-0.08)			(-0.11)			(0.82)			(-2.50)**		
	[-0.08]			[-0.11]			[0.82]			[-2.49]**		
$k=2$	0.009	0.001	0.038	0.009	0.001	0.038	0.772	0.005	0.038	-10.281	0.015	0.038
	(0.28)			(0.25)			(0.91)			(-2.37)**		
	[0.25]			[0.23]			[0.81]			[-2.20]**		
$k=3$	0.025	0.005	0.042	0.025	0.004	0.042	0.901	0.006	0.042	-18.665	0.039	0.041
	(0.63)			(0.61)			(0.82)			(-3.18)***		
	[0.60]			[0.58]			[0.78]			[-3.09]***		
$k=4$	0.045	0.014	0.043	0.044	0.013	0.043	1.313	0.012	0.043	-15.838	0.026	0.043
	(0.94)			(0.93)			(0.97)			(-2.92)***		
	[0.91]			[0.89]			[0.93]			[-2.83]***		
$k=5$	0.055	0.019	0.045	0.054	0.018	0.045	1.790	0.020	0.045	-14.617	0.020	0.045
	(0.99)			(0.97)			(1.06)			(-1.68)*		
	[0.95]			[0.93]			[1.03]			[-1.68]*		
$k=6$	0.075	0.032	0.047	0.075	0.031	0.047	2.400	0.032	0.047	-10.575	0.010	0.048
	(1.20)			(1.18)			(1.19)			(-0.95)		
	[1.17]			[1.16]			[1.17]			[-0.95]		
Panel C: Innovation growth of superstar firms												
$k=1$	0.004	0.000	0.026	0.005	0.000	0.026	0.283	0.002	0.026	0.767	0.000	0.026
	(0.16)			(0.20)			(0.41)			(0.25)		
	[0.15]			[0.19]			[0.41]			[0.25]		
$k=2$	0.011	0.001	0.038	0.011	0.001	0.038	1.315	0.018	0.036	0.005	0.000	0.038
	(0.28)			(0.29)			(1.11)			(0.00)		
	[0.24]			[0.24]			[0.95]			[0.00]		
$k=3$	0.022	0.004	0.041	0.019	0.003	0.041	2.201	0.037	0.041	1.727	0.000	0.041
	(0.39)			(0.32)			(1.31)			(0.31)		
	[0.35]			[0.29]			[1.19]			[0.30]		
$k=4$	0.040	0.010	0.044	0.038	0.009	0.044	2.762	0.050	0.044	-2.111	0.000	0.044
	(0.55)			(0.51)			(1.32)			(-0.34)		
	[0.52]			[0.48]			[1.22]			[-0.34]		
$k=5$	0.058	0.018	0.049	0.056	0.016	0.049	2.745	0.040	0.049	2.893	0.001	0.050
	(0.68)			(0.63)			(1.13)			(0.42)		
	[0.66]			[0.61]			[1.09]			[0.41]		
$k=6$	0.075	0.026	0.053	0.074	0.025	0.053	3.290	0.051	0.051	6.101	0.003	0.053
	(0.80)			(0.75)			(1.23)			(0.70)		
	[0.79]			[0.74]			[1.21]			[0.67]		

t-ratio of Hodrick (1992) with $k-1$ lags in parentheses. t-ratio of Newey–West (1987) with $k-1$ lags in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the results of the following predictive regression:

$$R_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k},$$

where $I_{t:t+k}$ is the innovation growth in k quarters. The predictive variables z_t are the aggregate misallocation and the MPK spread. The predictive variables z_t are the aggregate misallocation and the MPK spread. The columns show results for $k = 1, 2, 3, 4, 5$ and 6 quarters. I construct the innovation proxy I_t as the natural logarithm of one plus the number of patent applications divided by the firm market cap. The number of patent is from Kogan, Papanikolaou, Seru, and Stoffman (2017). The sample runs from 1975:Q1 to 2022:Q4.

show predictive power and their changes do no carry a significantly negative price of risk. Therefore, only the MPK spread is a good candidate state variable.

2.4.1 Consistency with the ICAPM

In this section, I examine whether changes in capital misallocation between superstar and non-superstar firms as a risk factor satisfy the restrictions under the ICAPM framework. Maio and Santa-Clara (2012) proposes three restrictions associated with the ICAPM. The first restriction concerns the forecasting power of state variables for investment opportunities. Specifically, the state variables must forecast the first moment (expected returns) or second moment (market volatility) of aggregate stock returns.

Subsequently, I examine whether aggregate capital misallocation or its component predicts future stock market excess returns. Table 2.8 reports the results of the following predictive regression:

$$R^e_{mkt,t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k} \tag{2.13}$$

where $R^e_{mkt,t:t+k} = R^e_{mkt,t+1} + \cdots + R^e_{mkt,t+k}$ is continuously compounded market excess return, using the CRSP value-weighted returns in excess of the US one-month Treasury Bills rate, from the end of quarter t to the end of quarter $t+k$.

I construct out-of-sample predictability following Campbell and Thompson (2008). I estimate the Equation (2.13) using all available returns up to quarter t with a minimum of 80 quarters. Then I use the estimates $\hat{\alpha}_t$ and $\hat{\beta}_t$, using data from the start of the sample period to quarter t, to forecast the k-quarter excess return from quarter t to $t+k$:

$$\hat{R}^e_{mkt,t:t+k} = \hat{\alpha}_t + \hat{\beta}_t z_t, \tag{2.14}$$

The out-of-sample R-squared for the predictive regressions uses the historical aver-

Table 2.8: **Predicting aggregate stock returns**

	$\text{Misall}_{\text{total}}$				$\text{Misall}_{\text{rest}}$				$\text{Misall}_{\text{top}}$				MPK spread			
	$\hat{\beta}$	R^2_{IS}	R^2_{OOS}	RMSE	$\hat{\beta}$	R^2_{IS}	R^2_{OOS}	RMSE	$\hat{\beta}$	R^2_{IS}	R^2_{OOS}	RMSE	$\hat{\beta}$	R^2_{IS}	R^2_{OOS}	RMSE
$k=1$	0.676	0.008	0.018	6.622	0.729	0.009	0.019	6.620	-1.056	0.003	0.015	6.639	-1.084	0.018	0.002	6.590
	(1.44)				(1.49)				(-0.76)				(-2.43)**			
	[1.43]				[1.48]				[-0.76]				[-2.42]**			
$k=4$	1.304	0.013	0.023	10.082	1.396	0.014	0.024	10.078	-1.398	0.002	0.014	10.137	-1.654	0.018	0.003	10.058
	(1.95)*				(2.02)**				(-0.60)				(-2.66)***			
	[1.61]				[1.67]*				[-0.51]				[-2.24]**			
$k=8$	1.660	0.013	0.020	12.822	1.777	0.014	0.021	12.817	-1.561	0.002	0.011	12.896	-2.613	0.028	-0.005	12.728
	(1.51)				(1.57)				(-0.43)				(-2.53)**			
	[1.33]				[1.38]				[-0.39]				[-2.26]**			
$k=12$	2.038	0.015	0.017	14.973	2.181	0.016	0.018	14.967	-2.028	0.002	0.014	15.068	-2.815	0.024	-0.011	14.906
	(1.28)				(1.32)				(-0.41)				(-1.87)*			
	[1.18]				[1.22]				[-0.38]				[-1.75]*			

t-ratio of Hodrick (1992) with k-1 lags in parentheses. *t*-ratio of Newey-West (1987) with k-1 lags in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the results of the following predictive regression:

$$R^e_{mkt,t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k},$$

where $R^e_{mkt,t:t+k}$ is the stock market excess returns in k quarters. Stock market returns are the value-weighted CRSP returns in excess of the risk-free rate. The predictive variables z_t are aggregate misallocation, misallocation among non-superstars, misallocation among superstars, and the MPK spread. The sample runs from 1975:Q1 to 2022:Q4.

age excess market return as a benchmark as follows:

$$R^2_{OOS} = 1 - \frac{\sum_{t=80}^{T-k}(R_{mkt,t:t+k} - \hat{R}_{mkt,t:t+k})^2}{\sum_{t=80}^{T-k}(R_{mkt,t:t+k} - k\bar{R}_{mkt,1:t})^2}, \tag{2.15}$$

where $\bar{R}_{mkt,1:t}$ is the average excess market return up to quarter t, and T represents the length of the return series. The summation covers all quarters starting in quarter 241. The out-of-sample R-squared can be negative if the predictive variable has poor out-of-sample predictability.

As a result, the MPK spread negatively predicts the future stock market returns in all quarters. The coefficient is significant at the 5% level for 1 quarter and 1% level for 4 quarters ahead. The in-sample R-squared is approximately 1.8% for 1 and 4 quarters and 2.8% for 8 quarters. The out-of-sample R-squared is positive and is approximately 0.2% and 0.5%. Thus, the MPK spread satisfies the first restriction for the ICAPM although the goodness of fit is relatively low.

The second restriction concerns the relationship between the predictive power of the state variable and the risk premium of the risk factor. Specifically, if a state variable

forecasts positive (negative) expected returns, then the innovation or changes in the state variable as a risk factor should earn a positive (negative) price of risk. Previous findings show that the MPK spread negatively predicts investment opportunities and changes in the MPK spread are negatively priced in the cross-sectional tests. Thus, these findings satisfy the second restriction for the ICAPM.

Intuitively, an asset negatively exposed to innovations in a state variable also negatively covaries with future expected returns. Such an asset provides a hedge against reinvestment risk, as it delivers higher returns when aggregate returns are expected to be lower. Because it offers protection during economic downturns, a risk-averse investor would be willing to hold this asset even if it offers a lower expected return. Therefore, a negative covariance with the innovation of the state variable results in a negative price of risk for the factor, implying that investors accept lower returns for the benefit of hedging against future economic downturns.

The third restriction requires that the estimated market price of risk, which reflects the risk-aversion coefficient of the representative investor, must be economically plausible. The ICAPM in unconditional form with a hedging risk factor and the market has the form

$$\mathbb{E}(R^e_{i,t+1}) = \gamma_m \, \mathrm{Cov}(R^e_{i,t+1}, MKT_{t+1}) + \gamma_z \, \mathrm{Cov}(R^e_{i,t+1}, \Delta z_{t+1}) \tag{2.16}$$

Intuitively, for an asset that does not provide a hedge against changes in current aggregate wealth, as it pays in good times (periods with high returns on wealth), a risk-averse investor would be willing to hold such an asset only if it offers a premium over the risk-free rate. Following Maio and Santa-Clara (2012), I estimate the price of risk of each factor γ using the GMM system with $N + 2$ moment conditions:

$$g_T(\mathbf{b}) \equiv \frac{1}{T}\sum_{t=0}^{T-1}\begin{pmatrix} R^e_{i,t+1} - \gamma_m R^e_{i,t+1}(MKT_{t+1} - \mu_m) \\ -\gamma_z R^e_{i,t+1}(\Delta z_{t+1} - \mu_z) \\ MKT_{t+1} - \mu_m \\ \Delta z_{t+1} - \mu_z \end{pmatrix} = 0, \quad i = 1, \ldots, N. \tag{2.17}$$

The first factor in this model is the market with unconditional mean μ_m. The second factor is the changes in the state variable $z \in \{\text{Misall}_{\text{total}}, \text{Misall}_{\text{rest}}, \text{Misall}_{\text{top}}, \text{MPK spread}\}$. To test the goodness of fit, I construct the mean absolute pricing error (MAE)

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{\alpha}_i|, \tag{2.18}$$

where $\hat{\alpha}_i$, $i = 1, \ldots, N$ represents the pricing errors associated with the N test assets. The second measure is the cross-sectional OLS R-squared:

$$R^2_{\text{OLS}} = 1 - \frac{\text{Var}_N(\hat{\alpha}_i)}{\text{Var}_N(\bar{R}_i)}, \tag{2.19}$$

where $\bar{R}_i = \frac{1}{T}\sum_{t=0}^{T-1} R^e_{i,t+1}$ is the average excess return for asset i. R^2_{OLS} measures the fraction of the cross-sectional variance in average excess returns explained by the model.

Table 2.9 reports the estimation of the first-stage GMM using equally weighted errors, using the same test portfolio as in Maio and Santa-Clara (2012). In particular, the coefficient of relative risk aversion (RRA) should fall within a reasonable range, typically between 1 and 10 (Mehra and Prescott, 1985). In all models, the risk price of the market is significantly positive and has a magnitude of approximately 5. However, only Regression (4) shows a significant negative price of risk of the changes in the state variable, particularly the MPK spread. This result is consistent with the results in the Fama-Macbeth regressions. Thus, these findings satisfy the third restriction.

In conclusion, the MPK spread predicts changes in economic growth and aggregate stock returns. Changes in the MPK spread, exacerbating capital misallocation, repre-

Table 2.9: **Factor risk premiums for empirical risk factors**

	(1)	(2)	(3)	(4)
Panel A: 25 size×book-to-market portfolios				
MKT				
γ	5.945***	5.823***	5.945***	5.343***
	(10.49)	(9.95)	(10.81)	(8.85)
$\Delta\text{Misall}_{\text{total}}$				
γ	0.098*			
	(1.81)			
$\Delta\text{Misall}_{\text{rest}}$				
γ		0.115**		
		(2.03)		
$\Delta\text{Misall}_{\text{top}}$				
γ			-0.066*	
			(-1.75)	
ΔMPK spread				
γ				-0.256***
				(-5.38)
R^2_{OLS}	0.661	0.663	0.666	0.490
MAE	0.021	0.020	0.022	0.018
Panel B: 25 size×momentum portfolios				
MKT				
γ	5.409***	5.596***	5.902***	5.399***
	(7.82)	(8.39)	(9.71)	(9.18)
$\Delta\text{Misall}_{\text{total}}$				
γ	0.154**			
	(2.52)			
$\Delta\text{Misall}_{\text{rest}}$				
γ		0.129**		
		(2.34)		
$\Delta\text{Misall}_{\text{top}}$				
γ			0.012	
			(0.22)	
ΔMPK spread				
γ				-0.273***
				(-7.28)
R^2_{OLS}	0.659	0.663	0.466	0.610
MAE	0.018	0.020	0.022	0.020

GMM robust t-statistics in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the estimation from the first-stage GMM with equally weighted errors. The test portfolio includes 25 size×book-to-market portfolios in Panel A and 25 size×momentum portfolios in Panel B. The coefficient γ represents the price of risk for the corresponding factor. The sample runs from 1975:Q1 to 2022:Q4.

sent bad news for investors whose marginal utility depends on long-run consumption growth. Hence, consistent with the ICAPM framework, the MPK spread is a state variable, and changes in the MPK spread are a priced risk factor.

2.5 Robustness tests

In this section, I check the robustness of the main findings. The main results are robust to alternative definitions of superstars, additional test portfolios in the Fama-MacBeth regressions, industry classifications, value-weighted capital misallocation, types of capital, subsample periods, and annual frequency data.

2.5.1 Alternative definitions of superstar firms

To identify superstar firms, I adopt a widely used approach that defines them as the largest firms based on market capitalization or sales (Bae, Bailey, and Kang, 2021; Schlingemann and Stulz, 2022). While the number of listed firms in the U.S. stock market increased substantially before the 2000s and has since declined, the concentration of superstar firms within the top decile of market capitalization and sales distribution has remained relatively stable.

Table 2.10 shows that the pricing results are robust to defining superstar firms as the 50 firms sorted by market cap and sales. The MPK spread is negatively priced, although the significance is relatively weaker. Yes, it is consistent with the main findings as the top 50 superstar firms sorted by market cap or sales are nearly identical to superstar firms identified by markup shares.

The main mechanism driving the MPK spread in this paper is the markup of superstar firms. Ayyagari, Demirgüç-Kunt, and Maksimovic (2024) find that markups are positively related to a greater probability of being a star firm. However, another potential friction contributing to capital misallocation could be financial constraints. To investigate this alternative, I estimate firm-level financial constraints using two standard proxies: the Kaplan and Zingales (1997) index (KZ index) and the Whited and Wu

Table 2.10: **Pricing 35 portfolios by alternative definition of superstars**

			Top 50 firms sorted on market cap			Top 50 firms sorted on sale		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	12.090	10.792	10.978	9.870	14.968	10.838	12.166	14.435
	(3.67)***	(3.25)***	(3.28)***	(2.89)***	(4.51)***	(3.21)***	(3.74)***	(4.51)***
	[3.67]***	[2.99]***	[3.18]***	[2.33]**	[3.49]***	[3.09]***	[3.70]***	[3.53]***
MKT	-0.257	-0.057	-0.057	0.073	-1.089	-0.027	-0.342	-0.931
	(-0.25)	(-0.05)	(-0.05)	(0.07)	(-1.10)	(-0.03)	(-0.34)	(-0.95)
	[-0.22]	[-0.04]	[-0.05]	[0.05]	[-0.77]	[-0.02]	[-0.29]	[-0.67]
$\Delta\text{Misall}_{\text{total}}$		-0.435						
		(-0.99)						
		[-0.90]						
$\Delta\text{Misall}_{\text{rest}}$			-1.189			-1.267		
			(-0.65)			(-0.69)		
			[-0.62]			[-0.65]		
$\Delta\text{Misall}_{\text{top}}$				-0.182			-0.034	
				(-1.55)			(-0.54)	
				[-1.24]			[-0.51]	
ΔMPK spread					-0.147			-0.154
					(-2.30)**			(-2.67)***
					[-1.76]*			[-2.05]**
R^2	0.012	0.064	0.038	0.140	0.129	0.040	0.035	0.144
Adj. R^2	-0.018	0.006	-0.022	0.086	0.074	-0.020	-0.026	0.091
RMSE	2.701	2.669	2.707	2.560	2.576	2.704	2.712	2.553

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 size $\times$ book-to-market and 10 momentum portfolios. The sample runs from 1975:Q1 to 2023:Q4. Returns and risk premia are reported in percent per year (quarterly percentages multiplied by four). For robustness, I identify superstars as the top 50 firms sorted on market cap or sales each quarter.

(2006) index (WW index). Higher values of these indices correspond to greater financial constraints. Accordingly, I redefine superstar firms as the bottom 5% within their industries based on financial constraints. Results reported in Table A5 indicate that the price of risk of the MPK spread is not significant under this alternative definition, suggesting that financial constraints do not drive my asset pricing results.

Additionally, I test the robustness of these findings by splitting the sample into top and bottom halves based on markup. When I randomly select 50 firms from each half and repeat this process 500 times, the results indicate that only 4.8% of the changes in the MPK spread are statistically significant at the 5% level (see Table A1). This simulation supports the conclusion that essentially superstar firms' markup drives the MPK spread.

2.5.2 Predictability of risk factors

Consistently, if a state variable predicts future economic growth and stock market returns, then changes in the state variable should also predict these variables since they carry the same information. Thus, I run the same predictive regressions using changes in the aggregate capital misallocation and changes in MPK spread.

Panel A of Table 2.11 reports the results of predicting per capita real consumption growth (nondurable and services) in k quarters. The predictive variables zt are changes in the aggregate misallocation and the MPK spread. The columns show results for $k = 1, 4, 8, 12$ and 20 quarters. In all quarters forward, higher changes in the MPK spread predict lower changes in per capita real consumption growth. The coefficient is the most significant at 8 and 12 quarters forward, with an adjusted R-squared of 6-7% in Regression (6) and (8). The changes in aggregate misallocation on the other hand have a statistically significant power in 4, 8, and 12 quarters but the economic significance is almost close to zero and has a counter-intuitive sign since higher misallocation as in the literature predicts lower economic growth hence consumption growth (David, Hopenhayn, and Venkateswaran, 2016).

Panel B reports the results of predicting industrial production growth. In 8,12, and

Table 2.11: **Predicting economic growth using risk factors**

	$\Delta\text{Misall}_{\text{total}}$			$\Delta\text{Misall}_{\text{rest}}$			$\Delta\text{Misall}_{\text{top}}$			ΔMPK spread		
	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE
Panel A: Per-capita real consumption growth												
$k=1$	-0.193	0.037	0.992	-0.172	0.030	0.996	-0.349	0.067	0.976	-0.191	0.021	1.000
	(-0.83)			(-0.77)			(-1.37)			(-2.38)**		
	[-0.82]			[-0.76]			[-1.36]			[-2.37]**		
$k=4$	0.305	0.028	1.830	0.322	0.031	1.827	-0.207	0.007	1.850	-0.687	0.081	1.779
	(1.91)*			(2.07)**			(-1.18)			(-2.59)***		
	[2.12]**			[2.27]**			[-1.27]			[-2.39]**		
$k=8$	0.648	0.069	2.420	0.681	0.076	2.411	-0.272	0.007	2.499	-0.855	0.070	2.418
	(2.61)***			(2.86)***			(-0.81)			(-2.86)***		
	[2.59]**			[2.84]***			[-0.79]			[-2.90]***		
$k=12$	0.840	0.071	2.951	0.885	0.079	2.939	-0.420	0.010	3.046	-0.935	0.058	2.972
	(2.32)**			(2.60)***			(-0.88)			(-2.90)***		
	[2.35]**			[2.64]***			[-0.87]			[-2.98]***		
Panel B: Industrial production growth												
$k=1$	-0.255	0.021	1.752	-0.219	0.016	1.757	-0.605	0.066	1.711	0.014	0.000	1.771
	(-0.79)			(-0.70)			(-1.69)*			(0.09)		
	[-0.78]			[-0.70]			[-1.68]*			[0.09]		
$k=4$	0.189	0.002	4.309	0.272	0.004	4.305	-1.266	0.049	4.206	-0.217	0.001	4.310
	(0.41)			(0.57)			(-3.64)***			(-0.35)		
	[0.39]			[0.54]			[-3.52]***			[-0.34]		
$k=8$	0.698	0.012	6.360	0.868	0.019	6.338	-2.261	0.072	6.163	-1.523	0.034	6.288
	(1.16)			(1.56)			(-2.63)***			(-2.14)**		
	[1.16]			[1.57]			[-2.55]**			[-2.14]**		
$k=12$	0.603	0.006	7.638	0.833	0.011	7.618	-3.041	0.086	7.324	-1.518	0.024	7.567
	(0.72)			(1.10)			(-2.85)***			(-2.00)**		
	[0.71]			[1.10]			[-2.82]***			[-2.04]**		
Panel C: Employment growth												
$k=1$	-0.221	0.050	0.972	-0.206	0.044	0.975	-0.265	0.040	0.977	-0.021	0.000	0.997
	(-1.05)			(-1.03)			(-1.14)			(-0.21)		
	[-1.05]			[-1.02]			[-1.14]			[-0.21]		
$k=4$	0.059	0.002	1.543	0.073	0.002	1.542	-0.230	0.013	1.534	-0.217	0.012	1.535
	(0.36)			(0.43)			(-1.96)**			(-0.84)		
	[0.36]			[0.44]			[-2.32]**			[-0.80]		
$k=8$	0.412	0.038	2.089	0.442	0.044	2.083	-0.323	0.013	2.116	-0.632	0.053	2.073
	(2.04)**			(2.20)**			(-1.41)			(-2.22)**		
	[2.08]**			[2.26]**			[-1.37]			[-2.22]**		
$k=12$	0.547	0.046	2.413	0.594	0.054	2.403	-0.524	0.025	2.440	-0.654	0.043	2.417
	(2.41)**			(2.82)***			(-1.54)			(-2.41)**		
	[2.42]**			[2.84]***			[-1.52]			[-2.47]**		

t-ratio of Hodrick (1992) with $k-1$ lags in parentheses; Newey–West (1987) in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the results of the following predictive regression: $Q_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k}$, where $Q \in \{CG, IP, E, R^e_{mkt}\}$. The macroeconomic data to construct this measure are obtained from the Bureau of Economic Analysis. Stock market returns are the value-weighted CRSP returns in excess of risk-free rate. The predictive variables z_t are changes in the aggregate misallocation and the MPK spread.

20 quarters forward, higher changes in the MPK spread also predict lower changes in industrial production growth. Panel C reports the results of predicting employment growth. Although the changes in aggregate misallocation have predictive power in 8 and 12 quarters ahead, the sign is not economically intuitive. Changes in the MPK spread on the other hand predict negative changes in employment growth.

2.5.3 Additional test portfolios

This section shows the robustness of the main results by using a different set of test portfolios. Table 2.12 reports the pricing results in the cross-section of 202 portfolios used in Giglio and Xiu (2021): 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolios sorted by size and momentum.

Table 2.12: **Pricing Giglio and Xiu (2021)'s 202 portfolios**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Models including MKT factor									
Constant	10.479	9.847	9.950	9.652	10.988	9.789	9.738	10.633	10.465
	(4.25)***	(3.93)***	(3.99)***	(3.80)***	(4.49)***	(3.90)***	(3.87)***	(4.34)***	(4.22)***
	[4.25]***	[3.73]***	[3.83]***	[3.17]***	[3.10]***	[3.14]***	[3.17]***	[3.09]***	[3.07]***
MKT	0.092	0.169	0.166	0.029	-0.336	-0.033	0.005	-0.275	-0.291
	(0.10)	(0.19)	(0.19)	(0.03)	(-0.39)	(-0.04)	(0.01)	(-0.32)	(-0.34)
	[0.09]	[0.15]	[0.15]	[0.02]	[-0.24]	[-0.03]	[0.00]	[-0.20]	[-0.22]
$\Delta\text{Misall}_{\text{total}}$		-0.345				-0.076	-0.108	-0.027	
		(-1.68)*				(-0.47)	(-0.66)	(-0.16)	
		[-1.51]				[-0.35]	[-0.51]	[-0.11]	
$\Delta\text{Misall}_{\text{rest}}$			-0.304			0.010			0.103
			(-1.53)			(0.06)			(0.64)
			[-1.38]			[0.05]			[0.44]
$\Delta\text{Misall}_{\text{top}}$				-0.491			-0.515		-0.311
				(-2.16)**			(-2.19)**		(-1.48)
				[-1.77]*			[-1.77]*		[-1.06]
ΔMPK spread					-0.960			-0.889	-0.808
					(-3.51)***			(-3.70)***	(-4.02)***
					[-2.39]**			[-2.59]**	[-2.84]***
R^2	0.001	0.064	0.046	0.209	0.340	0.238	0.211	0.357	0.386
Adj. R^2	-0.004	0.055	0.037	0.201	0.333	0.226	0.199	0.347	0.374
RMSE	2.524	2.449	2.472	2.252	2.057	2.216	2.255	2.036	1.994

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolios sorted by size and momentum. The sample runs from 1975:Q1 to 2023:Q4.

Table 2.12: **Pricing Giglio and Xiu (2021)'s 202 portfolios - continued**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel B: Models including MKT, SMB, HML, and UMD factors									
Constant	7.550	7.314	7.279	7.555	7.541	7.077	7.154	7.371	7.107
	(3.82)***	(3.88)***	(3.88)***	(3.73)***	(3.81)***	(3.74)***	(3.80)***	(3.93)***	(3.77)***
	[3.59]***	[3.63]***	[3.62]***	[3.51]***	[3.45]***	[3.48]***	[3.54]***	[3.55]***	[3.35]***
MKT	0.478	0.537	0.545	0.477	0.461	0.591	0.574	0.505	0.564
	(0.61)	(0.70)	(0.71)	(0.60)	(0.59)	(0.77)	(0.74)	(0.66)	(0.73)
	[0.46]	[0.52]	[0.53]	[0.46]	[0.43]	[0.57]	[0.56]	[0.48]	[0.53]
SMB	0.630	0.625	0.624	0.630	0.628	0.616	0.619	0.624	0.614
	(1.66)*	(1.64)	(1.64)	(1.66)*	(1.65)	(1.62)	(1.63)	(1.64)	(1.62)
	[1.16]	[1.14]	[1.14]	[1.16]	[1.13]	[1.12]	[1.13]	[1.12]	[1.09]
HML	0.966	0.994	0.994	0.967	0.937	0.992	0.993	0.959	0.956
	(1.96)*	(2.02)**	(2.02)**	(1.96)*	(1.90)*	(2.02)**	(2.02)**	(1.95)*	(1.94)*
	[1.39]	[1.43]	[1.43]	[1.39]	[1.32]	[1.42]	[1.43]	[1.35]	[1.34]
UMD	1.575	1.618	1.619	1.576	1.622	1.614	1.614	1.650	1.646
	(2.82)***	(2.89)***	(2.89)***	(2.82)***	(2.90)***	(2.88)***	(2.88)***	(2.95)***	(2.94)***
	[1.97]*	[2.01]**	[2.01]**	[1.97]*	[1.98]**	[2.00]**	[2.00]**	[2.01]**	[1.99]**
$\Delta\text{Misall}_{\text{total}}$		0.082				0.084	0.083	0.100	
		(0.51)				(0.52)	(0.51)	(0.63)	
		[0.44]				[0.44]	[0.44]	[0.52]	
$\Delta\text{Misall}_{\text{rest}}$			0.091			0.100			0.126
			(0.56)			(0.62)			(0.79)
			[0.49]			[0.53]			[0.65]
$\Delta\text{Misall}_{\text{top}}$				-0.022			-0.060		-0.116
				(-0.18)			(-0.53)		(-1.07)
				[-0.16]			[-0.45]		[-0.87]
ΔMPK spread					-0.325			-0.320	-0.348
					(-2.29)**			(-2.23)**	(-2.51)**
					[-1.91]*			[-1.86]*	[-2.06]**
R^2	0.528	0.532	0.533	0.528	0.542	0.534	0.533	0.544	0.546
Adj. R^2	0.519	0.520	0.521	0.516	0.530	0.519	0.519	0.530	0.529
RMSE	1.748	1.744	1.744	1.752	1.727	1.746	1.748	1.727	1.728

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.12 reports results. The price of risks of the changes in aggregate misallocation loses significance when included in the CAPM and Fama and French (1992b, 1993b) three-factor model. Yet, in all models, only the changes in the MPK spread display a negative price of risk, with a statistical significance of less than 1%. Although all intercepts are significant, changes in aggregate misallocation and its components as macroeconomic risk factors are non-return based and nontradable, so we cannot emphasize the significance of the intercepts unless they are implied from a theoretical model.

2.5.4 Industry classification

First, to examine whether the main results are driven by firms in leading industries, I re-estimate the pricing model after excluding technology firms, identified by the two-digit SIC code "73." The results, reported in Table A6, indicate that the MPK spread remains significantly negatively priced in all models. These findings confirm that superstar firms are not necessarily concentrated in leading industries such as technology. Thus, the effects documented in the paper are not driven solely by the dominance of firms in these industries.

Second, in Table A7, I assess the robustness of the main results using the 300-industry classification from Hoberg and Phillips (2016), which provides a more granular delineation of industry boundaries based on product market similarities. The results show that the MPK spread remains significantly negatively priced, consistent with the main findings. This suggests that the main conclusions are not sensitive to the choice of industry classification.

2.5.5 Value-weighted capital misallocation

The Appendix derives the general formula for the subsample decomposition of the value-weighted variance. The value-weighted capital misallocation has the form

$$\sigma^2_{mpk,w} = \underbrace{\frac{N}{N-1}\sum_{k=1}^{K}\frac{N_k-1}{N_k}\Omega_k\sigma^2_{mpk,w,k}}_{\text{Within-group misallocation}} + \underbrace{\frac{N}{N-1}\sum_{k=1}^{K}\Omega_k(\mu_{mpk,w,k}-\mu_{mpk,w})^2}_{\text{Between-group misallocation}} \tag{2.20}$$

where $\Omega_k = \sum_{i\in k} w_i$ is the total weight for each portfolio, and w_i is the weight of each stock. I normalize the weights so that $\sum_{k=1}^{K}\Omega_k = 1$. Furthermore, $\mu_{mpk,w} = \sum_{i=1}^{N} w_i mpk_i$ is the value-weighted mean MPK in the whole sample, and $\mu_{mpk,w,k} = \sum_{i\in k}\frac{w_i}{\Omega_k}mpk_i$ is the value-weighted mean MPK in each portfolio.

In the case when $k = 2$, the value-weighted capital misallocation has the form

$$\sigma^2_{mpk,w} = \underbrace{\frac{N}{N-1}\frac{N_0-1}{N_0}\Omega_0\sigma^2_{mpk,w,0}}_{\text{Misallocation among non-superstars}} + \underbrace{\frac{N}{N-1}\frac{N_*-1}{N_*}\Omega_*\sigma^2_{mpk,w,*}}_{\text{Misallocation among superstars}} + \underbrace{\frac{N}{N-1}\left[\Omega_0(\mu_{mpk,w,0}-\mu_{mpk,w})^2+\Omega_*(\mu_{mpk,w,*}-\mu_{mpk,w})^2\right]}_{\text{MPK spread}} \tag{2.21}$$

Table 2.13 reports results using Giglio and Xiu (2021)'s 202 test portfolios. When the model includes the market, changes in the aggregate capital misallocation are significantly and negatively priced. However, the significance vanishes when the model includes the Fama and French (1992b, 1993b) three factors and momentum factor. Consistently with the main finding, changes in the MPK spread carry a negative price of risk. Thus, the pricing power of changes in misallocation between superstar firms and other firms is robust to the cross-sectional measure for aggregate capital misallocation.

Table 2.13: **Pricing Giglio and Xiu (2021)'s 202 portfolios using value-weighted misallocation**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Models including MKT factor									
Constant	10.479	10.494	14.534	13.911	13.919	13.948	13.411	13.374	14.375
	(4.25)***	(4.22)***	(5.83)***	(5.55)***	(5.65)***	(5.84)***	(5.68)***	(5.55)***	(5.97)***
	[4.25]***	[3.34]***	[4.87]***	[4.66]***	[3.50]***	[4.78]***	[4.64]***	[3.63]***	[4.22]***
MKT	0.092	-0.149	-1.223	-1.097	-1.146	-1.139	-0.976	-1.039	-1.310
	(0.10)	(-0.17)	(-1.47)	(-1.30)	(-1.35)	(-1.38)	(-1.19)	(-1.23)	(-1.58)
	[0.09]	[-0.12]	[-1.05]	[-0.94]	[-0.77]	[-0.97]	[-0.83]	[-0.73]	[-0.99]
$\Delta\text{Misall}_{\text{total}}$		-0.769				-0.434	-0.509	-0.326	
		(-2.50)**				(-1.94)*	(-2.19)**	(-1.33)	
		[-1.95]*				[-1.53]	[-1.73]*	[-0.85]	
$\Delta\text{Misall}_{\text{rest}}$			0.609			0.297			0.384
			(2.26)**			(1.46)			(1.86)*
			[1.84]*			[1.15]			[1.28]
$\Delta\text{Misall}_{\text{top}}$				-0.613			-0.646		-0.361
				(-2.29)**			(-2.37)**		(-1.43)
				[-1.88]*			[-1.89]*		[-0.99]
ΔMPK spread					-1.266			-1.048	-0.876
					(-3.57)***			(-3.57)***	(-3.84)***
					[-2.19]**			[-2.31]**	[-2.64]***
R^2	0.001	0.127	0.210	0.256	0.300	0.300	0.263	0.343	0.376
Adj. R^2	-0.004	0.118	0.202	0.249	0.293	0.289	0.251	0.334	0.363
RMSE	2.524	2.366	2.250	2.183	2.118	2.124	2.179	2.056	2.011

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolios sorted by size and momentum. The sample runs from 1975:Q1 to 2023:Q4.

Table 2.13: **Pricing Giglio and Xiu (2021)'s 202 portfolios using value-weighted misallocation - continued**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Models including MKT, SMB, HML, and UMD factors									
Constant	7.550	7.473	7.272	7.330	7.838	7.322	7.130	7.783	7.485
	(3.82)***	(3.71)***	(3.92)***	(3.87)***	(3.87)***	(3.97)***	(3.85)***	(3.86)***	(4.05)***
	[3.59]***	[3.48]***	[3.68]***	[3.63]***	[3.29]***	[3.72]***	[3.59]***	[3.29]***	[3.45]***
MKT	0.478	0.372	0.412	0.395	0.253	0.405	0.443	0.270	0.333
	(0.61)	(0.47)	(0.54)	(0.51)	(0.32)	(0.53)	(0.58)	(0.34)	(0.43)
	[0.46]	[0.36]	[0.41]	[0.39]	[0.23]	[0.40]	[0.43]	[0.24]	[0.31]
SMB	0.630	0.548	0.561	0.566	0.622	0.555	0.570	0.617	0.635
	(1.66)*	(1.45)	(1.49)	(1.52)	(1.66)*	(1.48)	(1.53)	(1.65)	(1.71)*
	[1.16]	[1.01]	[1.04]	[1.05]	[1.09]	[1.03]	[1.05]	[1.08]	[1.12]
HML	0.966	0.932	0.917	0.933	0.916	0.917	0.893	0.899	0.865
	(1.96)*	(1.87)*	(1.84)*	(1.87)*	(1.84)*	(1.84)*	(1.80)*	(1.82)*	(1.75)*
	[1.39]	[1.32]	[1.30]	[1.32]	[1.23]	[1.30]	[1.26]	[1.21]	[1.17]
UMD	1.575	1.646	1.641	1.660	1.731	1.635	1.624	1.709	1.690
	(2.82)***	(2.93)***	(2.92)***	(2.95)***	(3.07)***	(2.91)***	(2.89)***	(3.04)***	(3.00)***
	[1.97]*	[2.04]**	[2.03]**	[2.06]**	[2.02]**	[2.02]**	[2.01]**	[2.00]**	[1.98]**
$\Delta\text{Misall}_{\text{total}}$		-0.136				-0.132	-0.119	-0.058	
		(-0.76)				(-0.74)	(-0.67)	(-0.32)	
		[-0.67]				[-0.65]	[-0.59]	[-0.26]	
$\Delta\text{Misall}_{\text{rest}}$			-0.051			-0.052			0.083
			(-0.27)			(-0.28)			(0.46)
			[-0.24]			[-0.25]			[0.37]
$\Delta\text{Misall}_{\text{top}}$				0.040			0.014		0.024
				(0.23)			(0.08)		(0.14)
				[0.20]			[0.07]		[0.11]
ΔMPK spread					-0.530			-0.509	-0.500
					(-2.89)***			(-2.76)***	(-2.72)***
					[-2.33]**			[-2.23]**	[-2.20]**
R^2	0.528	0.528	0.528	0.529	0.550	0.529	0.533	0.552	0.556
Adj. R^2	0.519	0.515	0.516	0.517	0.539	0.514	0.518	0.538	0.540
RMSE	1.748	1.753	1.752	1.751	1.711	1.756	1.748	1.712	1.709

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.5.6 Tangible capital versus intangible capital

With a rising trend of intangible capital among superstars, in this section, I inspect whether any specific type of capital could drive capital misallocation and may affect the main results. In the same cross-section with Table 2.1, when I estimate the changes in aggregate misallocation and its components using physical capital, the magnitude of the price of risk inflates, but the statistical significance remains weak. When the changes in misallocation only consider intangible capital, then the price of risk is no longer priced, but the negative size reserves. Importantly, in both types of capital, the price of risk of changes in the MPK spread remains negative and strongly significant. These results imply that the dispersion in the mean tangible and intangible MPK between superstars and non-superstars is equally important. Thus, no particular type of capital drives the main findings.

2.5.7 Pre-2000s and post-2000s subsamples

Two trends that may affect superstars differently between these two periods are (1) the composition of superstars and (2) the decline in number of listed firms. Before the 2000s, superstar firms are mainly firms in the manufacturing industries with high tangible capital. After the 2000s, many superstar firms are in the tech and service industry with high intangible capital (Schlingemann and Stulz, 2022). Thus, there is an increasing market concentration towards firms in riskier industries in the later subsample.

On the market-wide, the number of listed firms has declined after the 2000s (Doidge, Karolyi, and Stulz, 2017). This trend may also affect the markup of superstars in the public market so I inspect the pricing power of changes in capital misallocation in each subsample.

