Essays in Asset Management

Anthony Limburg

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2026

This thesis studies the real economic consequences of large-scale investment diversification and indexing in financial markets. In the first chapter, I show that diversified ownership causally reduces firm-level breakthrough innovation. As investors shift from individual stocks to diversified funds, their demand for skewed returns declines, leading firms to undertake fewer breakthrough innovation projects. The second chapter examines self-indexed ETFs, showing that these funds charge higher fees without delivering superior performance, consistent with conflicts of interest in advisory distribution channels. The third chapter tests whether firms adjust their behaviour to gain index inclusion using Russell index reconstitutions. I find no evidence of manipulation, supporting the validity of Russell index inclusion as a quasi-natural experiment. Overall, the thesis highlights that while diversification yields private benefits, it has important unintended effects on innovation, industry structure, and firm behaviour.

Supervisors This thesis was supervised by Martin Schmalz and Bige Kahraman.

Word Count The thesis contains approximately 46,000 words.

Statement of authorship The research contained in this thesis is exclusively carried out by me as a DPhil student at the University of Oxford between September 2022 and June 2026. No part of the thesis has been submitted for a degree at any university outside the University of Oxford or at any other university by any other person. Chapter 2 is co-authored with two other academics but I am responsible for the majority of the work. Chapters 1 and 3 are entirely my own work.

Acknowledgment of funding I acknowledge funding from the Clarendon Fund, the Templeton Education and Charity Trust (TECT), and the Saïd Business School Foundation.

Essays in Asset Management

Anthony Limburg

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2026

In memory of my father, Thomas.

Acknowledgments

My research journey has been strongly shaped by a few people to whom I would like to extend my deepest gratitude.

I appreciate the support of my advisors and co-authors for guiding my research process and providing essential feedback whenever I required it. Martin, your enthusiasm for research that is unconventional and complex has tremendously shaped my approach to economic research – and is clearly reflected in all three thesis chapters. Sharing your research ideology has unlocked most of the fun I have had conducting research over the years. Bige, your hands-on approach to supervision allowed me to develop essential empirical and story-telling skills. Our collaboration on the ETF paper started early during my PhD and defined my research process for other projects thereafter. It is fantastic to see that our project with Sida has gained traction among academics and practitioners alike.

My DPhil colleagues and friends at the Business School have also shaped my time in Oxford. Your informal feedback on ideas, discussions in the office, and light-heartedness have humanised the otherwise isolating research experience. In an alternate universe I would have loved to co-author work with some of you. Some but not all of you are Olivia Deuschle, Fiona Kasperk, Julian Corj, Mingze Ma, Aditya Khemka, Adam Koling, Fuaad Coovadia, Taiwu Liu and of course, Jasper Knyphausen. Jasper, I cannot imagine my time at Oxford without your friendship. Your unconditional support was essential in bridging the lows and celebrating the highs of my journey in Oxford, making it an unforgettable experience.

I am grateful to my family, who supported and encouraged me throughout my studies in a different country. I also owe so much to my loving partner, Sophie. Thank you for your patience and understanding over the years, and for always providing stability when research was volatile.

Abstract

In this thesis, I study the real economic consequences of large-scale diversification and indexing that go beyond the private benefits of a single private investor. In the first chapter I show that diversified firm ownership of public firms causally reduces firm-level breakthrough innovation. The channel I propose for this result rests on an established literature describing investors' skewness preferences: when diversification is cheap, investors move from single stock holdings to diversified funds, reducing their utility gain from skewed returns of individual firms. Firms cater to this by reducing the skewness of projects they undertake. The findings highlight a previously unknown consequence of diversification with potentially important societal consequences. The second chapter studies the rise of self-indexed exchange-traded funds (ETFs), in which asset managers design and track their own indices. Contrary to the common view that this vertical integration reduces costs and allows rent-sharing with end-investors, I show that self-indexed ETFs charge higher fees without delivering superior performance. The evidence is consistent with conflicts of interest, whereby issuers distribute these products through advisory channels, leading to welfare losses for end-investors. The third chapter examines whether firms strategically adjust their behaviour to gain inclusion in major stock indices. Using the transparent reconstitution rules of the Russell indices, I test a broad set of potential managerial responses, including earnings management, payout policies, and financing decisions. I find no evidence that firms manipulate outcomes to influence index assignment, supporting the validity of index inclusion as a quasi-natural experiment in the literature. This thesis shows that while diversification delivers clear private benefits, its large-scale adoption has important and previously overlooked effects on innovation, industry structure, and firm behaviour.

Contents

Introduction **1**

1 The Innovation Cost of Diversification **7**

1.1 Introduction . . . 8
1.2 Data . . . 15
1.3 Motivating evidence . . . 18
1.3.1 Revealed skewness preferences . . . 18
1.3.2 Cross-sectional tests . . . 24
1.3.3 Causal evidence . . . 26
1.4 Model . . . 31
1.4.1 Environment . . . 31
1.4.2 Investor preferences and choices . . . 33
1.4.3 Market equilibrium and firm choice . . . 34
1.4.4 Parameters . . . 34
1.4.5 Solution strategy . . . 35
1.4.6 Discussion . . . 37
1.4.7 Results . . . 38
1.5 Empirical model validation . . . 43
1.6 The governance channel . . . 48
1.6.1 Direct governance mechanisms . . . 49
1.6.2 Monitoring intensity . . . 50
1.6.3 Passive ownership . . . 54
1.7 Real effects of diversified ownership . . . 56
1.8 Conclusion . . . 61

2 Index Disruption: The Promise and Pitfalls of Self-Index ETFs **63**
2.1 Introduction . . . 64
2.2 Data and sample construction . . . 70
2.2.1 Descriptive statistics . . . 74
2.3 Do self-indexed ETFs provide financial benefits to investors? . . . 80
2.3.1 Differences in annual fees . . . 80
2.3.2 Differentiated products? . . . 83
2.4 Self-preferencing by issuer-advisors? . . . 92
2.4.1 Investment advisors and self-indexed ETFs . . . 92
2.4.2 Institutional self-ownership of self-indexed ETFs . . . 94
2.4.3 Model portfolio recommendations and self-indexed ETFs . . . 97
2.4.4 Policy discussion . . . 100
2.5 Conclusion . . . 101

3 Do Firms Self-Select Into Russell Indices? **103**
3.1 Introduction . . . 104
3.2 Institutional background: Russell 1000 Index reconstitution . . . 109
3.3 Data and variable construction . . . 112
3.4 Managerial actions for index inclusion . . . 117
3.4.1 Earnings management . . . 118
3.4.2 Earnings guidance . . . 122
3.4.3 Issue and repurchase announcements, and dividend payout . . . 123
3.5 Concerns and cross-sectionality . . . 128
3.6 Conclusion . . . 130

Conclusion **133**

Bibliography **ix**

Appendices **xix**
Appendix A . . . xxii
Appendix B . . . xxxiii
Appendix C . . . xli

List of Figures

Chapter 1: The Innovation Cost of Diversification

1.1 Idiosyncratic return skewness and institutional ownership. 9
1.2 Revealed skewness preferences. 22
1.3 Aggregate revealed skewness preferences for stocks. 23
1.4 Timeline of measurement windows for causal estimation. 29
1.5 Optimal skewness for (k, B) pairs 39
1.6 Equilibrium outcomes for (k, B) pairs. 41
1.7 Under-diversification for (k, B) pairs. 42
1.8 Equilibrium price for (k, B) pairs. 43
1.9 Median skewness for $SkewPref$-sorted portfolios. 45
1.10 Skewness premium over time. 47

Chapter 2: Index Disruption: The Promise and Pitfalls of Self-Index ETFs

2.1 Number of self- and public-indexed ETFs over time. 76
2.2 Share of issuer types over time. 77
2.3 ETF networks based on cosine similarity of stock holdings. 86

Chapter 3: Do Firms Self-Select Into Russell Indices?