In general, the long-short portfolio yields a negative expected return and the decreasing trend in expected returns across the beta-sorted portfolios remains across the subsamples. One interesting result is that before the 2000s when superstars were mainly in the manufacturing industries, the significance of the results was weak. Furthermore,

Table 2.14: **Pricing tangible versus intangible misallocation**

		Tangible capital				Intangible capital			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	12.090	10.964	11.798	13.468	10.224	12.225	13.150	12.183	17.843
	(3.67)***	(3.11)***	(3.33)***	(4.01)***	(3.10)***	(3.66)***	(3.94)***	(3.69)***	(5.67)***
	[3.67]***	[3.05]***	[3.33]***	[3.03]***	[1.71]*	[3.65]***	[3.55]***	[2.63]***	[3.73]***
MKT	-0.257	-0.024	-0.194	-0.965	-0.102	-0.283	-0.497	-0.598	-1.777
	(-0.25)	(-0.02)	(-0.18)	(-0.94)	(-0.10)	(-0.27)	(-0.47)	(-0.59)	(-1.87)*
	[-0.22]	[-0.02]	[-0.15]	[-0.65]	[-0.05]	[-0.23]	[-0.38]	[-0.39]	[-1.13]
$\Delta\text{Misall}_{\text{total}}$		-0.203				0.069			
		(-0.49)				(0.13)			
		[-0.47]				[0.13]			
$\Delta\text{Misall}_{\text{rest}}$			-0.041				0.443		
			(-0.10)				(0.88)		
			[-0.09]				[0.78]		
$\Delta\text{Misall}_{\text{top}}$				-0.710				-0.897	
				(-2.21)**				(-2.00)**	
				[-1.66]*				[-1.42]	
ΔMPK spread					-1.168				-0.961
					(-3.52)***				(-2.82)***
					[-1.93]*				[-1.84]*
R^2	0.012	0.024	0.013	0.259	0.529	0.012	0.031	0.195	0.191
Adj. R^2	-0.018	-0.037	-0.049	0.213	0.499	-0.050	-0.029	0.145	0.141
RMSE	2.701	2.727	2.742	2.375	1.894	2.743	2.716	2.475	2.482

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 size $\times$ book-to-market and 10 momentum portfolios. I use sales as output and net property, plant, and equipment (ppentq) as physical capital plus intangible capital estimated from Eisfeldt and Papanikolaou (2013). The sample runs from 1975:Q1 to 2023:Q4. Returns and risk premia are reported in percent per year (quarterly percentages multiplied by four).

Table 2.15: **Exposure to changes in the MPK spread - subsample periods**

	Q1 (low)	Q2	Q3	Q4	Q5 (high)	Q5–Q1
Panel A: Pre-2000s						
R^e	10.971**	10.566***	9.888***	9.995***	7.050*	-3.922*
	(2.56)	(2.85)	(2.90)	(2.95)	(1.83)	(-1.86)
α_{CAPM}	0.212	0.768	0.889	1.316	-3.161	-3.373
	(0.10)	(0.64)	(0.73)	(0.95)	(-1.65)	(-1.06)
α_{FF3}	2.004	0.551	1.126	0.414	-1.249	-3.253
	(0.95)	(0.39)	(0.86)	(0.26)	(-0.67)	(-1.05)
α_{FF5}	2.977	-0.409	-0.807	-2.250	-3.554	-6.531*
	(1.28)	(-0.27)	(-0.66)	(-1.56)	(-1.66)	(-1.80)
Panel B: Post-2000s						
R^e	11.099**	9.968***	10.548***	8.465**	5.755	-5.344***
	(2.36)	(2.64)	(2.97)	(2.25)	(1.28)	(-2.83)
α_{CAPM}	2.060	2.538*	3.371***	1.127	-2.876	-4.936**
	(1.00)	(1.93)	(3.31)	(0.71)	(-1.46)	(-2.44)
α_{FF3}	1.876	2.416*	3.531***	1.318	-2.856	-4.731***
	(0.96)	(1.84)	(3.67)	(0.88)	(-1.45)	(-2.57)
α_{FF5}	3.657**	1.477	2.370**	-1.174	-3.357	-7.013**
	(2.06)	(1.10)	(2.32)	(-0.82)	(-1.52)	(-2.28)

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports value-weighted average excess returns and alphas in annual percentage for portfolios sorted on exposure to changes in the MPK spread. Panel A reports the results for the subsample 1975:Q1–2000:Q4. Panel B reports the results for the subsample 2001:Q1–2023:Q4. For each stock, I regress the quarterly excess returns either on changes in misallocation or each component by a rolling window of 20 quarters (with a minimum of 12 quarters available). Each quarter, I sort stocks into quintiles based on their misallocation-beta, lagging by one quarter. I hold and rebalance the portfolio every quarter.

the economic magnitude is also lower, compared to the expected returns in the post-2000s subsample. These results cast on the rising importance of superstars in driving the price of risk of changes in misallocation.

2.5.8 Annual frequency

The main results in the paper use the quarterly frequency. Table A9 reports replication of the sample using the annual frequency. There is a stronger pricing power for the price of risk of the changes in aggregate misallocation. The pricing of changes in the MPK spread is robust to the annual frequency. Interestingly, the Fama-French model could explain 52% of the variation in the portfolio returns, indicated by the adjusted R-squared. When including changes in the MPK spread, the adjusted R-squared improves to 63.7%. Table A8 in the Appendix shows the same set of results for annual frequency, using intangible capital from Peters and Taylor (2017) instead of the estimated method from Eisfeldt and Papanikolaou (2013). These results also confirm the robustness of our main findings that the MPK spread drives the price of risk of capital misallocation, implying that superstars influence asset prices via the channel of capital misallocation.

2.6 Conclusion

This paper shows the asset pricing consequence of superstar firms. Specifically, I propose a new state variable in the ICAPM framework: capital misallocation between superstar and non-superstar firms. Using a measure of cross-sectional MPK dispersion, I decompose capital misallocation into three components: the misallocation within the superstar portfolio, the misallocation within the non-superstar portfolio, and the MPK spread. The MPK spread captures the dispersion in mean MPKs across the two portfolios.

In the cross-section of stock returns, I find that changes in the MPK spread are significantly and negatively priced. Stocks with negative exposure to the changes in this component carry higher risk premia, while stocks with positive exposure carry lower risk premia. Besides, adding the MPK spread-mimicking portfolio also improves the mean-variance efficiency of other portfolios. In the time series, the MPK spread negatively predicts long-run consumption growth, industrial production growth, employment growth, and future stock market returns. Whereas, in the short run, the MPK spread predicts lower aggregate innovation growth and innovation growth among non-superstars.

These findings are economically intuitive. In an efficient market, capital flows towards its most productive uses. Yet, if superstar firms' MPKs are driven by market distortions, such as market power that raises barriers to competitor entry, the higher MPK reflects an inefficiency rather than a pure productivity advantage. Such misallocation can dampen economic growth by restricting capital access to smaller, potentially innovative competitors. Therefore, capital misallocation between superstar and non-superstar firms emerges as a state variable, with changes in this misallocation capturing a priced risk factor.

While the literature highlights the rise of superstars and the rise in capital misallocation, this paper highlights the role of superstars in shaping the price of risk associated with capital misallocation. One direction for future work is to model and test the inno-

vation channel through which superstar firms deter innovation growth of other firms, resulting in the pricing power and predictability power of changes in misallocation between superstar and non-superstar firms.

A1 Appendix: Equally weighted capital misallocation

In each quarter t, capital misallocation is given by the dispersion in firm-level (log) MPKs in the economy

$$\sigma^2_{mpk,t} = \frac{1}{N-1}\sum_{i=1}^{N}(mpk_{it} - \mu_{mpk,t})^2 \tag{2.22}$$

where $\sigma^2_{mpk,t}$ and $\mu_{mpk,t} = \frac{1}{N}\sum_{i=1}^{N} mpk_{it}$ denote the variance and the mean MPK of the whole sample. Dropping the time subscript for convenience and expanding the variance yield

$$\sigma^2_{mpk} = \frac{1}{N-1}\sum_{i=1}^{N}(mpk_i - \mu_{mpk})^2 \tag{2.23}$$

$$= \frac{1}{N-1}\sum_{i=1}^{N} mpk_i^2 - 2\frac{N}{N-1}\mu^2_{mpk} + \frac{N}{N-1}\mu^2_{mpk} \tag{2.24}$$

Hence,

$$\sigma^2_{mpk} + \frac{N}{N-1}\mu^2_{mpk} = \frac{1}{N-1}\sum_{i=1}^{N} mpk_i^2 \tag{2.25}$$

$$\iff (N-1)\sigma^2_{mpk} + N\mu^2_{mpk} = \sum_{i=1}^{N} mpk_i^2 \tag{2.26}$$

Assume that firms are sorted into K portfolios, in which we allow for different numbers of stocks in each portfolio and denote it as N_k, where $\sum_{k=1}^{K} N_k = N_1 + N_2 + \cdots + N_K = N$. We can derive Equation (2.26) analogously for each portfolio with the corresponding variance $\sigma^2_{mpk,k}$ and mean $\mu_{mpk,k}$. Furthermore, we can decompose the right-hand side of Equation (2.25) into

$$\sigma^2_{mpk} + \frac{N}{N-1}\mu^2_{mpk} = \frac{1}{N-1}\left[\sum_{i=1}^{N_1} mpk_i^2 + \sum_{i=1}^{N_2} mpk_i^2 + \ldots + \sum_{i=1}^{N_K} mpk_i^2\right] \tag{2.27}$$

$$\iff \sigma^2_{mpk} = \frac{1}{N-1}\sum_{k=1}^{K}(N_k - 1)\sigma^2_{mpk,k} + \frac{1}{N-1}\left[\sum_{k=1}^{K} N_k\mu^2_{mpk,k} - N\mu^2_{mpk}\right] \tag{2.28}$$

Since the total sample mean equals the weighted average of the subsample means, weighted by the number of observations,

$$\mu_{mpk} = \frac{N_1\mu_{mpk,1} + N_2\mu_{mpk,2} + \ldots + N_K\mu_{mpk,K}}{N_1 + N_2 + \ldots + N_k} \tag{2.29}$$

$$= \frac{1}{N}\sum_{k=1}^{K} N_k\mu_{mpk,k} \tag{2.30}$$

$$\Rightarrow N\mu^2_{mpk} = \frac{1}{N}\left(\sum_{k=1}^{K} N_k\mu_{mpk,k}\right)^2 \tag{2.31}$$

Hence,

$$\sigma^2_{mpk} = \frac{1}{N-1}\sum_{k=1}^{K}(N_k - 1)\sigma^2_{mpk,k} + \frac{1}{N-1}\left[\sum_{k=1}^{K} N_k\mu^2_{mpk,k} - \frac{1}{N}\left(\sum_{k=1}^{K} N_k\mu_{mpk,k}\right)^2\right] \tag{2.32}$$

$$= \frac{1}{N-1}\sum_{k=1}^{K}(N_k - 1)\sigma^2_{mpk,k} + \frac{N}{N-1}\left[\underbrace{\frac{1}{N}\sum_{k=1}^{K} N_k\mu^2_{mpk,k}}_{\mathbb{E}(\mu^2_{mpk,k})} - \underbrace{\left(\frac{1}{N}\sum_{k=1}^{K} N_k\mu_{mpk,k}\right)^2}_{\mathbb{E}(\mu_{mpk,k})^2}\right] \tag{2.33}$$

$$\tag{2.34}$$

Recognizing that the second and the third terms capture the second and the (squared) first moment or the expected value of the subsample mean MPKs, we can further

rewrite the aggregate misallocation as

$$\sigma^2_{mpk} = \underbrace{\sum_{k=1}^{K} \frac{N_k - 1}{N-1} \sigma^2_{mpk,k}}_{\text{Within-group misallocation}} + \underbrace{\frac{N}{N-1} \text{Var}(\mu_{mpk,k})}_{\text{Between-group misallocation}} \tag{2.35}$$

where $\text{Var}(\mu_{mpk})$ denotes the variance of the portfolio mean MPKs. Thus, we can decompose the aggregate capital misallocation into the portfolio-specific misallocation, i.e., the misallocation among the firms in each portfolio and the dispersion in the *mean* MPKs across portfolios.

To separate superstars, I sort the sample into two portfolios each quarter: the superstar portfolio which includes superstar firms ($k = *$), and the non-superstar portfolio which includes the remaining firms ($k = 0$). The second term of Equation (2.35) simplifies to the case when $K = 2$. Specifically,

$$\begin{aligned}
\text{Var}(\mu_{mpk,k}) &= \frac{1}{N} \sum_{k=1}^{2} N_k \mu^2_{mpk,k} - \left(\frac{1}{N} \sum_{k=1}^{2} N_k \mu_{mpk,k} \right)^2 \\
&= \frac{N_0(N - N_0)}{N^2} \mu^2_{mpk,0} + \frac{N_*(N - N_*)}{N^2} \mu^2_{mpk,*} - 2\frac{N_0 N_*}{N^2} \mu_{mpk,0} \mu_{mpk,*} \\
&= \frac{N_0 N_*}{N^2} (\mu_{mpk,0} - \mu_{mpk,*})^2
\end{aligned}$$

Hence, we can rewrite Equation (2.35) in this case as

$$\sigma^2_{mpk} = \underbrace{\frac{N_0 - 1}{N-1} \sigma^2_{mpk,0}}_{\text{Misallocation among non-superstars}} + \underbrace{\frac{N_* - 1}{N-1} \sigma^2_{mpk,*}}_{\text{Misallocation among superstars}} + \underbrace{\frac{N_0 N_*}{N(N-1)} (\mu_{mpk,0} - \mu_{mpk,*})^2}_{\text{MPK spread}} \tag{2.36}$$

A2 Appendix: Value-weighted capital misallocation

The general formula for the value-weighted variance, assuming no observation with zero weight, has the form

$$s_w^2 = \frac{N}{N-1} \sum_{i=1}^{N} w_i (x_i - \bar{x}_w)^2 \tag{2.37}$$

where N is the number of observations, w_i is the weight for the observation x_i, $\bar{x}_w = \sum_{i=1}^{N} w_i x_i$ is the value-weighted mean of the sample.

Assume the sample splits into K portfolios, each with N_k observations and weights w_i such that $\sum_{i \in k} w_i = \Omega_k$ for each portfolio k. I normalize the weights so that $\sum_{i=1}^{N} w_i = 1$ and

$$\sum_{k=1}^{K} \Omega_k = \sum_{k=1}^{K} \sum_{i \in k} w_i = \sum_{i=1}^{N} w_i = 1. \tag{2.38}$$

For each portfolio k, the weighted variance $s_{w,k}^2$ and the weighted mean $\bar{x}_{w,k}$ are given by

$$s_{w,k}^2 = \frac{N_k}{N_k - 1} \sum_{i \in k} \frac{w_i}{\Omega_k} (x_i - \bar{x}_{w,k})^2 \tag{2.39}$$

$$\bar{x}_{w,k} = \sum_{i \in k} \frac{w_i}{\Omega_k} x_i \iff \sum_{i \in k} w_i x_i = \Omega_k \bar{x}_{w,k} \tag{2.40}$$

Then total weighted mean $\bar{x}_w$ is of the form

$$\bar{x}_w = \sum_{i=1}^{N} w_i x_i = \sum_{k=1}^{K} \sum_{i \in k} w_i x_i = \sum_{k=1}^{K} \Omega_k \bar{x}_{w,k}$$

From the total weighted variance, we have

$$\begin{aligned} s_w^2 &= \frac{N}{N-1}\sum_{i=1}^{N} w_i(x_i - \bar{x}_w)^2 \\ &= \frac{N}{N-1}\sum_{k=1}^{K}\sum_{i\in k} w_i(x_i - \bar{x}_w)^2 \\ &= \frac{N}{N-1}\sum_{k=1}^{K}\sum_{i\in k} w_i\left[(x_i - \bar{x}_{w,k})^2 + 2(x_i - \bar{x}_{w,k})(\bar{x}_{w,k} - \bar{x}_w) + (\bar{x}_{w,k} - \bar{x}_w)^2\right] \end{aligned}$$

Consider the cross-term

$$\begin{aligned} \sum_{k=1}^{K}\sum_{i\in k} w_i(x_i - \bar{x}_{w,k})(\bar{x}_{w,k} - \bar{x}_w) &= \sum_{k=1}^{K}(\bar{x}_{w,k} - \bar{x}_w)\sum_{i\in k} w_i(x_i - \bar{x}_{w,k}) \\ &= \sum_{k=1}^{K}(\bar{x}_{w,k} - \bar{x}_w)(\sum_{i\in k} w_i x_i - \bar{x}_{w,k}\sum_{i\in k} w_i) \\ &= \sum_{k=1}^{K}(\bar{x}_{w,k} - \bar{x}_w)(\Omega_k \bar{x}_{w,k} - \Omega_k \bar{x}_{w,k}) \\ &= 0 \end{aligned}$$

Thus,

$$s_w^2 = \frac{N}{N-1}\left[\underbrace{\sum_{k=1}^{K}\sum_{i\in k} w_i(x_i - \bar{x}_{w,k})^2}_{\text{Within-group variances}} + \underbrace{\sum_{k=1}^{K}\sum_{i\in k} w_i(\bar{x}_{w,k} - \bar{x}_w)^2}_{\text{Between-group variance}}\right]$$

The first term represents the within-group variances, and the second term represents the between-group variance, i.e. the weighted variance of the subsample means from the total mean. We can write the within-group variances as

$$\begin{aligned} \sum_{k=1}^{K}\sum_{i\in k} w_i(x_i - \bar{x}_{w,k})^2 &= \sum_{k=1}^{K}\Omega_k \sum_{i\in k}\frac{w_i}{\Omega_k}(x_i - \bar{x}_{w,k})^2 \\ &= \sum_{k=1}^{K}\frac{N_k - 1}{N_k}\Omega_k s_{w,k}^2 \end{aligned}$$

where $\frac{w_i}{\Omega_k}$ is the weight within each portfolio such that $\sum_{i \in k} \frac{w_i}{\Omega_k} = 1$. For the between-group variance,

$$\begin{aligned}\sum_{k=1}^{K} \sum_{i \in k} w_i (\bar{x}_{w,k} - \bar{x}_w)^2 &= \sum_{k=1}^{K} (\bar{x}_{w,k} - \bar{x}_w)^2 \sum_{i \in k} w_i \\ &= \sum_{k=1}^{K} \Omega_k (\bar{x}_{w,k} - \bar{x}_w)^2\end{aligned}$$

Combining both terms, we have:

$$s_w^2 = \frac{N}{N-1} \left[\sum_{k=1}^{K} \frac{N_k - 1}{N_k} \Omega_k s_{w,k}^2 + \sum_{k=1}^{K} \Omega_k (\bar{x}_{w,k} - \bar{x}_w)^2 \right]$$

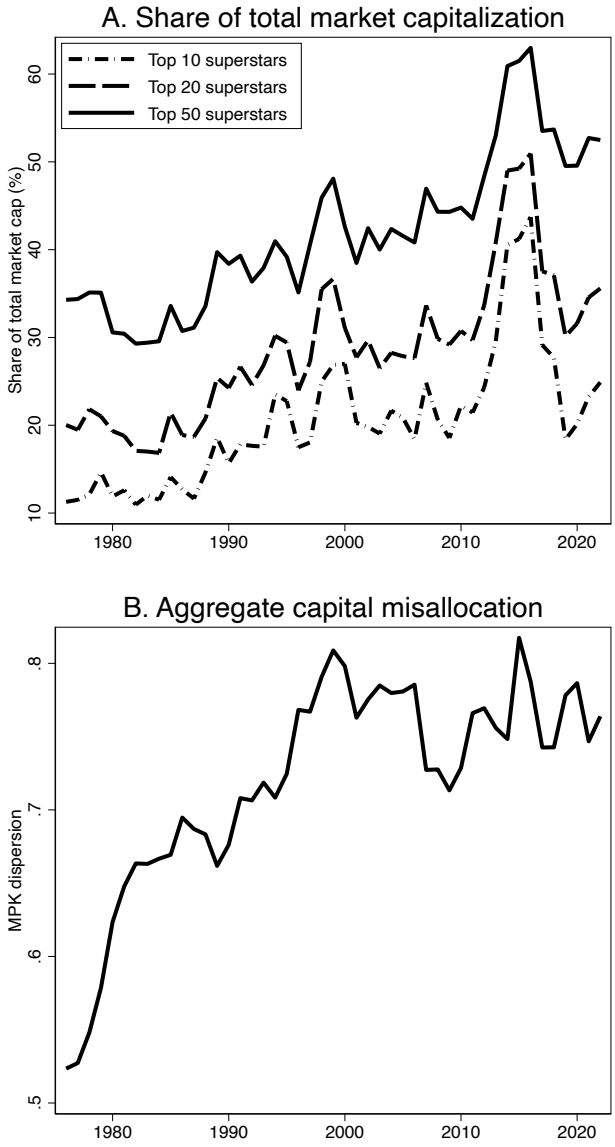
where $\Omega_k = \sum_{i \in k} w_i$, $\bar{x}_w = \sum_{i=1}^{N} w_i x_i$, and $\bar{x}_{w,k} = \sum_{i \in k} \frac{w_i}{\Omega_k} x_i$.

Given $K = 2$, the formula simplifies to

$$s_w^2 = \frac{N}{N-1} \left[\frac{N_1 - 1}{N_1} \Omega_1 s_{w,1}^2 + \frac{N_2 - 1}{N_2} \Omega_2 s_{w,2}^2 + \Omega_1 (\bar{x}_{w,1} - \bar{x}_w)^2 + \Omega_2 (\bar{x}_{w,2} - \bar{x}_w)^2 \right].$$

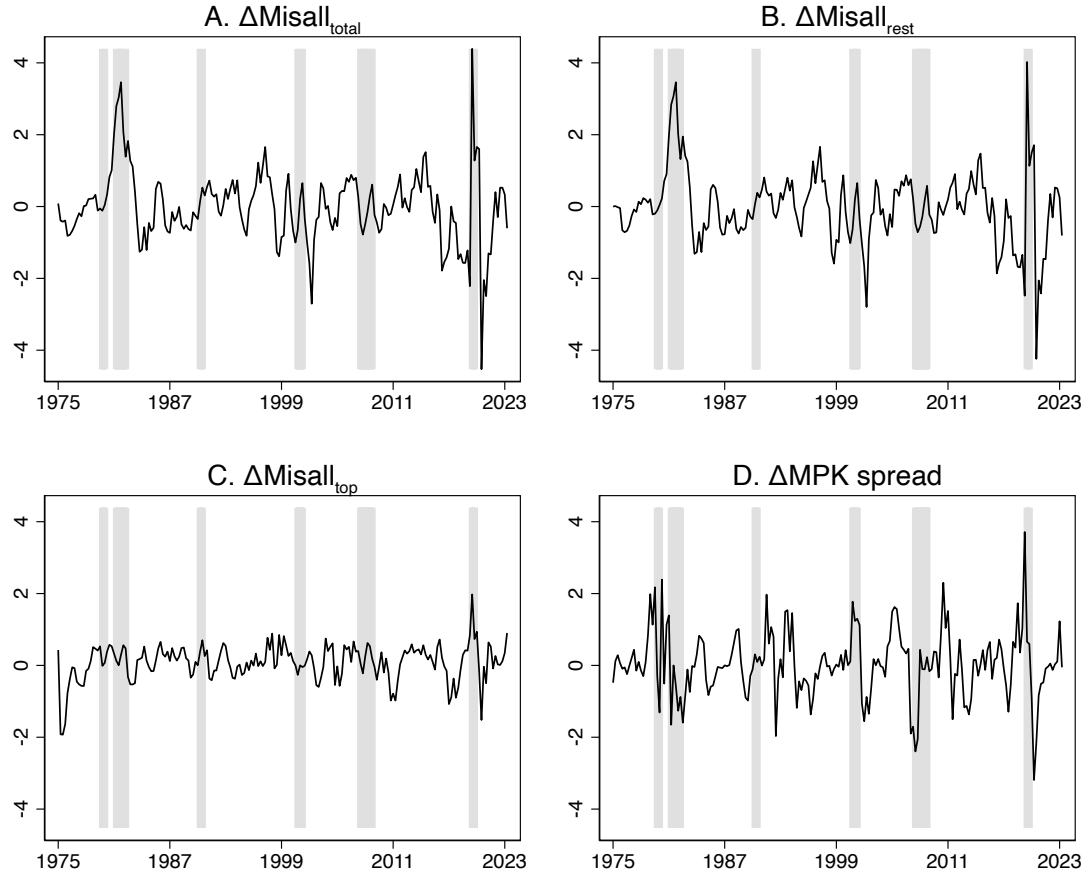
A3 Appendix: Additional results

Figure A1: **Market concentration and capital misallocation**



Description. This figure shows the rising trend in market concentration and aggregate capital misallocation. Panel A plots the market cap of top 10, 20, and 50 superstar firms by size and market power over the total market cap in percentage. Panel B plots the dispersion in MPK across firms, where MPK is the log output-to-capital (measured by sale/cogs). The sample is annual from 1975 to 2023.

Figure A2: **Changes in capital misallocation against NBER recessions**



Description. This figure plots the changes in aggregate misallocation $\Delta\sigma^2_{mpk,t}$, changes in misallocation in non-superstars $\Delta\sigma^2_{mpk,t0}$, changes in misallocation in superstars $\Delta\sigma^2_{mpk,t*}$, and changes in the MPK spread $\Delta \operatorname{Var}(\mu_{mpk,k})$. The sample is from 1975:Q1 to 2023:Q4. In each quarter, capital misallocation $\sigma^2_{mpk,t}$ is the cross-sectional dispersion of MPK across firms. The changes in misallocation are the annual changes in capital misallocation:

$$\Delta\sigma^2_{mpk,t} = \sigma^2_{mpk,t} - \sigma^2_{mpk,t-4},$$

whose level can be decomposed into:

$$\underbrace{\sigma^2_{mpk}}_{\text{Total misallocation (Misall}_{\text{total}})} = \underbrace{\frac{N_0-1}{N-1}\sigma^2_{mpk,0}}_{\text{Misallocation within non-superstars (Misall}_{\text{rest}})} + \underbrace{\frac{N_*-1}{N-1}\sigma^2_{mpk,*}}_{\text{Misallocation within superstars (Misall}_{\text{top}})} + \underbrace{\frac{N_0 N_*}{N(N-1)}(\mu_{mpk,0}-\mu_{mpk,*})^2}_{\text{MPK spread}}$$

Table A1: **Significance of ΔMPK spread - simulation results**

	Frequency	Percent
*** $p < 0.01$	2	.4
** $p < 0.05$	24	4.8
* $p < 0.10$	36	7.2
* $p < 1$	438	88
Total	500	100

Description. This table reports the significance of the price of risk of changes in the MPK spread in the second-stage Fama-MacBeth regression. Each simulation selects randomly 50 firms to the superstar portfolio.

Table A2: **Exposure to changes in the MPK spread (equally weighted results)**

	Q1 (low)	Q2	Q3	Q4	Q5 (high)	Q5–Q1
Panel A: Expected return						
R^e	16.965***	13.699***	12.935***	12.607***	14.025***	-2.940**
	(4.25)	(4.35)	(4.26)	(3.93)	(3.68)	(-2.56)
Panel B: CAPM						
MKT	1.403***	1.138***	1.115***	1.171***	1.332***	-0.072*
	(18.73)	(21.34)	(21.07)	(21.50)	(19.49)	(-1.92)
α_{CAPM}	4.712**	3.464**	2.990**	2.357	2.466	-2.245*
	(2.32)	(2.28)	(2.09)	(1.57)	(1.28)	(-1.91)
Panel C: FF3 + UMD						
MKT	1.091***	0.952***	0.941***	0.966***	1.029***	-0.062
	(21.56)	(30.84)	(32.38)	(33.46)	(27.10)	(-1.42)
SMB	1.189***	0.825***	0.772***	0.854***	1.144***	-0.045
	(14.76)	(16.25)	(16.24)	(18.39)	(18.21)	(-0.60)
HML	-0.013	0.147***	0.136***	0.109***	0.007	0.020
	(-0.18)	(4.20)	(4.36)	(2.82)	(0.13)	(0.36)
UMD	-0.158*	-0.097**	-0.089**	-0.126**	-0.177**	-0.020
	(-1.80)	(-2.52)	(-2.02)	(-2.30)	(-2.47)	(-0.42)
$\alpha_{FF3+UMD}$	6.504***	3.955***	3.442***	3.257***	4.323***	-2.180*
	(4.37)	(4.75)	(4.44)	(3.65)	(3.18)	(-1.82)
Panel D: FF5						
MKT	1.108***	0.988***	0.986***	1.021***	1.065***	-0.043
	(19.57)	(31.72)	(34.85)	(34.28)	(24.84)	(-0.97)
SMB	1.187***	0.860***	0.818***	0.907***	1.171***	-0.015
	(15.67)	(17.31)	(18.57)	(23.47)	(17.71)	(-0.22)
HML	-0.045	0.087*	0.052	0.006	-0.032	0.013
	(-0.45)	(1.68)	(1.08)	(0.12)	(-0.42)	(0.15)
RMW	-1.286***	0.126	0.471**	0.364**	-0.672**	0.614
	(-2.87)	(0.64)	(2.11)	(2.11)	(-2.40)	(1.45)
CMA	1.017*	0.717***	0.809***	1.085***	0.964**	-0.054
	(1.94)	(2.78)	(3.20)	(3.72)	(2.09)	(-0.13)
α_{FF5}	6.104***	2.397***	1.430*	0.861	2.933**	-3.172**
	(4.37)	(2.85)	(1.79)	(1.15)	(2.51)	(-2.39)

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports equally weighted average excess returns and alphas in annual percentage for portfolios sorted on exposure to changes in the MPK spread. For each stock, I regress the quarterly excess returns either on changes in misallocation or on each component by a rolling window of 20 quarters (with a minimum of 12 quarters available). Each quarter, I sort stocks into quintiles based on their misallocation-beta, lagging by one quarter. I hold and rebalance the portfolio every quarter. The sample runs from 1975:Q1 to 2023:Q4.

Table A3: **Characteristics of stocks in the MPK spread-mimicking portfolios**

	Low	Q2	Q3	Q4	High	High–Low	t(High–Low)
$\beta_{\Delta\text{MPK spread}}$	-0.592	-0.198	-0.022	0.169	0.513	1.105	(2.43)
Market cap	17,492.180	38,355.184	49,901.367	29,264.824	25,961.316	8,469.136	(3.36)
Markup ratio	1.785	1.680	1.682	1.825	1.934	0.148	(4.54)
Book-to-market	0.501	0.490	0.474	0.478	0.461	-0.039	(-1.94)
Innovation	19.908	29.970	36.518	18.421	18.042	-1.866	(-1.73)
Duration	62.301	80.425	81.964	79.883	64.740	2.439	(2.35)
Investment	2,179.251	6,231.840	12,381.994	6,100.768	4,392.167	2,212.916	(4.32)
Physical capital	3,754.690	9,973.484	9,830.880	7,848.565	4,537.075	782.385	(2.29)
Intangible capital	1,812.195	2,899.301	2,930.864	2,621.818	2,197.016	384.821	(4.26)

Description. This table reports the value-weighted average characteristics of stocks in each MPK spread-mimicking portfolio. The sample runs from 1975:Q1 to 2023:Q4.

Table A4: **Predicting aggregate patent ratio**

	$\Delta\text{Misall}_{\text{total}}$			$\Delta\text{Misall}_{\text{rest}}$			$\Delta\text{Misall}_{\text{top}}$			ΔMPK spread		
	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE	$\hat{\beta}$	R^2	RMSE
$k=1$	-0.039	0.003	0.021	0.022	0.001	0.021	-0.373	0.002	0.021	-6.769	0.154	0.019
	(-0.61)			(0.36)			(-0.39)			(-4.27)***		
	[-0.60]			[0.36]			[-0.38]			[-4.19]***		
$k=2$	0.048	0.003	0.027	0.012	0.000	0.027	1.092	0.010	0.027	-8.501	0.145	0.025
	(0.53)			(0.13)			(0.84)			(-4.05)***		
	[0.49]			[0.13]			[0.90]			[-4.52]***		
$k=3$	-0.074	0.005	0.033	-0.104	0.009	0.033	-0.734	0.003	0.033	-4.824	0.032	0.033
	(-0.58)			(-0.69)			(-0.58)			(-2.70)***		
	[-0.57]			[-0.67]			[-0.53]			[-2.65]**		
$k=4$	-0.033	0.001	0.038	-0.006	0.000	0.038	1.724	0.013	0.038	-5.373	0.030	0.038
	(-0.25)			(-0.03)			(1.13)			(-2.20)**		
	[-0.24]			[-0.03]			[1.14]			[-2.60]**		
$k=5$	-0.008	0.000	0.045	-0.077	0.002	0.045	0.357	0.000	0.045	-4.901	0.018	0.045
	(-0.04)			(-0.33)			(0.21)			(-1.86)*		
	[-0.04]			[-0.33]			[0.21]			[-2.02]**		
$k=6$	-0.248	0.016	0.050	-0.264	0.018	0.050	0.792	0.001	0.051	-15.007	0.021	0.050
	(-0.86)			(-0.91)			(0.45)			(-2.34)**		
	[-0.84]			[-0.90]			[0.47]			[-2.06]**		

t-ratio of Hodrick (1992) with $k-1$ lags in parentheses; Newey–West (1987) in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the results of the following predictive regression:

$$I_{t:t+k} = \alpha + \beta z_t + \epsilon_{t:t+k},$$

where $I_{t:t+k}$ is the innovation growth in k quarters. The predictive variables z_t are changes in the aggregate misallocation and the MPK spread. The columns show results for $k = 1, 4, 8, 12$ and 20 quarters. Following Bae, Bailey, and Kang (2021), I construct the innovation proxy I_t as the natural logarithm of one plus the number of patent applications divided by the population. The number of patents granted each year is from the US Patent Trademark Office (USPTO) and the US population is from the U.S. Bureau of Economic Analysis (BEA). The aggregate patent ratio equals = log(1 + # patents/population). The sample is at an annual frequency from 1975 to 2020.

Table A5: **Pricing 35 portfolios - Decomposition using financial constraints**

			Kaplan-Zingales (KZ) Index			Whited-Wu (WW) Index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	12.090	10.792	10.914	18.735	11.921	10.539	13.710	11.410
	(3.67)***	(3.25)***	(3.33)***	(5.02)***	(3.19)***	(3.10)***	(3.71)***	(3.43)***
	[3.67]***	[2.99]***	[2.98]***	[3.72]***	[3.19]***	[2.85]***	[3.46]***	[2.86]***
MKT	-0.257	-0.057	-0.124	-2.297	-0.208	0.015	-0.823	-0.109
	(-0.25)	(-0.05)	(-0.12)	(-2.01)**	(-0.19)	(0.01)	(-0.75)	(-0.11)
	[-0.22]	[-0.04]	[-0.10]	[-1.37]	[-0.17]	[0.01]	[-0.62]	[-0.08]
$\Delta\text{Misall}_{\text{total}}$		-0.435						
		(-0.99)						
		[-0.90]						
$\Delta\text{Misall}_{\text{rest}}$			-0.513			-0.445		
			(-1.22)			(-0.97)		
			[-1.08]			[-0.88]		
$\Delta\text{Misall}_{\text{top}}$				0.879			-0.307	
				(2.17)**			(-1.10)	
				[1.59]			[-1.01]	
ΔMPK spread					-0.027			0.475
					(-0.08)			(1.35)
					[-0.07]			[1.12]
R^2	0.012	0.064	0.093	0.351	0.013	0.060	0.092	0.048
Adj. R^2	-0.018	0.006	0.036	0.310	-0.049	0.001	0.036	-0.011
RMSE	2.701	2.669	2.629	2.224	2.741	2.676	2.629	2.692

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 portfolios sorted by size and book-to-market ratio and 10 portfolios sorted by momentum. Superstar firms are in the bottom 5% in their industries by their financial constraints. As proxies for financial constraint, I use Kaplan and Zingales (1997)'s (KZ) index and Whited and Wu (2006)'s (WW) index. The higher the KZ index or the WW index, the higher the likelihood of financial constraint.

Table A6: **Pricing 35 portfolios excluding tech industries**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	12.090	10.746	10.676	11.536	13.470	11.529	11.251	13.618	12.718
	(3.67)***	(3.26)***	(3.23)***	(3.53)***	(4.07)***	(3.75)***	(3.65)***	(4.23)***	(4.13)***
	[3.67]***	[2.99]***	[2.93]***	[3.38]***	[2.16]**	[3.55]***	[3.51]***	[2.21]**	[1.97]*
MKT	-0.257	-0.063	-0.032	-0.301	-0.901	-0.314	-0.225	-0.927	-0.627
	(-0.25)	(-0.06)	(-0.03)	(-0.30)	(-0.89)	(-0.32)	(-0.23)	(-0.92)	(-0.65)
	[-0.22]	[-0.05]	[-0.02]	[-0.25]	[-0.45]	[-0.26]	[-0.19]	[-0.46]	[-0.30]
$\Delta\text{Misall}_{\text{total}}$		-0.441				-0.052	-0.164	0.284	
		(-1.12)				(-0.19)	(-0.62)	(0.72)	
		[-1.01]				[-0.17]	[-0.57]	[0.38]	
$\Delta\text{Misall}_{\text{rest}}$			-0.480			-0.018			-0.198
			(-1.10)			(-0.06)			(-0.72)
			[-0.99]			[-0.06]			[-0.34]
$\Delta\text{Misall}_{\text{top}}$				-0.235			-0.207		0.163
				(-1.08)			(-0.95)		(0.75)
				[-1.01]			[-0.89]		[0.35]
ΔMPK spread					-1.171			-1.188	-1.259
					(-3.63)***			(-4.20)***	(-4.40)***
					[-1.92]*			[-2.19]**	[-2.09]**
R^2	0.012	0.080	0.076	0.081	0.474	0.091	0.083	0.475	0.513
Adj. R^2	-0.018	0.023	0.018	0.024	0.442	0.003	-0.005	0.424	0.448
RMSE	2.701	2.647	2.653	2.645	2.001	2.673	2.684	2.032	1.988

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 portfolios sorted by size and book-to-market ratio and 10 portfolios sorted by momentum. Superstar firms are defined by the top 5% in their industries by the markup share, using two-digit SIC codes. In this analysis, I exclude the tech industries with the two-digit SIC code of "73."

Table A7: **Pricing Giglio and Xiu (2021)'s 202 test portfolios - using Hoberg and Phillips (2016)'s 300 industry classification**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	10.628	12.403	11.509	11.335	10.464	14.040	14.125	11.862	12.459
	(2.84)***	(3.14)***	(2.96)***	(3.04)***	(2.80)***	(3.54)***	(3.51)***	(3.05)***	(3.36)***
	[2.83]***	[2.84]***	[2.84]***	[2.87]***	[2.21]**	[2.88]***	[2.82]***	[2.44]**	[2.70]***
MKT	-0.569	-1.072	-0.810	-0.918	-1.204	-1.583	-1.769	-1.564	-1.696
	(-0.40)	(-0.74)	(-0.56)	(-0.65)	(-0.84)	(-1.11)	(-1.23)	(-1.07)	(-1.18)
	[-0.31]	[-0.55]	[-0.43]	[-0.49]	[-0.56]	[-0.76]	[-0.84]	[-0.73]	[-0.81]
ΔMisall$_{\text{total}}$		-0.480				-0.631	-0.589	-0.227	
		(-1.80)*				(-2.18)**	(-2.09)**	(-0.98)	
		[-1.49]				[-1.67]*	[-1.58]	[-0.72]	
ΔMisall$_{\text{rest}}$			-0.264			-0.468			-0.215
			(-1.29)			(-2.01)**			(-1.14)
			[-1.10]			[-1.53]			[-0.83]
ΔMisall$_{\text{top}}$				-0.123			-0.121		-0.136
				(-1.81)*			(-1.79)*		(-2.01)**
				[-1.43]			[-1.25]		[-1.41]
ΔMPK spread					-0.766			-0.643	-0.584
					(-2.99)***			(-2.90)***	(-2.66)***
					[-2.20]**			[-2.11]**	[-1.94]*
R^2	0.038	0.165	0.099	0.169	0.467	0.292	0.401	0.543	0.555
Adj. R^2	0.033	0.156	0.090	0.161	0.462	0.282	0.391	0.536	0.546
RMSE	3.440	3.214	3.338	3.205	2.567	2.966	2.730	2.382	2.357

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and beta, and 25 portfolios sorted by size and momentum. Superstar firms are defined by the top 5% in their industries by the markup share, using 300 industry classification from Hoberg and Phillips (2016). The sample runs from 1975 to 2022.