3.1 Russell 2000 exclusion rank and index size. 112
3.2 Distance to index switch for Russell 3000 members. 115

Appendices

A.1 Time trend of idiosyncratic return skewness for alternative measures. xxii
A.2 Statistical significance of demand system estimation. xxiii
A.3 AUM-weighted taste estimates. xxiv

A.4 Distribution of firm-quarter instrument $z_t(n)$. xxv
A.5 Skewness premium over time (equal-weight). xxvi
B.1 Density plots of issuer characteristics by issuer type. xxxiii
B.2 ETF networks based on pairwise ETF return correlation. xxxiv

List of Tables

Chapter 1: The Innovation Cost of Diversification

1.1 Summary statistics. . . . 18
1.2 Skewness, institutional ownership and owners' skewness preferences. . . . 25
1.3 Causal relationship between institutional ownership and skewness . . . 30
1.4 Model parameters. . . . 35
1.5 Portfolio tests. . . . 46
1.6 Causal relationship between diversified ownership and governance mechanisms 51
1.7 Causal relationship between diversified ownership, monitoring and skewness. . 53
1.8 Russell inclusion study. . . . 55
1.9 Skewness and breakthrough innovation. . . . 58
1.10 Causal link between institutional ownership and breakthrough innovation. . . . 60

Chapter 2: Index Disruption: The Promise and Pitfalls of Self-Index ETFs

2.1 Number of ETFs by style and indexing group. . . . 74
2.2 Summary statistics. . . . 79
2.3 Explaining net expense ratios. . . . 81
2.4 Explaining ETF differentiation. . . . 87
2.5 Alphas from portfolio tests with monthly and daily returns. . . . 89
2.6 Performance regressions. . . . 90
2.7 Explaining other ETF characteristics. . . . 91
2.8 Exploring the role of investment advisors. . . . 93
2.9 Sentiment-flow sensitivity. . . . 95
2.10 Explaining institutional self ownership. . . . 97
2.11 Explaining affiliated model portfolio membership. . . . 99

Chapter 3: Do Firms Self-Select Into Russell Indices?

3.1 Fictional example of the Russell 1000 Index reconstitution methodology since 2007. 111
3.2 Summary statistics. . . . 117
3.3 Predicting unexpected discretionary expenses. . . . 119
3.4 Predicting unexpected cashflow from operations. . . . 120
3.5 Predicting discretionary accruals. . . . 121
3.6 Predicting managerial forecast error (MFE). . . . 124
3.7 Predicting issuance announcements (logit). . . . 125
3.8 Predicting repurchase announcements (logit). . . . 126
3.9 Predicting a dividend payout dummy (logit). . . . 127

Appendices

A.1 Skewness and profitability. . . . xxvii
A.2 Skewness and institutional ownership with overlapping windows. . . . xxviii
A.3 Robustness on causal skewness evidence. . . . xxix
A.4 Causal evidence on longer horizons. . . . xxx
A.5 Robustness to causal link between institutional ownership and breakthrough innovation. . . . xxxi
A.6 Causal evidence on innovation channels. . . . xxxii
B.1 Largest issuers of ETFs by AUM as of December 2020. . . . xxxv
B.2 Explaining the issuance of self-indexed ETFs at the issuer-month level. . . . xxxvi
B.3 Robustness tests. . . . xxxvii
B.4 Performance of active funds. . . . xxxviii
B.5 Sample of issuers classified as investment advisors and fund specialists. . . . xxxix
B.6 Variable definitions . . . xl
C.1 Predicting real earnings management when firms are consistently close to the index threshold. . . . xlii
C.2 Predicting discretionary accruals when firms are consistently close to the index threshold. . . . xliii
C.3 Predicting managerial forecast error (MFE) when firms are consistently close to the index threshold. . . . xliv
C.4 Predicting real earnings management when CEO horizon is short. . . . xlv

C.5 Predicting discretionary accruals when CEO horizon is short. xlvi
C.6 Predicting managerial forecast error (MFE) when CEO horizon is short. xlvii
C.7 Predicting real earnings management when CEOs are poorly compensated. . . .xlviii
C.8 Predicting discretionary accruals when CEOs are poorly compensated. xlix
C.9 Predicting managerial forecast error (MFE) when CEOs are poorly compensated. l
C.10 Predicting real earnings management when CEOs have high performance-pay. . li
C.11 Predicting discretionary accruals when CEOs have high performance-pay. lii
C.12 Predicting managerial forecast error (MFE) when CEOs have high performance-pay. liii
C.13 Predicting real earnings management when CEO stock ownership is low. liv
C.14 Predicting discretionary accruals when CEO stock ownership is low. lv
C.15 Predicting managerial forecast error (MFE) when CEO stock ownership is low. . lvi
C.16 Predicting real earnings management for firms near Russell 1000 inclusion. . . . lvii
C.17 Predicting discretionary accruals for firms near Russell 1000 inclusion. lviii
C.18 Predicting managerial forecast error (MFE) for firms near Russell 1000 inclusion. lix

Introduction

Diversification is the dominant organizing force in modern asset management. Since Markowitz (1952) first laid out the benefits of diversification, its promises have reshaped investor demand, financial product design, and the relationship between capital markets and the real economy.

Under mean-variance preferences, investors should hold the diversified combination of risky assets which eliminates idiosyncratic risk. Under the Capital Asset Pricing Model (CAPM), this portfolio is represented by the market portfolio (Sharpe, 1964; Lintner, 1965; Mossin, 1966). Over the past four decades, a sweeping institutional and technological transition implemented the vision of the CAPM in practice. Data from the U.S. show that the share of households owning mutual fund shares rose from 6% in 1980 to 54% in 2024 (Federal Reserve Bank of St. Louis, 1993). This trend was driven first by active funds, and since 2000, much of the growth in delegated asset management has come from index funds which conceptually offer the closest purchasable approximation of the CAPM's market portfolio. Now, index funds have become the primary vehicle for large-scale diversification: in 2023, they surpassed active funds in popularity, managing more than half of global fund assets (Sabban, 2024).

A substantial and growing literature studies the important consequences of the stark rise in diversification, indexing, and related passive institutional ownership. For example, a literature spearheaded by Shleifer (1986) finds that indexing has price impacts because passive demand is inelastic, i.e., the demand curve for stocks is downward sloping (e.g., Greenwood and Sammon (2025); Gabaix and Koijen (2021)). Other research focuses on whether an increase in inelastic demand adversely affects the price informativeness of stocks (e.g., Sammon (2025); Brogaard et al. (2018)) or whether the rise of passive ownership affects firm governance (e.g., Heath et al. (2022); Schmidt and Fahlenbrach (2017)). Another group of papers examines the anticompetitive effects of common ownership, defined as firm ownership that is diversified across competing firms (e.g., Azar et al. (2018); Antón et al. (2023)). Recent work also links passive investing to the rise in market concentration driven by the proliferation of mega-firms (Jiang et al., 2020). Other papers show that the market for indexing is highly oligopolistic, which might

have welfare-reducing effects on end-investors (An et al., 2023; Huang et al., 2023). This list of topics and papers is by no means exhaustive. Rather, it serves to show that the rise of index investing and cheap diversification are important topics with far-reaching economic implications. This PhD thesis adds to the academic discourse by studying three previously unexplored effects of the modern manifestation of diversification. Each of the three chapters includes a detailed discussion of the relevant academic literature.

In the first chapter, the thesis shows that access to diversification alters investors' tastes for positively skewed firm-level returns – and that this causally affects firms' innovation choices. The chapter starts by addressing a puzzle: since the 1980s, idiosyncratic return skewness in U.S. stocks has steadily declined (Gomez et al., 2020), despite the persistence of human gambling preferences that favour positively skewed payoffs (Kumar, 2009). This disappearance coincides with the rise in institutional ownership driven by cheap access to diversified funds. In this chapter, called *The Innovation Cost of Diversification*, I argue that these trends are intertwined: when diversification becomes cheap, investors move from single stocks into funds, which diversifies away idiosyncratic return skewness. This, in turn, reduces their aggregate willingness to pay for skewness in individual stocks. Firms respond by selecting fewer projects with positively skewed outcomes. Using institutional ownership data and building on recent methods to estimate investor preferences (Koijen et al., 2024), I document that owners' revealed demand for skewness has indeed fallen over time, and that firms with more diversified ownership exhibit lower future return skewness. I then use a granular instrumental variable (GIV) approach (Gabaix and Koijen, 2024) to show that increases in diversified ownership causally reduce subsequent idiosyncratic return skewness. To clarify the mechanism behind these results, I develop an equilibrium model where firms endogenously choose project skewness. The model shows how cheap diversification suppresses skewness even when all investors inherently love skewness and without any governance frictions. Finally, I examine the real effects of this mechanism. I find that breakthrough innovation, which generates highly skewed out-

comes, is negatively causally associated with diversified ownership. The findings suggest that one overlooked social cost of diversification is the suppression of transformative innovation in public firms.

The second chapter turns from demand to supply: the rise of indexing as the default expression of diversification gave rise to a new industry of index provision, which operates alongside the funds that track these benchmarks. Traditionally, passive funds track indices designed and maintained by independent third-party providers such as S&P Dow Jones, FTSE Russell, or MSCI. In return, these providers are paid licensing fees, typically as a fraction of fund size. An et al. (2023) show that index providers have substantial market power; the four largest providers serve 75% of the market and their brands carry value for investors (Kostovetsky and Warner, 2025). As a result, it is estimated that two thirds of licensing fees are markups. In large part for this reason, the industrial organization of indexing has recently evolved considerably. An increasing number of funds track indices that are built in-house, allowing fund issuers to save on licensing costs. These types of funds are called self-indexed funds. The popular narrative suggests that this practice benefits end-investors as the rents gained from the elimination of licensing fees can be shared with end-investors.[1] This chapter, which is based on joint work with Bige Kahraman and Sida Li, previously circulated as *Index Disruption: The Promise and Pitfalls of Self-Indexed ETFs*,[2] asks whether the practice of self-indexing ultimately benefits end-investors. Using panel data from passive ETFs in the U.S. from 2012 to 2020, we show that self-indexed funds are *more*, and not less expensive than otherwise similar products. We provide evidence that the higher prices are not driven by indexing strategies which are more active or distinct, and that our results are also not driven by sector-, style-, and time-specific effects. We further show that these increased fees are ultimately welfare-reducing to end-investors as self-indexed ETFs do not provide performance or risk-return benefits. We find only one story

[1] See, for example, https://www.ft.com/content/e886b2d2-e852-3071-85c1-c9a57113d8a5.
[2] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4943383.

that can reconcile the survival and significant growth of these funds: fund issuers that are also wealth advisors market their expensive self-indexed funds to advisory clients. Indeed, we find that only self-indexed ETFs issued by wealth advisors are more expensive than their peers. Furthermore, these expensive products offered by wealth advisors are *less* differentiated than the average licensed index ETF. Our paper is the first to document this previously overlooked conflict of interest and directly speaks to the policy discussion on self-indexed products.

In chapter three, which was previously circulated as *Do Firms Self-Select Into Russell Indices?*,[3] I expand our knowledge on the firm-level effects of index investing. A growing literature shows that index inclusion, and the related changes to firm ownership, matter for firms as they experience positive abnormal returns (Shleifer, 1986), changes in analyst coverage (Yu, 2008), governance (Heath et al., 2022), and access to capital (Brisker et al., 2013). Seeing as these effects change firm value and managers frequently celebrate index inclusions,[4] it is natural to ask whether firms and their managers take deliberate actions to join specific stock indices. In this chapter I use the Russell 1000 and Russell 2000 index cut-off to answer this question. The question matters for two key reasons. First, there is limited evidence that the possibility of index inclusion alters firm behaviour. So far, Hong et al. (2021) and Li et al. (2026) consider this possibility using the S&P 500 Index inclusion, which is determined by an index committee whose rules and decision processes are opaque. Within the Russell Index universe, inclusion rules are clearly defined which may increase incentives to game the system. While Cho et al. (2017) and Guo (2019) entertain this idea by looking at earnings management, my research substantially expands on this by encompassing all possible opportunistic actions firm managers could take. Second, much of the most influential research on the effects of passive investing relies on the Russell 1000/2000 index assignment as a setting for a quasi-natural experiment (e.g., Heath et al. (2022); Appel et al. (2018, 2016); Boone and White (2015)). It is assumed that near

[3] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4816038.

[4] See, for example, https://ir.astranahealth.com/news-events/press-releases/detail/31/apollo-medical-holdings-joins-russell-3000-index.

the index threshold, the sorting of firms into either the Russell 1000 or 2000 index is random. Because the Russell 2000 index is more popular with investors, firms which are assigned into that index experience significant increases in institutional, passive ownership. If firms were to self-select into Russell indices, much of the research on passive ownership would have to be seriously reconsidered. I therefore test a broad set of factors that could change the odds of index inclusion: real earnings management and discretionary accruals; guidance and dividend policy; and the timing of equity offerings and stock repurchases. I find no evidence that firms near the threshold manage outcomes to influence index assignment. These results hold even when managerial incentives are most acute, i.e. when index switching is likely to increase managerial pay or reduce board oversight. Crucially, the findings in this chapter can be understood as an important and overdue methods contribution for the literature on passive institutional ownership.

Read together, the three chapters in this thesis tell a coherent story about financial markets with growing demand for diversification: the diffusion of cheap diversification changes what investors value (tastes for return distributions), how the industry is organized (who designs benchmarks and where rents accrue), and what firms choose to do (the direction of innovative effort). The welfare calculus of index investing must look beyond the private benefits of low-cost diversification. Modelling and quantifying the externalities of diversification in equilibrium settings are the logical next step for this research agenda.

Chapter 1

The Innovation Cost of Diversification[†]

Abstract Idiosyncratic return skewness in U.S. equities has declined sharply since the 1980s, despite persistent investor preferences for positively skewed payoffs. Over the same period, falling diversification costs allowed investors to move from individual stocks into diversified funds. I show that this shift reduces investors' aggregate willingness to pay for firm-level skewness, weakening firms' incentives to pursue positively skewed projects. Consistent with this mechanism, aggregate revealed skewness preferences have fallen over time. Exploiting exogenous shocks to firm-level diversified ownership using a granular instrumental variable approach, I show that higher diversification levels causally reduce future idiosyncratic return skewness, consistent with firms catering to their owners. A stylized equilibrium model formalizes this mechanism and shows how cheap diversification can suppress skewness even when all investors are inherently skewness-loving. Finally, I show that diversified ownership causally reduces firm-level breakthrough innovation – an activity generating positively skewed outcomes. An unintended consequence of diversification is the reduction of transformative innovation.