Table A8: **Cross-sectional asset pricing tests - alternative measure for intangible capital**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	20.904	19.552	19.635	20.890	19.656	20.172	20.208	17.806	18.174
	(6.96)***	(7.43)***	(7.47)***	(7.24)***	(6.75)***	(7.57)***	(7.61)***	(7.31)***	(7.64)***
	[5.49]***	[6.14]***	[6.18]***	[5.69]***	[4.44]***	[5.65]***	[5.81]***	[4.53]***	[4.81]***
MKT	-11.685	-9.675	-9.761	-10.864	-9.621	-10.051	-10.137	-7.748	-8.049
	(-2.93)***	(-2.69)**	(-2.72)***	(-2.86)***	(-2.54)**	(-2.78)***	(-2.81)***	(-2.26)**	(-2.37)**
	[-2.09]**	[-1.95]*	[-1.97]*	[-2.02]*	[-1.54]	[-1.86]*	[-1.91]*	[-1.29]	[-1.36]
SMB	2.648	2.375	2.381	2.544	2.855	2.560	2.497	2.840	2.862
	(1.62)	(1.45)	(1.45)	(1.55)	(1.73)*	(1.56)	(1.52)	(1.72)*	(1.73)*
	[1.02]	[0.93]	[0.93]	[0.97]	[0.96]	[0.94]	[0.93]	[0.91]	[0.93]
HML	2.787	2.578	2.545	2.383	2.133	2.739	2.713	2.598	2.650
	(1.13)	(1.05)	(1.03)	(0.95)	(0.85)	(1.11)	(1.10)	(1.05)	(1.08)
	[0.71]	[0.67]	[0.67]	[0.60]	[0.47]	[0.67]	[0.68]	[0.56]	[0.58]
ΔMisall$_{\text{total}}$		-0.260				-0.381	-0.371	-0.144	
		(-0.79)				(-1.11)	(-1.08)	(-0.46)	
		[-0.61]				[-0.78]	[-0.78]	[-0.27]	
ΔMisall$_{\text{rest}}$			-0.223			-0.352			-0.172
			(-0.70)			(-1.05)			(-0.55)
			[-0.53]			[-0.74]			[-0.33]
ΔMisall$_{\text{top}}$				-0.101			-0.133		-0.048
				(-0.77)			(-1.01)		(-0.35)
				[-0.53]			[-0.68]		[-0.20]
ΔMPK spread					-0.463			-0.492	-0.458
					(-3.83)***			(-3.92)***	(-3.64)***
					[-2.36]**			[-2.30]**	[-2.16]**
R^2	0.568	0.559	0.557	0.563	0.629	0.591	0.578	0.672	0.673
Adj. R^2	0.542	0.524	0.522	0.528	0.599	0.549	0.535	0.639	0.632
RMSE	1.814	1.851	1.854	1.842	1.698	1.800	1.829	1.612	1.626

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 size $\times$ book-to-market, 10 momentum, 25 size and operating profitability, and 25 size and investment portfolios. I obtain intangible capital from Peters and Taylor (2017), accessed via WRDS. The sample runs from 1975 to 2022.

Table A9: **Cross-sectional asset pricing tests - annual frequency**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	21.248	21.022	20.335	20.741	24.741	19.322	20.304	24.218	23.201
	(7.74)***	(7.62)***	(7.60)***	(6.72)***	(8.15)***	(7.59)***	(6.73)***	(8.06)***	(7.03)***
	[6.05]***	[4.75]***	[4.56]***	[5.03]***	[4.20]***	[4.87]***	[5.00]***	[4.15]***	[3.86]***
MKT	-12.531	-12.336	-11.645	-13.558	-16.363	-10.691	-13.068	-15.715	-16.189
	(-3.54)***	(-3.45)***	(-3.31)***	(-3.46)***	(-4.31)***	(-3.14)***	(-3.39)***	(-4.16)***	(-3.93)***
	[-2.52]**	[-2.02]**	[-1.87]*	[-2.38]**	[-2.14]**	[-1.88]*	[-2.32]**	[-2.06]**	[-2.07]**
SMB	2.223	2.214	2.096	2.805	2.270	1.889	2.798	2.209	2.608
	(1.53)	(1.51)	(1.43)	(1.67)	(1.55)	(1.30)	(1.66)	(1.51)	(1.55)
	[0.94]	[0.80]	[0.74]	[1.00]	[0.71]	[0.70]	[1.00]	[0.69]	[0.75]
HML	2.976	3.269	3.449	3.030	3.540	3.689	3.128	3.788	3.714
	(1.68)*	(1.83)*	(1.94)*	(1.50)	(1.99)*	(2.07)**	(1.56)	(2.14)**	(1.86)*
	[1.04]	[0.97]	[1.00]	[0.91]	[0.91]	[1.12]	[0.94]	[0.98]	[0.90]
$\Delta\text{Misall}_{\text{total}}$		0.945				0.829	0.231	0.566	
		(2.96)***				(2.75)***	(0.81)	(1.97)*	
		[1.80]*				[1.71]*	[0.57]	[0.99]	
$\Delta\text{Misall}_{\text{rest}}$			1.087			0.902			0.119
			(2.95)***			(2.85)***			(0.42)
			[1.74]*			[1.78]*			[0.22]
$\Delta\text{Misall}_{\text{top}}$				-0.185			-0.177		-0.078
				(-0.81)			(-0.79)		(-0.36)
				[-0.56]			[-0.54]		[-0.19]
ΔMPK spread					-1.311			-1.028	-1.037
					(-4.85)***			(-4.40)***	(-4.08)***
					[-2.44]**			[-2.20]**	[-2.17]**
R^2	0.494	0.505	0.512	0.550	0.543	0.525	0.551	0.555	0.631
Adj. R^2	0.465	0.465	0.473	0.514	0.506	0.476	0.505	0.510	0.584
RMSE	2.162	2.161	2.146	2.060	2.077	2.139	2.080	2.069	1.905

Fama-Macbeth t-statistics in parentheses. Shanken t-statistics in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the prices of risk with Fama and MacBeth (1973a) and Shanken t-statistics for the 25 size $\times$ book-to-market, 10 momentum, 10 investment, and 10 operating profitability portfolios. The sample runs from 1975 to 2023. Returns and risk premia are reported in percent per year.

Chapter 3

Intermediary Asset Pricing Through the Lens of a Demand System[1]

Abstract

How does broker-dealers' leverage shift asset prices when their assets under management are relatively small? We explore the channel through which broker-dealers extend leverage to hedge funds using an asset demand system. During recessions, broker-dealers face binding leverage constraints and reduce their borrowing to hedge funds. Then, hedge funds are forced to sell assets and deleverage, creating demand pressures that shift asset prices. We find that hedge funds exert the largest price impact across all stocks among institutions, and their price impact increases substantially during crises as they need to deleverage. On average, a 1% increase in broker-dealers' leverage raises the price impact of the average hedge fund by 2% in the following year. These impacts are more pronounced among illiquid stocks. Our findings suggest that hedge funds significantly influence stock prices when broker-dealers' leverage constraints are binding.

Keywords: Intermediary Asset Pricing, Demand System, Broker-Dealers, Hedge Funds.

JEL codes: D92; E24; G32; J41; L25.

[1] This chapter is co-authored with Dongryeol Lee. We are grateful to Aleksandar Andonov, Esther Eiling, Enrico Perotti, Moto Yogo, Simon Rottke, Ralph Koijen, Tyler Muir, and participants at the UvA Finace Brownbag Seminar for helpful comments and discussions.

3.1 Introduction

Since the 2008 Global Financial Crisis (GFC), there has been a growing interest in the roles of financial intermediaries in financial markets. The Intermediary Asset Pricing literature shows that the financial health of intermediaries, typically proxied by broker-dealers' leverage, is crucial for understanding risk premia (e.g., He and Krishnamurthy, 2013; Adrian, Etula, and Muir, 2014; He, Kelly, and Manela, 2017; Haddad and Muir, 2021).

However, broker-dealers represent a small portion of the stock market, with their holdings accounting for less than 2%. So, how can broker-dealers' leverage affect asset pricing when their assets under management are so small? We explore this question through the channel in which broker-dealers extend leverage to hedge funds. Hedge funds rely on broker-dealers to provide financing and facilitate trades through prime brokerage services. During recessions, broker-dealers face binding leverage constraints and reduce borrowing to hedge funds. Then, hedge funds are forced to sell assets, creating short-term demand pressures that shift asset prices. Thus, broker-dealers' (BD) leverage could impact financial markets through its effects on hedge funds.

We test this mechanism using stock-level 13F holding data from 2000:Q1 to 2023:Q3. We start by documenting that brokers-dealers reduce their lending to hedge funds when their leverage increases. A 1% increase in broker-dealer leverage is associated with a 4% decline in hedge fund leverage, measured as borrowing from prime brokers relative to net assets. As hedge funds' funding conditions tighten, they reduce their holdings, which implies they are forced to sell their assets and deleverage.

To examine how hedge funds' deleveraging impacts stock markets, we use the demand system of asset pricing (Koijen and Yogo, 2019). We find that hedge funds exert the largest price impact across all stocks among institutions, and their price impact increases substantially during crises. A 10% demand shock to hedge funds shifts stock prices by an average of 0.2% across all stocks, with the effect increasing to 0.3% during the 2008 financial crisis. In contrast, broker-dealers exhibit negligible price impact,

suggesting their more inelastic demand and role as liquidity providers rather than price movers. Moreover, other institutional investors, aside from hedge funds and broker-dealers, also show smaller price impacts than hedge funds. Given that these institutions collectively hold more than 60% of the stock market, our results highlight the pivotal role hedge funds play in the stock market.

Finally, we find that higher broker-dealer leverage predicts greater price impacts. A 1% increase in BD leverage raises the price impact of hedge funds by 2% in the following year, a magnitude comparable to that of all other institutions combined. Notably, when measuring price impact per million dollars of holdings, a 1% increase in BD leverage corresponds to a 23.8% price impact for hedge funds, compared to just 0.5% for other institutions. This finding further validates our claim that BD leverage influences the stock market through hedge funds.

This paper provides several insights into the role of broker-dealer leverage in shaping asset prices through hedge funds' holdings. While existing research has emphasized the importance of intermediaries in determining risk premia, this study highlights the channel through which broker-dealers can influence stock prices, even when their assets under management are relatively small. Specifically, we show that broker-dealers' leverage constraints during downturns amplify hedge funds' price impact, as hedge funds are forced to sell assets in response to tightened leverage conditions of broker-dealers. By quantifying the price impact of hedge fund demand shocks and linking these shocks to broker-dealer leverage, the paper provides empirical evidence that connects financial intermediary health to cross-sectional variations in stock returns.

Related literature. Our paper contributes to three strands of literature. First, the Intermediary Asset Pricing literature emphasizes the importance of intermediaries' financial health in understanding asset prices and risk premia. For example, He and Krishnamurthy (2013, 2018) show that when intermediaries face tighter leverage constraints, their ability to provide liquidity to the market decreases, leading to higher risk premia. Brunnermeier and Pedersen (2009) show that margin constraints on interme-

diaries lead to a feedback loop between asset prices and market liquidity, magnifying price fluctuations during periods of stress. However, little is known about what type of intermediaries influence asset prices, since institutional investors are heterogeneous. Our paper looks at broker-dealers, hegde funds, and other types of institutions.

Broker-dealers' laverage or capital constraint is typically proxied for the intermediaries' finanacial health. Adrian, Etula, and Muir (2014) find that the leverage of security broker-dealers reflects the tightness of their funding constraints. Shocks to the leverage construct an intermediary SDF as these shocks are priced in the cross-section of asset returns. He, Kelly, and Manela (2017) show that the capital constraints of primary dealers construct a priced risk factor. Besides, Kargar (2021) find that the wealth share of broker-dealers predicts stock market excess returns and shocks to the relative wealth share of broker-dealers explain the cross-section of stock returns. Alternatively, our paper highlights the link of leverage constraints and asset pricing through via the channel of hedge funds' demand for assets. Particularly, the leverage constraint influences hedge funds' price impact during economic downsturns.

Besides, our findings are consistent with Haddad and Muir (2021) that intermediaries' financial constraints have a more pronounced impact on asset classes where households are less active. Stocks are an important asset class, with institutions holding of at least 70% of total assets under management, while households own less than 30% by the end of 2020.

Second, the literature has documented the relationship between hedge funds and broker-dealers. Hedge funds, as large institutional investors, often rely on broker-dealers to provide financing and facilitate trades through prime brokerage services. Broker-dealers, in turn, use their balance sheets to extend leverage to hedge funds, making this relationship crucial for market liquidity.

Adrian and Shin (2010) show that during periods of financial stress, when broker-dealers reduce their leverage due to capital constraints, hedge funds are forced to deleverage, exacerbating market volatility. Our paper adds to the evidence of this relationship by looking at their demand for assets and price impacts. While existing

research tends to focus on the systemic risks of leverage during crises, our study quantifies how broker-dealer leverage constraints affect hedge funds' price impact on stocks and cross-sectional asset pricing.

Also, studies focusing on hedge funds reinforce our findings. For example, Boyson, Stahel, and Stulz (2010) show that large adverse shocks to liquidity, such as poor dealer stock performance, increase the likelihood of contagion (clustering of extreme negative returns) across hedge funds. Similarly, Ben-David, Franzoni, and Moussawi (2012) demonstrates that hedge fund flows and leverage-driven trades can create price pressures. Using the demand system, we estimate the price impact of hedge funds on stocks. We find that higher leverage of broker-dealers predicts higher price impact of hedge funds on stocks.

Third, our paper contributes to understanding the role of institutional investors' demand on asset pricing. The demand-based approach to asset pricing introduced by Koijen and Yogo (2019) allow a more granular understanding of how different types of institutional investors nad household affect prices based on their portfolio choices. Our paper focuses on the interaction between broker-dealers and hedge funds. By showing that hedge funds, rather than broker-dealers, exert the largest price impact due to their higher demand elasticity, and that this impact is magnified during periods of financial stress when broker-dealer leverage is constrained, our findings provides insights into how institutional investors through their balance sheet constraints can affect asset pricing.

The remainder of the paper is organized as follows. Section 3.2 explains how we construct the data and main variables. Section 3.3 shows evidence on the channel that broker-dealers extend leverage to hedge funds. Section 3.4 explores the impact of this channel on the stock market using the demand system of asset pricing. Finally, Section 3.5 concludes.

3.2 Data construction

3.2.1 Institutional holdings and stock characteristics

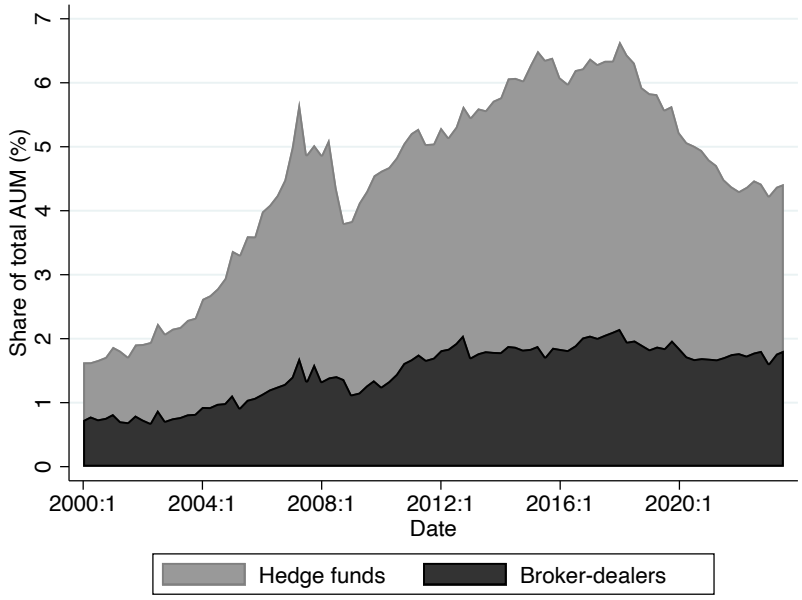
We obtain the direct stock holdings of institutional investors from FactSet, sourced from regulatory 13F filings. Asset managers who own 13F securities of more than $100 million in market value must report their long positions on the Securities and Exchange Commission Form 13F. In each quarter, the holding for each stock held by each institution equals shares multiplied by price. The assets under management (AUM) of each institution are then the sum of all holdings within each quarter. We follow Koijen and Yogo (2019) to assign zero holding when a stock from the previous 11 quarters is no longer in the portfolio since the investment universe of the median investor is persistent over 11 quarters.

We match the information on stocks from the Center for Research in Security Prices (CRSP) and Compustat. Our data set contains NYSE/AMEX/NASDAQ stocks with common share codes 10, 11, 12, or 18. Stock returns and shares outstanding come from CRSP Monthly Stock File. Accounting data are from CRSP/Compustat Merged Fundamentals Annual and Quarterly. We compute market equity as the price times shares outstanding. Since the holdings data are on a quarterly basis, we compound the returns over the quarter and keep other characteristics at quarter-end.

3.2.2 Investor classification

We classify investors into three groups: Broker-dealers, hedge funds, and other types of institutions, using FactSet's classification of investor types. The market clearing condition requires the shares outstanding of each stock to equal the total shares held by all investors in that stock. Hence, we assign the residual between shares outstanding and the total shares held by 13F investors to the holdings of households (i.e. individual investors). Any small investors who own less than $10 million in AUM are

Figure 3.1: **Asset Under Management of Broker-Dealers and Hedge Funds**



Description. This figure plots the share of total asset under management (AUM) by each investor type. In each quarter, we aggregate the AUM share by each investor group. The AUM share is the percentage of AUM of each investor over the total share AUM in each quarter. AUM is the total holdings of each investor in each quarter. Holdings are reported in 13F institutional holdings from FactSet Ownership. The sample period runs from 2000:Q1 to 2023:Q3.

also considered as households since small investors are not required to file Form 13F[2]. In case that the total holdings across investors for a stock exceed its market value, we scale the holdings of each investor proportionally, so that all holdings add up to the market value, following Koijen and Yogo (2019).

Figure 3.1 shows the share of total AUM for broker-dealers and hedge funds. The main observation is that AUM share of broker-dealers has remained relatively small and stable throughout the period. Starting at around 1% in 2000, broker-dealers maintained a consistent share of AUM, with slight fluctuations, but overall their holdings do not exceed 2% of the total AUM. Even during periods of downturns, such as the 2008 GFC, broker-dealers' AUM share showed only a modest increase, implying their limited role as asset holders.

[2] All small investors under this filter own a negligible fraction of the total market cap so they do not affect the findings if we include some of them as institutions.

Table 3.1: **Summary of Holdings**

	(1) # Institutions	(2) % Mkt Held	(3) Median AUM	(4) 90th AUM	(5) Median # Stocks	(6) 90th # Stocks
Panel A: All institutions						
2000-2004	2,010	58.60	376	6,273	91	551
2005-2009	2,730	67.68	345	6,195	78	546
2010-2014	3,230	70.59	339	6,373	73	532
2015-2019	4,287	73.21	311	6,064	70	522
2020-2023	5,444	66.24	232	4,730	67	449
Panel B: Hedge funds						
2000-2004	288	1.30	254	1,318	45	169
2005-2009	580	3.04	259	1,973	32	181
2010-2014	652	3.65	292	2,454	26	179
2015-2019	834	4.25	341	2,932	20	214
2020-2023	869	2.88	271	2,585	18	196
Panel C: Broker-dealers						
2000-2004	18	0.80	1,533	22,266	586	2,263
2005-2009	30	1.28	1,685	26,672	556	2,254
2010-2014	44	1.73	917	26,994	453	2,331
2015-2019	53	1.93	787	28,951	321	2,144
2020-2023	61	1.74	1,003	46,858	310	1,899
Panel D: Others						
2000-2004	1,705	56.49	404	7,339	100	598
2005-2009	2,120	63.36	370	8,181	93	612
2010-2014	2,534	65.21	351	8,041	88	566
2015-2019	3,400	67.03	302	7,173	85	555
2020-2023	4,514	61.63	223	5,375	77	470

Description. This table reports the average of each summary statistics within the given period. The table reports the number of institutions, the percentage of the market held, the median and 90th percentile AUM, and the median and 90th percentile number of stocks held. The sample period is from 2000:Q1 to 2023:Q3.

In contrast, the AUM share of hedge funds starts at a modest level of approximately 2% in 2000 but grows significantly over the next two decades. Hedge funds' share spikes around the 2008 financial crisis, reflecting their ability to accumulate assets during times of market volatility and stress. At their peak in the mid-2010s, hedge funds held over 6% of total AUM, a much larger share than broker-dealers. After reaching their peak, hedge funds' AUM share gradually declined, settling around 4-5% by 2023.

Table 3.1 reports the summary of holdings by each investor group over five periods from 2000 to 2023. In Panel A, which aggregates all institutions, the number of institutions and their percentage of market share steadily increases from 2,010 institutions holding 58.60% of the total AUM between 2000-2004, to a peak of 4,287 institutions

holding 73.21% of the market between 2015-2019. However, the trend slightly reverses in the most recent period (2020-2023), with 5,444 institutions holding a reduced 66.24% market share. This shift indicates that while the number of institutions continues to grow, the average market share per institution has diminished slightly.

Specifically for hedge funds in Panel B, their influence on the market has also grown steadily. The percentage of market held by hedge funds increases from 1.30% in the early 2000s (2000-2004) to 4.25% by 2015-2019, before declining to 2.88% in the most recent period. This decline in market share in the latest period contrasts with their median and 90th percentile AUM figures, which show relatively high asset concentration. Hedge funds' 90th percentile AUM, for instance, rises to 2,932 million between 2015-2019 before slightly declining to 2,585 million by 2020-2023. These figures suggest the concentration of assets within a few major players.

In Panel C, broker-dealers maintain a relatively small but concentrated presence in the market. Despite only holding a small percentage of the market (from approximately 0.80% to 1.93% over the periods), the broker-dealers in the 90th percentile consistently manage large amounts of assets, with 46,858 million in AUM for the top 10% in the 2020-2023 period. This concentration highlights the role of large broker-dealers in managing significant amounts of capital despite their relatively small numbers in the market.

3.2.3 Broker-dealers' leverage

We follow Adrian, Etula, and Muir (2014) to use the leverage ratio measured by assets relative to the equity of all security broker-dealers, denoted as BD leverage:

$$lev_t^{BD} = \frac{\text{Total financial assets}_t^{BD}}{\text{Total financial assets}_t^{BD} - \text{Total liability}_t^{BD}} \tag{3.1}$$

where the total financial assets and the total liability of security broker-dealers are from the U.S. Flow of Funds. The leverage factor is then given by the seasonally adjusted log

the leverage

$$levfac_t = \left[\Delta \ln \left(lev_t^{BD}\right)\right]^{SA}. \tag{3.2}$$

We replicate the leverage factor to obtain the values up to 2023:Q3. The data for total assets and total liability are available on the FRED (respectively the series BOGZ1FL664090005Q and BOGZ1FL664190005Q).

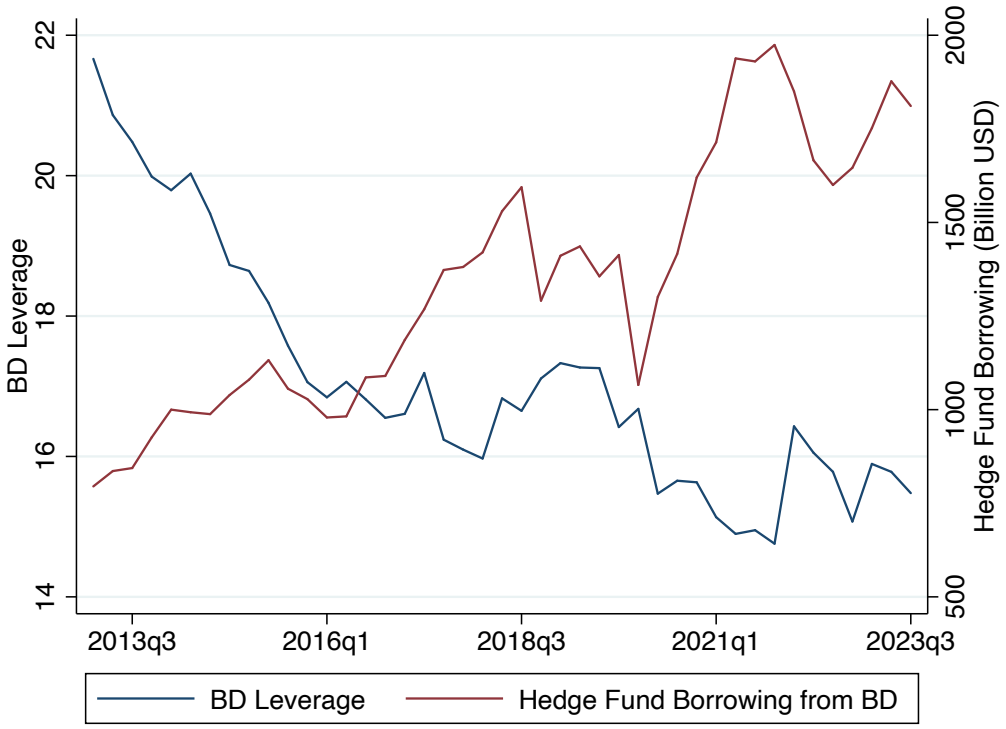
3.3 Broker-dealers extending leverage to hedge funds

In this section, we show evidence that broker-dealers extend leverage to hedge funds. Figure 3.2 plots the time series of broker-dealers' leverage (BD Leverage) and hedge fund borrowing from prime brokers. The figure suggests hedge fund reduces their borrowing from prime brokers once BD leverage increases. This implies that broker-dealers do not extend leverage to hedge funds once their funding conditions tighten.

To further examine how hedge funds reduce their borrowing from broker-dealers, we regress their leverage on BD leverage. We first compute hedge funds' leverage as the ratio of their borrowing from prime brokers to net assets (*Net Leverage*) or net equity holdings (*Net Equity Leverage*). Table 3.2 documents the results. The table suggests a 1% increase in BD leverage is associated with approximately 4% decline in hedge funds' leverage. This implies that broker-dealers reduce their borrowing substantially to hedge funds when they face leverage constraints.

Given that broker-dealers reduce their lending to hedge funds once their funding conditions tighten, we test whether hedge funds liquidate their positions in the stock market as they cannot borrow from broker-dealers. Table 3.3 regresses hedge funds' holdings in the stock market on BD leverage. The table shows hedge funds reduce their stock holdings as BD leverage increases. Column (1) suggests a 1% increase in BD leverage is associated with $34 billion decrease in hedge funds' stock holdings. We also examine their share of holdings to evaluate whether their reductions exceed those

Figure 3.2: **Broker-Dealers' Leverage and Hedge Fund Borrowing**



Description. This figure plots the time series of broker-dealers' leverage (BD Leverage) and hedge fund borrowing from prime brokers.

Table 3.2: **Broker-Dealers' and Hedge Funds' Leverage**

	Net Leverage	Net Equity Leverage
	(1)	(2)
BD Leverage	-0.038***	-0.022***
	(-5.281)	(-6.216)
Observations	43	43
R^2	0.354	0.611

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table regresses hedge funds' leverage on broker-dealers' leverage (BD). *Net (Equity) Leverage* refers to the ratio of hedge funds' borrowing from prime brokers to their net assets (net equity holdings). *t*-statistics reported in the parentheses are based on Newey-West standard errors with 4 lags.

Table 3.3: **Broker-Dealers' and Hedge Funds' Holdings**

	Holdings (Billion $)	Share
	(1)	(2)
BD Leverage	-33.849***	-0.048**
	(-4.152)	(-2.100)
Observations	95	95
R^2	0.496	0.169

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table regresses hedge funds' stock holdings on broker-dealers' leverage (BD). *Holding (Billion $)* refers to the dollar amount of holdings in billion $ and *Share* refers to its share of holdings in the stock market. *t*-statistics reported in the parentheses are based on Newey-West standard errors with 4 lags.

of other investors. Column (2) confirms this result, showing that a 1% increase in BD leverage corresponds to a 5% decrease in hedge funds' stock holdings. This implies hedge funds actively reduce their positions in the stock market and deleverage once they face borrowing constraints.

3.4 Broker-cealers' leverage and price impacts

The previous section characterizes how an increase in broker-dealers' leverage leads hedge funds to reduce their borrowing from brokers and deleverage. This section explores its impact on the stock market using the demand system of asset pricing (Koijen and Yogo (2019)).

3.4.1 Characteristic-based demand estimation

During recessions, broker-dealers face binding leverage constraints and cannot extend leverage to hedge funds. Then, hedge funds are forced to sell assets, creating short-term demand pressures that shift asset prices. Thus, we outline the characteristic-based demand model used to investigate the demand elasticity of institutional investors, specifically hedge funds and broker-dealers.

Using the demand system asset pricing methodology from Koijen and Yogo (2019), we can model the demand of investor i for stock n as,

$$\frac{w_{it}(n)}{w_{it}(0)} = \exp\left\{\beta_{0,it}\text{me}_t(n) + \boldsymbol{\beta}'_{\mathbf{1},it}\boldsymbol{x_t}(n) + \alpha_{it}\right\}\epsilon_{it}(n), \tag{3.3}$$

where $w_{it}(n)$ denotes investor i's portfolio weight for stock n at quarter t, $w_{it}(0)$ denotes portfolio weight on the outside asset, me_t is the market cap, X_t is a vector of characteristics, and ϵ_{it} is the latent demand. We include the same characteristics as in Koijen and Yogo (2019): an instrument for market equity, book equity, profitability, investment, dividends, and market beta. We estimate Equation (3.3) using the non-linear GMM method under the moment condition

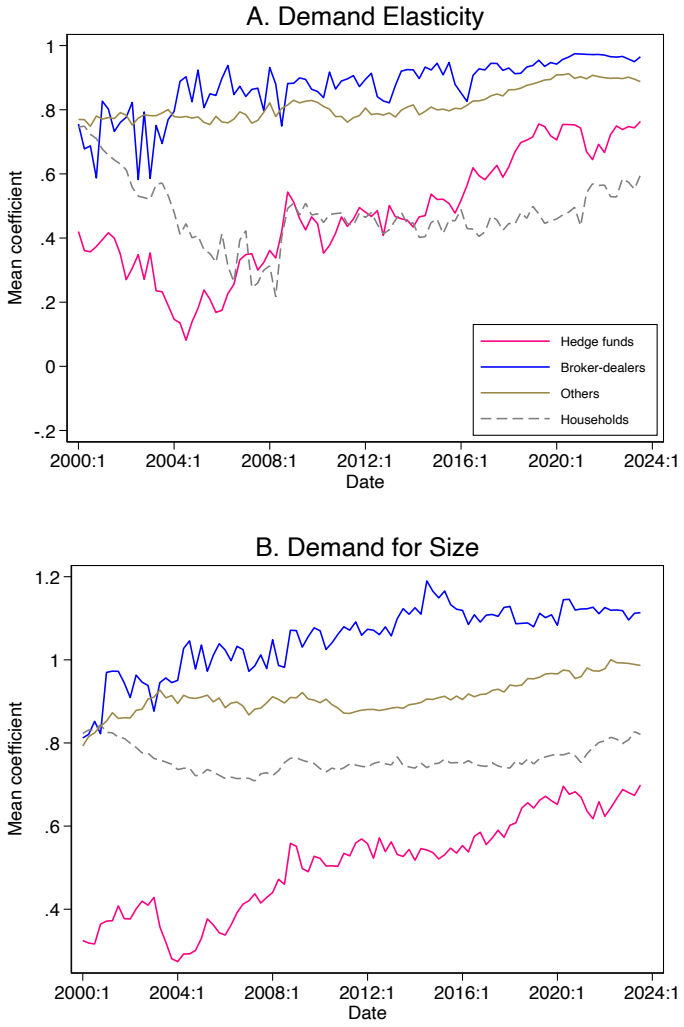
$$\mathbb{E}_t[\epsilon_{it}(n)|\widehat{\text{me}}_{it}(n), \boldsymbol{x}_t(n)] = 1 \tag{3.4}$$

where $\widehat{\text{me}}_{it}(n)$ denotes the instrument for market equity. For institutions with more than 1,000 holdings, we can estimate the model at the investor level. For those institutions with fewer than 1,000 holdings, we group them by type and then quantiles of AUM within each type to run a pooled estimation.

Figure 3.3 shows the cross-sectional mean of the estimated coefficients on characteristics by each investor type. Panel A reports the coefficients on log maket-to-book equity for hedge funds, broker-dealers, households, and other institutions. A higher coefficient (up to one) implies that the institution do not deviate away from their benchmark portfolio when the price changes. Thus, a higher coefficient implies a more inelastic demand.

We observe that the demand elasticity for hedge funds is lower than that of broker-dealers throughout the sample period, reflecting the more elastic demand of hedge funds relative to broker-dealers who face leverage constraints. Broker-dealers consistently show the highest demand elasticity, which remains relatively stable over time, while hedge funds experience a decline in elasticity until around 2009, after which their demand become more ineleastic.

Figure 3.3: **Demand Elasticities**



Description. This figure reports the cross-sectional mean of the estimated coefficients by institution type, weighted by assets under management. The sample period runs from 2000:Q1 to 2023:Q3. Panel A reports the coefficients on log market-to-book equity. A higher coefficient (close to 1) implies higher rice elasticity of demand. Panel B reports the coefficients on book equity. A higher coefficient implies portfolio tilting towards large-cap stocks.

Panel B reports the coefficients on book equity, which measures the preference for size by institutional type. Broker-dealers have the highest coefficient on book equity, implying that they tilt their portfolio towards large-cap stocks. The result is consistent with banks in Koijen and Yogo (2019)'s sample, since banks are largely the parent holding institutions of broker-dealers. Hedge funds, in contrast, tilt their portfolio towards smaller stocks, although the trend in demand for larger stocks have increased over time.

3.4.2 Price impacts

As an implication from the deman system, we can estimate the price impact on each stock for each type of institution. Then, we can examine how the price impact by each type of institution comove with BD leverage. Following Koijen and Yogo (2019), the market clearing condition suggests for each stock n,

$$ME_t(n) = \sum_{i=1}^{I} A_{it} w_{it}(n, me, x, \epsilon), \tag{3.5}$$

where A_i is investor i's assets under management (AUM). Taking the log of the market clearing condition gives,

$$\boldsymbol{p}_t = \boldsymbol{f}(\boldsymbol{p}_t) = \log\left(\sum_i A_{it} \boldsymbol{w}_{it}(\boldsymbol{p}_t)\right) - \boldsymbol{s}_t, \tag{3.6}$$

where $s = \log(S)$ is the log of the number of shares outstanding. Thus, we can obtain an implicit function for the log price as follows,

$$\boldsymbol{p}_t = \boldsymbol{g}(\boldsymbol{s}_t, \boldsymbol{x}_t, \boldsymbol{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_t). \tag{3.7}$$

The equation implies that asset prices are fully determined by shares outstanding (s), characteristics (x), the wealth distribution (A), the coefficients on the characteristics (β), and the latent demand (ϵ). The price impact of a change in investor i's latent demand

for stock n is given as the n-th diagonal element of the matrix has the form

$$PI_t^i := \frac{\partial \boldsymbol{p}_t}{\partial \log(\boldsymbol{\varepsilon}_{i,t})'} = \left(\boldsymbol{I} - \sum_{j=1}^{I} A_{j,t} \beta_{0,j,t} \boldsymbol{H}_t^{-1} \boldsymbol{G}_{j,t} \right)^{-1} A_{i,t} \boldsymbol{H}_t^{-1} \boldsymbol{G}_{i,t}. \tag{3.8}$$

where $\beta_{0,j,t}$ is the coefficient of investor j on market-to-book, and matrices

$$\boldsymbol{H} := \operatorname{diag}\left(\sum_i A_i \boldsymbol{w}_i \right) = \sum_i A_i \operatorname{diag}(\boldsymbol{w}_i) \tag{3.9}$$

$$\boldsymbol{G}_i := \operatorname{diag}(\boldsymbol{w}_i) - \boldsymbol{w}_i \boldsymbol{w}_i'. \tag{3.10}$$

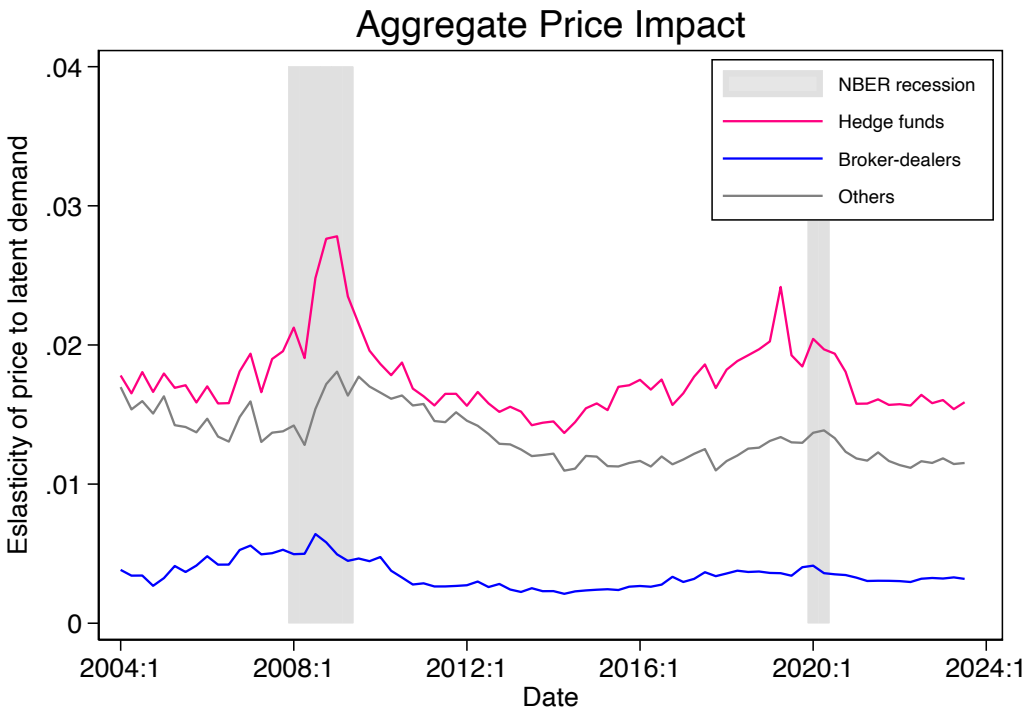
Equation (3.8) represents the coliquidity matrix for investor i, which measures the *elasticity of asset price changes* (denoted by $\boldsymbol{p}_t$) with respect to idiosyncratic shocks to investor i's latent demand (denoted by $\boldsymbol{\varepsilon}_{i,t}$). The inverse of this matrix is the aggregate demand elasticity. The n-th diagonal element of the term outside the inverse in Equation (3.8) has the form

$$\frac{A_{i,t} w_{i,t}(n) \left[1 - w_{i,t}(n)\right]}{\sum_{j=1}^{J} A_{j,t} w_{j,t}(n)}. \tag{3.11}$$

which accounts for how sensitive investors' portfolio weights ($\boldsymbol{w}_{j,t}$) are to changes in asset prices. Thus, Equation (3.8) estimates the *price impact of demand shocks* for investor i. Following Koijen and Yogo (2019), we estimate the price impact for each stock and institution via the diagonal elements of $\frac{\partial \boldsymbol{p}_t}{\partial \log(\boldsymbol{\varepsilon}_{i,t})'}$. For each stock, we estimate the price impact for each institution with a 10% demand shocks on that stock. We then aggregate the price impact on each stock by institutional types, weighted by their holding in the stock. Finally, we obtain the value-weighted price impact across all stocks for each institutional type, weighted by the stock market cap. This time-series captures the aggregate price impact of each type of institution on the overall stock market.