[†]I thank Martin Schmalz, Bige Kahraman, Joel Shapiro, Jiri Knesl, Thomas Noe, Ken Okamura, Jasper Knyphausen and Mingze Ma for helpful comments and acknowledge support from the Clarendon Fund, the Templeton Education and Charity Trust (TECT), and the Saïd Business School Foundation.

1.1 Introduction

Gambling preferences are a persistent feature of human behaviour (France, 1902). In financial markets, these preferences manifest in investors' attraction to securities with positively skewed payoffs – assets that offer a small chance of very large returns. Many investors willingly accept lower returns in exchange for extreme upside. Coincidentally, most U.S. equities' idiosyncratic return distributions have been positively skewed, aligning with the deep-rooted preferences of investors. That is, until the 1980s, when return skewness started to disappear (Gomez et al., 2020).

The disappearance of skewness is puzzling given the persistence of gambling preferences. If human nature has not changed and firms are governed by their owners, why have stock level payoffs become more symmetric? Why would investors not govern firms to take on projects with more upside? In this paper, I aim to answer these questions by unveiling a novel economic mechanism.

To motivate this mechanism, I present Figure 1.1, which shows that a contemporary trend to the decrease in idiosyncratic return skewness is the increase in institutional ownership of U.S. stocks. This upward trend is largely fuelled by cheap and accessible stock market diversification, offered first by mutual funds and then by ETFs. As such, the share of equity held directly by households has declined since 1980, while the share held through intermediaries has risen.

I claim that the disappearance of skewness and rise of institutional ownership are intertwined trends: when diversification is cheap, investors diversify so as to eliminate idiosyncratic firm shocks in their portfolios. To diversify, investors rely on intermediaries offering funds, which in turn increases institutional ownership. As such, institutional ownership serves as a proxy for the extent to which investors hold diversified portfolios. When buying funds, investors also diversify away the very skewness that satisfies their gambling preferences. There is a trade-off between the risk benefits of diversification and the drawbacks of reductions in

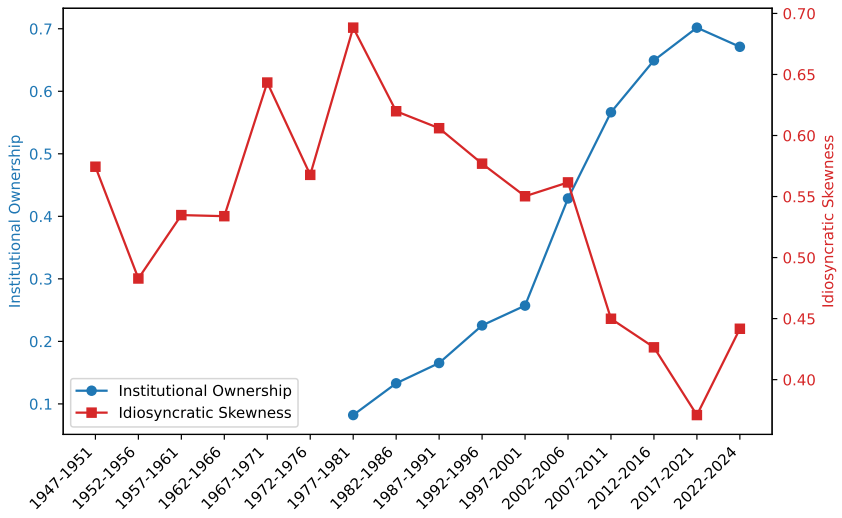


Figure 1.1. Idiosyncratic return skewness and institutional ownership.
This figure plots the median fraction of shares outstanding held by 13F institutional owners for 5-year bucket for all common stocks listed in the U.S. (left axis) and their median idiosyncratic skewness as calculated from residuals from 5-year rolling CAPM regressions (right axis).

skewness. When diversification becomes cheaper – as it has over the last decades – each investor is more likely to forego the upsides from single-stock investing. Since a diversified portfolio neutralizes firm-level skewness, investors' aggregate willingness to pay for skewness in any individual firm declines when diversification becomes more popular. And if investors no longer reward skewness, firms have weaker incentives to pursue projects that generate it.

To highlight this mechanism I first investigate time trends of revealed skewness preferences of U.S. investors. Unlike inherent skewness preferences, which are rooted in human psychology, revealed skewness preferences reflect the interaction of inherent preferences with available investment opportunities. To this end, I build on Koijen et al. (2024) and estimate investor-level preferences for firm characteristics using institutional ownership data from 13F filings. I find that retail investors consistently overweight stocks with positively skewed returns, while institutional investors exhibit either no or negative revealed preferences for skewness. These revealed preferences are stable over time. For each stock I then calculate a quarterly measure of owners' revealed skewness preferences by value-weighting the investor-specific preferences

by the shares outstanding they hold. In line with the increase in institutional ownership, I find that aggregate revealed skewness preferences for stocks have decreased over the last decades.

I then turn to the question whether revealed skewness preferences and diversification levels of a firm's owners predict future return skewness. In a frictionless world, firm policies perfectly reflect their owners' preferences. When owners have lower revealed preferences for skewness it therefore seems plausible that firms take on fewer projects with positive skew. To assess the validity of this claim, I proxy how diversified a firm's owners are using institutional ownership data from 13F filings. Since most benefits of diversification already accrue at portfolio sizes of 10 to 20 stocks, it is reasonable to assume that assets managed through intermediaries, including active funds, are part of diversified portfolios.

I find that lagged diversified ownership is negatively associated with future return skewness, whereas the lagged measure of investors' revealed skewness preference is positively associated with subsequent skewness. These relations remain economically and statistically significant after controlling for a comprehensive set of time-varying firm characteristics, as well as time and industry fixed effects. I also present causal evidence for this claim. Using a granular instrumental variable (GIV) approach exploiting idiosyncratic shocks to portfolio holdings of institutional investors (Gabaix and Koijen, 2024; Zhang, 2024), I show that a ten percentage point increase in diversified ownership causally reduces idiosyncratic return skewness over the next five years by approximately 0.23 units, corresponding to 33% of the sample mean and 25% of one standard deviation. This result, and all other findings on idiosyncratic return skewness in this paper, hold whether I use the CAPM, Fama French or Carhart factor models to calculate idiosyncratic returns.

To better understand the channel through which diversified ownership causally reduces idiosyncratic return skewness, I present a stylised, frictionless model of firms and investors. The model, which loosely builds on Schmalz and Zame (2024), features heterogenous investors who are risk averse and skewness-loving and firms which endogenise the skewness of their

idiosyncratic returns through a firm value maximization problem. The firm's profit function exhibits a linear cost in skewness, in line with empirical evidence. Investors can choose to invest in a riskless bond, a diversified fund which comes at a cost, and a single stock. The fund diversifies away all idiosyncratic risk, but offers no skewness. The firm's optimal choice for skewness, prices, and asset allocations are determined in equilibrium where demand matches supply of the stock.

I estimate the model numerically and find that the firm's choice of skewness is increasing in both the fund fee and the investors' preferences for skewness. When both the fund fee and skewness preferences are very low, the fund dominates the stock and the firm caters only to fund owners who prefer higher returns as opposed to skewness. As fund fees increase, the stock becomes more attractive to investors and the firm caters to these under-diversified owners by increasing return skewness. These predictions match the previously described trends and causal evidence in the U.S. stock market. The model reveals that diversified ownership can causally reduce firm-level skewness without any changes in governance or other frictions. Firms are understood to 'cater' to investors by chasing premia as in Baker and Wurgler (2004) and Polk and Sapienza (2008). The model also captures how investors can exhibit low *revealed* skewness preferences despite such preferences being inherent to human nature.

I also test other predictions that the model makes about the relationship between skewness, prices, ownership and preferences. First, I find empirical evidence for the model's claim that stocks are pricier and generate lower returns when skewness preferences are high. Second, in line with the model's catering mechanism, I find that the skewness premium has decreased significantly since the 1980s. Third, consistent with the model's predictions, I find that idiosyncratic return skewness is more sensitive to decreasing diversification costs when skewness preferences are high. Fourth, the model predicts that when skewness preferences are high, investors are under-diversified compared to a mean-variance framework because they derive benefits from holding the single stock. This matches previous work analysing individual bro-

kerage accounts (Kumar, 2009). One concern about the disappearance of skewness is that it makes stocks less attractive to investors, raising the cost of capital for firms (Gomez et al., 2020). In my model, the disappearance of skewness originates in the change of diversification costs. The model predicts that as fund fees shrink, prices *increase* instead of decrease. As such, these concerns do not materialize.

The catering mechanism formalised in the model requires no governance intervention – but governance could produce similar empirical patterns if institutional investors increase disciplining and monitoring activities. I examine this alternative by exploiting the fact that the two channels generate distinguishable predictions. Under governance, the effect of institutional ownership on skewness should be traceable to specific governance instruments, should be strongest where monitoring gains are largest, and should depend on who the institutional owners are. Under catering, the effect operates through prices and should be uniform across governance environments. The data favours catering on all three counts: exogenous increases in diversified ownership do not affect CEO pay convexity, leverage, or dividend policy; the effect on skewness does not vary with monitoring intensity; and an exogenous shock to passive ownership that leaves total diversified ownership unchanged has no effect on skewness.

Finally, I consider the real effects of the disappearance of skewness. If the reduction in diversification costs decreases the propensity of firms to undertake positively skewed projects, it is important to consider which kind of projects are affected in practice. I show that one type of positively skewed project is breakthrough innovation. Based on U.S. patent data, and using the granular instrumental variable approach, I find that diversified ownership causally reduces firm-level breakthrough innovation. At the same time, firms decrease R&D spending and both the number and topical diversity of patents registered. The results suggest that a negative consequence of cheap diversification is the reduction of breakthrough innovation in public firms.

Related Literature

This paper builds directly on the literature on idiosyncratic return skewness.[1] A group of papers extends asset pricing theory by introducing skewness preferences and investigate potential implications such as under-diversification (Simkowitz and Beedles, 1978; Kraus and Litzenberger, 1976; Barberis and Huang, 2008; Mitton and Vorkink, 2007). Another strand of literature examines the idiosyncratic return skewness risk premium, where the general consensus reflects a positive risk premium and lower future returns for stocks with high idiosyncratic return skewness (e.g., Boyer et al. (2010); Conrad et al. (2013); Langlois (2020); Driessen et al. (2025)). Kumar (2009) analyses socioeconomic and macroeconomic determinants of skewness preferences using individual account data. Del Viva et al. (2017) link return skewness to accounting measures of growth options. Finally, Gomez et al. (2020) are the first to point out the secular decline in idiosyncratic return skewness since the 1980s but do not offer an explanation for it. This paper adds to the literature by providing a new mechanism linking firm ownership to return skewness. I posit that cheap diversification might be one driver for the disappearance of skewness.[2]

My work also contributes to the rapidly growing literature on the real effects of diversification. Bond and Garcia (2022); Bednarek (2023); Schmalz and Zame (2024); Baruch and Zhang (2022) theoretically study the effects of diversification on information efficiency, welfare, volatility and prices in equilibrium settings. Much work studies these effects also empirically, typically using index inclusions as exogenous shock to firm-level passive ownership (e.g., Sammon (2025); Ben-David et al. (2018)). The weakness of these studies is that their findings are limited to studying cross-sectional (and not equilibrium) effects of passive (and not diversified)

[1] Note that a large literature studies co-skewness/systemic skewness (e.g., Harvey and Siddique (2000); Anghel et al. (2022)). These measures link return skewness to market states and are entirely distinct from measures of idiosyncratic skewness.

[2] This paper does not aim to provide a resolution of the puzzle in Gomez et al. (2020). Instead, the goal is to identify a previously overlooked mechanism that plausibly accounts for part of the observed decline.

ownership. My paper is the first to study outcomes of increased diversification when investors also have skewness preferences. Another innovation is that my model endogenises firm project selection via the choice of skewness of idiosyncratic returns. This feature allows me to link diversification costs to firm behaviour. Under the assumption that institutional ownership is a good proxy for diversified ownership, this paper also innovates by using a granular instrumental variable approach to causally study the effects of diversification.

Finally, this paper contributes to a growing literature on institutional ownership and innovation. The seminal paper linking institutional ownership to increased innovation (Aghion et al., 2013) has recently been challenged (Simeth and Wehrheim, 2024), and the mechanisms remain under debate. Both Luong et al. (2017) and Simeth and Wehrheim (2024) provide convincing evidence that foreign institutional ownership specifically increases firm innovation efforts. Other papers study the effects of passive ownership on innovation. For example, Liu et al. (2019) use Russell inclusion to show that passive ownership increases patent quantity and quality. Few papers analyse breakthrough innovation specifically. Tian and Yang (2023) also use the Russell shock and find that while the index inclusion increases patent quantity, firms shift their efforts to incremental instead of explorative innovation. My paper adds to this literature by providing causal evidence and an equilibrium mechanism operating through catering rather than governance for why firms engage in fewer breakthrough innovation efforts when their owners are more diversified.

Two papers on private firms provide independent support for this mechanism. Bernstein (2015) finds that going public causally reduces patent novelty, and Ebrahimian and Ljungqvist (2025) show that private firms have significantly increased their share of innovation value creation compared to public firms since the 1980s. Both findings are consistent with the claims made in this paper: private firms have less diversified ownership, and their owners therefore retain stronger preferences for the positively skewed payoffs that breakthrough innovation generates. The fact that the public-private innovation gap has widened over the same period that

diversification costs have fallen lends external validity to the catering channel identified in this paper.

1.2 Data

I obtain quarterly data for this study from several sources. Stock market data is taken from the sample of common stocks with positive prices in CRSP and any non-price data from CRSP is winsorized at 2.5 and 97.5%. Firm fundamental data is from Compustat and institutional ownership from 13F filings is obtained via Thomson Reuters. All continuous firm fundamental and ownership variables are winsorized at 1 and 99%. Classifications of 13F investors into different types are obtained from Brian Bushee's personal website[3] and Fama-French factors are from Kenneth R. French's website[4]. Any other data sources will be mentioned when applicable in the paper.