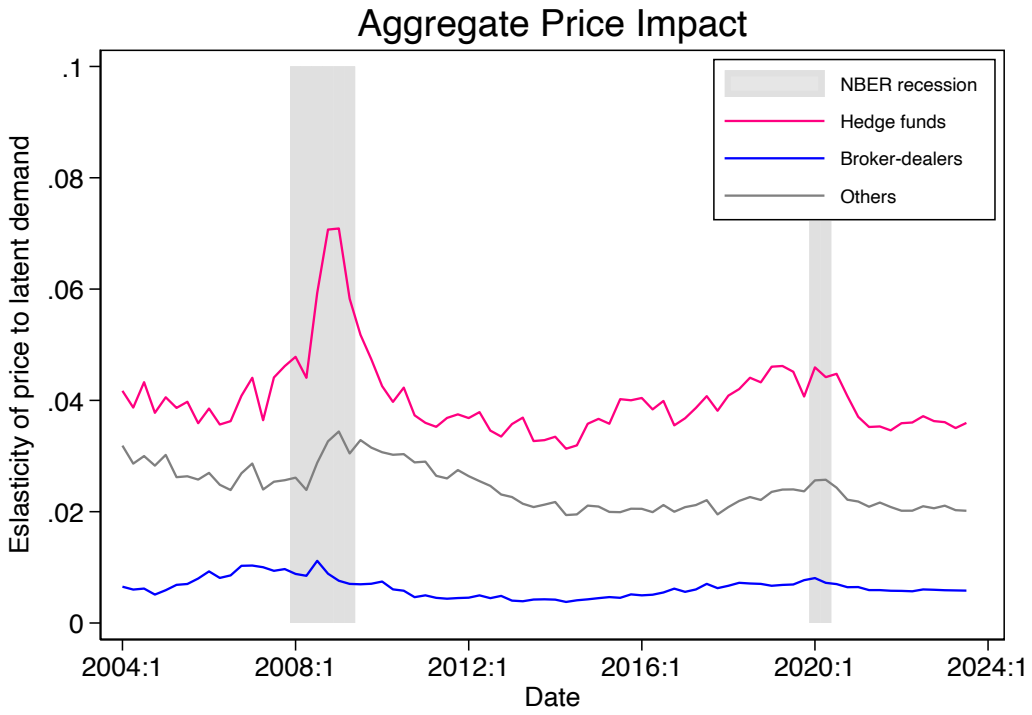
Figure 3.4 plots the price impact across all stocks over the sample period from 2004:Q1 to 2023:Q3. The price impact is represented as an elasticity, where an estimate of 0.1 implies that the price increases by 1% for a 10% demand shock. In Panel A, which

Figure 3.4: **Price Impact Across All Stocks**



Description. This figure reports the cross-sectional distribution of price impact across all stocks for the average hedge funds, broker-dealers, and other investors. Price impact is defined as an elasticity, where an estimate of 0.1 implies that the price increases by 1% for a 10% demand shock. For each stock, we estimate the price impact for each institution with a 10% demand shocks on that stock, following Koijen and Yogo (2019). We then aggregate the price impact on each stock by institutional types, weighted by their holding in the stock. Finally, we take the value-weighted price impact across all stocks for each institutional type, weighted by the stock market cap. The sample period is from 2000:Q1 to 2023:Q3.

Figure 3.5: **Price Impact Across Least Liquid Stocks**



Description. This figure reports the cross-sectional distribution of price impact across least liquid stocks stocks for the average hedge funds, broker-dealers, and other investors. Price impact is defined as an elasticity, where an estimate of 0.1 implies that the price increases by 1% for a 10% demand shock. Least liquid stocks are defined based on their position in the cross-sectional distribution of price impact. Least liquid stocks as those at the 90th percentile of the distribution of price impacts across stocks. For each stock, we estimate the price impact for each institution with a 10% demand shocks on that stock, following Koijen and Yogo (2019). We then aggregate the price impact on each stock by institutional types, weighted by their holding in the stock. Finally, we take the value-weighted price impact across all stocks for each institutional type, weighted by the stock market cap. The sample period is from 2000:Q1 to 2023:Q3.

depicts the price impact on all stocks, hedge funds consistently show a higher price impact than broker-dealers throughout the sample period. A 10% positive demand shock by the average hedge fund increases stock prices by 0.2% across all stocks.

Broker-dealers, in contrast, exhibit a relatively low price impact across the entire period. This finding indicates that hedge funds, with a more elastic demand, exert greater influence on the prices of average stocks than broker-dealers, who have a more inelastic demand. The stable yet low impact from broker-dealers could reflect their more passive role in price setting across all stocks.

Figure 3.5 focuses on the price impact on the least liquid stocks. Least liquid stocks as those at the 90th percentile of the distribution of price impacts across all stocks[3]. These stocks exhibit higher price sensitivity to demand shocks, meaning that changes in investor demand have a larger effect on their prices. During financial crises or periods of financial stress, liquidity often dries up, especially for the least liquid stocks.

Panel B shows that hedge funds have a larger impact on price movements in least liquid stocks. A 10% positive demand shock by the average hedge fund increases stock prices by 0.4% across these stocks. During recessions, particularly in the 2008 GFC, hedge funds' price impact spikes dramatically, with a 10% demand shock leading to an increase of up to 7% in stock prices. Importantly, the price impact from the hedge funds are economically large. Koijen and Yogo (2019) reports the price impact for the average investment advisor with a 10% demand shock on the least liquid stocks is approximately 0.22% in 2017:Q2. We find that the price impact for the average hedge fund with a 10% demand shock on the least liquid stocks is approximately 0.4% in 2017:Q2, doubled the average investment advisor.

This sharp increase reflects the vulnerability of illiquid stocks to shifts in hedge fund demand, especially during periods of market stress when liquidity is scarce. Broker-dealers, on the other hand, maintain a consistently low price impact on illiquid stocks, suggesting that they do not play as significant a role in driving price changes, even during periods of downturns.

Overall, among all institutions, we find that hedge funds exert the largest price impact across all stocks, and their price impact comoves with BD leverage. Although often broker-dealers are considered as marginal investors, broker-dealers themselves have a relatively muted price impact across all stocks, consistent with their more inelastic demand and their role as liquidity provxiders rather xice movers. This result implies that hedge funds' demand and price pressure on stocks comove with leverage. Providing leverage to hedge funds is an indirect way of how BD leverage matters for

[3] Liquidity is associated with how quickly and easily a stock's price adjusts to changes in supply and demand, with less liquid stocks having a higher price impact for a given demand shock Koijen and Yogo (2019).

asset pricing.

Table 3.4 reports the main findings of this paper. This table reports the results from the single predictive regressions that examine how BD leverage predicts future price impacts for hedge funds and broker-dealers, over time horizons ranging from one to eight quarters ahead. The regression has the following form:

$$PI^{I}_{t:t+q} = \alpha + \beta lev^{BD}_{t} + \epsilon_{t:t+q}. \tag{3.12}$$

In Column (1)-(3), the dependent variable is the average price impact $PI^{I}_{t:t+q}$. In Column (4)-(6), the dependent variable is the average price impact scaled by total AUM of each type (in million USD). The regressions use BD leverage as the key predictor. We report t-statistics based on Newey and West (1987) corrections, respectively.

In Panel A, for hedge funds, we observe that a 1% increase in BD leverage leads to an increase in the aggregate price impact by 1.5% to 1.7% at in the followig quarter and the followig quarter year. When we measure the price impact per million dollars of holdings, the effect of BD leverage on hedge funds becomes much more pronounced: a 1% increase in BD leverage leads to a substantial price impact of 23.8% in the following quarter, decreasing to 18.0% over an two-year horizon (q=8). These findings highlight the significant sensitivity of hedge funds' price impacts to BD leverage, suggesting that hedge funds respond strongly to changes in BD leverage, especially in terms of their holdings' price impact per dollar invested.

Panels B and C reveal smaller price impacts for broker-dealers and other institutions. In Panel B, the aggregate price impact for broker-dealers is much lower, with coefficients around 0.8% across different quarters, and the price impact per million dollars of holdings is between 9.9% and 11.2%. In Panel C, other institutions exhibit similar aggregate price impact coefficients; yet, the price impact per million holdings is minimal (between 0.5% and 0.6%). These results suggest that the effect of BD leverage is particularly strong for hedge funds, significantly amplifying their price impacts compared to broker-dealers and other institutions. The relatively higher R^2 values in Panel B indicate that BD leverage explains a substantial portion of the variation in broker-

Table 3.4: **Predicting Price Impacts on All Stocks**

Dependent variable:	Aggregate price impact			Price impact per $1M holdings		
	(1) q=1	(2) q=4	(3) q=8	(4) q=1	(5) q=4	(6) q=8
Panel A: Hedge funds						
BD leverage	0.015	0.017	0.014	0.238	0.225	0.180
	[3.80]***	[2.75]***	[2.31]**	[4.30]***	[3.02]***	[3.12]***
R^2	0.094	0.175	0.171	0.135	0.162	0.170
RMSE	0.004	0.003	0.003	0.055	0.047	0.036
Panel B: Broker-dealers						
BD leverage	0.008	0.008	0.007	0.112	0.112	0.099
	[9.25]***	[7.78]***	[4.16]***	[6.35]***	[4.70]***	[4.99]***
R^2	0.309	0.391	0.387	0.267	0.307	0.340
RMSE	0.001	0.001	0.001	0.017	0.016	0.013
Panel C: Others						
BD leverage	0.019	0.018	0.018	0.005	0.006	0.005
	[4.33]***	[3.23]***	[4.49]***	[5.39]***	[3.92]***	[4.87]***
R^2	0.128	0.191	0.324	0.241	0.312	0.387
RMSE	0.004	0.003	0.002	0.001	0.001	0.001

t-ratio of Newey-West (1987) with k-1 lags in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the coefficient of regressing future price impacts for each type of institutions on BD leverage: $PI^I_{t:t+q} = \alpha + \beta lev^{BD}_t + \epsilon_{t:t+q}$. For each stock, we estimate the price impact for each institution with a 10% demand shocks on that stock, following Koijen and Yogo (2019). We then aggregate the price impact on each stock by institutional types, weighted by their holding in the stock. $PI^I_{t:t+q}$ is the value-weighted price impact across all stocks for each instituional type, weighted by the stock market cap. In Column (1)-(3), the dependent variable is the aggregate price impact $PI^I_{t:t+q}$. In Column (4)-(6), the dependent variable is the aggregate price impact scaled by total AUM of each type (in million USD). The sample period is from 2000:Q1 to 2023:Q3.

dealers' price impacts, althoughthough the magnitude of the effect remains lower than for hedge funds.

Overall, the table indicates that hedge funds are the primary channel through which BD leverage affects stock market price impacts. The sensitivity of hedge funds' price impacts, especially when measured per million dollars, underscores their role in transmitting the effects of BD leverage to the broader market. This implies hat BD leverage influences the market primarily through hedge funds.

Table 3.5 examines how BD leverage affects the price impacts of the least liquid stocks held by hedge funds, broker-dealers, and other institutions. In Panel A, for hedge funds, a 1% increase in BD leverage significantly raises the aggregate price impact on illiquid stocks by 4.3% to 5.0%, with a large increase in price impact per million dollars of holdings, from 58.1% at a one-quarter horizon to 44.1% over eight quarters. This indicates that hedge funds' positions in less liquid stocks are highly sensitive to changes in BD leverage.

In contrast, Panels B and C show much smaller effects for broker-dealers and other institutions. Broker-dealers see an increase of about 1.2% in aggregate price impact and 15.9%-18.2% in price impact per million dollars. Other institutions experience minimal sensitivity, with only a 1.0%-1.1% increase in price impact per million holdings. These findings suggest that BD leverage primarily affects illiquid stock price impacts through hedge funds rather than through broker-dealers or other institutions.

In summary, BD leverage is a significant predictor of future price impacts for both hedge funds and broker-dealers. Interestingly, hedge funds exhibit a larger response to changes in BD leverage. These findings highlight the crucial role that intermediary leverage plays in shaping market dynamics, particularly by influencing the future price impacts of institutional investors.

Table 3.5: **Predicting Price Impacts on Least Liquid Stocks**

Dependent variable:	Aggregate price impact			Price impact per $1M holdings		
	(1) q=1	(2) q=4	(3) q=8	(4) q=1	(5) q=4	(6) q=8
Panel A: Hedge funds						
BD leverage	0.043	0.050	0.042	0.581	0.554	0.441
	[3.91]***	[2.92]***	[2.62]**	[4.22]***	[3.01]***	[3.13]***
R^2	0.106	0.206	0.210	0.129	0.158	0.168
RMSE	0.012	0.009	0.007	0.139	0.117	0.090
Panel B: Broker-dealers						
BD leverage	0.012	0.011	0.009	0.182	0.180	0.159
	[8.68]***	[4.78]***	[2.32]**	[7.36]***	[5.63]***	[5.29]***
R^2	0.280	0.344	0.249	0.323	0.391	0.427
RMSE	0.002	0.001	0.001	0.024	0.021	0.017
Panel C: Others						
BD leverage	0.042	0.039	0.037	0.011	0.011	0.010
	[4.36]***	[3.19]***	[4.32]***	[5.22]***	[3.77]***	[4.67]***
R^2	0.123	0.176	0.296	0.218	0.280	0.355
RMSE	0.010	0.008	0.005	0.002	0.002	0.001

t-ratio of Newey-West (1987) with k-1 lags in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the coefficient of regressing future price impacts for each type of institutions on BD leverage: $PI^{I}_{t:t+q} = \alpha + \beta lev^{BD}_{t} + \epsilon_{t:t+q}$. For each stock, we estimate the price impact for each institution with a 10% demand shocks on that stock, following Koijen and Yogo (2019). We then aggregate the price impact on each stock by institutional types, weighted by their holding in the stock. $PI^{I}_{t:t+q}$ is the value-weighted price impact across least liquid stocks for each instituional type, weighted by the stock market cap. Least liquid stocks are defined at 90th percentile cross-sectional distribution of price impact. In Column (1)-(3), the dependent variable is the aggregate price impact $PI^{I}_{t:t+q}$. In Column (4)-(6), the dependent variable is the aggregate price impact scaled by total AUM of each type (in million USD). The sample period is from 2000:Q1 to 2023:Q3.

3.5 Conclusion

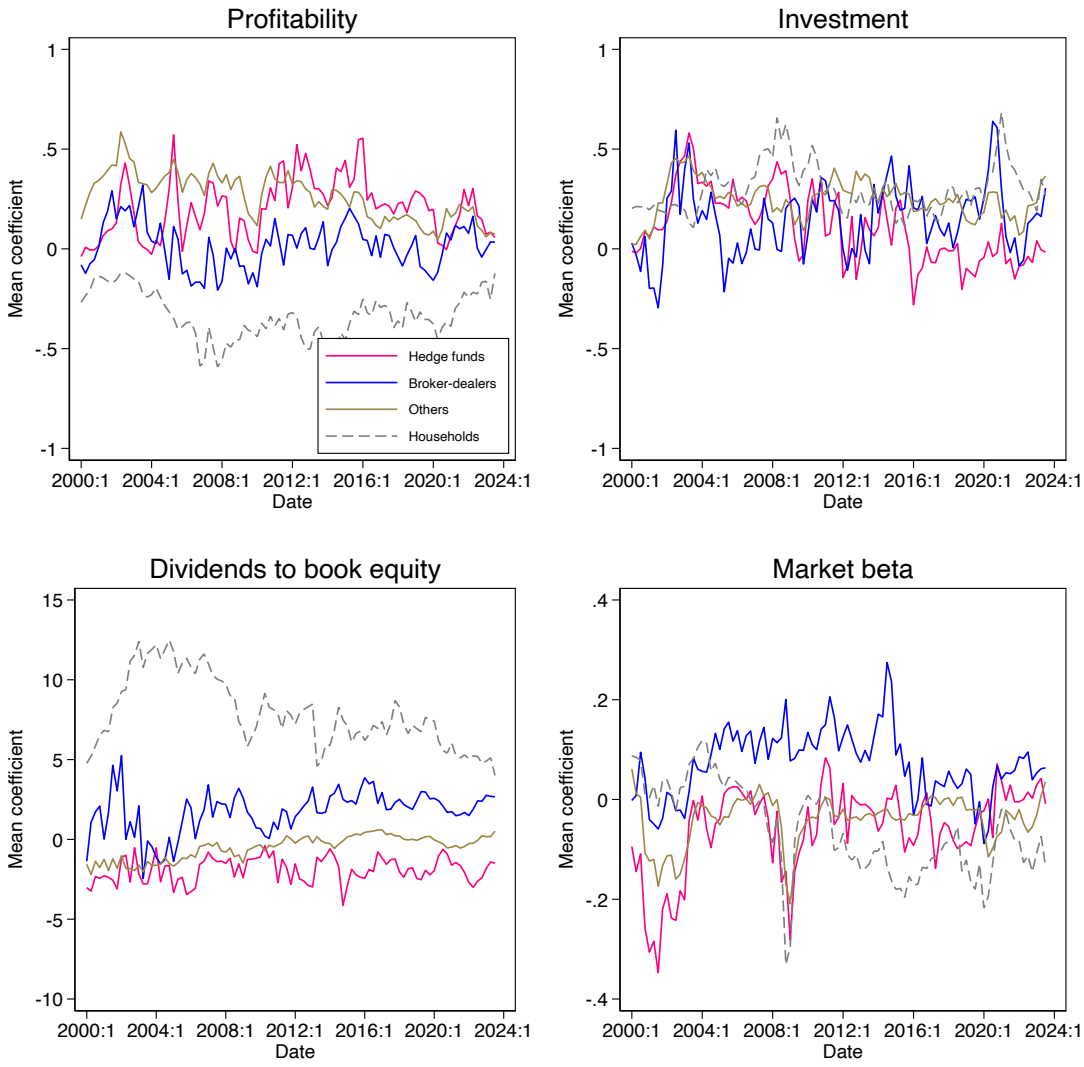
This paper highlights the role of broker-dealers and hedge funds in asset pricing. Our findings imply that broker-dealers, despite their smaller share of total assets under management, shape the leverage for hedge funds. During periods of financial stress, when broker-dealers face tighter leverage constraints, hedge funds are forced to deleverage, leading to significant price impacts across stocks. This effect is particularly pronounced for less liquid stocks and those highly sensitive to leverage shocks, confirming the importance of intermediary health in driving cross-sectional variation in stock returns.

Using a characteristic-based demand system, we quantify the elasticity of demand for hedge funds and broker-dealers. Hedge funds, with their higher demand elasticity, show a stronger direct impact on prices, while broker-dealers influence asset prices indirectly through their role in providing leverage. This interplay between broker-dealers' leverage constraints and hedge funds' demand help us understand how financial health of intermediaries affect asset pricing, especially during times of economic downturns.

Future research could examine the impact of regulatory changes targeting broker-dealer leverage, such as capital requirements on asset pricing through shifts in hedge fund behavior. For example, we could assess how regulations like the Dodd-Frank Act or Basel III, which impose stricter capital standards on broker-dealers, alter hedge funds' ability to leverage and maintain positions, particularly in high-volatility.

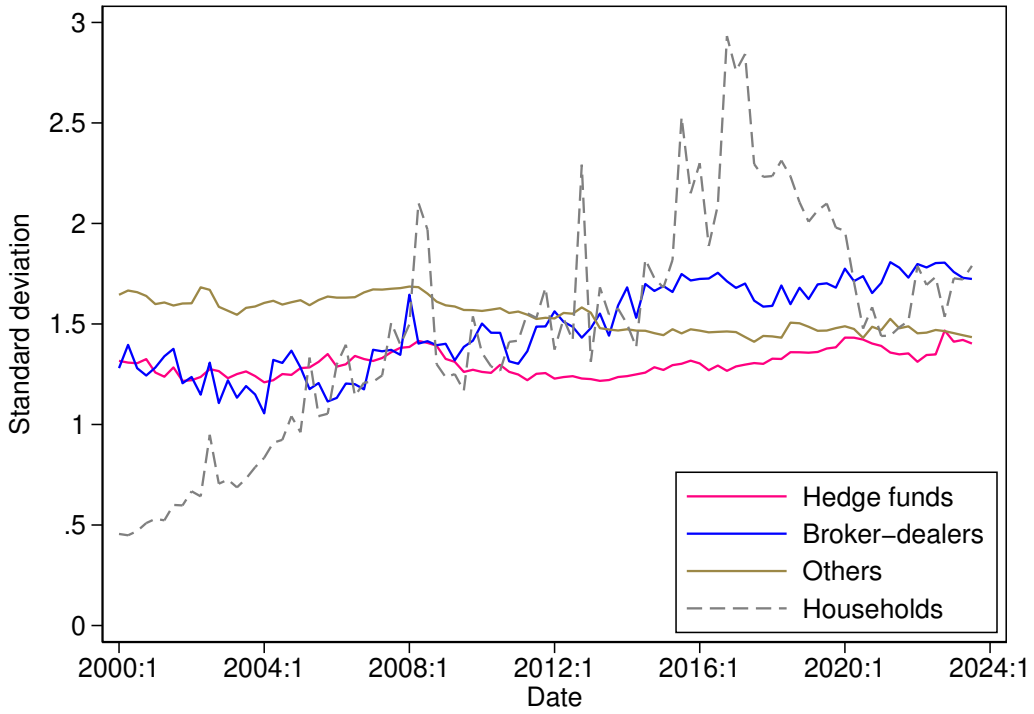
A1 Appendix: Additional results

Figure A1: **Coefficient on Characteristics**



Description. This figure reports the cross-sectional mean of the estimated coefficients by institution type, weighted by assets under management. The sample period runs from 2000:Q1 to 2023:Q3. Panel A reports the coefficients on profitability. Panel B reports the coefficients on investment. Panel C reports the coefficients on dividend-to-book equity. Panel D reports the coefficients on market beta.

Figure A2: **Standard Deviation of Latent Demand**



Description. This figure reports the cross-sectional standard deviation of log latent demand by institution type, weighted by assets under management. The sample period is from 2000:Q1 to 2023:Q3. A higher standard deviation implies more extreme portfolio weights tilted away from observed characteristics.

Table A1: **Predicting Price Impacts across All Stocks using Primary-Dealer Capital Ratio**

Dependent variable:	Aggregate price impact			Price impact per $1M holdings		
	(1) q=1	(2) q=4	(3) q=8	(4) q=1	(5) q=4	(6) q=8
Panel A: Hedge funds						
PD capital ratio	0.111	0.098	0.094	1.904	1.642	1.375
	(3.39)***	(3.11)***	(4.50)***	(5.48)***	(3.52)***	(2.84)***
	[3.35]***	[2.77]***	[4.40]***	[5.42]***	[3.11]***	[2.73]***
R^2	0.233	0.282	0.374	0.389	0.400	0.478
RMSE	0.004	0.003	0.002	0.047	0.040	0.029
Panel B: Broker-dealers						
PD capital ratio	0.038	0.035	0.036	0.582	0.587	0.566
	(4.73)***	(3.78)***	(6.19)***	(5.18)***	(3.58)***	(3.81)***
	[4.68]***	[3.44]***	[5.99]***	[5.12]***	[3.19]***	[3.66]***
R^2	0.309	0.361	0.513	0.328	0.388	0.535
RMSE	0.001	0.001	0.001	0.016	0.015	0.011
Panel C: Others						
PD capital ratio	0.147	0.106	0.083	0.026	0.024	0.023
	(4.69)***	(2.84)***	(2.47)**	(4.46)***	(2.83)***	(2.44)**
	[4.64]***	[2.51]**	[2.36]**	[4.42]***	[2.49]**	[2.34]**
R^2	0.358	0.315	0.348	0.256	0.259	0.333
RMSE	0.004	0.003	0.002	0.001	0.001	0.001

t-ratio of Hodrick (1992) with k-1 lags in parentheses.

t-ratio of Newey-West (1987) with k-1 lags in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the coefficient of regressing future price impacts for each type of institutions on capital ratio: $PI^I_{t:t+q} = \alpha + \beta\eta_t + \epsilon_{t:t+q}$. For each stock, we estimate the price impact for each institution with a 10% demand shocks on that stock, following Koijen and Yogo (2019). We then aggregate the price impact on each stock by institutional types, weighted by their holding in the stock. $PI^I_{t:t+q}$ is the value-weighted price impact across all stocks for each institutional type, weighted by the stock market cap. In Column (1)-(3), the dependent variable is the aggregate price impact $PI^I_{t:t+q}$. In Column (4)-(6), the dependent variable is the aggregate price impact scaled by total AUM of each type (in million USD). The sample period is from 2000:Q1 to 2023:Q3.

Table A2: **Predicting Price Impacts across Least Liquid Stocks using Primary-Dealer Capital Ratio**

Dependent variable:	Aggregate price impact			Price impact per $1M holdings		
	(1) q=1	(2) q=4	(3) q=8	(4) q=1	(5) q=4	(6) q=8
Panel A: Hedge funds						
PD capital ratio	0.295	0.263	0.252	4.716	4.063	3.362
	(3.19)***	(2.87)***	(4.54)***	(5.32)***	(3.45)***	(2.84)***
	[3.15]***	[2.56]**	[4.44]***	[5.26]***	[3.06]***	[2.74]***
R^2	0.220	0.267	0.369	0.382	0.394	0.472
RMSE	0.011	0.009	0.007	0.117	0.100	0.071
Panel B: Broker-dealers						
PD capital ratio	0.060	0.049	0.051	0.914	0.849	0.829
	(4.65)***	(4.04)***	(3.20)***	(5.44)***	(3.95)***	(4.72)***
	[4.60]***	[3.73]***	[3.05]***	[5.38]***	[3.56]***	[4.55]***
R^2	0.303	0.307	0.392	0.366	0.401	0.564
RMSE	0.002	0.001	0.001	0.023	0.021	0.015
Panel C: Others						
PD capital ratio	0.347	0.256	0.194	0.057	0.052	0.048
	(4.89)***	(3.00)***	(2.60)***	(4.69)***	(2.94)***	(2.47)**
	[4.83]***	[2.66]***	[2.49]**	[4.64]***	[2.60]**	[2.38]**
R^2	0.380	0.351	0.389	0.287	0.288	0.357
RMSE	0.009	0.007	0.005	0.002	0.002	0.001

t-ratio of Hodrick (1992) with k-1 lags in parentheses.

t-ratio of Newey-West (1987) with k-1 lags in square brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the coefficient of regressing future price impacts for each type of institutions on capital ratio: $PI^I_{t:t+q} = \alpha + \beta\eta_t + \epsilon_{t:t+q}$. For each stock, we estimate the price impact for each institution with a 10% demand shocks on that stock, following Koijen and Yogo (2019). We then aggregate the price impact by institutional types, weighted by their holding in the stock. $PI^I_{t:t+q}$ is the value-weighted price impact on each stock across least liquid stocks for each institutional type, weighted by the stock market cap. Least liquid stocks are defined at 90th percentile cross-sectional distribution of price impact. In Column (1)-(3), the dependent variable is the aggregate price impact $PI^I_{t:t+q}$. In Column (4)-(6), the dependent variable is the aggregate price impact scaled by total AUM of each type (in million USD). The sample period is from 2000:Q1 to 2023:Q3.

Chapter 4

Overlapping Factors[1]

Abstract

Characteristic-based factors, such as value, momentum, and low volatility, are constructed from the same universe of stocks. Therefore, they can contain some of the same stocks. Although the degree of overlap among these factors is not excessive, I find that the overlap matters. Stocks included simultaneously in the same leg across multiple factors, so-called overlapping stocks, earn an average abnormal return of 66 basis points per month. In contrast, pure factor stocks included in a single factor earn on average, across all factors, only 5 basis points per month, despite similar industry composition, institutional ownership, and portfolio turnover. Further, the portfolio of overlapping stocks is priced in the cross-section of stock returns, while the portfolio of all pure factor stocks is not. Exposures to macroeconomic and liquidity risks cannot explain the return difference. Only in the short leg, overlapping stocks are more exposed to investor sentiment and have a higher short interest than pure factor stocks.

Keywords: Factor models, portfolio construction, cross-sectional asset pricing.

JEL codes: G11, G12, G14, G23.

[1] I am grateful to my supervisors, Aleksandar Andonov and Esther Eiling, for their continuous support. I thank Martijn Boons, Elisa Pazaj, Tarun Ramadorai, Simon Rottke, Esad Smajlbegovic, Amar Soebhag, Xiao Xiao, and participants in the Frontiers of Factor Investing Conference 2022 and the Finance Group Brownbag Seminar at the University of Amsterdam for helpful comments and discussions.

4.1 Introduction

Despite the large number of asset pricing factors proposed in the literature, investors typically focus on a small set of well-known characteristic-based factors, such as value (e.g., Fama and French, 1992a), momentum (e.g., Jegadeesh and Titman, 1993), and idiosyncratic volatility (e.g., Ang, Hodrick, Xing, and Zhang, 2006). These factors are known to have low correlations. For example, the correlation between momentum and low volatility is 0.13 and the correlation between value and momentum is even negative at -0.22 (see also Asness et al., 2013)[2]. At the same time, characteristics-based factors are all constructed from the same universe of stocks. Therefore, stocks can be selected simultaneously across multiple factor portfolios, even for seemingly unrelated characteristics.

This paper shows that the low correlation coefficients hide important overlap in performance drivers across factor-based investment strategies. Although commonly used long-short factors do not overlap excessively compared to a random-sorting benchmark, the overlap is important for investors and for asset pricing. Particularly, I find that the subset of stocks that are included simultaneously in the same leg across multiple factors, so-called overlapping stocks, drives nearly all factor performance. Overlapping factor stocks earn on average an abnormal return of 66 basis points per month. In sharp contrast, pure factor stocks that are included only in a single factor display no abnormal returns. The average abnormal return of pure factor stocks across all factors is only 5 basis points per month.

This finding is surprising given that these pure factor stocks have similar size, industry composition, and turnover to overlapping factor stocks. Importantly, overlapping stocks are not in the extreme tails of the original factor portfolios since the characteristic spreads of the overlapping stocks are not higher than those of the pure factor stocks. The overlapping stocks are not in the extreme tails of any factor portfolio since the characteristic spreads are not larger than those of the original factor portfolios. For

[2] Based on monthly returns between 1952:07 and 2020:12.

example, the value spread, which measures the difference in the book-to-market ratio between value and growth stocks, of the total high-minus-low (HML) portfolio is 0.95, while the value spread of the overlapping HML stocks is 0.80. Pure HML stocks have a higher value spread of 1.07.

I focus on five well-known factors, starting with value (HML), momentum (MOM), and idiosyncratic volatility (IVOL). These factors have been shown to robustly predict returns in the cross-section (e.g., Ang, Hodrick, Xing, and Zhang, 2006; Asness, Moskowitz, and Pedersen, 2013; Kelly, Pruitt, and Su, 2019; Ehsani and Linnainmaa, 2022) and these strategies account for a large share of factor investment products offered by the asset management industry[3]. Furthermore, these three factors are constructed based on seemingly unrelated information (book-to-market ratio for HML, past returns for MOM, and past idiosyncratic volatility for IVOL) that are identified as independent characteristics among non-micro cap stocks (Green, Hand, and Zhang, 2017). I extend my analysis by considering the profitability factor (RMW) and the investment factor (CMA), as in Fama and French (2015b).

I form factor portfolios by sorting stocks into quintiles while controlling for size[4]. I then decompose each factor into overlapping stocks that are sorted *into the same leg* of multiple factors and pure factor stocks that are uniquely sorted into a single factor. I combine all overlapping stocks into a value-weighted long-short portfolio and all pure factor stocks into another value-weighted long-short portfolio. I document three main results.

First, I examine the degree of overlap between the long-short portfolios of different factors. Compared to a benchmark that is based on random sorting stocks into quintile portfolios, I do not find excessive overlap. The benchmark predicts that 10.4% stocks will be selected into the same quintile of two or three factors. This is similar to the data, where the fraction of overlap in the long leg (i.e., top quintile) is 9.9% and in the short leg (i.e., bottom quintile) it is 13.2%.

[3] Yet Lettau, Ludvigson, and Manoel (2018) show that not all value funds systematically tilt towards value stocks.

[4] All quintile portfolios contain an equal proportion of large caps, medium caps, and small caps and I always report value-weighted returns.

Second, I document important differences in the performance of overlapping stocks versus pure factor stocks. The portfolio that contains all overlapping stocks generates an average return of 66 basis points per month. In sharp contrast, the average monthly return of all pure factor stocks (across the three factors) is only 5 basis points. Thus, almost the entire return performance of these factors is driven by the same overlapping subset of stocks[5]. This result is robust when I use a different set of factors that includes RMW, CMA, and IVOL.

Third, I show that overlapping stocks are priced in the cross-section of stock returns, while pure factor stocks do not carry a significant price of risk. I estimate Fama-MacBeth's (1973) regressions on different cross-sections of stock returns where I include both the value-weighted long-short portfolio of all pure factor stocks and that of all overlapping stocks as two separate factors[6]. I find that the price of risk for the overlapping portfolio is highly significant and greater than one in all specifications. In contrast, the portfolio with only pure factor stocks (across all three factors) is not significantly priced when both the overlapping and the pure factor portfolios are included as factors.

These results also shed new light on the well-known negative correlation between HML and MOM (see e.g., Asness et al., 2013). I find that this negative correlation is driven by the pure factor stocks. However, the stocks that overlap in the same leg (i.e., value stocks that are also past winners and growth stocks that are also past losers) are the ones that deliver the risk premium. Pure HML stocks and pure MOM stocks are not priced in the cross-section of returns when I control for overlapping stocks.

Finally, I seek to explain the difference in performance between overlapping stocks and pure factor stocks. An advantage of my empirical approach is that I can identify exactly which stocks are in the overlapping and pure factor portfolios. Despite the vastly different performance, the overlapping and pure factor stocks are similar across many

[5] In addition, the mean-variance spanning tests show that overlapping stocks always span the pure factor stocks, but not vice versa.

[6] I combine 25 size×value, 25 size×momentum, 25 size×volatility as the test portfolios. In a robustness test, I also use different test portfolios such as a combination of 25 size×net share issues and 25 size×accruals portfolios.

dimensions. These stocks have similar sizes (since I control for size in my portfolios), industry composition (stocks in the manufacturing industry account for the most in both portfolios), and portfolio turnover (the average turnover rate is 26% for the pure factor portfolio and 39% for the overlapping portfolio).

I then explore several potential risk-based explanations, by comparing the exposures of overlapping and pure factor stocks with respect to various measures of consumption risk, a recession indicator, liquidity risk measures, the term spread, and the default premium. Overall, I find no important differences.

Next, I move on to mispricing-based explanations. My results suggest that they play a more significant role in overlapping stocks than for pure factor stocks when focusing on the short leg. Specifically, the short leg of the overlapping portfolio loads more negatively on investor sentiment as in Baker and Wurgler (2007) than the short leg of the pure factor portfolio. Also, the short interest is higher for overlapping stocks than for pure factor stocks after the 2000s, which implies higher anomaly returns to the short leg of the overlapping portfolios. At the same time, I find that institutional ownership is similar for overlapping stocks and pure factor stocks. Importantly, the return difference between overlapping and pure factor stocks is not only due to the short legs but in part due to differences in the long legs of these portfolios as well.

I perform several robustness tests. First, I consider an alternative set of factors that includes CMA, RMW, and IVOL. My methodology and results are robust to this alternative set of factors as the overlapping stocks again strongly outperform pure factor stocks. Second, my findings are robust across different subsample periods. Even though traditional factors have low performance in the period 2001-2020, I still observe the strong outperformance of the overlapping portfolio compared to the pure factor portfolio in this subsample period. Third, my findings are robust to sorting stocks into characteristic quintiles using NYSE breakpoints instead of assigning an equal number of stocks to each quintile. Further, my results are robust when portfolios are rebalanced every quarter. Given lower trading costs with quarterly rebalancing, this robustness test suggests that focusing on overlapping stocks is a feasible strategy for investors.

My findings have important implications for asset managers investing in smart-beta or factor-based products. Many institutional investors such as pension funds allocate a substantial share of capital to these strategies (e.g., Canada Pension Plan Investment Board, Norwegian Sovereign Wealth Fund). My results suggest that rather than removing overlap between factors or hiring separate managers for each factor product, investors should focus more on this overlap, as these stocks are the ones driving factor returns. My results also highlight the importance of academic research to provide a unified explanatory framework for cross-sectional anomalies rather than focusing on explaining each factor separately.

Related literature. During the past decades, the asset pricing literature has uncovered hundreds of factors linked to the cross-section of stock returns. Many factors are based on firm characteristics (e.g. among many others, Fama and French, 1992a, 1993a, 1996; Cochrane, 2011; Nagel, 2013). Analyzing this factor zoo, Harvey, Liu, and Zhu (2016) report that many do not survive a multiple testing framework. McLean and Pontiff (2016) show that many factor premiums decline after publications of the factors, while Hou, Xue, and Zhang (2020) find that their economic significance is often weaker than originally reported. These findings suggest that many seemingly distinct factors could simply reflect the same information set (Jensen, Kelly, and Pedersen, 2023; Feng, Giglio, and Xiu, 2020). This is in line with my result that the portfolio and asset pricing performance of five well-known factors are driven by exactly those stocks that appear simultaneously in multiple factors. Although I only study five well-known factors, it is straightforward to extend this method to more factors.

Many papers have tried to summarize the information or extract the commonality of factors. Stambaugh and Yuan (2017) form two mispricing factors by averaging stock rankings of anomalies that exhibit high correlations. In other words, the paper combines signals from similar factors. By contrast, I combine signals from seemingly unrelated factors, so there is no high correlation among the factors and no similar construction of the chosen factors. Chen and Velikov (2023) study the expected returns of the average anomaly. When averaging across signals, a weaker signal for one factor

could be offset against strong signals for other factors. Instead, my paper focuses only on stocks with strong signals and I compare stocks with a single signal versus multiple signals.

Bryzgalova, Huang, and Julliard (2023) apply a Bayesian approach to extract the common information among factors. In my paper, I show that it is not just a common driver but actually the same subset of stocks included in different factors driving factor premia. Bessembinder (2018) finds that the long-term market returns are driven by a small subset of stocks. Consistently, I find that returns on five well-known factors are also driven by a subset of stocks, namely those that appear simultaneously in multiple factors.

Another strand of literature looks at the interaction effect between factors. Ehsani and Linnainmaa (2022) find that factors displaying momentum themselves help explain the performance of multiple factors. I also find that momentum stocks are present in the set of overlapping stocks. However, the outperformance of this set of overlapping stocks is not just due to factor rotation, since the set includes those stocks that appear in momentum and other factors at the same time. Green, Hand, and Zhang (2017) find many independent characteristics among non-micro cap stocks, which include namely book-to-market, return volatility, and momentum. My results show that factor performance of the corresponding characteristic sorted portfolios relies heavily on stocks that score high (or low) on a multiple of these characteristics.

Stambaugh, Yu, and Yuan (2015) find that, among overpriced stocks (stocks in the short leg), the IVOL characteristic is negatively related to the abnormal return. My results are consistent as I find that, in the short leg, higher IVOL implies more overlapping stocks, and then more negative returns. Asness, Moskowitz, and Pedersen (2013) finds the negative correlation between value and momentum. Ali, Hwang, and Trombley (2003) finds that the book-to-market ratio is greater for stocks with higher IVOL. Zhang (2006) reports that momentum is concentrated among stocks with high IVOL. In my paper, I find all these relationships hold in the total factor portfolios and the portfolios containing pure factor stocks but not in the portfolios containing overlapping

stocks. Yet, only the portfolios of overlapping stocks earn abnormal returns and are significantly priced in the cross-section.