Skewness. I measure Pearson's Skewness, which is defined as the standardized third moment of a distribution: for a random variable X with mean μ and standard deviation σ, it is characterised by

$$skew = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]. \tag{1.1}$$

With n observations of draws x_i, I use the unbiased estimator for skewness (Joanes and Gill, 1998):

$$\widehat{skew} = \frac{\sqrt{n(n-1)}}{n-2}\,\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{3/2}}. \tag{1.2}$$

As in Gomez et al. (2020), I measure return skewness as the skewness of a firm's monthly

[3] https://accounting-faculty.wharton.upenn.edu/bushee/
[4] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

idiosyncratic returns from a 5-year rolling window ($n = 60$) where a firm's idiosyncratic returns are the CAPM residuals. The firm-level skewness measure is calculated for each of these windows separately. To show that none of my results are driven by choice in factor model or by the used measure of skewness, I consider alternative methods. In particular, I also calculate the idiosyncratic returns using the Fama French three-factor model as well as the Carhart four-factor model, and try the Bowley and Kelly measures of skewness as well as L-skewness (Hosking, 1990).[5] All results in this paper, including the downward trend in idiosyncratic return skewness since the 1980s, are robust to such alternative specifications.[6] I plot the trends in Appendix Figure A.1.

Diversified ownership. In this paper, institutional ownership serves as proxy for diversified ownership. As I will explain in more detail below, most of the diversification benefit can already be achieved in portfolios of 20 or fewer stocks. It is therefore plausible to assume that most investments managed by intermediaries, including active mutual funds and hedge funds, are diversified. I refine the measure of diversified institutional ownership by removing holdings from asset managers which are unlikely to benefit from diversification benefits. To this end I exclude all 13F filers with fewer than 20 unique stocks in a given quarter.[7]

Using this proxy necessitates that non-institutional ownership is not diversified. To motivate this assumption, I use facts from the survey of consumer finances (Kelly, 1995; Goetzmann and Kumar, 2008) and from recent work using family office account data of ultra-wealthy investors (Balloch and Richers, 2023). First, data from consumer finance surveys from 1983 (1996)

[5]The Bowley measure is defined as $skew^B = \frac{(Q_{75}-Q_{50})-(Q_{50}-Q_{25})}{Q_{75}-Q_{25}}$ and the Kelly measure as $skew^K = \frac{(Q_{90}-Q_{50})-(Q_{50}-Q_{10})}{Q_{90}-Q_{10}}$. One benefit of the Bowley measure is that it does not incorporate extreme returns. Therefore, it alleviates the concern that the asymmetry of the idiosyncratic return distribution is driven by the inherent asymmetry coming from the -100% downside limit of stock returns.

[6]I am happy to provide results using any of the factor models upon request.

[7]Note that these 'concentrated' 13F filers account for less than 1% of aggregate institutional ownership in an average quarter. I therefore find that this adjustment, as well as other portfolio threshold choices, do not alter any of the findings presented in this paper.

show that the median U.S. direct stock-owner owns a portfolio of only one (three) unique stocks. Rich households, which have more pricing power, could however still be diversified. I find this is not the case: between 2016 and 2022, even the richest U.S. households with wealth levels of more than $100mn. were severely underdiversified. Despite holding more than 100 unique stocks on average, they had higher exposure to idiosyncratic equity risk than the median U.S. stock owner due to concentrated bets. Underdiversification of the rich is also present in my early sample: the median household among the top 2% in the income distribution in 1983 holds only 10 unique stocks. Second, the diversification level of the U.S. stock owner is only weakly correlated with pension or mutual fund investments. This also suggests that diversified investing through intermediaries is not simply a substitute for diversification in individual brokerage accounts.

Because 13F data is only available after 1981, all results using institutional ownership in this study are estimated using data between 1981 and 2024. For any work which estimates or uses investors' skewness preferences the analyses begin in the year 2000 due to power considerations.

I present the summary statistics of key variables in 5-year intervals in Table 1.1. This reflects the data used in most of the paper's empirical tests. For the demand-system estimation in Section 1.3 I use data at quarterly frequency.

The summary statistics reveal interesting features in the data. Most notably, idiosyncratic return skewness has a mean of 0.705 and 80% of observations are positive. This reflects the general understanding that the distribution of idiosyncratic returns is positively skewed. The average profitability is around 5%, market-to-book ratio 14% and institutional ownership 38%. The variables related to patents will be discussed when applicable in Section 1.7.

Table 1.1. Summary statistics.
Summary statistics of key variables taken from snapshots in 5-year intervals. $Skewness$ is the idiosyncratic return skewness from monthly returns in 5-year windows, using the CAPM. $InstitutionalOwnership$ is 5-year lagged institutional ownership. $DividendstoBookEquity$ are dividends divided by book equity. $Log(Mcap)$ is the log of firm market cap. $Log(Assets)$ is log of total assets. $Leverage$ is the book value of total debt divided by the book value of total assets. $Market-to-Book$ is the market value of equity divided by the book value of equity. $Log(Age)$ is the log of firm age in years since IPO. $CapitalExpenditure$ is capital expenditures scaled by total assets. $Investment$ is the annual change in total assets scaled by total assets. $PatentValue(5y)$ is the 5-year sum of a firms patent values (from Kogan et al. (2017)), divided by the firm market capitalisation. $MaxPatentCitations(5y)$ is the logarithm of number of citations from the firms most cited patent in a five year window. $PatentImportance(5y)$ is the 5-year median patent importance for each firm obtained from Kelly et al. (2021).

Variable	N	Mean	Std. Dev.	5th Perc.	25th Perc.	Median	75th Perc.	95th Perc.
Skewness	61388	0.705	0.896	-0.472	0.129	0.558	1.11	2.44
Log (Book Equity)	47798	4.53	2.15	1.04	3.01	4.44	6	8.24
Log (Market Cap)	65658	18.3	2.4	14.7	16.6	18.1	19.9	22.6
Profitability	43866	0.053	0.156	-0.256	0.014	0.0622	0.109	0.335
Investment	44075	0.0878	0.287	-0.326	-0.0248	0.0653	0.177	0.583
Dividends to Book Equity	36223	0.0201	0.0426	0	0	0	0.0256	0.0903
Market-to-Book Ratio	47545	14.3	0.975	12.8	13.7	14.3	14.9	16.1
Institutional Ownership	45156	0.385	0.321	0.00477	0.0867	0.31	0.657	0.961
Leverage	47646	0.558	0.277	0.119	0.349	0.553	0.751	0.949
Log (Age)	51680	2.12	0.914	0.408	1.51	2.2	2.77	3.58
Capital Expenditure	39101	0.0144	0.0205	0	0.00185	0.00747	0.018	0.0547
Patent Value (5y)	67113	0.00197	0.0166	0	0	0	0.000843	0.00952
Patent Max Citations (5y)	67113	0.888	1.41	0	0	0	1.93	3.81
Patent Importance (5y)	67113	0.108	0.25	0	0	0	0.0503	0.776

1.3 Motivating evidence

1.3.1 Revealed skewness preferences

An obvious first step when investigating the disappearance of skewness is to analyse revealed investor preferences for skewness. Without frictions, firms only adopt projects with positively skewed returns if owners have preferences for such projects. While there are already some papers analysing skewness preferences (e.g., Kumar (2009); Lin and Liu (2018)), to the best of my knowledge there are none that capture the time-varying trends which are essential to this study. To this end I closely follow the approach in Koijen et al. (2024).

I model institutional and household equity holdings through an asset demand system that relates portfolio weights to stock characteristics and prices. The framework departs from standard asset-pricing models by directly estimating how investors allocate capital across individual stocks, imposing market clearing to recover equilibrium prices, and using observed heterogeneity in holdings to infer investor-level preferences. I refer to the original paper for the theoretical grounding.

Investor i allocates wealth $A_{i,t}$ in quarter t across a set of assets $n \in \mathcal{N}_i$.[8] The optimal portfolio weight on stock n is assumed to take the logit form:

$$w_{i,t}(n) = \exp\Big(\alpha_{i,t} + \beta_{0,i,t}\, mb_t(n) + \boldsymbol{\beta}'_{1,i,t}\mathbf{x}_t(n)\Big)\, \varepsilon_{i,t}(n), \tag{1.3}$$

where $w_{i,t}(n)$ is the share of investor i's wealth held in stock n, $mb_t(n) = \log(ME_t(n)/BE_t(n))$ denotes log market-to-book equity. The vector $\mathbf{x}_t(n)$ collects other firm characteristics: 5-year idiosyncratic return skewness, size, profitability, dividend ratio, investment and beta. The key assumption here is that the firm characteristics $\mathbf{x}_t(n)$ are exogenous.

The coefficient $\beta_{0,i,t}$ captures the price elasticity of demand: lower values imply more elastic demand, i.e., investors reduce holdings more strongly when prices rise. The vector $\boldsymbol{\beta}'_{1,i,t}$ captures the sensitivity of holdings to observable firm characteristics. The residual $\varepsilon_{i,t}(n)$ represents latent demand, encompassing beliefs, sentiment, or unobserved constraints that are not captured by observables.

A challenge in this estimation is that log market-to-book equity $mb_t(n)$ and latent demand $\varepsilon_{i,t}(n)$ are jointly endogenous, since investors' demand can affect equilibrium prices. For example, large investors could have price impacts from trading. To address this simultaneity, I use an instrumental variable approach as proposed in Koijen et al. (2024), exploiting predetermined

[8] As opposed to Koijen et al. (2024), I include all stocks in the U.S. investment universe in the estimation, leaving no outside asset. This deliberate choice ensures that I do not fail to capture investor preferences particularly associated with small stocks which are more popular with retail investors (Barber and Odean, 2000).

investment mandates.

Let $\mathbb{1}_i(n)$ be an indicator equal to one if asset n belongs to investor i's investment universe $\mathcal{N}_i$, defined as the set of stocks currently or recently held (within the past 11 quarters). The instrument for log market equity is then constructed as

$$z_{e,i,t}(n) = \log\left(\sum_{j \neq i} A_{j,t} \frac{\mathbb{1}_j(n)}{|\mathcal{N}_j|}\right), \tag{1.4}$$

which corresponds to the counterfactual log market equity of stock n if all other investors j were to hold equal-weighted portfolios within their investment universes. This instrument varies across stocks because different investors face distinct exogenous universes due to mandates or benchmarks, and across investors because each investor excludes its own holdings. To provide evidence that investment universes matter for asset allocation, I calculate the share of unique stocks owned by each manager each quarter, which was already owned by the manager in the previous three years. The average share is 86% in the sample, suggesting that most investment decisions fall within stable choice sets. Finally, the identifying assumption is that investment universes and the distribution of wealth across investors are predetermined and exogenous to contemporaneous latent demand shocks.

Under this instrument, equation (1.3) can be estimated via a two-stage least squared regression approach:

$$mb_{e,t}(n) = \tau_{i,t} + \pi_{0,i,t}\, z_{e,i,t}(n) + \boldsymbol{\pi}'_{1,i,t}\mathbf{x}_t(n) + \nu_{i,t}(n), \tag{1.5}$$

$$\log(w_{i,t}(n)) = \alpha_{i,t} + \beta_{0,i,t}\, \widehat{mb}_{e,i,t}(n) + \boldsymbol{\beta}'_{1,i,t}\mathbf{x}_t(n) + \log(\varepsilon_{i,t}(n)) \tag{1.6}$$

I estimate the first stage over the full sample of managers, stocks and time periods, resulting in more than 1.3 million regressions. To assess the validity of the instrument, I plot the share of F-statistics below 10 as well as the share of absolute t-statistics on the instrument's regression

coefficients $\widehat{\pi}_{0,i,t}$ below 1.69 in Figure A.2 in the Appendix. The share of F statistics below 10 drop from 40% at the start of the sample to 10% by the end, while the share of absolute t statistics below 1.69 firmly ranges between 4 and 10%, suggesting the instrument is strong. For the second stage I follow Koijen et al. (2024) to accommodate sparse portfolios and small cross-sectional samples. First, I pool investors in the same type category with similar assets under management into groups with at least 2000 observations of holdings per quarter. Second, I use a two-step shrinkage procedure to estimate the model. In particular, I first estimate equation (1.6) separately for each investor type and quarter to obtain type-quarter target coefficients $\tilde{\beta}_{\text{type},t}$. I then run pooled manager-level shrinkage estimations with demeaned within-manager data (serves as individual manager fixed effects) as in equation (1.7). As in the original paper, I set the ridge hyperparameters $\lambda_0 = 120$ and $\xi = 0.7$ and force the ridge estimation to cap the price elasticity coefficient at 0.999 to ensure downward sloping demand. The resulting coefficients capture time-varying investor preferences for prices and characteristics at the pooled manager-level g.

$$\beta_{g,t} = \arg\min_{\beta} \|y_{g,t} - X_{g,t}\beta\|^2 + \lambda_g \left\|\beta - \tilde{\beta}_{\text{type},t}\right\|^2, \tag{1.7}$$

where $\lambda_g = \frac{\lambda_0}{N_g^{\xi}}$ scales the ridge penalty with the number of holdings N_g.

To visualize these evolving preferences, I compute wealth-weighted averages of each coefficient across investors within type groups and plot 3-year rolling averages from 2000 until 2024. Appendix Figure A.3 plots the demand coefficients also present in the original Koijen et al. (2024) study. The estimates and trends closely resemble their findings. For example, institutional owners have stronger preferences for large firms and lower preferences for dividend stocks than households. The results reveal more interesting patterns among different types of institutional owners. I omit a discussion about these as this paper's focus lies more so in the difference in tastes between institutional investors as a single group and individual investors.

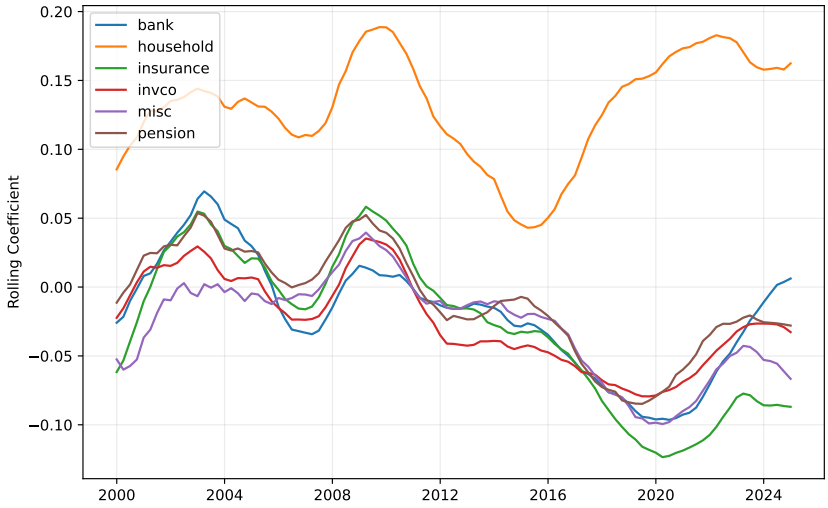


Figure 1.2. Revealed skewness preferences.
This figure plots the 3-year rolling AUM-weighted skewness taste estimates for different types of investors between 2000 and 2024. The taste coefficient is estimated from the Koijen et al. (2024) framework, adjusted to account for idiosyncratic return skewness.