Müller and Schmickler (2021) consider a large set of anomalies and find evidence for many anomaly interactions. Here, I start from a portfolio perspective and focus on a smaller set of factors that are frequently implemented in practice. I compare different multi-factor strategies and decompose the factor portfolios into pure factor stocks and overlapping factor stocks. Favilukis and Zhang (2019) find that, by sorting portfolios conditionally, momentum is high among stocks in the short leg of many anomalies, including growth, high IVOL, low operating profitability, and high investment. My methodology is different in the way that I sort my portfolios independently on each characteristic. Since my paper focuses on overlap, my method is transparent in what is in the overlap.

Recently, studies in Machine Learning have developed frameworks that incorporate information from these interaction effects (see, for example, Moritz and Zimmermann, 2016; Bryzgalova, Pelger, and Zhu, 2019; Chen and Velikov, 2023; Freyberger, Neuhierl, and Weber, 2020; Kozak, Nagel, and Santosh, 2020; Avramov, Cheng, Metzker, and Voigt, 2023). In my paper, my portfolio strategy is transparent and aligns closely with the conventional approach to factor construction. Furthermore, my method enables the precise identification of stocks that overlap in factor portfolios, allowing for a detailed examination of their characteristics.

The rest of the paper is organized as follows. Section 4.2 describes the data and methodology. Section 4.3 analyzes the performance and asset pricing tests of overlapping stocks and pure factor stocks. Section 4.4 discusses potential explanations for the difference in performance. Section 4.5 presents several practical implications and robustness checks. Section 4.6 concludes.

4.2 Data and methodology

4.2.1 Data

I compute the monthly firm characteristics for HML, MOM, and IVOL factors from July 1952 to December 2020. My main data source is the CRSP/Compustat U.S. stock universe. I filter all common stocks (with share code 10 or 11) traded on NYSE, AMEX, or Nasdaq (with exchange codes 1, 2, or 3).

HML is formed on the book-to-market ratio. I follow closely the standard method described in Kenneth R. French's data library[7]. MOM is formed on prior 12-to-2 returns. I follow Jegadeesh (1990) and Jegadeesh and Titman (1993) by using monthly returns from the past months $t-2$ to $t-12$, skipping the most recent month to avoid market microstructure issues. IVOL is formed on residual volatility, which is the standard deviation of residuals in the Fama-French three-factor model as in Ang, Hodrick, Xing, and Zhang (2006).

$$R_t^i = \alpha^i + \beta_{MKT}^i MKT_t + \beta_{SMB}^i SMB_t + \beta_{HML}^i HML_t + \varepsilon_t^i.$$

For each stock, I regress the excess returns on the Fama-French three factors using a window of 60 days (minimum of 20 days). I then compute the residual volatility by the standard deviation of all residuals in each month for a given stock[8].

I choose the HML, MOM, and IVOL factors for the main analysis since these factors robustly predict cross-sectional returns and have received the most attention from prior academic research. Asness, Moskowitz, and Pedersen (2013) find evidence for value and momentum return premiums across markets and asset classes. Ang, Ho-

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

[8] I use a window of 60 days (minimum of 20 days) instead of 20 days (minimum of 15 days) as in Ang, Hodrick, Xing, and Zhang (2006). 60 days give more observations for each rolling regression, which can yield more reliable estimates. In comparison with the paper, I find a similar pattern that the short portfolio, containing stocks with highest idiosyncratic volatilities, yields the lowest expected return. The difference in average returns between the long and the short portfolios is 0.25% per month. Ang et al. report 1.06% per month (Table VI, Panel B), which is based on a shorter sample period (1963-2000) and a shorter window for rolling regressions (20 days).

drick, Xing, and Zhang (2006) find that the returns on stocks with high idiosyncratic volatility cannot be explained by the size, value, and momentum factors. Kelly, Pruitt, and Su (2019) find that value, momentum, and idiosyncratic volatility are three among ten firm characteristics that necessarily explain the variation in stock returns[9]. Besides, these three factors seem to be unrelated in terms of the type of information used to extract a signal for expected returns. HML is purely based on accounting data. MOM relies on past performance. Whereas, IVOL is formed on residual volatility of the asset pricing model.

For robustness checks, I replicate the CMA and RMW factors. CMA is formed on the change in total assets from the fiscal year ending in $t-2$ to the fiscal year ending in $t-1$, divided by the total assets in $t-1$. RMW is formed on operating profitability. In each June, operating profitability is the annual revenue less cost of goods sold, interest expense and selling general, and administrative expenses, divided by the book equity in the previous fiscal year.

In my analysis, I either include HML or I include CMA and RMW in the set of factors. Fama and French (2015b) find that HML is redundant when adding CMA and RMW, and they are based on related information. The construction of HML requires book equity which in turns requires total assets. CMA and RMW also require the book equity and total assets, resulting in a higher correlation between these factors and a potentially larger degree of overlap. Table A2 shows that none of the five characteristic values is on average highly correlated with each other. The correlation between momentum and low volatility is 0.13 and the correlation between value and momentum is negative at -0.22. The return correlations show that CMA is highly correlated with HML (0.65 on average); however, I do not include these two factors in the same set. My conclusions are highly robust to alternative choices of factors.

[9] Kelly, Pruitt, and Su (2019) investigate 36 firm characteristics and find ten of them remain statistically significant at 1%.

4.2.2 Factor decomposition methodology

I decompose the universe of stocks into overlapping stocks and pure factor stocks. I also decompose each factor into so-called pure factor stocks and so-called overlapping stocks that also appear in the same (long or short) leg of other factors. This section discusses my empirical approach.

First, I construct each factor portfolio based on the characteristic ranking of stocks while controlling for size. Specifically, in each month, I first sort stocks into the 30^{th} and 70^{th} percentile of market cap, using NYSE breakpoints. Then, within each size tercile, I rank stocks by the characteristic value and split them *equally* into quintiles. Thus, each characteristic quintile portfolio contains the same amount of stocks. Portfolios are value-weighted and rebalanced every month.

This deviation is a mild difference from the traditional approach by Fama and French, where stocks are sorted into characteristic quintiles by using NYSE breakpoints. Here, I make sorting univariately distributed so that I can derive the benchmark. Since the number of stocks is the same in every quintile, the probability that a stock appears in any portfolio is simply 1/5. With this probability, I can calculate the benchmark fraction that each factor overlaps with each other. Yet my robust results show that the conclusions remain the same when I sort all characteristics by using NYSE breakpoints.

For another technical convenience, I do not sort stocks from the lowest to the highest characteristic values, but from the least attractive to the most attractive manners of characteristic. In other words, for every factor, Quintile 5 or the top quintile denotes the long portfolio, while Quintile 1 or the bottom quintile denotes the short portfolio. Henceforth, I reverse the sort from high to low values for idiosyncratic volatility and investment. The final data set is the characteristic scores from one to five for each stock in each month[10].

I define each time that a stock appears in a top or bottom characteristic quintile as the stock receiving a signal for its positive or negative return. I do not count stocks in the intermediate quintiles (Quintiles 2, 3, and 4) as stocks receiving a signal, so in my

[10] The correlation between my returns and French's data is above 97% for all factors.

methodology, I only consider strong signals[11]. To measure the fraction of overlap, for each stock in each month, I count how many signals the stock receives for a given top or bottom quintile. I denote frequency k in Quintile j as the number of times that a stock has score j, for $j \in \{1, 5\}$, among all factors in the strategy. Thus, k ranges from 0 to S signals, where S denotes the number of factors such as $S = 3$ in the strategy that invests in HML, MOM, and IVOL. I identify the number of signals by three cases. Namely in Quintile j, a stock can receive

1. no signal ($k = 0$): The stock exhibits no signal when it does not appear in Quintile j of any characteristic. The stock is not sorted into the quintile of my interest so I do not investigate this case.

2. a single signal ($k = 1$): The stock exhibits a single signal when it appears in Quintile j of only one factor but not in this quintile of any other factors. This stock is defined as a pure factor stock. For example, if a stock scores 3 for momentum and volatility but 5 for value, then this stock only exhibits a positive signal for high expected returns for value. Thus, this stock receives a pure value signal.

3. multiple signals ($k \geq 2$): The stock exhibits multiple signals when it appears in Quintile j of at least two factors. In other words, at least two factors contain the same stock in the same leg. This stock is defined as an overlapping stock. In an extreme case, if a stock scores 5 in all three factors, then it exhibits multiple signals for high expected returns for all three factors.
   Note that I only consider signals *in the same leg* of multiple factors as multiple signals. Although a stock sorted in the top quintile of one factor can also be in the bottom quintile of another factor, these two signals cancel out in my equal-weighted multi-factor strategy, so I do not consider this stock as an overlapping stock.

Signals in the extreme quintiles indicate the decision to go long, if stocks appear in

[11] Stambaugh and Yuan (2017) study average signals from the percentile ranks of 11 anomalies. Here, I only count strong signals that are from the top 20% and the bottom 20% of the characteristic rankings.

the top portfolio, or to go short if stocks appear in the bottom portfolio. With this measure, I can decompose our universe of stocks into stocks receiving a single signal (so-called pure factor stocks) and stocks receiving multiple signals (so-called overlapping stocks). The overlapping portfolio contains overlapping stocks that overlap between factors: HML & MOM, HML & IVOL, MOM & IVOL, or HML & MOM & IVOL, illustrated as the shaded regions in Figure 4.1. The pure factor portfolio contains pure factor stocks across the three factors, so this portfolio contains all pure HML, pure MOM and pure IVOL stocks.

This measure also allows me to decompose each factor into the pure factor portfolio that contains stocks receiving only a signal for that factor and the overlapping portfolio that contains stocks receiving the same signal for that factor and some other factors. For example, in the long leg of HML, the pure HML portfolio contains pure value stocks, while the overlapping HML portfolio contains value & winner, value & low-vol, and value & winner & low-vol stocks.

4.2.3 Analysis of degree of overlap

Since all long-short factors are constructed from the same universe of stocks, by chance there can be overlap. In this section, I compare the degree of overlap based on a random-sorting benchmark to the actual degree of overlap I observe in the data.

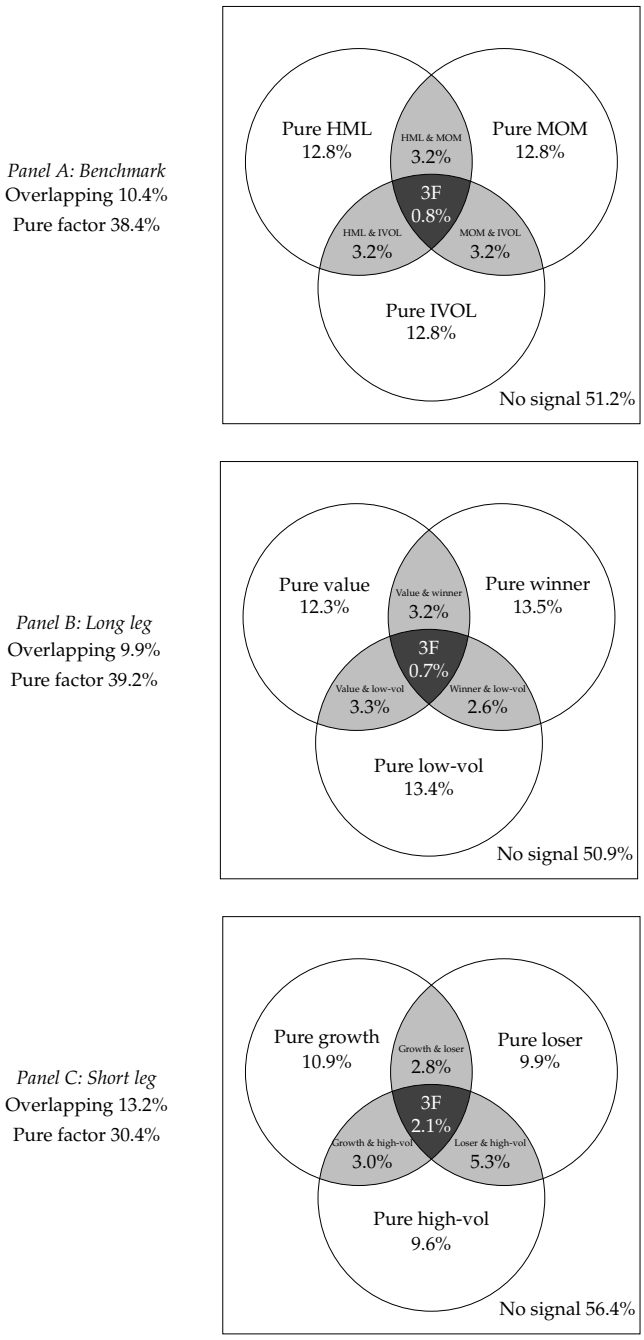
I derive a benchmark fraction of overlap assuming all signals follow a uniform distribution. When initially sorting on size, the probability that a stock appears in the i^{th} size percentile equals

$$\Pr(size = i) = \frac{1}{m}, \quad i = 1, \ldots, m.$$

Then within each size percentile, the probability that a stock has a characteristic score j equals

$$\Pr(char = j | size = i) = \frac{1}{5}, \quad i = 1, \ldots, m; j \in \{1, 5\}.$$

Figure 4.1: **Fractions of overlapping stocks and pure factor stocks**



Description. This figure presents a Venn diagram of HML, MOM, and IVOL portfolios. Within the portfolios, the shaded areas represent overlapping stocks and the non-shaded areas represent pure factor stocks. The areas outside the portfolios represent stocks that are not sorted into the portfolio of any factor. The benchmark is the same for the long leg and short leg.

The unconditional probability is hence given by

$$\begin{aligned} p := \Pr(char = j) &= \sum_{i=1}^{m} \Pr(char = j | size = i) \Pr(size = i) \\ &= m \times \frac{1}{5} \times \frac{1}{m} = \frac{1}{5}. \end{aligned}$$

This likelihood represents the probability that a stock exhibits a signal to a given quintile portfolio. By independent sorting, the size controls do not affect the signal likelihood p. Let $f_S(k)$ be the fraction of signal k in an S-factor strategy. Then k follows the Binomial(S, p) distribution and $f_S(k)$ is the probability that takes the form

$$\begin{aligned} f_S(k) &= \binom{S}{k} p^k (1-p)^{S-k} \\ &= \binom{S}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{S-k}, \quad k = 0, \ldots, S. \end{aligned}$$

Under this notation, the fraction of a single signal is then $f_S(1)$, and the fraction of multiple signals is given by $\sum_{k=2}^{S} f_S(k)$, the total probability that the stock has from two to S signals in the given factor portfolio.

Table 4.1: **Fraction of pure factor stocks and overlapping stocks**

	Long leg	Short leg	Benchmark
Panel A: In the universe of stocks			
Pure factor	0.392	0.304	0.384
Pure HML	0.123	0.109	0.128
Pure MOM	0.135	0.099	0.128
Pure IVOL	0.134	0.096	0.128
Overlapping	0.099	0.132	0.104
HML & MOM	0.032	0.028	0.032
HML & IVOL	0.033	0.030	0.032
MOM & IVOL	0.026	0.053	0.032
HML & MOM & IVOL	0.007	0.021	0.008
Panel B: In each factor portfolio			
HML	1.000	1.000	1.000
Pure HML	0.628	0.579	0.640
Overlapping HML	0.372	0.421	0.360
HML & MOM	0.165	0.148	0.160
HML & IVOL	0.171	0.159	0.160
HML & MOM & IVOL	0.036	0.114	0.040
MOM	1.000	1.000	1.000
Pure MOM	0.674	0.491	0.640
Overlapping MOM	0.326	0.509	0.360
HML & MOM	0.161	0.140	0.160
MOM & IVOL	0.130	0.262	0.160
HML & MOM & & IVOL	0.035	0.107	0.040
IVOL	1.000	1.000	1.000
Pure IVOL	0.669	0.481	0.640
Overlapping IVOL	0.331	0.519	0.360
HML & IVOL	0.166	0.150	0.160
MOM & IVOL	0.130	0.262	0.160
HML & MOM & IVOL	0.035	0.107	0.040

Description. This table reports the average fraction of pure factor/ overlapping stocks in the data versus the benchmark fraction in the strategy that invests equally in HML, MOM, and IVOL. The benchmark fractions are given by $f_S(k) = \binom{S}{k}\left(\frac{1}{5}\right)^k\left(\frac{4}{5}\right)^{S-k}, k = 0,\ldots,S$ for $S = 3$. The benchmark is the same for the long leg and the short leg. The empirical fractions are the average of fractions in each month per portfolio.

Table 4.1 reports the average fractions of signals in our universe of stocks, in each factor portfolio, with the corresponding benchmark fractions. Figure 4.1 also visualizes in the Venn diagram these factions in our universe of stocks, in comparison to the benchmark. I find that on average the fraction of overlapping stocks and the fraction of pure factor stocks in all portfolios are almost the same as the benchmark. The difference is economically small, approximately by a few percent per month.

For example, while the random-sorting benchmark predicts that the fraction of overlapping stocks is 10.4%, I find that my data contains on average 9.9% stocks with multiple signals in the long leg and 13.2% stocks with multiple signals in the short leg (Panel A). Within the HML portfolio, the long leg contains 62.8% pure value stocks and 37.2% overlapping value stocks, while the short leg contains 57.9% pure growth stocks and 42.1% overlapping growth stocks (Panel B). The benchmark predicts that each leg contains 64% pure value stocks and 36% overlapping value stocks, so the difference in each leg is relatively small. Hence, in the data, there is no excessive overlap in either our universe of stocks or in each individual factor portfolios. This natural overlap occurs because, given the same universe of stocks, I choose the top 20% and the bottom 20% of stocks in each characteristic. In overall, I find that there is overlap between factors but the degree of overlap is not excessive relative to the random-sorting benchmark. Yet it is surprising when I show that these overlapping stocks nearly drive all the factor performance.

4.3 Performance and asset pricing test

4.3.1 Portfolio performance

In this section, I study the performance of each decomposed portfolio to examine whether the overlap between factors matters and which subset of stocks drives the performance of the multi-factor strategy. I start with a multi-factor strategy that equally invests 1/3 in HML, 1/3 in MOM and 1/3 in IVOL. I then disentangle stocks that appear in the same leg of two or more factors at the same time (overlapping stocks)

and stocks that appear in only one factor (pure factor stocks).

Table 4.2: **Portfolio performance of the HML-MOM-IVOL strategy**

	Return				Sharpe	Distribution				CAPM		Average characteristic			
Portfolio	Mean (%)	Min. (%)	Max. (%)	Volatility (%)	ratio	Positive (%)	Negative (%)	Skewness	Kurtorsis	$\hat{\alpha}$	$\hat{\beta}_{MKT}$	Log-mkt. cap	B/M	Prior-ret (%)	Resid. vol. (%)
Market	0.651	-22.883	16.530	4.288	0.152	60.827	39.173	-0.485	1.881			4.880	0.538	17.696	1.304
Overlapping portfolio															
Long	0.954	-20.860	14.176	4.141	0.230	61.922	38.078	-0.565	2.229	0.401***	0.848***	4.989	0.780	33.685	1.105
Short	0.298	-27.080	30.384	6.547	0.045	52.190	47.810	0.000	2.418	-0.595***	1.372***	4.371	0.429	4.980	1.899
Long–Short	0.656	-31.374	25.001	4.712	0.139	58.637	41.363	-0.618	6.672	0.997***	-0.524***	0.618	0.352	28.705	-0.794
Pure factor portfolio (across all factors)															
Long	0.730	-22.175	15.688	4.228	0.173	61.557	38.443	-0.478	1.881	0.104***	0.963***	4.921	0.555	25.058	1.236
Short	0.680	-21.057	19.802	4.595	0.148	59.611	40.389	-0.397	1.553	0.001	1.043***	4.867	0.421	16.842	1.371
Long–Short	0.050	-7.880	11.717	1.443	0.035	49.757	50.243	0.395	8.897	0.102**	-0.081***	0.055	0.133	8.216	-0.135
EW. multi-factor	0.287	-21.225	21.032	2.611	0.110	59.124	40.876	-0.466	14.057	0.462***	-0.268***	0.323	0.216	18.849	-0.417
HML	0.101	-15.148	19.781	3.557	0.028	51.582	48.418	0.168	2.630	0.181	-0.123***	-0.638	0.947	-3.025	-0.078
MOM	0.596	-35.169	25.824	5.033	0.118	58.029	41.971	-0.828	7.718	0.670***	-0.114***	0.710	-0.297	68.153	-0.116
IVOL	0.165	-19.845	26.885	4.264	0.039	52.798	47.202	0.324	5.411	0.534***	-0.566***	0.897	-0.004	-8.580	-1.056

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the summary statistics of overlapping/ pure factor portfolio value-weighted returns, in excess of the risk-free rate, comparing to the market and factor returns. The sample is from 1952:07 to 2020:12 (822 monthly returns). Market is the excess returns of the CRSP U.S. total market index. The risk-free rate is from Kenneth R. French's Data Library. The last four columns present the averages over time of the value-weighted characteristic of stocks in each portfolio. The EW. multi-factor portfolio is the equal-weighted combination of the three factors. Note that the pure factor portfolio contains pure HML, pure MOM, and pure IVOL stocks.

Table 4.2 reports the summary statistics of each portfolio in comparison to the market and the equal-weighted multi-factor portfolio. Market is the value-weighted average of all stocks in the sample, giving an excess return of approximately 65.1 basis points per month. The multi-factor portfolio, which contains both the pure factor and overlapping stocks, is the average of all three factor returns, so the mean return of this portfolio is between the long-short overlapping and the long-short pure factor portfolios ($5.0 < 28.7 < 65.6$ basis points).

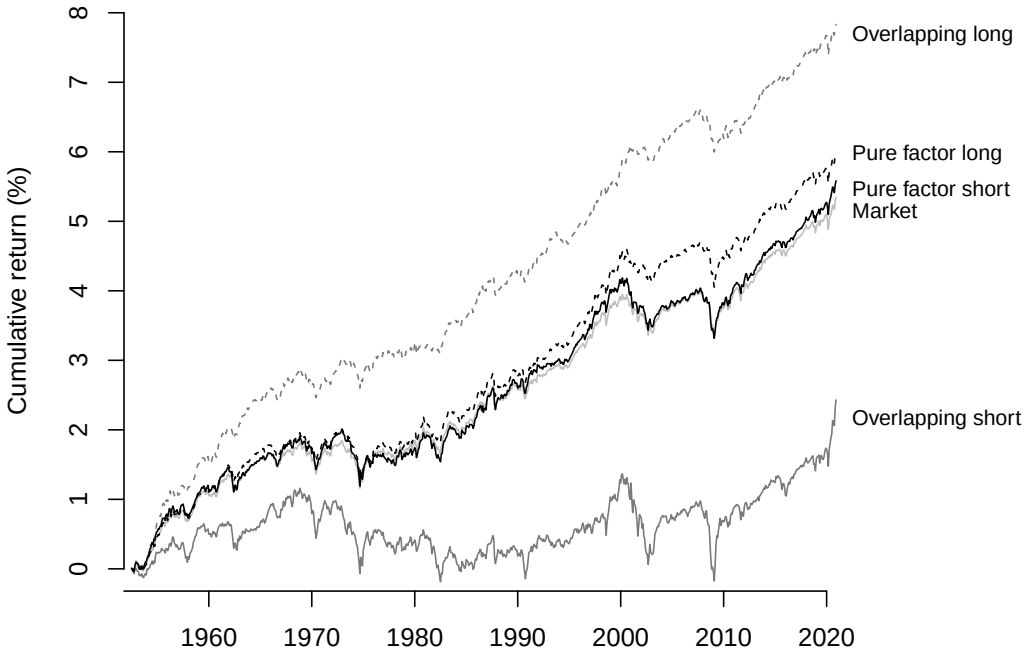
I observe the outperformance of overlapping stocks in several aspects. First, only the long-short overlapping portfolio performs closely to the market ($65.6 \approx 65.1$ basis points). This implies that a subset of stocks overlapped in multiple factor portfolios can generate the same average return as the market portfolio which invests in the entire universe of stocks. The long-short pure factor portfolio, by contrast, generates almost no average return compared to the market (5.0 basis points).

Second, the long-short overlapping portfolio yields a higher expected return so it also results in higher volatility, specifically 4.71% per month. However, this portfolio yields a higher Sharpe ratio than the equal-weighted multi-factor portfolio and the long-short pure-factor portfolio ($0.14\% > 0.11\% > 0.04\%$). Third, the long-short overlapping portfolio yields the largest and most significant CAPM alpha ($1.0 > 0.46 > 0.10$). Especially in the short leg, only overlapping stocks deliver a negative alpha of -0.60. It implies that only stocks overlapped in the short leg of multiple factors can deliver shorting benefits.

Overall, the average returns on pure factor stocks are very close to zero (even though I combine pure factor stocks of three different factors). In contrast, overlapping stocks appear to drive the entire average return across seemingly unrelated factor strategies.

Figure 4.2 visualizes the performance of each portfolio by plotting the cumulative excess returns. The figure supports the previous finding and suggests including especially overlapping stocks in the short leg. I observe that only the overlapping short portfolio underperforms the market cumulatively. It is important that not only the short leg but also the long leg of the overlapping portfolio outperforms the market over time.

Figure 4.2: **Cumulative returns of the overlapping and pure factor portfolios**



Description. This figure plots the cumulative log-returns of the overlapping and pure factor portfolios in each leg, in comparison to the market portfolio. The sample contains the monthly excess returns from 1952:07 to 2020:12 (822 months). The short portfolios are in solid and the long portfolios are in dashed lines. The pure factor portfolios are in black and overlapping portfolios are in grey.

The combination of the long and short legs of overlapping stocks accumulates the most returns, as the gap between these two portfolios is the widest. Moreover, this gap increases over time.

Next, I decompose each factor portfolio to study the characteristics of the stocks that drive each factor return. I classified stocks in each factor portfolio by two types: (1) those appearing in this factor only (pure factor stocks with a single signal to this factor), and (2) those appearing in this factor and other factors (overlapping stocks with the same signal to multiple factors). The first subset of stocks forms a pure factor portfolio and the second subset forms an overlapping portfolio.

Table 4.3: **Individual factor decomposition**

	Total	Pure	Over-	Decomposition of overlapping portfolio			
	Factor	Factor	lapping	HML & MOM	HML & IVOL	MOM & IVOL	3F
Panel A: HML							
Return (%)							
Long	1.077	0.983	1.205	1.256	1.054		1.387
Short	0.976	1.101	0.708	0.630	0.958		0.313
Long–Short	0.101	-0.118	0.498	0.625	0.095		1.074
Log-market cap							
Long	4.486	4.264	4.865	4.561	5.118		4.941
Short	5.124	5.414	4.732	4.726	5.103		4.159
Long–Short	-0.638	-1.150	0.134	-0.165	0.015		0.781
Avg. weight	1.000	0.598	0.402	0.152	0.180		0.071
Characteristic spreads (Long–Short)							
Book-to-market ratio	0.947	1.068	0.801	0.643	0.947		0.679
Prior 12-to-2 return (%)	-3.025	-17.015	16.730	71.181	-32.460		65.654
Residual volatility (%)	-0.078	0.220	-0.692	0.060	-1.101		-1.200
Panel B: MOM							
Return (%)							
Long	1.328	1.363	1.365	1.256		1.340	1.387
Short	0.732	0.924	0.475	0.630		0.372	0.313
Long–Short	0.596	0.439	0.891	0.625		0.968	1.074
Log-market cap							
Long	5.133	5.231	4.918	4.561		5.371	4.941
Short	4.423	4.724	4.153	4.726		3.822	4.159
Long–Short	0.710	0.507	0.765	-0.165		1.549	0.781
Avg. weight	1.000	0.604	0.396	0.148		0.181	0.069
Characteristic spreads (Long–Short)							
Book-to-market ratio	-0.297	-0.510	0.051	0.643		-0.607	0.679
Prior 12-to-2 return (%)	68.153	68.771	64.789	71.181		60.744	65.654
Residual volatility (%)	-0.116	0.184	-0.604	0.060		-1.018	-1.200
Panel C: IVOL							
Return (%)							
Long	1.021	0.938	1.246		1.054	1.340	1.387
Short	0.856	1.042	0.660		0.958	0.372	0.313
Long–Short	0.165	-0.104	0.585		0.095	0.968	1.074
Log-market cap							
Long	5.213	5.234	5.194		5.118	5.371	4.941
Short	4.316	4.376	4.269		5.103	3.822	4.159
Long–Short	0.897	0.858	0.925		0.015	1.549	0.781
Avg. weight	1.000	0.574	0.426		0.176	0.182	0.069
Characteristic spreads (Long–Short)							
Book-to-market ratio	-0.004	-0.127	0.259		0.947	-0.607	0.679
Prior 12-to-2 return (%)	-8.580	-24.732	13.825		-32.460	60.744	65.654
Residual volatility (%)	-1.056	-0.942	-1.129		-1.101	-1.018	-1.200

Description. We decompose each factor into pure factor stocks and multi-signal stocks. This table reports the summary statistics of decomposed portfolios for each factor: HML (Panel A), MOM (Panel B), and IVOL (Panel C), where 3F denotes HML & MOM & IVOL. The sample is from 1952:07 to 2020:12 (822 monthly returns). Portfolios are value-weighted of stocks returns and rebalanced every month. The characteristic spreads are the long–short averages over time of the value-weighted characteristics of the stocks in each portfolio.

Table 4.3 reports the average returns and characteristics of each decomposed factor portfolio. In all factors, I find that the overlapping portfolios drive almost the entire performance of the factors. The long-short overlapping portfolios generate higher average returns than the total factor portfolios ($49.8 > 10.1$ basis points in HML, $89.1 > 59.6$ basis points in MOM, and $58.5 > 16.5$ basis points in IVOL). In contrast, the return of the pure factor portfolio falls below that of the total factor. The average long-short returns for the pure MOM portfolio is 43.9 basis points, while the returns are even negative for the pure HML portfolio (-11.8 basis points) and for the pure IVOL portfolio (-10.4 basis points).

Interestingly, I find that the value premium is driven by those value stocks that are also past winners (positive momentum) and those growth stocks that are also past losers (negative momentum). The return outperformance of HML & MOM stocks over the pure HML stocks comes from both legs ($1.26\% > 0.98\%$ in the long leg and $0.63\% < 1.10\%$ in the short leg). The difference makes the long-short return of the HML & MOM portfolio on average higher than that of the pure HML portfolio.

This result might at first seem to contradict Asness, Moskowitz, and Pedersen (2013)'s finding that HML and MOM are negatively correlated. I observe the same in that the total HML portfolio has a negative prior return of -3.03% per month. The pure HML portfolio has a large negative prior return, but the HML & MOM portfolio has a large positive prior return. Therefore, the negative correlation between HML and MOM seems to come from the negative comovement between pure HML and pure MOM stocks. I find that the correlation between the pure HML portfolio returns and the pure MOM portfolio returns is indeed negative (-0.54, not shown in the table).

However, what drives the performance of multiple factors are overlapping stocks. If I look at the decomposition of the overlapping HML portfolio, the HML & MOM portfolio not only generates a high average return but also yields the highest past performance. The high average prior returns indicate that the past performance of the HML portfolio is driven by MOM when HML stocks are also sorted into the MOM portfolio. I find that only uniquely sorted stocks in either HML or MOM make these

two factors negatively correlated, but they contribute almost no returns to the factors, especially to HML. Hedging HML stocks against MOM stocks implies removing any overlapping stocks that are both these two factors, which weigh for 15.2% in the HML portfolio and 14.8% in the MOM portfolio, that drive the factor premiums.

I find no peculiar characteristics in the overlapping portfolios. Since I control for size, the long leg and short leg in each factor have approximately the same log-market cap. Importantly, the value spread (i.e., the average book-to-market ratio of value stocks minus that of the growth stocks) is smaller for the overlapping HML portfolio (0.80) than for the total HML portfolio (0.95) and than for the pure HML stocks (1.068). In all other overlapping portfolios, the characteristic spreads are roughly the same as in the total factor portfolio, so stocks in the overlapping portfolios do not appear in the extreme tails of the factors.

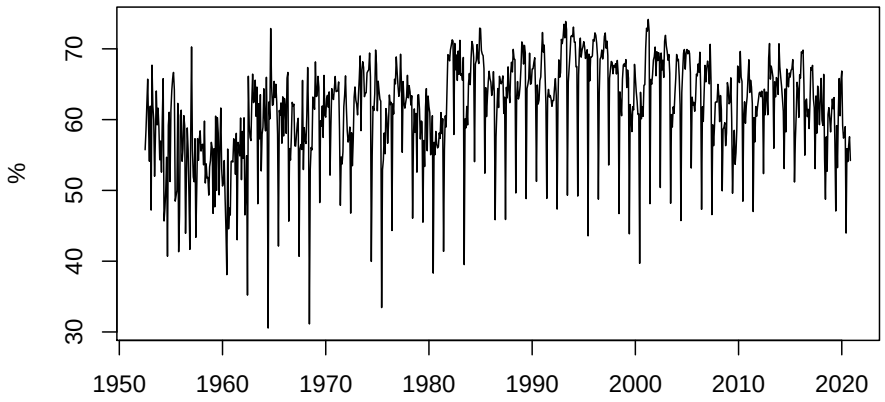
I also check the portfolio turnover and find that the turnover rate of the overlapping portfolio is reasonable. Figure 4.3 plots the proportion of stocks in a portfolio this month staying in the same portfolio in the following month. In most months, the proportion of stocks staying in the same portfolio is high and stable at an average of 61.5%. Thus, the average turnover rate in the overlapping portfolio is about 38.5% per month. The average turnover rate is 26% for the pure factor portfolio. This result implies that for practitioners, the strategy that invests in overlapping stocks is feasible. Also, it infers a stability link for future research to study the portfolio turnover of the subset of stocks driving the factor performance[12].

In the MOM factor, the long-short portfolio of HML & MOM stocks also yields a higher average return and a higher past performance than that of pure momentum stocks. The outperformance of the overlapping MOM portfolio over the pure MOM portfolio seems to come from the short leg, with an average return of 0.48% per month. However, the prior return gap in the overlapping MOM portfolio is not considerably different from that in the pure MOM or the whole MOM portfolio. I find similar con-

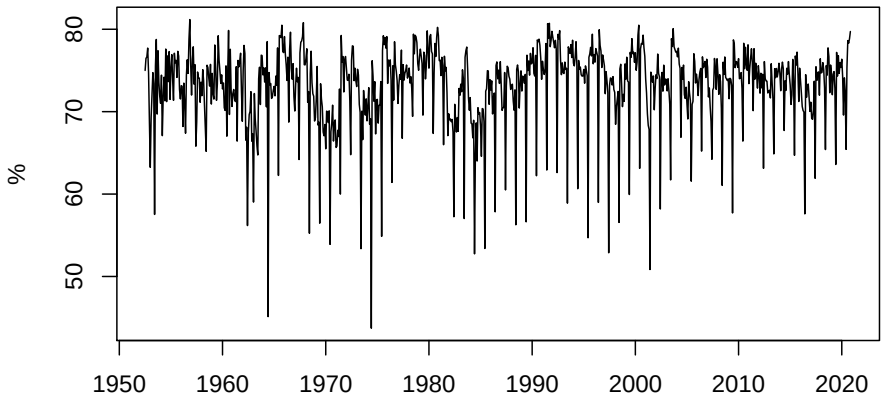
[12] The turnover rate is possibly lower if portfolios are rebalanced quarterly. In my robustness check for quarterly rebalancing, I still find abnormal returns in the overlapping stocks. Therefore, focusing on the overlapping stocks would be a feasible strategy for investors.

Figure 4.3: **Proportion of stocks staying in the portfolio next month**

(a) *Overlapping portfolio*



(b) *Pure factor portfolio*



Description. This figure plots the proportions of stocks in the overlapping and the pure factor portfolios that stays in the same portfolio in the following month. The sample contains the monthly excess returns from 1952:07 to 2020:12 (822 months). The portfolio is rebalanced every month.

clusions for the IVOL factor. The average return of the MOM & IVOL portfolio is higher than that of other decomposed portfolios. The outperformance of the overlapping IVOL and the MOM & IVOL portfolio comes from both legs. Indeed, MOM stocks seem to drive other factor performance, which is in line with the finding of Ehsani and Linnainmaa (2022). However, I find that what drives multiple factor performance is not just factor rotation, but it is a subset of stocks appearing in multiple factors at the same moment in time.

Figure 4.4 plots the cumulative returns of the decomposed overlapping portfolios in each factor. I observe that all portfolios containing overlapping stocks perform as least as well as the total factor portfolio, in all panels. In contrast, the pure HML, pure MOM, and pure IVOL portfolios underperform the total factor portfolios cumulatively over time.

4.3.2 Cross-sectional asset pricing tests

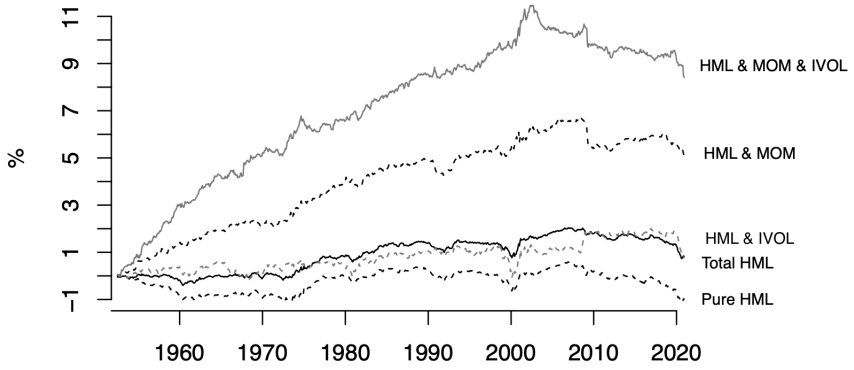
The previous section shows the difference in portfolio performance when I compare pure factor stocks to overlapping stocks. This section analyzes the asset pricing implications. I run the Fama and MacBeth (1973b) regressions for different cross-sections of stock returns.

The test portfolios combine 25 portfolios formed on size and value, 25 portfolios formed on size and momentum, and 25 portfolios formed on size and idiosyncratic volatility. Alternatively, I also test on a combination of 25 portfolios formed on size and net stock issues and 25 portfolios formed on size and accruals. In each specification, I include the MKT factor, SMB factor, the three original factors (HML, MOM and IVOL), the long-short overlapping portfolio as a factor, and the long-short pure factor portfolio as a factor. Note that the pure factor portfolio contains pure factor stocks across all three factors. In the first stage, I run a time-series regression for each test portfolio i, yielding estimates for the factor loadings β_i for each factor portfolio f_t

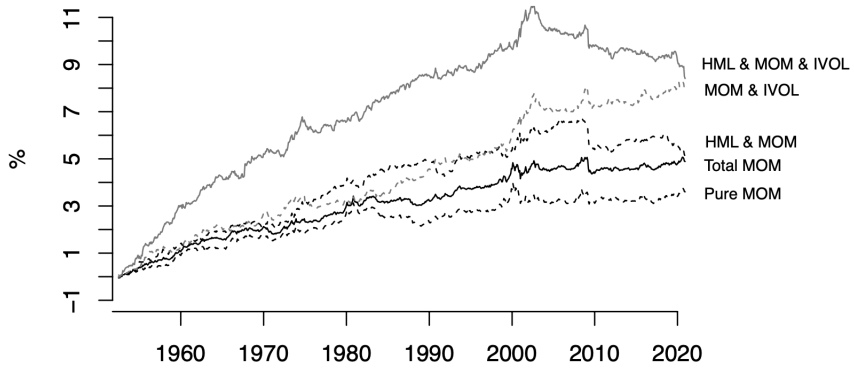
$$R_{i,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t}, \quad i = 1, ..., N.$$

Figure 4.4: **Cumulative returns of decomposed factor portfolios**

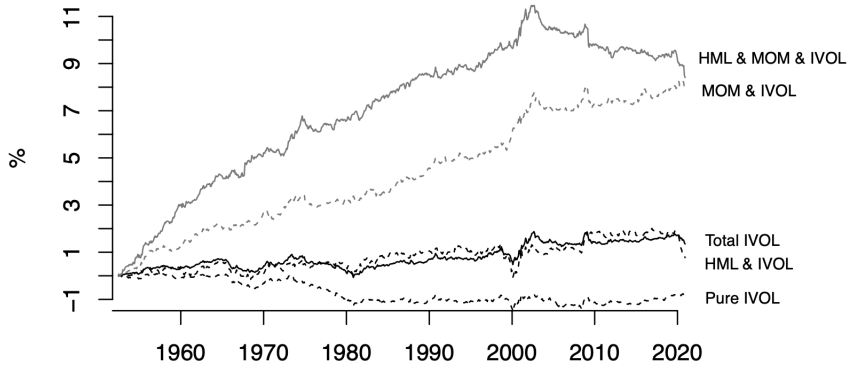
(a) *HML*



(b) *MOM*



(c) *IVOL*



Description. This figure plots the cumulative log-returns of long-short portfolios in each factor: HML (Panel A), MOM (Panel B), and IVOL (Panel C). The sample contains the monthly excess returns from 1952:07 to 2020:12 (822 months)

In the second stage, I run a cross-sectional regression of estimated factor loadings $\hat{\beta}_i$ for each month t

$$R_{i,t} = \lambda_0 + \lambda_t \hat{\beta}_i + \epsilon_{i,t}, \quad t = 1, ..., T.$$

The risk premium for each factor portfolio λ is the time average of the estimated $\hat{\lambda}_t$. Table 4.4 reports the time-series mean and t-statistic of the cross-sectional regression coefficients.

Table 4.4: **Fama–MacBeth cross-sectional regressions**

	(1)	(2)	(3)	(4)	(5)
Panel A: Cross-section of 25 size×value, 25 size×momentum, and 25 size×volatility portfolios					
λ_{MKT}	-1.038***	0.753**	0.354	-0.203	0.559**
	(-6.21)	(2.18)	(1.36)	(-0.94)	(1.97)
λ_{SMB}	0.134***	0.216***	0.165***	0.141***	0.181***
	(2.59)	(3.81)	(3.20)	(2.73)	(3.45)
λ_{HML}		0.248**			
		(2.52)			
λ_{MOM}		0.685***			
		(6.88)			
λ_{IVOL}		0.344***			
		(3.63)			
$\lambda^{3F}_{overlap}$			1.231***		1.525***
			(9.15)		(7.19)
λ^{3F}_{pure}				0.601***	-0.194
				(6.89)	(-0.96)
λ_0	1.748***	-0.117	0.303	0.869***	0.099
	(10.45)	(-0.32)	(1.14)	(3.91)	(0.34)
Adjusted R^2	0.309	0.773	0.694	0.560	0.733
Panel B: Cross-section of 25 size×net share issues and 25 size×accruals portfolios					
λ_{MKT}	-0.704**	0.379	0.269	-0.131	0.167
	(-2.25)	(0.81)	(0.56)	(-0.34)	(0.33)
λ_{SMB}	0.168***	0.111	0.139**	0.086	0.107
	(2.82)	(1.34)	(2.30)	(1.27)	(1.50)
λ_{HML}		0.530**			
		(1.96)			
λ_{MOM}		2.350**			
		(2.49)			
λ_{IVOL}		0.209			
		(0.78)			
$\lambda^{3F}_{overlap}$			1.312***		1.193***
			(3.17)		(2.73)
λ^{3F}_{pure}				0.728***	0.497
				(2.82)	(1.39)
λ_0	1.354***	0.259	0.365	0.769**	0.468
	(4.31)	(0.55)	(0.75)	(2.00)	(0.93)
Adjusted R^2	0.325	0.662	0.511	0.548	0.551

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the second-stage cross-sectional regressions by Fama and MacBeth (1973b) on the returns of 75 portfolios formed on size×value, size×momentum, and size×volatility (Panel A) and 50 portfolios formed on size×net share issues and size×accruals (Panel B). Double sorted portfolios are retrieved from Kenneth R. French's data library.

I find strong evidence that a value-weighted long-short portfolio with all overlapping stocks is priced in the cross-section of stock returns. When the test portfolios include size with value, momentum, and volatility (in Panel A), the estimated price of risk of this overlapping portfolio is significant and positive at 1.23, with a t-value of 9.15 (Column 3). This is much larger and more significant than each factor risk premium when I include all three total factors in the model. The price of risk estimate of the pure factor portfolio is significant and positive (Column 4), but the unpriced risk λ_0 is also significantly different from zero. Note that the pure factor portfolio is exposed to all three factors as it contains all pure HML, all pure MOM, and all pure IVOL stocks. When I include both the overlapping portfolio and the pure factor portfolio as two separate factors in the model (Column 5), only the overlapping portfolio remains significantly priced.

Indeed, the overlapping portfolio is priced in almost all cross-sections of stock returns. When I use the test portfolios sorted on completely different characteristics such as net share issues and accruals (Panel B), the total IVOL factor is no longer significantly priced. However, the estimated price of risk of the overlapping portfolio is still greater than one and highly significant. I find the same evidence that the overlapping stocks are priced in every scenario but the pure factor stocks are not priced when the model includes both returns of the overlapping and the pure factor portfolios as separate factors. Also, when I only include the overlapping portfolio, the intercept becomes insignificant, while it is significant and positive when I only include the pure factor portfolio.

Finally, I study the asset pricing implications of the overlapping stocks contained in each factor separately. I run the Fama-Macbeth cross-sectional regressions of returns of 25 portfolios formed on size and the characteristic of each individual factor, the overlapping and pure factor portfolios. Table 4.5 reports the estimated risk premiums for value (Panel A), momentum (Panel B), and volatility (Panel C).

Table 4.5: **Fama–MacBeth cross-sectional regressions on decomposed individual factors**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Cross-section of 25 size×value portfolios								
λ_{MKT}	-1.237***	-0.656	-0.294	-0.814	-0.135	-0.602	-0.854	-0.065
	(-2.61)	(-1.11)	(-0.39)	(-1.34)	(-0.17)	(-0.94)	(-1.55)	(-0.06)
λ_{SMB}	0.083	0.122	0.133	0.111	0.135	0.123	0.117	0.129
	(1.06)	(1.48)	(1.57)	(1.34)	(1.60)	(1.48)	(1.42)	(1.54)
λ_{HML}		0.275***						
		(2.79)						
$\lambda^{3F}_{overlap}$			1.076***		1.350**			
			(2.67)		(2.10)			
λ^{3F}_{pure}				0.308	-0.021			
				(1.61)	(-0.06)			
$\lambda^{HML}_{overlap}$						0.722***		1.364
						(2.87)		(1.18)
λ^{HML}_{pure}							0.308	-0.183
							(1.35)	(-0.28)
λ_0	1.967***	1.314**	0.951	1.493**	0.797	1.264*	1.518***	0.739
	(4.19)	(2.14)	(1.19)	(2.36)	(0.94)	(1.88)	(2.64)	(0.65)
Adjusted R^2	0.397	0.543	0.531	0.449	0.528	0.508	0.488	0.505
Panel B: Cross-section of 25 size×momentum portfolios								
λ_{MKT}	-1.511***	-0.395	0.629	0.070	1.376*	-0.153	-0.597*	0.547
	(-4.98)	(-1.09)	(1.29)	(0.17)	(1.75)	(-0.39)	(-1.72)	(0.68)
λ_{SMB}	0.190**	0.227**	0.184*	0.132	0.270**	0.202**	0.212**	0.184*
	(2.01)	(2.40)	(1.94)	(1.38)	(2.28)	(2.13)	(2.23)	(1.91)
λ_{MOM}		0.723***						
		(6.95)						
$\lambda^{3F}_{overlap}$			1.188***		1.945***			
			(6.78)		(3.00)			
λ^{3F}_{pure}				0.716***	-0.768			
				(6.33)	(-1.01)			
$\lambda^{MOM}_{overlap}$						1.168***		1.771***
						(7.16)		(2.84)
λ^{MOM}_{pure}							0.984***	0.432
							(5.33)	(1.10)
λ_0	2.208***	1.173***	0.067	0.646	-0.713	0.917**	1.390***	0.162
	(6.99)	(3.21)	(0.14)	(1.51)	(-0.88)	(2.34)	(3.97)	(0.19)
Adjusted R^2	0.375	0.879	0.874	0.833	0.893	0.859	0.834	0.870
Panel C: Cross-section of 25 size×volatility portfolios								
λ_{MKT}	-0.702***	0.931**	0.690	0.391	0.616	0.721	0.424	0.689
	(-2.81)	(2.05)	(1.61)	(0.91)	(1.41)	(1.62)	(0.98)	(1.55)
λ_{SMB}	0.080	0.158	0.224**	0.224**	0.198*	0.183*	0.065	0.370**
	(0.77)	(1.50)	(2.04)	(1.97)	(1.74)	(1.71)	(0.62)	(2.50)
λ_{IVOL}		0.398***						
		(3.66)						
$\lambda^{3F}_{overlap}$			1.693***		1.658***			
			(4.72)		(4.58)			
λ^{3F}_{pure}				1.352***	-0.419			
				(3.28)	(-0.53)			
$\lambda^{IVOL}_{overlap}$						1.249***		2.753***
						(4.80)		(3.20)
λ^{IVOL}_{pure}							0.993***	-1.443
							(3.72)	(-1.64)
λ_0	1.428***	-0.355	-0.097	0.211	-0.006	-0.113	0.221	-0.101
	(5.90)	(-0.74)	(-0.22)	(0.46)	(-0.01)	(-0.24)	(0.49)	(-0.22)
Adjusted R^2	0.150	0.683	0.608	0.416	0.612	0.575	0.419	0.665

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the second-stage cross-sectional regressions by Fama and MacBeth (1973b) on the returns of 25 portfolios formed on size×value (Panel A), size×momentum (Panel B), and size×volatility (Panel C). Double sorted portfolios are retrieved from Kenneth R. French's data library.

I find that overlapping stocks are also priced in the cross-section of each characteristic test portfolio. The lambda estimate of the overlapping portfolio is more than one and highly significant whenever I add the returns of this portfolio to the model (Columns 3 and 5 in all panels). Whereas, the lambda estimate of pure factor stocks loses its significance and even flips the sign when I add the betas with respect to both portfolios to the model (from Column 4 to Column 5 in all panels). When decomposing each factor into the portfolio that contains overlapping stocks and the one that contains pure factor stocks, I draw the same conclusions. In all panels, the OLS adjusted R-squared shows that the goodness of fit of the model increases by a bit when I add the overlapping portfolio to the FF2 model. Although λ_0 is not close to zero in all regressions, they are all insignificant when the model includes the overlapping portfolios.

In brief, these findings imply that factor performance is driven by the performance of a subset of stocks that makes factor portfolios overlap. The overlapping stocks are priced in almost all cross sections. In contrast, when I add a combination of all pure factor stocks of the three factors, the pure factor stocks are not significantly priced.

Next, I study the diversification benefits of each portfolio. I test whether the overlapping portfolio adds to the mean-variance efficiency of the pure factor portfolio, and vice versa. I regress the returns of the test assets on the market, size factors, and the returns of the benchmark assets. If the test assets exactly price the benchmark assets, then the intercept alphas should equal zero. This is known as the Jensen measure. Under the null hypothesis, the benchmark assets span the test assets. If the Jensen measure is significantly different from zero, then adding the test assets to the benchmark improves the mean-variance efficiency. Table 6 reports the results of spanning tests.

Table 4.6: **Portfolio spanning tests**

	HML & MOM & IVOL		HML		MOM		IVOL	
Panel A: Dependent variable = returns to the overlapping portfolio								
	FF2	FF2 + pure factor	FF2	FF2 + pure HML	FF2	FF2 + pure MOM	FF2	FF2 + pure IVOL
α	1.003***	0.895***	0.819***	0.885***	1.147***	0.888***	1.040***	0.948***
	(7.11)	(6.84)	(5.25)	(6.15)	(6.31)	(6.48)	(7.24)	(7.04)
Panel B: Dependent variable = returns to the pure factor portfolio								
	FF2	FF2 + overlapping	FF2	FF2 + overlapping HML	FF2	FF2 + overlapping MOM	FF2	FF2 + overlapping IVOL
α	0.100**	-0.033	-0.153	-0.437***	0.398**	-0.368***	0.194*	-0.081
	(2.03)	(-0.70)	(-1.10)	(-3.35)	(2.16)	(-2.59)	(1.80)	(-0.78)

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the spanning test alphas and the corresponding t-statistics. Panel A shows the time-series regressions of the returns to the overlapping portfolios on the market, size factor (MKT + SMB := FF2), and the pure factor portfolios. Panel B shows the regressions of the returns to the pure factor portfolios on the FF2 and the overlapping portfolios. The sample is from 1952:07 to 2020:12 (822 monthly returns).

In panel A, the overlapping portfolios are the test assets while the counter-part pure factor portfolios are the benchmark assets. In the total factor strategy and in each factor portfolio, adding the pure factor portfolio leads to smaller alphas but all coefficients are still large and significantly positive, in comparison to when including only the market and the size factor. Therefore, the overlapping portfolio is not spanned by the pure factor portfolio.

In panel B, I run the same analysis but the overlapping portfolios are the benchmark assets while the counter-part pure factor portfolios are the test assets. Except for HML, all alphas in the regressions on the market and size factors are significantly positive. The smaller magnitude in comparison to the alphas in Panel A is consistent with my first result that the overlapping portfolios yield a higher Sharpe ratio than the pure factor portfolio, given relatively the same volatility. Furthermore, when adding the overlapping portfolio returns, the estimates for alphas become smaller and some even flip their signs (in the overlapping portfolio and the overlapping IVOL portfolio). I confirm that the overlapping portfolio spans the pure factor portfolios. However, the pure factor portfolios are not always spanned by overlapping portfolios. Overlapping stocks with multiple signals to their expected returns can improve the efficient frontier.

4.4 Potential explanations

In this final section, I seek to explain the difference in performance between overlapping stocks and pure factor stocks. This analysis is my work in progress. An advantage of my empirical approach is that I can identify exactly which stocks are in the overlapping and pure factor portfolios. Despite the vastly different performance, the overlapping and pure factor stocks are similar across many dimensions. These stocks have similar size (since I control for it in my portfolios), industry composition[13], and portfolio turnover. Importantly, overlapping stocks are not in the extreme tails of the original factor portfolios since the characteristic spreads of the overlapping stocks are

[13] Stocks in the manufacturing industry account for the most in both overlapping and pure factor portfolios.

not higher than those of the pure factor stocks. I look at several risk exposures that are either risk-based or mispricing.

4.4.1 Risk-based explanations

A natural explanation for the performance difference is consumption growth risk since it is an appropriate measure of systematic risk in the Consumption-based CAPM (e.g., Breeden, 1979; Mankiw and Shapiro, 1986).

I regress the excess returns on each leg of the overlapping and pure factor portfolios on the real consumption growth rate, per capita, CG_t. In each month, real consumption per capita is the total real consumption of nondurable goods and services, divided by the total population. All data to construct this measure are obtained from the Bureau of Economic Analysis. In the regressions, I also control for other macroeconomic variables including a recession indicator from NBER $I_{recession,t}$, term spread $TERM_t$ measured by 10Y government bond minus 3M Treasury Bills, and default premium DEF_t measured by BAA minus AAA rated corporate bonds.

$$R_t = a + bCG_t + cI_{recession,t} + dTERM_t + eDEF_t + \epsilon_t.$$

Panel B of Table 4.7 reports the estimates. I observe that, in the long leg, the overlapping portfolio has a significantly positive exposure to the real consumption growth per capita. A percent increase in the real consumption growth rate is associated with an average of 46.4 basis points increase in returns of overlapping stocks. Whereas, the pure factor portfolio has a lower exposure (14.7 basis points), and this coefficient is not significant. In other positions and other variables, I do not find evidence to explain the difference in performance of these two portfolios. Although significant, the recession indicator loads similarly on both overlapping and pure factor portfolios, which is reasonable because recessions affect the market as a whole so it affects both overlapping and pure factor stocks.

I find that the real consumption growth risk can explain some of the abnormal returns of overlapping stocks in the long leg. However, when I estimate the same expo-

sure to different measures such as Parker and Julliard (2005a)'s ultimate consumption risk, consumption growth rate on durable goods, Kroencke (2017)'s unfiltered NIPA consumption on quarterly rebalanced portfolios, I find no such significant result (see Table A3). Hence, consumption risk is not yet sufficient to explain the outperformance on a quarterly basis.

Table 4.7: **Economic risk exposures**

	Long		Short		Long–short	
	Overlapping	Pure factor	Overlapping	Pure factor	Overlapping	Pure factor
Panel A: Consumption risk and other macroeconomic risk exposures						
CG_t	0.464**	0.147	0.281	0.220	0.183	-0.073
	(2.04)	(0.62)	(0.74)	(0.85)	(0.67)	(-0.88)
$I_{Recession,t}$	-1.171**	-1.187**	-1.100	-1.269**	-0.070	0.082
	(-2.25)	(-2.18)	(-1.27)	(-2.15)	(-0.11)	(0.44)
$TERM_t$	0.143	0.227	0.101	0.189	0.043	0.038
	(1.08)	(1.63)	(0.46)	(1.25)	(0.27)	(0.78)
DEF_t	0.216	0.496	0.904	0.787*	-0.688	-0.291*
	(0.53)	(1.15)	(1.33)	(1.69)	(-1.40)	(-1.96)
a	0.492	-0.044	-0.676	-0.327	1.168**	0.284*
	(1.17)	(-0.10)	(-0.97)	(-0.69)	(2.33)	(1.87)
Observations	701	701	701	701	701	701
Adjusted R^2	0.012	0.009	-0.000	0.008	-0.001	0.001
Panel B: Liquidity measures						
LIQ_t	6.053*	8.061	6.738**	6.124*	-2.009	0.614
	(1.89)	(1.39)	(1.97)	(1.65)	(-0.43)	(0.47)
VIX_t	-0.183***	-0.195***	-0.186***	-0.182***	0.012	-0.004
	(-6.92)	(-4.07)	(-6.62)	(-5.98)	(0.33)	(-0.36)
TED_t	-0.271	-1.085	-0.184	-0.237	0.813	0.052
	(-0.56)	(-1.23)	(-0.35)	(-0.42)	(1.16)	(0.26)
a	4.932***	5.282***	4.875***	4.770***	-0.350	0.105
	(9.64)	(5.68)	(8.93)	(8.07)	(-0.47)	(0.50)
Observations	419	419	419	419	419	419
Adjusted R^2	0.177	0.083	0.164	0.136	-0.000	-0.006
Panel C: Exposures to investor sentiment						
S_{t-1}	-0.117	-0.149	-0.624**	-0.252	0.507***	0.103*
	(-0.72)	(-0.87)	(-2.34)	(-1.35)	(2.60)	(1.71)
a	0.767***	0.600***	0.116	0.533***	0.651***	0.067
	(4.75)	(3.51)	(0.44)	(2.86)	(3.34)	(1.11)
Observations	641	641	641	641	641	641
Adjusted R^2	-0.001	-0.000	0.007	0.001	0.009	0.003

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the estimated risk exposures to several variables. In all panels, the dependent variable R_t is the excess returns of the overlapping/ pure factor portfolios. In panel A, we regress the monthly excess returns on per capita real consumption growth (CG), NBER's recession dummy variable, term spread measured by 10Y government bond minus 3M Treasury Bills, and default premium measured by BAA minus AAA rated corporate bonds, from 1962:08 to 2020:12

$$R_t = a + bCG_t + cI_{recession,t} + dTERM_t + eDEF_t + \epsilon_t.$$

In panel B, we regress the monthly excess returns on Stambaugh, Yu, and Yuan (2012)'s liquidity factor (LIQ), the implied market volatility (VIX), and funding liquidity (TED spread), from 1986:02 to 2020:12

$$R_t = a + bLIQ_t + cVIX_t + dTED_t + \epsilon_t.$$

In panel C, we regress the monthly portfolio excess returns on the Baker and Wurgler (2007)'s investor sentiment index, lagged by one month, from 1965:07 to 2018:12

$$R_t = a + bS_{t-1} + \epsilon_t.$$

Besides, Panel B of Table 4.7 reports the estimates for risk exposure to several liquidity measures such as Stambaugh, Yu, and Yuan (2012)'s liquidity factor, the implied market volatility, and the funding liquidity measured by the TED spread. Generally, in the long legs, I observe that only consumption growth and market liquidity are significant. However, these variables affect the overlapping and the pure factor portfolios in a similar fashion, since the coefficients are similar in the sign and magnitude. When consumption grows and the market liquidity rises, stocks in the long leg of both the overlapping and pure factor portfolios receive higher positive returns.

4.4.2 Mispricing explanations

Next, I examine several mispricing-based explanations by looking at investor sentiment, short interest, and institutional ownership.

Investor sentiment

Investor sentiment seems to be a natural candidate to explain the short leg of overlapping stocks since intuitively investors are more sensitive to growth, past-loser, and high-volatility stocks. It is because these stocks might be more difficult to arbitrage and to value (Baker and Wurgler, 2007). In this analysis, I follow Stambaugh and Yuan (2017) by regressing the portfolio excess returns on the lag of investor sentiment as in Baker and Wurgler (2007)

$$R_t = a + bS_{t-1} + \epsilon_t$$

Panel C of Table 4.7 reports the estimates. In the short leg, I observe a strong result that investor sentiment loads significantly and more negatively on the overlapping portfolios: -62.4 basis points for the long-short returns of overlapping portfolios. Whereas, the investor sentiment does not load significantly on the pure factor portfolio. Consequently, the exposure of investor sentiment on overlapping stocks is five times that on pure factor stocks (50.77 basis points in the overlapping portfolio versus

10.3 basis points in the pure factor portfolio). My result is consistent with the result of Stambaugh, Yu, and Yuan (2012): High investor sentiment period follows by lower returns on the short leg. When investor sentiment is higher, stocks are more overpriced, so more overlap in the short leg.

Short interest

I look at the short leg of the portfolios to find whether the amount of short selling can explain the difference. I do so by constructing short interest at the portfolio level. Figure 4.5 plots the value-weighted average short interest in the short leg of the overlapping and the pure factor portfolios over time. Figure A2 reports equal-weighted values.

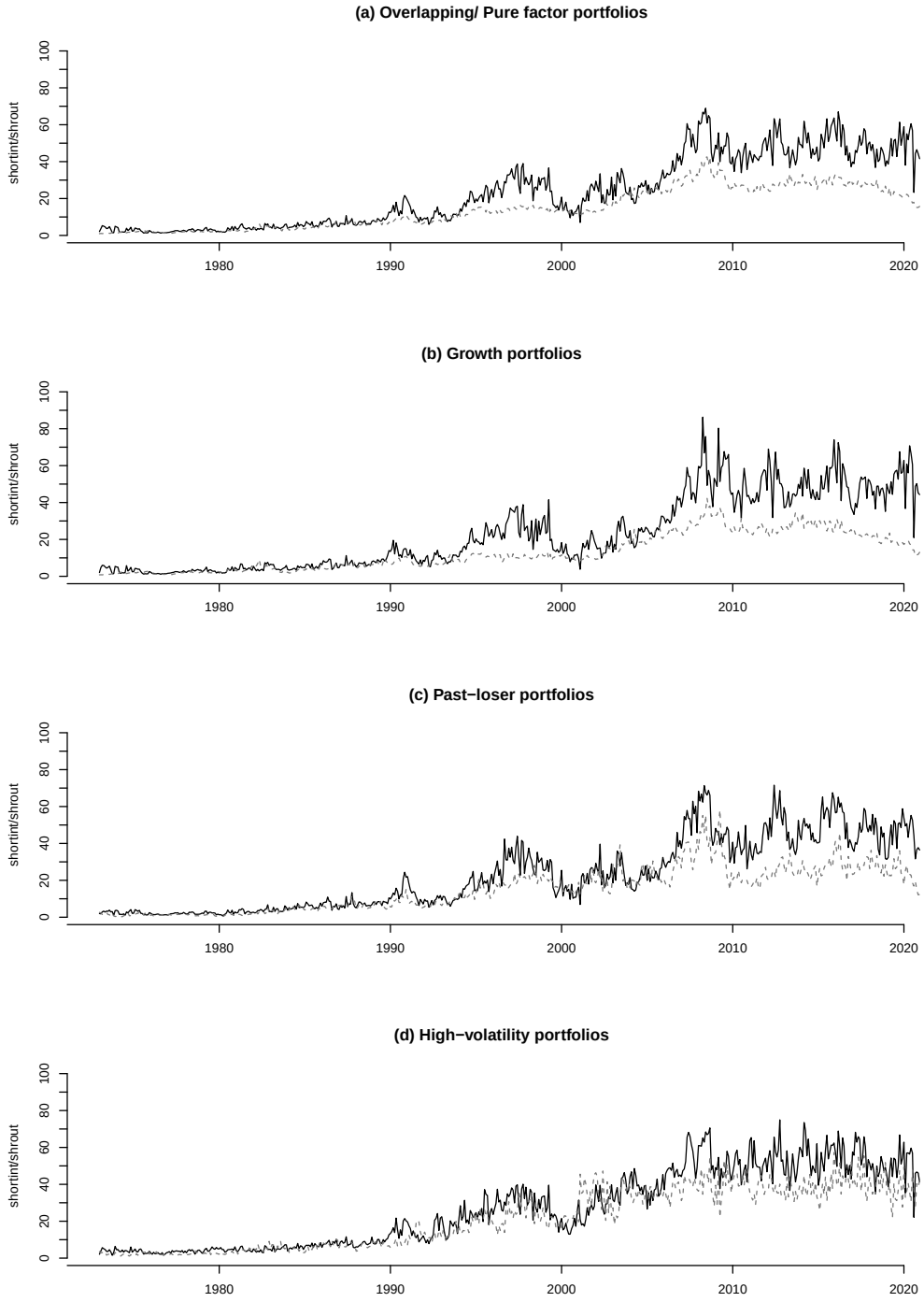
In all panels, I find that short interest is higher in the overlapping portfolios than in the pure factor portfolios. Especially in the last decade, the gap has become widen. This result implies that investor increases shorting selling in stocks that are overlapped in multiple factors. All these results somewhat explain the outperformance of overlapping stocks. Yet the outperformance of overlapping stocks comes from both legs.

Institutional ownership

Finally, I look into who owns overlapping stocks and pure factor stocks. I compute the institution ownership (IO) ratio, which is the total institutional ownership divided by adjusted total shares outstanding, using data from the SEC 13F. Figure 4.6 plots the value-weighted IO ratios for the long leg of the overlapping portfolio and the pure factor portfolio. Figure A3 reports equal-weighted IO ratios.

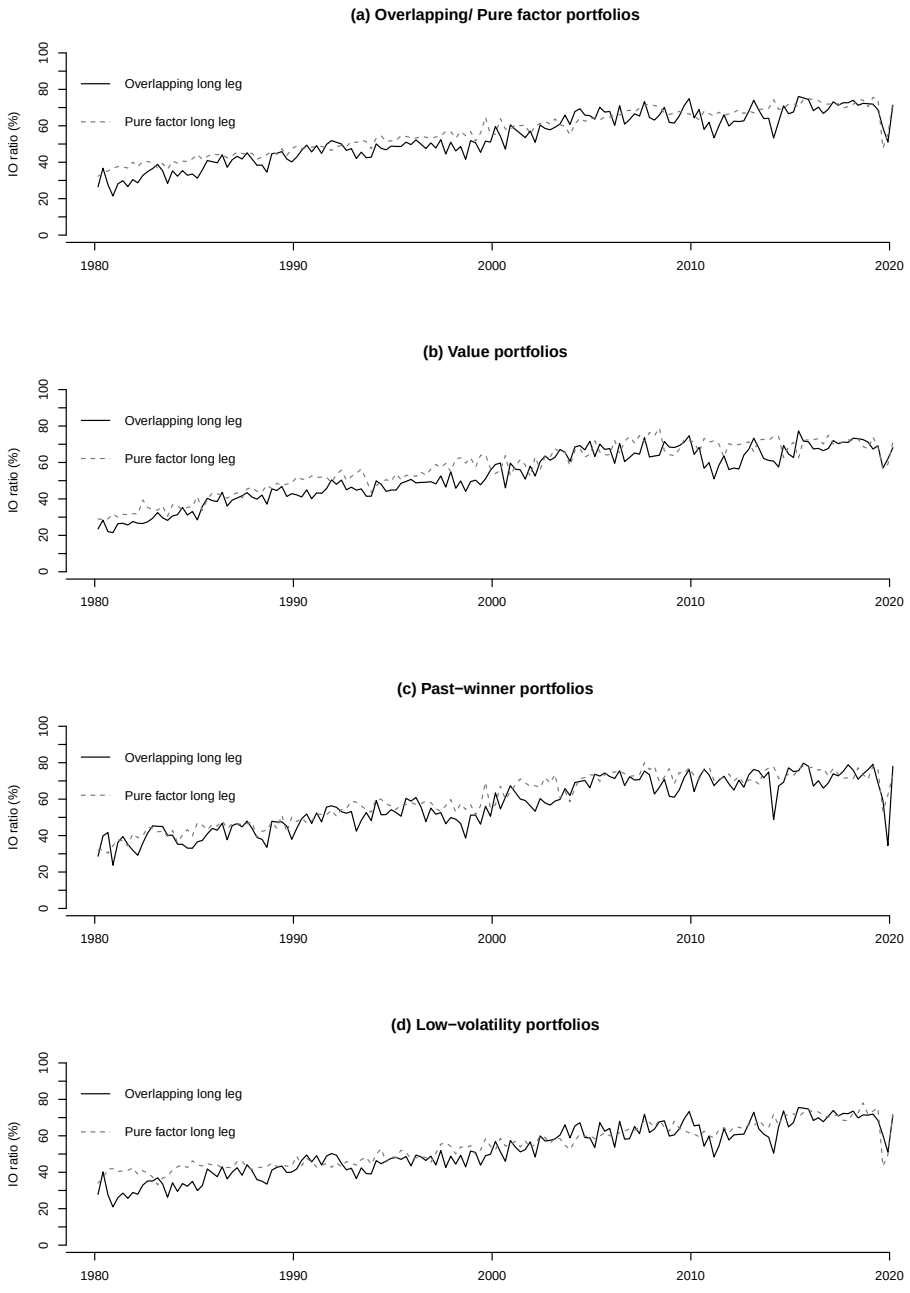
I find that institutional ownership has increased over the years in all portfolios. Until recently, the IO ratio has hit approximately 70%. Regardless of the trend, I find no difference in the ratio that institutional investors own in overlapping versus pure factor stocks. Perhaps there are times that the IO ratio in the overlapping portfolio is slightly lower than in the pure factor portfolio. However, this might be due to limits to arbitrage. As institutional investors are less willing and able to establish the overlapping

Figure 4.5: **Short interest in the short legs (value-weighted)**



Description. This figure plots the value-weighted short interest divided by shares outstanding in the short leg of multiple portfolios. I obtain short interest for each stock from Compustat Supplementary Short Interest File. The sample is from 1978:01 to 2020:12.

Figure 4.6: Institutional ownership ratio in the long legs (value-weighted)



This figure plots the value-weighted institutional ownership ratio in the long leg of several factor portfolios. The IO ratio, in percentage, is the total institutional ownership divided by the total shares outstanding (adjusted). I obtain the total IO from SEC 13F and the adjusted total shares outstanding from CRSP. I keep data at quarter dates end. The sample is from 1980:03 to 2020:03.

strategy, it is harder for them to trade these stocks. Therefore, it might be because investors find it more difficult to exploit these opportunities, although overlapping stocks outperform pure factor stocks in the long leg. In general, I find no difference in IO between the overlapping and the pure factor portfolios[14].

4.5 Practical implications and robustness checks

4.5.1 Alternative choice of factors

So far my multi-factor strategy focuses on HML, MOM and IVOL, controlling for size. I now show that my findings are robust when I use an alternative set of factors, consisting of investment (CMA), profitability (RMW), and IVOL. Fama and French (2015b) find that the HML factor is redundant with respect to the CMA and RMW factors. I therefore remove HML and consider CMA and RMA instead. Furthermore, I leave out MOM, as the previous analysis revealed an important role of overlapping momentum stocks. Table 4.8 shows the same setup of results for portfolio decomposition.

[14] I also find similar industry composition (unreported result).

Table 4.8: **Decomposing individual factor portfolios (CMA, RMW, and IVOL)**

	Total	Pure	Over-	Decomposition of overlapping portfolio			
	Factor	Factor	lapping	CMA & RMW	CMA & IVOL	RMW & IVOL	3F
Panel A: CMA							
Return (%)							
Long	1.071	1.046	1.074	1.146	0.991		1.248
Short	0.951	1.092	0.692	0.885	0.707		0.291
Long–short	0.120	-0.047	0.381	0.260	0.284		0.957
Log-market cap							
Long	4.566	4.351	5.231	5.151	5.274		5.660
Short	4.969	5.034	4.864	4.979	4.978		4.365
Long–short	-0.403	-0.683	0.368	0.172	0.296		1.295
Avg. weight	1.000	0.681	0.319	0.114	0.154		0.050
Characteristic spreads (long–short)							
Change in total assets	-0.496	-0.426	-0.665	-0.716	-0.489		-1.023
Operating profitability	-0.002	-0.263	0.431	1.189	-0.216		1.291
Residual volatility (%)	-0.179	0.141	-0.767	-0.013	-1.072		-1.394
Panel B: RMW							
Return (%)							
Long	1.095	1.114	1.093	1.146		1.033	1.248
Short	0.797	0.867	0.724	0.885		0.753	0.291
Long–short	0.299	0.247	0.369	0.260		0.279	0.957
Log-market cap							
Long	5.129	5.076	5.227	5.151		5.223	5.660
Short	4.504	4.861	4.102	4.979		3.693	4.365
Long–short	0.625	0.214	1.125	0.172		1.530	1.295
Avg. weight	1.000	0.620	0.380	0.116		0.214	0.051
Characteristic spreads (long–short)							
Change in total assets	-0.079	0.141	-0.420	-0.716		0.100	-1.023
Operating profitability	1.064	1.076	1.020	1.189		0.921	1.291
Residual volatility (%)	-0.214	0.154	-0.854	-0.013		-1.268	-1.394
Panel C: IVOL							
Return (%)							
Long	1.021	1.016	1.028		0.991	1.033	1.248
Short	0.856	1.055	0.649		0.707	0.753	0.291
Long–short	0.165	-0.040	0.379		0.284	0.279	0.957
Log-market cap							
Long	5.213	5.209	5.231		5.274	5.223	5.660
Short	4.316	4.407	4.231		4.978	3.693	4.365
Long–short	0.897	0.801	0.999		0.296	1.530	1.295
Avg. weight	1.000	0.582	0.418		0.155	0.213	0.050
Characteristic spreads (long–short)							
Change in total assets	-0.142	0.051	-0.362		-0.489	0.100	-1.023
Operating profitability	0.076	-0.162	0.446		-0.216	0.921	1.291
Residual volatility (%)	-1.056	-0.929	-1.186		-1.072	-1.268	-1.394

Description. We decompose each factor into a portfolio containing pure factor stocks and a portfolio containing overlapping stocks. This table reports the summary statistics of decomposed portfolios for each factor: CMA (Panel A), RMW (Panel B), and IVOL (Panel C), where 3F denotes CMA & RMW & IVOL. The sample is from 1952:07 to 2020:12 (822 monthly returns). Portfolios are value-weighted of stocks returns and rebalanced every month. The characteristic spreads are the long–short averages over time of the value-weighted characteristics of the stocks in each portfolio.

I find that the long-short return is the highest in the overlapping factor portfolio that contains stocks from the factor appearing in the same leg of one or two other factors. These stocks again have normal size and similar characteristic spreads to the total factor portfolio. Therefore, stocks in the overlapping factor portfolio are not small-cap stocks and they are not sorted into the extreme tails of the factors. CMA & RMW, CMA & IVOL, and RMW & IVOL stocks have similar long-short returns to each other, so there is not one category of overlapping stocks that drives the overall returns on overlapping stocks.

Table 4.9: **Fama–MacBeth cross-sectional regressions on alternative factors (CMA, RMW, and IVOL)**

Dependent variable: Cross-section of 25 size×investment, 25 size×profitability, and 25 size×volatility portfolios					
	(1)	(2)	(3)	(4)	(5)
λ_{MKT}	-0.759***	0.452	0.545	-0.246	0.534
	(-4.13)	(1.07)	(1.58)	(-0.88)	(1.55)
λ_{SMB}	0.126**	0.145**	0.151***	0.122**	0.159***
	(2.29)	(2.39)	(2.73)	(2.21)	(2.85)
λ_{CMA}		0.132			
		(1.25)			
λ_{RMW}		0.197*			
		(1.85)			
λ_{IVOL}		0.329***			
		(3.10)			
$\lambda^{3F}_{overlap}$			1.446***		1.489***
			(5.50)		(5.58)
λ^{3F}_{pure}				0.538***	-0.031
				(2.71)	(-0.13)
λ_0	1.445***	0.143	0.060	0.882***	0.082
	(7.91)	(0.33)	(0.17)	(2.98)	(0.23)
Adjusted R^2	0.283	0.642	0.599	0.369	0.610

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the second-stage cross-sectional regressions by Fama and MacBeth (1973b) on the returns of 75 portfolios formed on size×investment, size×profitability, and size×volatility. Double sorted portfolios are retrieved from Kenneth R. French's data library.

I also run the Fama-Macbeth regressions on the returns of 75 portfolios formed by sorting stocks on size and the characteristics of these alternative factors. when including the three individual total factors, I find that only IVOL is significantly priced (the lambda of RMW is marginally significant). However, the lambda of the overlapping portfolio is again positive and highly significant. The lambda of the pure factor portfolio loses its significance when I also add the overlapping portfolio to the model. In short, my findings are robust to this alternative choice of long-short factors.

4.5.2 Subsample periods

With respect to the portfolio performance of the multi-factor strategy consisting of HML, MOM, and IVOL, I replicate the results by splitting my main sample period into subsample periods. Panel A of Table A4 in the Online Appendix reports the portfolio statistics for the subsample period from July 1952 to December 1980. Panel B reports the statistics for the subsample period from January 1981 to December 2000. Panel C reports the statistics for the subsample period from January 2001 to December 2020.

In all subsample periods, the long-short overlapping portfolio has a higher mean excess return and a higher Sharpe ratio than the multi-factor portfolio. Whereas, the long-short pure factor portfolio shows the opposite. Factors are known to be mostly profitable during the 1980-2000 and die out after 2000. I observe the same trends since during the 1981-2000 period: the mean excess returns in all portfolios are high and the return distributions are positively skewed in all long-short portfolios. During 2001-2010, the mean excess returns are lower, but the positive CAPM alphas still survive. Therefore, my main findings are robust to different sample periods.

4.5.3 Sorting on NYSE breakpoints

In my main analysis, I sort stocks proportionally into the characteristic quintiles. By doing so, stocks are univariate distributed, so I can compare the fractions of overlapping stocks and pure factor stocks in my data with the random-sorting benchmark. For robustness check, I sort stocks into the characteristic quintiles by using NYSE break-

points as in the traditional Fama-French factor formation. Table A5 in the Online Appendix shows the average fraction of stocks in each quintile. I observe that the average fractions of overlap in this table are similar to the fractions when I sort stocks proportionally as shown in Table 4.1. The fraction of overlapping stocks is relatively higher in the short leg i.e. 20.4% compared to 13.2% in the main analysis. However, the performance of overlapping and pure factor portfolios is not different from the ones in my main analysis.

Table A6 in the Online Appendix reports the summary statistics of portfolios sorted using NYSE breakpoints. I observe little difference in the average returns, Sharpe ratio, and the average characteristic values in the overlapping and pure factor portfolios from when I sort stocks equally into the quintiles. Therefore, my previous conclusions regarding the performance of overlapping stocks versus pure factor stocks remain the same.

I also run the Fama-MacBeth regressions on the portfolios sorted by using NYSE breakpoints. Table A7 and Table A8 in the Online Appendix report the cross-sectional regressions. It still holds that the portfolio contains overlapping stocks that are priced in all cross sections. The estimated price of risk of this portfolio is significant and exceeds one, and the intercept of these regressions is insignificant and close to zero. In contrast, the the estimated price of risk of the pure factor portfolio when included alone has a significant and large intercept, while it is not priced when included with the overlapping portfolio. Of all, my results are robust to sorting by NYSE breakpoints.

4.5.4 Quarterly rebalancing

I perform another check for robustness as well as another practical implication for investors by rebalancing portfolios on a quarterly basis. For investors, quarterly rebalancing could be a more feasible strategy than monthly rebalancing in terms of lower trading costs. To construct the factor portfolios, I use the characteristic scores in the last month of the quarter. The portfolio returns are value-weighted and the portfolios are rebalanced every quarter.

Table A9 in the Online Appendix reports the portfolio statistics. All returns are expressed in percentages per quarter, so the values in this table can be divided by three if I want to compare them with the results of the monthly returns. I still observe the outperformance of the overlapping portfolio whose average long-short return is 1.39% per quarter while the pure factor portfolio only generates an average return of 0.19% per quarter. The Sharpe ratio of the overlapping portfolio is approximately triple that of the pure factor portfolio (0.35 versus 0.14). Again, the abnormal returns of the factors are driven by overlapping stocks within these factors. In comparison to investing in the multi-factor portfolio, the overlapping portfolio results in higher abnormal returns on average ($1.39\% > 0.75\%$ per quarter). As an implication for investors, focusing on overlapping stocks could be a feasible strategy.

Table A10 in the Online Appendix reports the Fama-Macbeth cross-sectional regressions on 75 test portfolios formed on size and value, momentum, and idiosyncratic volatility (Panel A) and 25 portfolios formed on size and net share issues and accruals (Panel B). Again, I observe that only the overlapping portfolio is significantly priced and the unpriced risk is not significant in the model that includes the returns on the overlapping portfolio and the pure factor portfolio (Column 5). The adjusted R-squared increases as I include the overlapping portfolio as a factor in the model. Hence, my results are robust to quarterly rebalancing. Although the sample size drops from 882 monthly returns to 274 quarterly returns and the trading costs are possibly lower on the quarterly rebalancing, I still find abnormal returns in the overlapping stocks.

4.6 Conclusion

Characteristics-based long-short factors are all constructed from the same universe of stocks. Already by chance alone, they can have some stocks in common. This paper shows that although the degree of overlap between HML, MOM, and IVOL is not excessive compared to a random-sorting benchmark, the overlap is important for investors and for asset pricing. I find strong evidence that these overlapping stocks drive the portfolio performance and the asset pricing performance of factors. This result also holds when I decompose each factor portfolio and when I consider CMA and RMW as alternative factors. Importantly, these overlapping stocks are not in the extreme tails of any factor, so I am not selecting outliers. In addition, I find that overlapping stocks are priced in all cross sections while pure factor stocks are not priced when added to the overlapping portfolio. Furthermore, including overlapping stocks improves the efficient frontier while it is not always true for pure factor stocks.

Despite the vast difference in performance, overlapping and pure factor stocks are surprisingly similar in many dimensions; in terms of size, sorting characteristics, turnover, industry composition, institutional ownership, and exposure to various macro-economic risks. While I have not found a single variable that can explain the return difference between the long-short portfolios, I find that in the long leg, overlapping stocks load more on ultimate consumption risk than pure factor stocks. In the short leg, overlapping stocks have higher short interest than pure factor stocks and they are more exposed to investor sentiment.

My paper provides some important implications for asset managers investing in smart-beta or factor-based products. My results suggest that rather than removing overlap between factors or hiring separate managers for each factor product, investors should focus more on these overlapping stocks as they are the dominant ones driving most of the factor returns.

A1 Appendix: Pricing decile portfolios

I regress the returns of decile portfolios sorted on book-to-market, prior 12-to-2 returns, and residual volatility against fmy different asset pricing models: CAPM with SMB (denoted as FF2), FF2 with each factor (HML, MOM, or IVOL), FF2 with returns on the overlapping factor portfolio (overlapping HML, overlapping MOM, or overlapping IVOL respectively), and FF2 with returns on the pure factor portfolio (pure HML, pure MOM, or pure IVOL respectively). Table A1 reports the mean excess returns, estimates for alphas and factor loadings in the top and bottom decile portfolios.

Table A1: **Pricing decile portfolios**

	Mean	FF2	FF2 + Factor		FF2 + Overlapping		FF2 + Pure factor	
Portfolio	return	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}_{Factor}$	$\hat{\alpha}$	$\hat{\beta}^{Factor}_{overlap}$	$\hat{\alpha}$	$\hat{\beta}^{Factor}_{pure}$
Panel A: Decile book-to-market portfolios (Factor = HML)								
High	0.889***	0.150	-0.165**	0.909***	-0.094	0.297***	0.232***	0.542***
	(4.26)	(1.35)	(-2.35)	(35.10)	(-0.92)	(13.20)	(2.88)	(26.78)
Low	0.597***	-0.075	0.079*	-0.446***	0.112**	-0.229***	-0.128***	-0.348***
	(3.46)	(-1.23)	(1.76)	(-26.89)	(2.21)	(-20.43)	(-3.41)	(-36.86)
High–Low	0.292*	0.225	-0.244***	1.355***	-0.206	0.526***	0.361***	0.890***
	(1.86)	(1.48)	(-2.95)	(44.52)	(-1.58)	(18.35)	(4.11)	(40.52)
Average $\lvert\hat{\alpha}\rvert$		0.116	0.083		0.076		0.116	
GRS *F*-stat		2.376	2.980		1.884		5.938	
GRS *p*-value		0.009	0.001		0.044		0.000	
Panel B: Decile 12-to-2 return portfolios (Factor = MOM)								
Up	1.224***	0.479***	0.028	0.591***	0.081	0.347***	0.298***	0.456***
	(5.94)	(4.63)	(0.43)	(36.27)	(0.96)	(21.96)	(4.89)	(39.49)
Down	-0.014	-0.956***	-0.238**	-0.940***	-0.200*	-0.659***	-0.729***	-0.569***
	(-0.05)	(-5.89)	(-2.38)	(-37.70)	(-1.78)	(-31.22)	(-5.86)	(-24.15)
Up–Down	1.238***	1.435***	0.266***	1.530***	0.281*	1.006***	1.027***	1.025***
	(5.26)	(6.16)	(2.62)	(60.48)	(1.90)	(36.17)	(7.48)	(39.41)
Average $\lvert\hat{\alpha}\rvert$		0.364	0.134		0.121		0.258	
GRS *F*-stat		5.939	4.366		3.171		6.713	
GRS *p*-value		0.000	0.000		0.001		0.000	
Panel C: Decile idiosyncratic volatility portfolios (Factor = IVOL)								
Lowvol	0.585***	0.200***	-0.002	0.342***	0.012	0.204***	0.145***	0.281***
	(4.22)	(3.65)	(-0.04)	(20.68)	(0.28)	(19.36)	(3.37)	(20.70)
Highvol	0.046	-0.943***	-0.348***	-1.009***	-0.479***	-0.503***	-0.852***	-0.463***
	(0.14)	(-7.19)	(-4.01)	(-31.01)	(-4.52)	(-20.36)	(-7.17)	(-12.36)
Lowvol–Highvol	0.539**	1.143***	0.346***	1.351***	0.492***	0.707***	0.997***	0.744***
	(1.98)	(7.16)	(3.91)	(40.69)	(4.26)	(26.31)	(7.55)	(17.87)
Average $\lvert\hat{\alpha}\rvert$		0.230	0.143		0.165		0.251	
GRS *F*-stat		8.006	5.274		5.459		7.874	
GRS *p*-value		0.000	0.000		0.000		0.000	
Observations		822	822		822		822	

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the time-series regressions of each set of the long and the short characteristic-sorted portfolios on CAPM with size (FF2 := CAPM + SMB), FF2 with factor returns, FF2 with returns on the long–short overlapping portfolio within the factor, and FF2 with returns on the long–short pure factor portfolio within the factor. The sample is 1952:07 to 2020:12 (822 monthly returns). Decile portfolios are retrieved from Kenneth R. French's data library. The first column displays the mean excess return of each portfolio.

On average, the long-short strategy earns positive excess returns. With portfolios sorted on book-to-market, the high-minus-low portfolio yields an average return of 0.29% per month. With portfolios sorted on prior returns, the loser portfolio generates an insignificant negative return of -0.01% per month. The winner portfolio generates a significant positive return of 1.22% per month. Thus, the winner-minus-loser portfolio earns an average return of 1.24% per month. With portfolios sorted on residual volatility, the difference between the low-volatility and high-volatility portfolios is 0.54% per month.

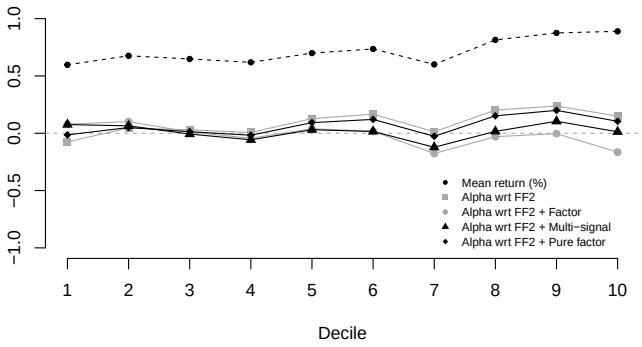
Adding each factor to the benchmark model (FF2 + Factor) significantly reduces the alpha. With portfolios sorted on book-to-market, the estimated alpha for the high-minus-low strategy reduces from 0.23% to -0.24%. The average absolute alpha over ten portfolios reduces from 11.6 to 8.3 basis points. With portfolios sorted on prior returns, the estimated alpha of the winner-minus-loser strategy reduces from 1.44% to 0.27% and the average absolute alpha reduces from 36.4 to 13.4 basis points. With portfolios sorted on residual volatility, the estimated alpha of the lowvol-minus-highvol strategy reduces from 1.14% to 0.35% and the average absolute alpha reduces from 23.0 to 14.3 basis points.

Consistently, adding the overlapping factor performance to the benchmark (FF2 + overlapping) improves pricing the decile portfolios. In the HML strategy, alpha becomes insignificant and reduces from 0.23 to -0.21. In the MOM strategy, alpha is less significant and it reduces by more than half of the benchmark ($1 - 0.28/1.44$). In the IVOL strategy, alpha also reduces by more than half of the benchmark ($1 - 0.49/1.14$). These results imply that less returns remain unexplained in the model after including the overlapping factor portfolio as a factor. I find that including the overlapping factor performance improves the alpha of the long-short strategy just as well as including each individual factor.

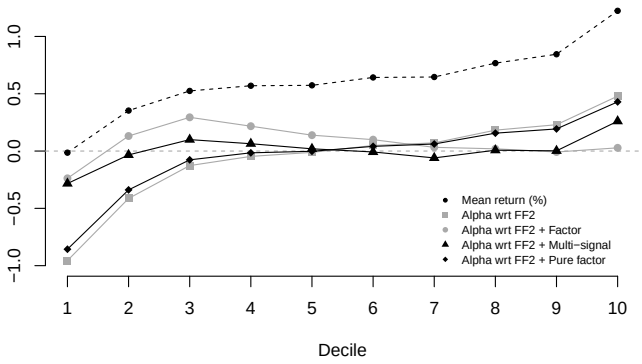
I also find that the overlapping factor portfolio prices each individual portfolio more modestly than the factor itself. Figure A1 plots the mean return and the estimated alpha of each decile portfolio. Across the deciles, the alpha in the FF2 + overlapping model

Figure A1: **Time-series alphas in decile portfolios**

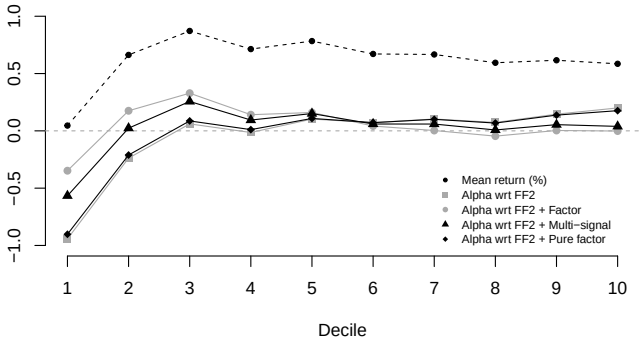
(a) *Book-to-market portfolios (Factor = HML)*



(b) *12-to-2 return portfolios (Factor = MOM)*



(c) *Residual volatility portfolios (Factor = IVOL)*



Description. This figure plots the mean returns and time-series alphas of decile portfolios formed on HML (Panel A), MOM (Panel B), and IVOL (Panel C). The time-series regressions of each set of decile portfolios are on CAPM with size (FF2 := CAPM + SMB), FF2 with factor returns, FF2 with returns on the overlapping factor portfolio, and FF2 with returns on the pure factor portfolio. The sample is from 1952:07 to 2020:12 (822 monthly returns).

closely follows the alpha in the FF2 + Factor model. Yet there is less overshooting in the intermediate portfolios, especially from Decile 2 to Decile 4 of prior return and residual volatility portfolios. I also find that FF2 + overlapping only yields significant alphas in the long and short portfolios and not in any intermediate decile portfolio.

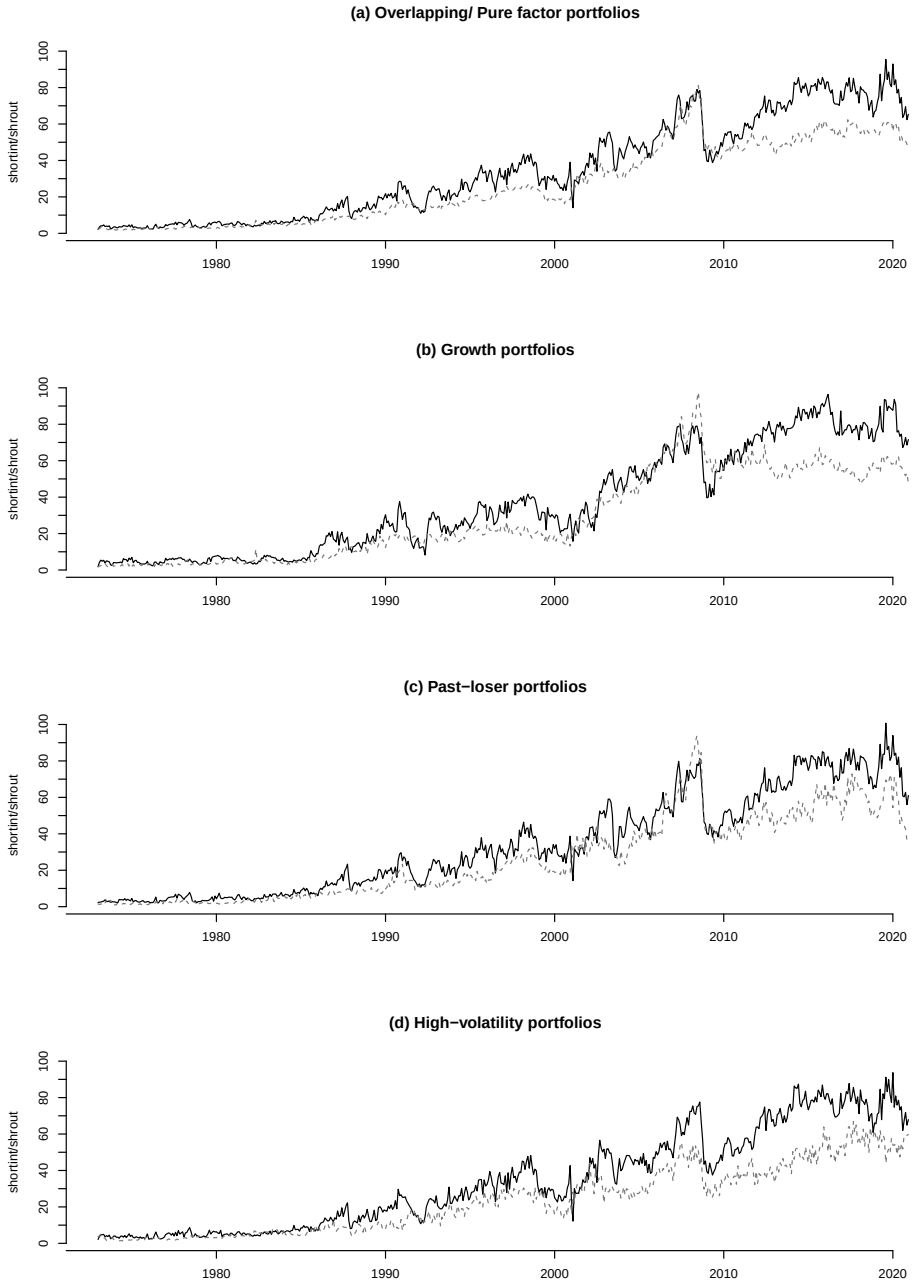
Formally, Table A1 reports the test for joint significance of alphas, introduced by Gibbons, Ross, and Shanken (1989). In all panels, I find that the GRS F-statistic to test the null hypothesis that alphas across all decile portfolios are zero remain low. Therefore, including the overlapping factor performance in the benchmark model reduces the significance of alpha across decile portfolios.

I find opposite conclusions when including the pure factor performance in the benchmark (FF2 + Pure factor). Adding this portfolio as a factor does not improve pricing the decile portfolios. Although the estimated betas of the pure factor strategy are significantly large in all three panels, the estimated alphas are not much different from the benchmark. The average absolute alphas and the GRS-test statistics are also just similar to the benchmark. Therefore, adding the pure factor returns as a factor to the benchmark model has no contribution in pricing the characteristic-sorted portfolios.

So far, I find that the overlapping factor portfolio is powerful in pricing single-sorted portfolios and it outperforms the factor in pricing intermediate portfolios. In contrast, the pure factor portfolio contributes no benefit to the cross-sectional multi-factor strategies, in comparison to the benchmark. The overlapping factor portfolio contains fewer stocks than the total factor portfolios, but this portfolio has approximately the same explanatory power as all three factors when included as a priced risk factor. The pure factor portfolio that contains stocks sorted only to that factor does not perform differently than the benchmark.

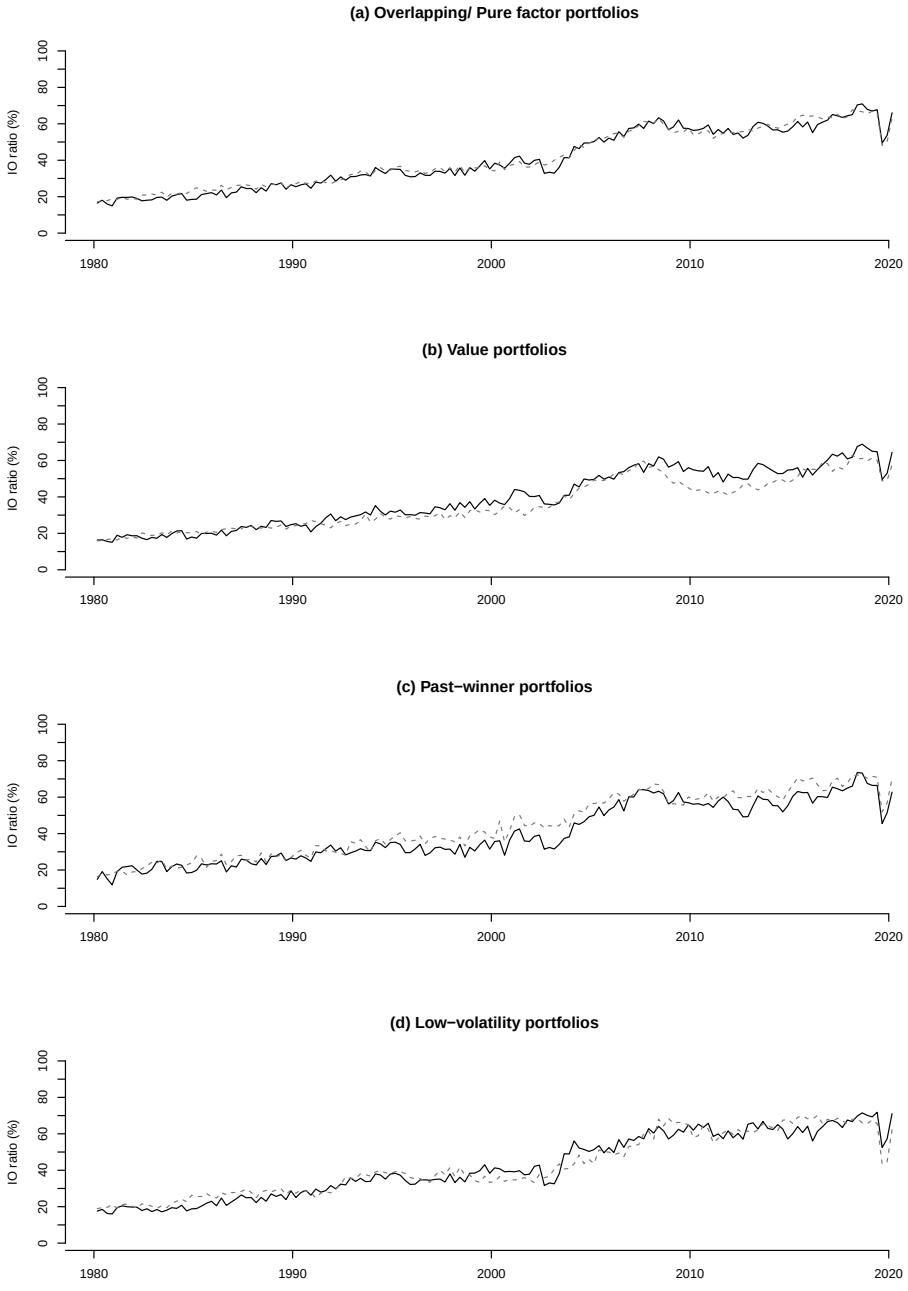
A2 Appendix: Additional results

Figure A2: **Short interest in the short legs (equal-weighted)**



Description. This figure plots the equal-weighted short interest divided by shares outstanding in the short leg of multiple portfolios. I obtain short interest for each stock from Compustat Supplementary Short Interest File. The sample is from 1952:07 to 2020:12 (822 monthly returns).

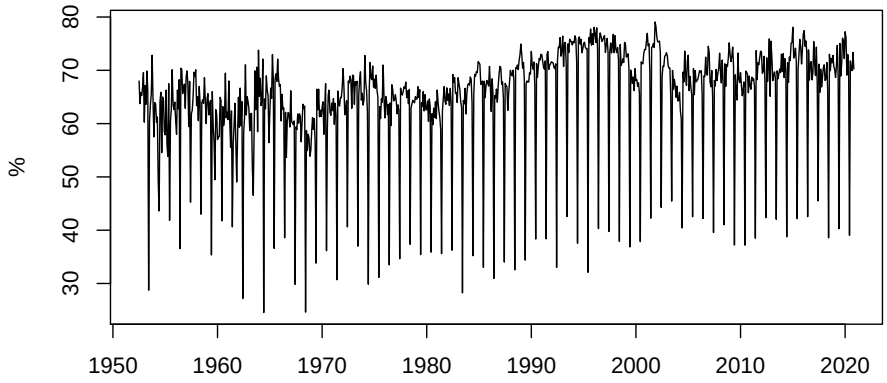
Figure A3: **Institutional ownership ratio in the long legs (equal-weighted)**



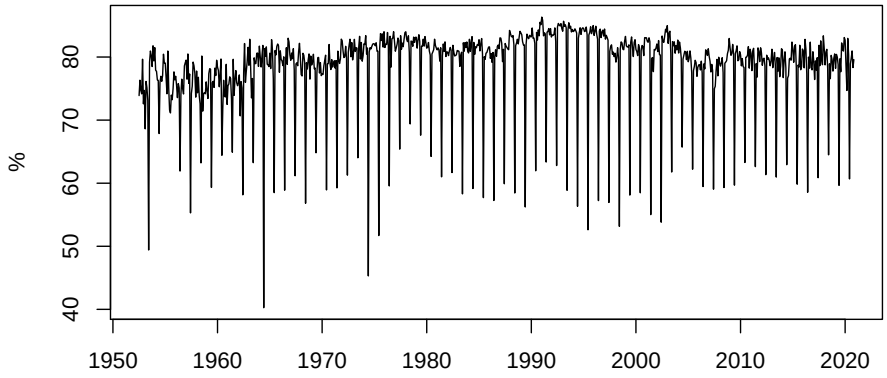
Description. This figure plots the equal-weighted institutional ownership ratio in the long leg of multiple factor portfolios. The IO ratio, in percentage, is the total institutional ownership divided by the total shares outstanding (adjusted). I obtain the total IO from SEC 13F and the adjusted total shares outstanding from CRSP. I keep data at quarter dates end. The sample is from 1980:03 to 2020:03.

Figure A4: **Proportion of stocks staying in the portfolio next month in the CMA-RMW-IVOL strategy**

(a) *Overlapping portfolio*



(b) *Pure factor portfolio*



Description. This figure plots the proportions of stocks in the overlapping and the pure factor portfolios that stays in the same portfolio in the following month. The sample is from 1952:07 to 2020:12 (822 monthly returns). The portfolio is rebalanced every month.

Table A2: **Correlations of firm characteristics and factor returns**

Panel A: Correlations between characteristics

	Mkt. cap	B/M	Op. prof.	Δ total assets	12-to-2 return	Resid vol.
Mkt. cap	1.00					
B/M	-0.09	1.00				
Op. prof.	0.02	-0.07	1.00			
Δ total assets	-0.00	-0.09	0.04	1.00		
12-to-2 return	0.03	-0.25	0.02	-0.02	1.00	
Resid vol.	-0.15	0.18	-0.08	0.02	-0.10	1.00

Panel B: Correlations between factor returns

	MKT	HML	RMW	CMA	MOM	IVOL
MKT	1.00					
HML	-0.14	1.00				
RMW	-0.20	-0.28	1.00			
CMA	-0.34	0.65	-0.06	1.00		
MOM	-0.10	-0.22	0.17	-0.08	1.00	
IVOL	-0.57	0.25	0.36	0.48	0.13	1.00

Description. This table shows the correlations between stock characteristics and between the factor returns from 1952:07 to 2020:12 (822 months). We construct each factor using five value-weighted portfolios from characteristic, controlling for size. The monthly size breakpoints are the 30th and 70th NYSE percentile. Stocks are equally divided into quintiles of characteristics. Panel A reports the average correlation between stock characteristics over time. Panel B shows the correlation between factor portfolio returns over time. Factor returns are the returns of the long-short portfolio (Q5-Q1).

Table A3: **Consumption risk exposures**

	Long		Short		Long–short	
	Overlapping	Pure factor	Overlapping	Pure factor	Overlapping	Pure factor
Panel A: Exposure to quarterly real consumption growth						
CG	0.159	-0.007	0.101	-0.011	0.166	0.112
	(0.59)	(-0.02)	(0.40)	(-0.04)	(0.63)	(1.29)
Constant	0.876***	-0.434	0.510**	0.374	1.310***	0.136
	(3.14)	(-1.12)	(1.99)	(1.34)	(4.85)	(1.53)
Observations	274	274	274	274	274	274
Adjusted R^2	-0.002	-0.004	-0.003	-0.004	-0.002	0.002
Panel B: Exposure to ultimate consumption risk						
UC	-0.032	-0.077	-0.058	-0.058	0.046	-0.000
	(-0.41)	(-0.70)	(-0.80)	(-0.72)	(0.58)	(-0.01)
Constant	1.220**	-0.011	0.973**	0.727	1.231**	0.247
	(2.41)	(-0.02)	(2.04)	(1.38)	(2.36)	(1.46)
Observations	262	262	262	262	262	262
Adjusted R^2	-0.003	-0.002	-0.001	-0.002	-0.003	-0.004
Panel C: Exposure to ultimate consumption risk and consumption growth rate on durable goods						
UC	-0.082	-0.143	-0.094	-0.108	0.061	0.014
	(-1.05)	(-1.27)	(-1.26)	(-1.32)	(0.75)	(0.53)
CG (DG)	1.254***	1.638**	0.879*	1.234**	-0.384	-0.355**
	(2.63)	(2.40)	(1.95)	(2.49)	(-0.78)	(-2.23)
Constant	0.864*	-0.475	0.724	0.377	1.340**	0.347**
	(1.67)	(-0.64)	(1.47)	(0.70)	(2.48)	(2.00)
Observations	262	262	262	262	262	262
Adjusted R^2	0.019	0.016	0.009	0.018	-0.004	0.011
Panel D: Exposure to unfiltered NIPA consumption (nondurables and services)						
Unf. NIPA (NG&S)	0.137	0.229	0.110	0.152	-0.093	-0.042
	(1.28)	(1.52)	(1.09)	(1.37)	(-0.85)	(-1.18)
Constant	0.711**	-0.927**	0.371	0.060	1.637***	0.311***
	(2.25)	(-2.07)	(1.24)	(0.18)	(5.09)	(2.97)
Observations	266	266	266	266	266	266
Adjusted R^2	0.002	0.005	0.001	0.003	-0.001	0.001
Panel E: Exposure to unfiltered NIPA consumption (nondurables only)						
Unf. NIPA (NG)	0.137	0.229	0.110	0.152	-0.093	-0.042
	(1.28)	(1.52)	(1.09)	(1.37)	(-0.85)	(-1.18)
Constant	0.711**	-0.927**	0.371	0.060	1.637***	0.311***
	(2.25)	(-2.07)	(1.24)	(0.18)	(5.09)	(2.97)
Observations	266	266	266	266	266	266
Adjusted R^2	0.002	0.005	0.001	0.003	-0.001	0.001

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the estimated risk exposures to several variables. In all panels, the dependent variable is the excess returns of the overlapping/ pure factor portfolios. CG is the quarterly real consumption growth rate on nodurables and services. UC is the Parker and Julliard (2005a)'s quarterly ultimate consumption risk. CG (DG) is the quarterly real consumption growth rate on durable goods. Unfiltered NIPA consumption measured are obtained from Kroencke (2017). The sample is from 1952:Q3 to 2020:Q4.

Table A4: **Portfolio performance on different subsample periods**

	Return				Sharpe	Distribution				CAPM		Average characteristic			
Portfolio	Mean (%)	Min. (%)	Max. (%)	Volatility (%)	ratio	Positive (%)	Negative (%)	Skewness	Kurtorsis	$\hat{\alpha}$	$\hat{\beta}_{MKT}$	Log-mkt cap	Value	Momentum (%)	Resid. vol. (%)
Panel A: Subsample 1952:07 - 1980:12															
Market	0.583	-12.685	16.530	4.099	0.142	58.772	41.228	-0.210	1.097			4.297	0.608	14.890	1.129
Overlapping portfolio															
Long	0.937	-18.341	14.176	4.183	0.224	60.819	39.181	-0.251	1.335	0.403***	0.915***	4.390	0.836	30.835	1.021
Short	0.151	-17.907	23.175	5.613	0.027	50.292	49.708	0.117	1.472	-0.587***	1.265***	4.086	0.443	1.712	1.534
Long–short	0.786	-14.331	10.492	3.525	0.223	59.649	40.351	-0.324	1.243	0.990***	-0.350***	0.304	0.392	29.123	-0.513
Pure factor portfolio (across all factors)															
Long	0.625	-13.652	12.623	3.953	0.158	60.526	39.474	-0.395	0.843	0.072*	0.949***	4.327	0.616	18.706	1.044
Short	0.562	-15.807	19.802	4.395	0.128	57.018	42.982	-0.216	1.444	-0.044	1.040***	4.302	0.473	14.067	1.163
Long–short	0.063	-7.853	4.169	1.279	0.049	51.170	48.830	-0.764	5.373	0.116*	-0.091***	0.025	0.143	4.639	-0.119
EW. multi-factor	0.383	-8.508	6.381	1.819	0.210	64.327	35.673	-0.489	3.910	0.491***	-0.186***	0.128	0.245	16.675	-0.277
HML	0.176	-13.505	12.394	3.247	0.054	54.094	45.906	0.066	1.740	0.239*	-0.107**	-0.581	1.015	0.905	0.059
MOM	0.969	-14.128	19.642	3.941	0.246	62.573	37.427	-0.182	2.410	0.963***	0.403	-0.259	57.056	-0.029	
IVOL	0.002	-13.661	11.557	3.272	0.001	53.801	46.199	-0.170	1.643	0.271	-0.461***	0.564	-0.023	-7.936	-0.860

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the summary statistics of overlapping/ pure factor portfolio value-weighted returns, in excess of the risk-free rate, comparing to the market and factor returns. Market is the excess returns of the CRSP U.S. total market index. The risk-free rate is from Kenneth R. French's Data Library. The last four columns present the averages over time of the value-weighted characteristic of stocks in each portfolio. The EW. multi-factor portfolio is the equal-weighted combination of the three factors. Note that the pure factor portfolio contains pure HML, pure MOM, and pure IVOL stocks.

Table A4: **Portfolio statistics on different subsample periods (continued)**

	Return				Sharpe	Distribution				CAPM		Average characteristic			
Portfolio	Mean (%)	Min. (%)	Max. (%)	Volatility (%)	ratio	Positive (%)	Negative (%)	Skewness	Kurtorsis	$\hat{\alpha}$	$\hat{\beta}_{MKT}$	Log-mkt cap	Value	Momentum (%)	Resid. vol. (%)
Panel B: Subsample 1981:01 - 2000:12															
Market	0.751	-22.883	12.524	4.371	0.172	60.417	39.583	-0.779	3.516			4.458	0.557	25.067	1.555
Overlapping portfolio															
Long	1.210	-20.860	12.566	4.023	0.301	63.750	36.250	-0.655	3.493	0.611***	0.797***	4.486	0.826	38.909	1.233
Short	0.206	-27.080	15.882	6.226	0.033	52.500	47.500	-0.397	1.510	-0.761***	1.288***	3.746	0.409	15.744	2.305
Long–short	1.004	-8.892	19.650	4.411	0.228	60.833	39.167	0.334	1.193	1.372***	-0.490***	0.740	0.417	23.164	-1.071
Pure factor portfolio (across all factors)															
Long	0.957	-22.175	12.507	4.544	0.211	61.250	38.750	-0.610	2.801	0.200***	1.008***	4.513	0.569	38.001	1.504
Short	0.843	-21.057	12.478	4.692	0.180	60.000	40.000	-0.506	1.657	0.059	1.044***	4.380	0.443	24.933	1.650
Long–short	0.114	-3.025	7.968	1.480	0.077	48.750	51.250	1.036	3.386	0.142	-0.036*	0.133	0.126	13.068	-0.146
EW. multi-factor	0.420	-5.729	7.993	2.284	0.184	60.000	40.000	0.096	0.688	0.583***	-0.217***	0.450	0.240	19.415	-0.549
HML	0.316	-15.148	13.736	3.902	0.081	53.750	46.250	0.105	1.518	0.569**	-0.337***	-0.620	0.935	-8.975	-0.268
MOM	0.498	-15.502	25.824	5.083	0.098	56.667	43.333	0.363	4.312	0.343	0.207***	0.956	-0.316	81.215	-0.130
IVOL	0.446	-15.752	25.139	4.366	0.102	57.083	42.917	0.527	4.746	0.836***	-0.519***	1.012	0.100	-13.996	-1.251

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: **Portfolio statistics on different subsample periods (continued)**

	Return				Sharpe	Distribution				CAPM		Average characteristic			
Portfolio	Mean (%)	Min. (%)	Max. (%)	Volatility (%)	ratio	Positive (%)	Negative (%)	Skewness	Kurtorsis	$\hat{\alpha}$	$\hat{\beta}_{MKT}$	Log-mkt cap	Value	Momentum (%)	Resid. vol. (%)
Panel C: Subsample 2001:01 - 2020:12															
Market	0.647	-17.174	13.660	4.480	0.145	64.167	35.833	-0.517	1.154			6.132	0.419	14.323	1.302
Overlapping portfolio															
Long	0.721	-19.639	11.730	4.201	0.172	61.667	38.333	-0.919	2.362	0.193	0.817***	6.345	0.655	32.522	1.095
Short	0.599	-25.000	30.384	7.964	0.075	54.583	45.417	0.078	2.240	-0.425*	1.581***	5.401	0.428	-1.130	2.013
Long–short	0.122	-31.374	25.001	6.217	0.020	55.000	45.000	-0.816	5.952	0.617*	-0.764***	0.944	0.228	33.652	-0.917
Pure factor portfolio (across all factors)															
Short	0.653	-17.076	15.688	4.289	0.152	63.333	36.667	-0.443	1.667	0.048	0.934***	6.177	0.453	21.169	1.242
Short	0.686	-19.194	15.692	4.785	0.143	62.917	37.083	-0.505	1.545	0.009	1.045***	6.157	0.326	12.708	1.389
Long–short	-0.033	-7.880	11.717	1.616	-0.020	48.750	51.250	0.743	13.851	0.039	-0.111***	0.020	0.127	8.461	-0.147
EW. multi-factor	0.019	-21.225	21.032	3.657	0.005	50.833	49.167	-0.370	10.959	0.288	-0.415***	0.474	0.150	21.382	-0.483
HML	-0.221	-14.939	19.781	3.610	-0.061	45.833	54.167	0.337	4.686	-0.260	0.060	-0.736	0.861	-2.675	-0.082
MOM	0.162	-35.169	16.686	6.216	0.026	52.917	47.083	-1.506	7.924	0.528	-0.565***	0.901	-0.330	70.903	-0.226
IVOL	0.116	-19.845	26.885	5.291	0.022	47.083	52.917	0.315	4.587	0.595**	-0.740***	1.258	-0.081	-4.082	-1.141

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: **Fraction of pure factor and overlapping stocks based on NYSE breakpoints**

	Long leg	Short leg
Panel A: In the universe of stocks		
Pure factor	0.381	0.346
Pure HML	0.105	0.137
Pure MOM	0.156	0.082
Pure IVOL	0.120	0.127
Overlapping	0.087	0.204
HML & MOM	0.031	0.034
HML & IVOL	0.024	0.061
MOM & IVOL	0.026	0.065
HML & MOM & IVOL	0.006	0.044
Panel B: In each factor portfolio		
HML	1.000	1.000
Pure HML	0.634	0.495
Overlapping HML	0.366	0.505
HML & MOM	0.188	0.122
HML & IVOL	0.145	0.222
HML & MOM & IVOL	0.033	0.161
MOM	1.000	1.000
Pure MOM	0.713	0.365
Overlapping MOM	0.287	0.635
HML & MOM	0.142	0.149
MOM & IVOL	0.120	0.289
HML & MOM & IVOL	0.025	0.197
IVOL	1.000	1.000
Pure IVOL	0.683	0.426
Overlapping IVOL	0.317	0.574
HML & IVOL	0.137	0.206
MOM & IVOL	0.149	0.219
HML & MOM & IVOL	0.031	0.149

Description. This table reports the average fraction of pure factor/ overlapping stocks in the data versus the benchmark fraction in the strategy that invests equally in HML, MOM, and IVOL. The benchmark fractions are given by $f_S(k) = \binom{S}{k}\left(\frac{1}{5}\right)^k\left(\frac{4}{5}\right)^{S-k}, k = 0,\ldots,S$ for $S = 3$. The benchmark is the same for the long leg and the short leg. The empirical fractions are the average of fractions in each month per portfolio.

Table A6: **Portfolio performance with sorting based on NYSE breakpoints**

	Return				Sharpe	Distribution				CAPM		Average characteristic			
Portfolio	Mean (%)	Min. (%)	Max. (%)	Volatility (%)	ratio	Positive (%)	Negative (%)	Skewness	Kurtorsis	$\hat{\alpha}$	$\hat{\beta}_{MKT}$	Log-mkt cap	B/M	Prior-ret (%)	Resid. vol. (%)
Market	0.651	-22.883	16.530	4.288	0.152	60.827	39.173	-0.485	1.881			4.880	0.538	17.696	1.304
Overlapping portfolio															
Long	0.944	-20.900	13.934	4.173	0.226	61.922	38.078	-0.521	2.123	0.389***	0.853***	5.035	0.770	34.108	1.094
Short	0.325	-27.572	29.855	6.291	0.052	52.068	47.932	-0.010	2.139	-0.547***	1.340***	4.273	0.428	6.662	1.895
Long–short	0.619	-28.638	23.481	4.435	0.140	58.516	41.484	-0.582	5.420	0.936***	-0.487***	0.762	0.342	27.446	-0.800
Pure factor portfolio (across all factors)															
Long	0.734	-22.879	15.179	4.244	0.173	61.436	38.564	-0.537	1.997	0.107***	0.965***	4.959	0.551	25.797	1.240
Short	0.728	-21.602	20.093	4.481	0.162	59.854	40.146	-0.386	1.599	0.065*	1.018***	4.866	0.418	17.890	1.340
Long–short	0.007	-8.107	9.995	1.317	0.005	50.000	50.000	0.115	6.647	0.042	-0.054***	0.093	0.133	7.906	-0.101
EW. multi-factor	0.276	-18.610	16.991	2.453	0.112	60.097	39.903	-0.742	10.404	0.441***	-0.254***	0.479	0.219	18.295	-0.423
HML	0.094	-15.486	16.656	3.446	0.027	52.068	47.932	0.046	2.311	0.165	-0.109***	-0.461	0.958	-2.658	-0.085
MOM	0.580	-34.247	20.597	4.691	0.124	58.637	41.363	-0.827	6.531	0.650***	-0.109***	0.754	-0.293	65.780	-0.124
IVOL	0.154	-19.823	21.667	4.076	0.038	53.406	46.594	0.023	4.013	0.508***	-0.544***	1.144	-0.008	-8.237	-1.060

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the summary statistics of overlapping/ pure factor portfolio value-weighted returns, in excess of the risk-free rate, comparing to the market and factor returns. The sample is from 1952:07 to 2020:12 (822 monthly returns). Market is the excess returns of the CRSP U.S. total market index. The risk-free rate is from Kenneth R. French's Data Library. The last four columns present the averages over time of the value-weighted characteristic of stocks in each portfolio. The EW. multi-factor portfolio is the equal-weighted combination of the three factors. Note that the uni-signal portfolio contains the total set of pure HML, pure MOM, and pure IVOL stocks.

Table A7: **Fama–MacBeth cross-sectional regressions with sorting by NYSE breakpoints**

	(1)	(2)	(3)	(4)	(5)
Panel A: Cross-section of 25 size×value, 25 size×momentum, and 25 size×volatility portfolios					
λ_{MKT}	-1.038***	0.753**	0.365	-0.646***	0.476*
	(-6.21)	(2.18)	(1.40)	(-3.43)	(1.75)
λ_{SMB}	0.134***	0.216***	0.167***	0.118**	0.182***
	(2.59)	(3.81)	(3.23)	(2.28)	(3.44)
λ_{HML}		0.248**			
		(2.52)			
λ_{MOM}		0.685***			
		(6.88)			
λ_{IVOL}		0.344***			
		(3.63)			
$\lambda^{3F}_{overlap}$			1.151***		1.255***
			(9.01)		(8.34)
λ^{3F}_{pure}				0.492***	-0.101
				(5.06)	(-0.72)
λ_0	1.748***	-0.117	0.279	1.374***	0.146
	(10.45)	(-0.32)	(1.04)	(7.35)	(0.51)
Adjusted R^2	0.309	0.772	0.698	0.422	0.724
Panel B: Cross-section of 25 size×net share issues and 25 size×accruals portfolios					
λ_{MKT}	-0.704**	0.379	0.183	-0.607*	0.202
	(-2.25)	(0.81)	(0.38)	(-1.89)	(0.42)
λ_{SMB}	0.168***	0.111	0.147**	0.117*	0.109
	(2.82)	(1.34)	(2.43)	(1.65)	(1.55)
λ_{HML}		0.530**			
		(1.96)			
λ_{MOM}		2.350**			
		(2.49)			
λ_{IVOL}		0.209			
		(0.78)			
$\lambda^{3F}_{overlap}$			1.093***		1.227***
			(2.95)		(3.12)
λ^{3F}_{pure}				0.458	0.402
				(1.49)	(1.30)
λ_0	1.354***	0.259	0.447	1.277***	0.444
	(4.31)	(0.55)	(0.92)	(4.00)	(0.92)
Adjusted R^2	0.325	0.662	0.486	0.380	0.525

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the second-stage cross-sectional regressions by Fama and MacBeth (1973b) on the returns of 75 portfolios formed on size×value, size×momentum, and size×volatility (Panel A) and 50 portfolios formed on size×net share issues and size×accruals (Panel B). Double sorted portfolios are retrieved from Kenneth R. French's data library.

Table A8: **Fama–MacBeth cross-sectional regressions on individual factors with sorting by NYSE breakpoints**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Cross-section of 25 size×value portfolios								
λ_{MKT}	-1.237***	-0.656	-0.367	-1.108**	-0.307	-0.621	-0.886	0.028
	(-2.61)	(-1.11)	(-0.51)	(-2.06)	(-0.42)	(-0.99)	(-1.62)	(0.03)
λ_{SMB}	0.083	0.122	0.133	0.090	0.137	0.124	0.114	0.139
	(1.06)	(1.48)	(1.57)	(1.13)	(1.61)	(1.49)	(1.39)	(1.63)
λ_{HML}		0.275***						
		(2.79)						
$\lambda^{3F}_{overlap}$			0.927***		1.039***			
			(2.70)		(2.67)			
λ^{3F}_{pure}				0.182	-0.091			
				(0.80)	(-0.32)			
$\lambda^{HML}_{overlap}$						0.655***		1.399
						(2.85)		(1.63)
λ^{HML}_{pure}							0.267	-0.358
							(1.14)	(-0.61)
λ_0	1.967***	1.314**	1.023	1.826***	0.959	1.281*	1.554***	0.638
	(4.19)	(2.14)	(1.35)	(3.34)	(1.26)	(1.95)	(2.74)	(0.66)
Adjusted R^2	0.397	0.543	0.534	0.384	0.537	0.512	0.477	0.543
Panel B: Cross-section of 25 size×momentum portfolios								
λ_{MKT}	-1.511***	-0.395	0.629	-0.387	0.680	-0.170	-0.636*	0.310
	(-4.98)	(-1.09)	(1.28)	(-1.05)	(0.80)	(-0.44)	(-1.85)	(0.42)
λ_{SMB}	0.190**	0.227**	0.179*	0.099	0.184	0.202**	0.207**	0.196**
	(2.01)	(2.40)	(1.89)	(1.03)	(1.62)	(2.13)	(2.18)	(2.06)
λ_{MOM}		0.723***						
		(6.95)						
$\lambda^{3F}_{overlap}$			1.153***		1.186**			
			(6.72)		(2.48)			
λ^{3F}_{pure}				0.750***	0.037			
				(6.02)	(0.07)			
$\lambda^{MOM}_{overlap}$						1.097***		1.527***
						(6.98)		(2.61)
λ^{MOM}_{pure}							0.918***	0.450
							(5.18)	(1.22)
λ_0	2.208***	1.173***	0.058	1.168***	0.002	0.931**	1.434***	0.406
	(6.99)	(3.21)	(0.12)	(3.16)	(0.00)	(2.38)	(4.13)	(0.51)
Adjusted R^2	0.375	0.878	0.865	0.833	0.859	0.855	0.829	0.857
Panel C: Cross-section of 25 size×volatility portfolios								
λ_{MKT}	-0.702***	0.931**	0.692	-1.148***	0.234	0.757*	0.433	0.377
	(-2.81)	(2.05)	(1.64)	(-3.99)	(0.42)	(1.71)	(0.89)	(0.77)
λ_{SMB}	0.080	0.158	0.207*	-0.031	0.132	0.197*	0.053	0.353***
	(0.77)	(1.50)	(1.91)	(-0.28)	(1.08)	(1.83)	(0.51)	(2.61)
λ_{IVOL}		0.398***						
		(3.66)						
$\lambda^{3F}_{overlap}$			1.521***		0.908			
			(4.79)		(1.58)			
λ^{3F}_{pure}				-1.654***	-0.688			
				(-3.08)	(-1.09)			
$\lambda^{IVOL}_{overlap}$						1.161***		2.226***
						(4.86)		(3.66)
λ^{IVOL}_{pure}							1.057***	-1.531*
							(3.12)	(-1.87)
λ_0	1.428***	-0.355	-0.108	1.777***	0.329	-0.150	0.257	0.173
	(5.90)	(-0.74)	(-0.24)	(6.67)	(0.59)	(-0.32)	(0.52)	(0.35)
Adjusted R^2	0.150	0.697	0.630	0.416	0.665	0.600	0.339	0.699

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the second-stage cross-sectional regressions by Fama and MacBeth (1973b) on the returns of 25 portfolios formed on size×value (Panel A), size×momentum (Panel B), and size×volatility (Panel C). Double sorted portfolios are retrieved from Kenneth R. French's data library.

Table A9: **Portfolio performance of the HML-MOM-IVOL strategy (quarterly rebalancing)**

	Return (percentage per quarter)				Sharpe	Distribution				CAPM		Average characteristic			
Portfolio	Mean (%)	Min. (%)	Max. (%)	Volatility (%)	ratio	Positive (%)	Negative (%)	Skewness	Kurtorsis	$\hat{\alpha}$	$\hat{\beta}_{MKT}$	Log-mkt. cap	B/M	Prior-ret (%)	Resid. vol. (%)
Makret	0.387	-12.887	10.745	3.857	0.100	59.489	40.511	-0.592	0.990			4.891	0.541	18.029	1.274
Overlapping portfolio															
Long	0.949	-19.639	13.550	4.134	0.230	63.869	36.131	-0.842	3.436	0.611***	0.936***	5.007	0.775	34.454	1.085
Short	-0.437	-22.622	20.505	5.743	-0.076	47.810	52.190	-0.488	2.152	-0.922***	1.344***	4.380	0.443	4.536	1.876
Long–Short	1.386	-16.181	16.024	4.000	0.346	63.869	36.131	0.453	2.765	1.533***	-0.408***	0.627	0.332	29.918	-0.791
Pure factor portfolio (across all factors)															
Long	0.556	-13.652	12.455	3.791	0.147	60.219	39.781	-0.381	1.374	0.213***	0.952***	4.942	0.554	25.273	1.201
Short	0.369	-15.807	11.622	4.138	0.089	56.569	43.431	-0.528	0.902	-0.003	1.030***	4.867	0.419	17.242	1.340
Long–Short	0.188	-4.259	4.325	1.325	0.142	50.730	49.270	0.407	0.948	0.216***	-0.078***	0.075	0.135	8.032	-0.138
EW. multi-factor	0.746	-7.460	11.777	2.158	0.346	65.693	34.307	1.056	4.501	0.828***	-0.227***	0.338	0.210	19.155	-0.423
HML	0.151	-15.148	13.736	3.488	0.043	53.650	46.350	-0.050	3.187	0.174	-0.064	-0.641	0.941	-2.072	-0.067
MOM	1.480	-14.122	21.489	4.384	0.338	66.788	33.212	0.475	3.223	1.507***	-0.074	0.757	-0.297	68.584	-0.157
IVOL	0.607	-12.885	17.680	3.709	0.164	58.029	41.971	0.726	3.939	0.802***	-0.542***	0.897	-0.014	-9.045	-1.045

Description. This table reports the summary statistics of overlapping/ pure factor portfolio value-weighted returns, in excess of the risk-free rate, comparing to the market and factor returns. The sample is from 1952:Q3 to 2020:Q4 (274 quaterly returns). Market is the excess returns of the CRSP U.S. total market index. The risk-free rate is from Kenneth R. French's Data Library. The last four columns present the averages over time of the value-weighted characteristic of stocks in each portfolio. The EW. multi-factor portfolio is the equal-weighted combination of the three factors. Portfolios are rebalanced every quarter, using the characteristic scores at quarter ends. Note that the pure factor portfolio contains pure HML, pure MOM, and pure IVOL stocks.

Table A10: **Fama–MacBeth cross-sectional regressions (quarterly rebalancing)**

	(1)	(2)	(3)	(4)	(5)
Panel A: Cross-section of 25 size×value, 25 size×momentum,					
and 25 size×volatility portfolios					
λ_{MKT}	-1.884***	1.338**	-0.160	-0.894***	-0.206
	(-7.40)	(2.43)	(-0.46)	(-3.05)	(-0.57)
λ_{SMB}	0.314***	0.414***	0.296***	0.281***	0.292***
	(3.73)	(4.65)	(3.54)	(3.34)	(3.48)
λ_{HML}		0.114			
		(0.74)			
λ_{MOM}		1.406***			
		(9.10)			
λ_{IVOL}		0.797***			
		(5.25)			
$\lambda^{3F}_{overlap}$			2.244***		2.156***
			(9.55)		(7.26)
λ^{3F}_{pure}				1.073***	0.397*
				(7.72)	(1.68)
λ_0	2.338***	-0.944*	0.410	1.263***	0.466
	(8.96)	(-1.67)	(1.12)	(4.15)	(1.21)
Adjusted R^2	0.409	0.919	0.736	0.693	0.738
Panel B: Cross-section of 25 size×net share issues					
and 25 size×accruals portfolios					
λ_{MKT}	-0.628	0.554	0.284	-0.507	0.254
	(-1.33)	(0.87)	(0.52)	(-1.07)	(0.44)
λ_{SMB}	0.399***	0.421***	0.324***	0.362***	0.324***
	(4.17)	(3.87)	(3.30)	(3.73)	(3.30)
λ_{HML}		0.383			
		(1.25)			
λ_{MOM}		1.730*			
		(1.74)			
λ_{IVOL}		0.408			
		(1.25)			
$\lambda^{3F}_{overlap}$			1.584***		1.576***
			(3.39)		(3.36)
λ^{3F}_{pure}				0.641**	0.268
				(2.49)	(0.89)
λ_0	1.068**	-0.128	0.115	0.917*	0.144
	(2.22)	(-0.20)	(0.21)	(1.89)	(0.25)
Adjusted R^2	0.513	0.752	0.697	0.624	0.699

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Description. This table reports the second-stage cross-sectional regressions by Fama and MacBeth (1973b) on the returns of 75 portfolios formed on size×value, size×momentum, and size×volatility (Panel A) and 50 portfolios formed on size×net share issues and size×accruals (Panel B). Double sorted portfolios are retrieved from Kenneth R. French's data library. The sample is from 1963:Q3 to 2020:Q4.

Bibliography

Adrian, T., E. Etula, and T. Muir (2014). Financial intermediaries and the cross-section of asset returns. *Journal of Finance 69*(6), 2557–2596.

Adrian, T. and H. S. Shin (2010). Liquidity and leverage. *Journal of Financial Intermediation 19*(3), 418–437.

Ali, A., L.-S. Hwang, and M. A. Trombley (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics 69*(2), 355–373.

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.

Asness, C. S., T. J. Moskowitz, and L. Pedersen (2013). Value and momentum everywhere. *Journal of Finance 68*(3), 929–985.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Avramov, D., S. Cheng, L. Metzker, and S. Voigt (2023). Integrating Factor Models. *Journal of Finance 78*(3), 1593–1646.

Ayyagari, M., A. Demirgüç-Kunt, and V. Maksimovic (2024). The Rise of Star Firms: Intangible Capital and Competition. *The Review of Financial Studies 37*(3), 882–949.

Bachmann, R. and C. Bayer (2014). Investment dispersion and the business cycle. *American Economic Review 104*(4), 1392–1416.

Bae, K.-H., W. Bailey, and J. Kang (2021). Why is stock market concentration bad for the economy? *Journal of Financial Economics 140*(2), 436–459.

Baker, M. and J. Wurgler (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives 21*(2), 129–152.

Baqaee, D. R. and E. Farhi (2020). Productivity and Misallocation in General Equilibrium*. *The Quarterly Journal of Economics 135*(1), 105–163.

Ben-David, I., F. Franzoni, and R. Moussawi (2012). Hedge fund stock trading in the financial crisis of 2007–2009. *The Review of Financial Studies 25*(1), 1–54.

Bessembinder, H. (2018). Do stocks outperform treasury bills? *Journal of Financial Economics 129*(3), 440–457.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Boyson, N. M., C. W. Stahel, and R. Stulz (2010). Hedge fund contagion and liquidity shocks. *Journal of Finance 65*(5), 1789–1816.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics 7*(3), 265–296.

Brunnermeier, M. and L. Pedersen (2009). Market liquidity and funding liquidity. *The Review of Financial Studies 22*(6), 2201–2238.

Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models. *Journal of Finance 78*(1), 487–557.

Bryzgalova, S., M. Pelger, and J. Zhu (2019). Forest through the trees: Building cross-sections of stock returns. *Wealth Management (Editor's Choice) eJournal*.

Campbell, J. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies 21*(4), 1509–1531.

Chen, A. Y. and M. Velikov (2023). Zeroing in on the expected returns of anomalies. *Journal of Financial and Quantitative Analysis 58*(3), 968–1004.

Cheng, S. F., D. Vyas, R. Wittenberg-Moerman, and W. Zhao (2024). Exposure to superstar firms and financial distress. *Review of Accounting Studies*.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(4), 1047–1108.

Covarrubias, M., G. Gutierrez, and T. Philippon (2020). From good to bad concentration? us industries over the past 30 years. *NBER Macroeconomics Annual 34*(1), 1–46.

David, J. and V. Venkateswaran (2019). The sources of capital misallocation. *American Economic Review 109*(7), 2531–67.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016). Information, Misallocation, and Aggregate Productivity. *The Quarterly Journal of Economics 131*(2), 943–1005.

David, J. M., L. Schmid, and D. Zeke (2022). Risk-adjusted capital allocation and misallocation. *Journal of Financial Economics 145*(3), 684–705.

De Loecker, J. and J. Eeckhout (2021). Global market power. Technical report, NBER.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications*. *The Quarterly Journal of Economics 135*(2), 561–644.

Doidge, C., G. A. Karolyi, and R. Stulz (2017). The U.S. listing gap. *Journal of Financial Economics 123*(3), 464–487.

Dou, W. W., Y. Ji, D. Tian, and P. Wang (2023). Misallocation and asset prices. Working paper, SSRN.

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

Eisfeldt, A. and D. Papanikolaou (2013). Organization capital and the cross-section of expected returns. *Journal of Finance 68*(4), 1365–1406.

Eisfeldt, A. L., E. Kim, and D. Papanikolaou (2020). Intangible value. Working Paper 28056, National Bureau of Economic Research.

Eisfeldt, A. L. and A. A. Rampini (2006). Capital reallocation and liquidity. *Journal of Monetary Economics 53*(3), 369–399.

Eisfeldt, A. L. and A. A. Rampini (2008). Managerial incentives, capital reallocation, and the business cycle. *Journal of Financial Economics 87*(1), 177–199.

Eisfeldt, A. L. and Y. Shi (2018). Capital reallocation. *Annual Review of Financial Economics 10*(1), 361–386.

Emery, L. and J. Koëter (2024). The size premium in a granular economy. Working paper, ssrn.

Fama, E. and J. D. MacBeth (1973a). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–36.

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

Fama, E. F. and K. R. French (1992b). The Cross-Section of Expected Stock Returns. *Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (1993a). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1993b). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

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

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

Fama, E. F. and J. D. MacBeth (1973b). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–36.

Favilukis, J. Y. and T. Zhang (2019). One anomaly to explain them all. *SSRN Electronic Journal*.

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

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies 33*(5), 2326–2377.

Gibbons, M. R., S. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*(5), 1121–52.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947 – 1990.

Gopinath, G., S. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017). Capital Allocation and Productivity in South Europe. *The Quarterly Journal of Economics 132*(4), 1915–1967.

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

Grullon, G., Y. Larkin, and R. Michaely (2019). Are US Industries Becoming More Concentrated? *Review of Finance 23*(4), 697–743.

Gutierrez, G. and T. Philippon (2017). Investmentless growth: An empirical investigation. *Brookings Papers on Economic Activity*, 89–169.

Gutierrez, G. and T. Philippon (2019). Fading stars. *AEA Papers and Proceedings 109*, 312–16.

Gutierrez, G. and T. Philippon (2020). Some facts about dominant firms. NBER Working Papers 27985, National Bureau of Economic Research, Inc.

Haddad, V. and T. Muir (2021). Do intermediaries matter for aggregate asset prices? *The Journal of Finance 76*(6), 2719–2761.

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., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *American Economic Review 103*(2), 732–70.

He, Z. and A. Krishnamurthy (2018). Intermediary Asset Pricing and the Financial Crisis. NBER Working Papers 24415, National Bureau of Economic Research, Inc.

Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy 124*(5), 1423 – 1465.

Hodrick, R. J. (1992). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. *The Review of Financial Studies 5*(3), 357–386.

Hou, K. and D. Robinson (2006). Industry concentration and average stock returns. *Journal of Finance 61*(4), 1927–1956.

Hou, K., C. Xue, and L. Zhang (2014, 09). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies 28*(3), 650–705.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Jegadeesh, N. (1990). Evident of predictable behavior of security returns. *Journal of Finance 45*, 881–98.

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., B. Kelly, and L. Pedersen (2023). Is there a replication crisis in finance? *The Journal of Finance 78*.

Kaplan, S. N. and L. Zingales (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The Quarterly Journal of Economics 112*(1), 169–215.

Kargar, M. (2021). Heterogeneous intermediary asset pricing. *Journal of Financial Economics 141*(2), 505–532.

Kehrig, M. (2015). The cyclical nature of the productivity distribution. *Microeconomics: Production*.

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.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics 132*(2), 665–712.

Koijen, R. S. J. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475 – 1515.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Kroencke, T. A. (2017). Asset Pricing without Garbage. *Journal of Finance 72*(1), 47–98.

Kung, H. and L. Schmid (2015). Innovation, growth, and asset prices. *Journal of Finance 70*(3), 1001–1037.

Kwon, S. Y., Y. Ma, and K. Zimmermann (2024). 100 Years of Rising Corporate Concentration. *American Economic Review 114*(7), 2111–2140.

Lettau, M., S. C. Ludvigson, and P. Manoel (2018). Characteristics of mutual fund portfolios: Where are the value funds? *NBER Working Paper Series*.

Maio, P. and P. Santa-Clara (2012). Multifactor models and their consistency with the icapm. *Journal of Financial Economics 106*(3), 586–613.

Mankiw, N. G. and M. D. Shapiro (1986). Risk and return: Consumption beta versus market beta. *The Review of Economics and Statistics 68*(3), 452–459.

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

Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *Journal of Monetary Economics 15*(2), 145–161.

Midrigan, V. and Y. Xu (2014). Finance and misallocation: Evidence from plant-level data. *American Economic Review 104*(2), 422–58.

Moritz, B. and T. Zimmermann (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Capital Markets: Market Efficiency eJournal*.

Müller, K. and S. Schmickler (2021). Interacting anomalies. *SSRN Electronic Journal*.

Nagel, S. (2013). Empirical cross-sectional asset pricing. *Annual Review of Financial Economics 5*(1), 167–199.

Neuhann, D. and M. Sockin (2024). Financial market concentration and misallocation. *Journal of Financial Economics 159*, 103875.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Parker, J. and C. Julliard (2005a). Consumption risk and the cross section of expected returns. *Journal of Political Economy 113*(1), 185–222.

Parker, J. A. and C. Julliard (2005b). Consumption risk and the cross section of expected returns. *Journal of Political Economy 113*(1), 185–222.

Peters, M. (2020). Heterogeneous Markups, Growth, and Endogenous Misallocation. *Econometrica 88*(5), 2037–2073.

Peters, R. and L. A. Taylor (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics 123*(2), 251–272.

Schlingemann, F. P. and R. M. Stulz (2022). Have exchange-listed firms become less important for the economy? *Journal of Financial Economics 143*(2), 927–958.

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. Special Issue on Investor Sentiment.

Stambaugh, R. F., J. Yu, and Y. Yuan (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance 70*(5), 1903–1948.

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

Su, D. (2022). Rise of superstar firms and fall of the price mechanism. Working paper, SSRN.

Whited, T. and G. Wu (2006). Financial constraints risk. *The Review of Financial Studies 19*(2), 531–559.

Zhang, X. F. (2006). Information uncertainty and stock returns. *Journal of Finance 61*(1), 105–137.

Summary

This dissertation, titled *"Market Concentration, Institutional Demand, and Risk Factors in Asset Pricing,"* explores several developments in asset pricing through three interconnected essays. Each chapter builds on a major post-20th-century economic trend: rising stock market concentration, the increase in institutional ownership, and the decline in the number of U.S.-listed firms.

In Chapter 2, *"Market Concentration, Capital Misallocation, and Asset Pricing,"* I examine the asset pricing implications of superstar firms. Superstar firms, which dominate stock markets through their large size and market power, can deter efficient capital allocation. Using quarterly data on U.S.-listed firms from 1975 to 2023, I use a proxy for capital misallocation as the cross-sectional dispersion in the marginal product of capital (MPK), and decompose this measure into three components: within superstar firms, within non-superstar firms, and between the two groups, the latter referred to as the "MPK spread." For asset pricing tests, I apply Fama-MacBeth regressions and construct factor-mimicking portfolios. I also run standard predictive regressions to examine the forecasting power of each component for multiple macroeconomic indicators.

Empirically, I find that only changes in the MPK spread, i.e., capital misallocation between superstar and non-superstar firms, are robustly priced in the cross-section of stock returns. Stocks with negative exposure to changes in the MPK spread earn annual returns approximately 4.8% higher than those with positive exposure. Moreover, higher MPK spread predicts lower future consumption growth, industrial production growth, employment growth, and innovation growth, particularly among non-superstar firms. These findings support the MPK spread as a key state variable con-

sistent with the Intertemporal Capital Asset Pricing Model (ICAPM), implying that changes in capital misallocation between superstar and non-superstar firms capture a priced macroeconomic risk factor. This chapter contributes to the literature by identifying a priced risk factor driven by market concentration.

Chapter 3, *"Intermediary Asset Pricing Through the Lens of a Demand System,"* co-authored with Dongryeol Lee, investigates how broker-dealers, though holding less than 2% of stock market assets, affect asset prices through their lending to hedge funds. Using institutional holdings from 2000 to 2023, we estimate a characteristic-based demand system to capture institution-specific demand elasticities. We then quantify how broker-dealer leverage shifts hedge funds' demand for assets and price impact.

We find that declines in broker-dealer leverage significantly reduce hedge fund asset demand, leading to price pressures. A one-standard-deviation decline in broker-dealer leverage reduces stock prices by approximately 1.6%, with the effect more pronounced in stocks held by hedge funds. Particularly, a 1% increase in broker-dealer leverage corresponds to a 2% rise in hedge funds' price impact in the following year, especially for illiquid stocks. These effects intensify during recessions, when broker-dealers face binding constraints and hedge funds are forced to deleverage. This chapter provides evidence that intermediary leverage constraints amplify asset return variation and affect stock prices through hedge funds.

Finally, Chapter 4, *"Overlapping Factors,"* analyzes the role of stocks that simultaneously load on multiple characteristic-based factors such as value, momentum, and volatility. I construct a stock-level measure of factor overlap, sort stocks into deciles based on overlap count, and evaluate return predictability and performance. I also perform cross-sectional pricing tests and various robustness checks.

Despite low correlations between traditional factor strategies, I find substantial overlap when sorting stocks into factor portfolios. Overlapping stocks–those that appear in the same leg across multiple factors–earn average monthly abnormal returns of 66 basis points, compared to just 5 basis points for pure factor stocks that appear in only one factor. These overlapping stocks are robustly priced in the cross-section, while pure factor

portfolios are not. Importantly, overlapping stocks are not extreme in factor characteristics or more exposed to macroeconomic or liquidity risk. In the short leg, overlapping stocks exhibit greater sensitivity to investor sentiment and elevated short interest. This chapter highlights the importance of accounting for overlap in factor construction, as overlapping stocks drive much of the return and pricing power in multi-factor strategies.

Together, these three essays provide insights into how several changes in financial markets, ranging from market concentration and intermediary constraints to the structure of factor investing strategies, shape asset prices and risk premia. The findings could provide important implications for investors, asset managers, and policymakers seeking to understand the underlying drivers of risk premia in the post-20th-century equity markets.

Samenvatting (Summary in Dutch)

Dit proefschrift, getiteld *"Market Concentration, Institutional Demand, and Risk Factors in Asset Pricing,"* onderzoekt risico en rendementsverhoudingen op financiële markten door middel in drie onderling verbonden essays. Elk hoofdstuk richt zich op een belangrijke economische trend sinds de 20e eeuw: de toenemende marktconcentratie, de toename van eigendom door institutionele beleggers en de afname van het aantal beursgenoteerde bedrijven in de VS.

In Hoofdstuk 2, *"Market Concentration, Capital Misallocation, and Asset Pricing,"* onderzoek ik de effecten van zogenaamde "superster" ondernemingen op het risico en verwacht rendement op beursgenoteerde aandelen. Supersterondernemingen, die aandelenmarkten domineren door hun grote omvang en grote marktmacht, kunnen een efficiënte kapitaalallocatie in de weg staan. Met behulp van kwartaalgegevens over Amerikaanse beursgenoteerde bedrijven van 1975 tot 2023 meet ik kapitaalmisallocatie op basis van de spreiding in het marginale product van kapitaal (MPK). Ik maak een onderscheid tussen drie componenten: binnen supersterbedrijven, binnen niet-supersterbedrijven, en tussen superster- en niet-supersterbedrijven, en duid de laatste component aan als de "MPK spread." Ik pas Fama-MacBeth regressies toe en construeer factor nabootsende portefeuilles om de de effecten op het risico en het verwachte rendement te meten. Ik gebruik ook regressies om de voorspellende kracht van elke component voor toekomstige fluctuaties in verschillende macro-economische indicatoren te beoordelen.

Mijn empirische resultaten laten zien dat alleen veranderingen in de MPK spread, hetgeen duidt op kapitaalmisallocatie tussen superster- en niet-supersterbedrijven, robu-

ust worden geprijsd in aandelenrendementen. Aandelen met een negatieve blootstelling aan veranderingen in de MPK spread behalen ongeveer 4,8% hogere jaarlijkse rendementen dan aandelen met een positieve blootstelling. Bovendien voorspelt een hogere MPK spread lagere toekomstige consumptiegroei, industriële productiegroei, werkgelegenheidsgroei en innovatiegroei, vooral voor niet-supersterbedrijven. Deze bevindingen tonen aan dat de MPK spread een belangrijke verklarende variabele is voor aandelenrendementen, hetgeen consistent is met het Intertemporal Capital Asset Pricing Model (ICAPM). Dit impliceert dat veranderingen in kapitaalmisallocatie tussen superster- en niet-supersterbedrijven een macro-economische risicofactor vormt. Dit hoofdstuk levert een bijdrage aan de literatuur door een geprijsde risicofactor te identificeren die gedreven wordt door toenemende marktconcentratie.

In Hoofdstuk 3, *"Intermediary Asset Pricing Through the Lens of a Demand System,"* geschreven in samenwerking met Dongryeol Lee, onderzoeken we hoe broker-dealers, ondanks het feit dat ze minder dan 2% van de aandelenmarkt bezitten, de aandelenprijzen beïnvloeden door leningen te verstrekken aan hedgefondsen. Met behulp van institutionele holdings data van 2000 tot 2023 schatten we een op characteristic-based demand system om instellingsspecifieke vraagelasticiteiten te schatten. Vervolgens kwantificeren we hoe de hefboomwerking van broker-dealers de vraag van hedgefondsen naar aandelen verandert.

We vinden dat een afname van de hefboomwerking van broker-dealers de vraag naar hedgefondsactiva aanzienlijk vermindert, wat leidt tot prijsdruk. Een daling van de leverage van broker-dealers met één standaarddeviatie verlaagt de aandelenkoersen met ongeveer 1,6%. Dit effect sterker is bij aandelen die worden aangehouden door hedgefondsen. Een stijging van de hefboomwerking van makelaars-handelaars met 1% komt overeen met een stijging van de invloed van hedgefondsen op de koersen in het volgende jaar met 2%, vooral voor niet-liquide aandelen. Deze effecten zijn sterker tijdens recessies, wanneer broker-dealers te maken krijgen met bindende beperkingen en hedgefondsen gedwongen worden om te deleveragen. Dit hoofdstuk levert structureel bewijs dat intermediaire hefboombeperkingen de variatie in aande-

lenrendementen versterken en de aandelenkoersen beïnvloeden via hedgefondsen.

Tenslotte analyseer ik in Hoofdstuk 4, *"Overlapping Factors,"* de rol van aandelen die tegelijkertijd in meerdere kenmerkende factoren voorkomen zoals waarde, momentum en volatiliteit. Per aandeel meet ik de mate van overlap, i.e., in hoeverre dit aandeel voorkomt in de long (of in de short) portfolio van meerdere factoren. Dan sorteer ik aandelen in op basis van de mate van overlap en test de invloed hiervan op het risico en gemiddeld rendement.

Ondanks de lage correlaties tussen traditionele factorstrategieën, vind ik aanzienlijke overlap in het sorteren van aandelen. Overlappende aandelen - aandelen die in dezelfde lag voorkomen van meerdere factoren - behalen een gemiddeld abnormaal maandelijks rendement van 66 basispunten, vergeleken met slechts 5 basispunten voor zuivere factoraandelen die slechts in één factor voorkomen. Deze overlappende aandelen zijn robuust geprijsd in de cross-sectie van gemiddelde aandelenrendementen, terwijl zuivere factorportefeuilles dat niet zijn. Belangrijk is dat overlappende aandelen niet extreem zijn in factorkarakteristieken of meer blootgesteld zijn aan macro-economische of liquiditeitsrisico's. In plaats daarvan vertonen ze op de korte termijn een grotere gevoeligheid voor het beleggerssentiment en een verhoogde short interest. Dit hoofdstuk benadrukt het praktische belang van rekening houden met overlap in factorconstructie, aangezien overlappende aandelen veel van het rendement en prijszettingsvermogen in multifactorstrategieën bepalen.

Samen bieden deze drie essays nieuwe inzichten in hoe diverse economische veranderingen, variërend van marktconcentratie en intermediaire beperkingen tot de ontwikkeling van factorbeleggingsstrategieën, de prijzen van activa en risicopremies beïnvloeden. De bevindingen kunnen belangrijke implicaties hebben voor beleggers, vermogensbeheerders en beleidsmakers die de onderliggende drijfveren van risicopremies op moderne aandelenmarkten willen begrijpen.

Gebaseerd op een vertaaling van DeepL.com

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. For a full list of PhD theses that appeared in the series we refer to List of PhD Theses – Tinbergen.nl. The following books recently appeared in the Tinbergen Institute Research Series:

826. Z. CSAFORDI, *Essays on Industry Relatedness*

827. B. VAN OS, *On Dynamic Models: Optimization-Based Methods and Practical Applications*

828. D. T. Ó CEALLAIGH, *Self-control failures and physical inactivity: Measuring, understanding and intervening*

829. S. B. DONOVAN, *Ties that bind and fray: Agglomeration economies and location choice*

830. A. SOEBHAG, *Essays in Empirical Asset Pricing*

831. H. YUAN, *Essays in Behavioral Economics*

832. A. LENGYEL, *Essays on government bond markets and macroeconomic stabilization*

833. S. KÜTÜK, *Essays on Risk Creation in the Banking Sector*

834. E. VLADIMIROV, *Essays on the Econometrics of Option Pricing*

835. R.E.K. PRUDON, *From the onset of illness to potential recovery. Empirical economic analysis of health, disability and work*

836. K. MOUSSA, *Signal Extraction by the Extremum Monte Carlo Method*

837. D. FAVOINO, *The Adaptation of Firms to Institutional Change*

838. B. WACHE, *Information Frictions in Financial Flows*

839. A. FEHÉR, *Essays in Law and Economics*

840. Q. WIERSMA, *Dynamic Models for Multi-Dimensional Time Series*

841. R. SILVESTRINI, *On the Importance of Firm Heterogeneity, Business Dynamism, and Market Power Dynamics in the Macroeconomy*

842. E.S.R. DIJK, *Innovative Start-Ups and Competition Policy - How to Reign in Big*

843. T.D. SCHENK, *Essays in Causal Inference with Panel Data*

844. S. TYROS, *Workers' Skills and (green) Technology Adoption*

845. C.J. GRASER, *Mechanisms for the Evolution of Prosociality*

846. K. IOANNIDIS, *On the role of information in strategic and individual decision making*

847. G.M. SZINI, *Advances in Panel and Network Econometrics*

848. L. VOOIS, *Empirical Studies of Health-related Expectations and Behaviors*

849. M. RÖSCH, *Multinational Firms and Local Workers*

850. E. MIHAYLOV, *Essays on Routine-Biased Technical Change, Job Tasks and Wages*

851. M. MAVUS KUTUK, *Essays on Currency Crash Risk, Carry Trade Returns, and Sovereign Bond Yields*

852. V. MOGHANI, *Essays on Health and Labor Economics*

853. T. VAN DER ZWAN, *Policy, Pricing and Prediction: Investigating Shock Dynamics, Equity Risk and Machine Learning*

854. S. WÖHRMÜLLER, *Essays in Macroeconomics with Household Heterogeneity*

855. D. OPSCHOOR, *On Asymmetries and Heterogeneities in Economic Modelling and Forecasting*

856. P. BOSE, *From Preferences to Policy: Essays in Political Economy*

857. J. KLOOSTER, *Robust Inference in Instrumental Variable Models*

858. K. HANEMAAIJER, *Widening the Divide: Barriers for Marginalised Groups in Education and Justice*

859. I.C. JOAO, *Dynamic Clustering Methods in Panel Data*

860. K. WACKER, *Analyses of Policies and Innovation in Banking*

861. S. VAN ALTEN, *Genetics, Human Capital Formation and the Intergenerational Transmission of Socioeconomic Status*

862. M. BASTIAANS, *Public Policies, Labor Market Shocks and Inequality*

863. M. MUSUMECI, *Three Essays on the Economics of Science and the Economics of Education: The Role of Barriers to Success in Academia and in School*

864. J. VAN SPRONSEN, *Sovereign Debt Markets and Resilience in a Heterogenous Union on the Interaction of European Economic and National Fiscal Policies*

865. A. HIRMAS, *Seeing differently, Doing differently: Essays on Visual Attention and Individual Decision-Making*

866. K. SOMMER, *Environmental Policy in Open Economies: Differences in Policy Stringency and the Effects on Leakage and Pollution Havens*

867. M. ARTEMOVA, *Studies in Observation-Driven Time Series Models: Theory, Methods, and Applications*

868. Y. LI, *Micro Studies on Self-Employment, Housing and Pensions*

869. S.J. DEN NIJS, *Investing in a Green Economy: Firm Perspectives and Spatial Economic Implications*

870. L. BREMER, *Technological Progress in the Transition to a Sustainable Economy*

871. D. MININA, *Essays on Expectation Formation and Learning*

872. J. ILCIUKAS, *Fertility and Family*

873. A. KURZ, *Competitiveness in the Green Transition*

874. A. TITTON, *Economic Consequences of Environmental Catastrophes*

875. D.A. CANDIA RIQUELME, *Assessing Innovative Transport Technologies and Policies: A Spatial General-Equilibrium Analysis*

876. M. HAASBROEK, *Trade and Industrial Policy in Emerging Economies*

877. Y. VAN DER STRATEN, *Financing Adaptation and the Green Transition: Climate Risk, Housing, and Capital Markets*