The preferences for skewness from 2000 until 2024 are displayed in Figure 1.2. The figure makes clear that individual investors (marked as 'household') have a positive revealed preference for stocks with skewed returns. Their demand coefficient ranges between 0.1 and 0.2. This implies that for households, a one standard deviation change in skewness (0.9) increases equity holdings by $(e^{0.15\times 0.9} - 1) = 14.4\%$. Other investors, which together make up 13F owners, have either zero or negative revealed preferences for this characteristic. These preferences are in line with the findings from Kumar (2009), who use data from a brokerage house between 1991 and 1996. My results show that these differences in revealed tastes have persisted since then.

It is important to note that these differences in revealed preferences need not reflect differences in inherent tastes. Institutional investors manage diversified portfolios in which idiosyncratic skewness is largely eliminated. For such investors, the marginal benefit of holding any individual skewed stock is small because the portfolio neutralises the lottery-like payoff that makes skewness attractive. Their low revealed preference for skewness is therefore a conse-

quence of their diversification, not evidence that the individuals behind these institutions are indifferent to skewness.

One can estimate the aggregate skewness preference for stock n by value-weighting the preferences $\hat{\beta}_{skew,i,t}$ of its owners using the shares outstanding held by investors. I call this firm-quarter variable $SkewPref_{n,t}$. Unsurprisingly, it has a strong negative correlation with institutional ownership of -0.63. Figure 1.3 plots the mean, median and market cap-weighted aggregate skewness preference for stocks over time. In line with low revealed skewness preferences of institutional owners and their increasing share of stock ownership, aggregate skewness preferences have significantly decreased since 2000.[9] In the next section I study this link cross-sectionally using panel regressions.

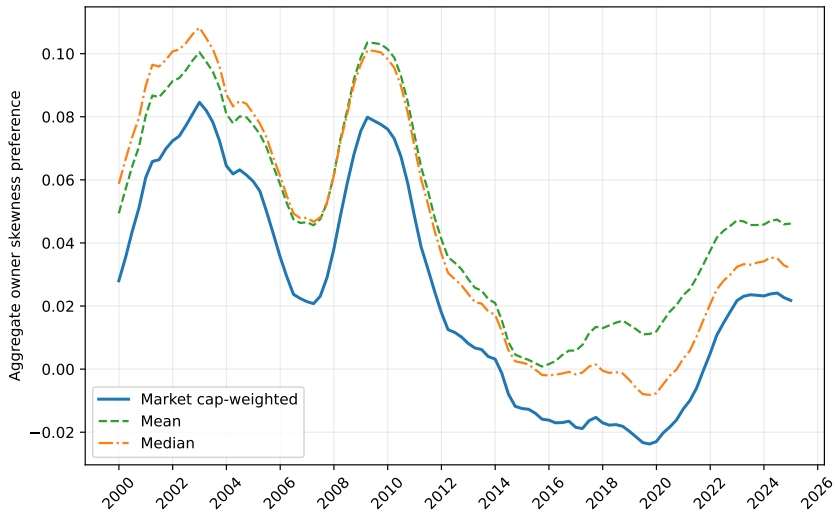


Figure 1.3. Aggregate revealed skewness preferences for stocks.
This figure plots the aggregate skewness preference $SkewPref$ for the average and median stock as well as the value-weighted stock market between 2000 and 2024. $SkewPref$ is defined as each firms owners' aggregate skewness preference, weighted by each owner's shares held.

[9] It is worth considering what explains the fluctuations in skewness preferences over time. In the short-term, stock over- and underpricing could have significant effects. For example, the plot shows that skewness preferences increase after market crashes. This could be the result of stocks with positive skewness being overpriced and hence suddenly gaining larger weights in investors' portfolios. Since this study is primarily concerned with long-term trends, I omit a closer analysis of the short-term determinants of skewness preferences.

1.3.2 Cross-sectional tests

In this section I provide cross-sectional evidence that institutional ownership and owners' skewness preferences are not only related to the skewness of a firm's previous returns, but also predict future realised return skewness. These findings are in line with an economic mechanism in which firms select projects according to their owners' preferences.

In Table 1.2 I report results from linear least-squared regressions where I predict return skewness from a 5-year window. Because skewness is measured at a 5-year rolling horizon, running quarterly regressions is problematic due to serial correlation in the error term. To address this issue I limit the sample to one observation per firm every five years.[10] I use a host of firm characteristics as control variables as well as time and industry fixed effects. Standard errors are clustered at the firm level.

The coefficients for lagged institutional ownership and owners' skewness preferences are both economically and statistically significant, even in models including firm-level control variables (columns 3 and 4). High institutional ownership and low skewness preferences of owners predict lower future return skewness. This could indicate that idiosyncratic return skewness is partly ownership-driven. All else equal, a 10 percentage point increase in lagged diversified institutional ownership is approximately associated with a relative drop in skewness of between 5 and 10%.

The evidence provided in the panel regressions is by no means causal. To find a causal relationship between (institutional) ownership characteristics and firm outcomes the use of an exogenous shock to diversified ownership would be the appropriate method. Ownership shocks frequently used in the finance literature shock either passive ownership only (e.g., Russell in-

[10] In Appendix Table A.2 I report results from overlapping regressions with firm fixed effects. The results are similar in magnitude and significance.

Table 1.2. Skewness, institutional ownership and owners' skewness preferences.
The dependent variable is the idiosyncratic return skewness from a 5-year window estimated with the CAPM. $InstOwn_{t-5y}$ is 5-year lagged institutional ownership and $SkewPref_{t-5y}$ is the 5-year lagged owner skewness preference estimate. $Log(Mcap)$ is the log of firm market cap. $Log(Assets)$ is log of total assets. $Leverage$ is the book value of total debt divided by the book value of total assets. $Market-to-Book$ is the market value of equity divided by the book value of equity. $Log(Age)$ is the log of firm age in years since IPO. $CapitalExpenditure$ is capital expenditures scaled by total assets. $Investment$ is the annual change in total assets scaled by total assets. $Spread$ is the average daily bid-ask spread scaled by price. Standard errors are clustered at the firm-level.

	Idiosyncratic Return Skewness			
	(1)	(2)	(3)	(4)
$InstOwn_{t-5y}$	−1.136***		−0.548***	
	(0.024)		(0.030)	
$SkewPref_{t-5y}$		6.111***		2.788***
		(0.154)		(0.168)
$Log(MCap)$			−0.144***	−0.156***
			(0.026)	(0.027)
$Log(Assets)$			0.010	0.009
			(0.027)	(0.027)
$Leverage$			−0.124*	−0.141*
			(0.075)	(0.076)
$Log(Age)$			−0.002	−0.002
			(0.013)	(0.013)
$Investment$			0.200***	0.216***
			(0.029)	(0.030)
$Market-to-Book$			0.164***	0.176***
			(0.027)	(0.027)
$CapEx$			−1.341***	−1.216***
			(0.442)	(0.448)
$Spread$			2.237***	2.434***
			(0.416)	(0.420)
Num. Obs.	27 439	26 024	21 303	21 055
R^2	0.15	0.12	0.21	0.21
FE: Quarter	X	X	X	X
FE: Industry	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

clusion)[11] or common ownership (e.g., asset manager mergers)[12]. Both of these shocks are not suitable for this purpose. In the next section I propose an alternative identification method.

[11] See Gormley and Kim (2025) for more details on ownership changes around Russell Index reconstitutions.

[12] E.g., Azar et al. (2018); Koch et al. (2021).

1.3.3 Causal evidence

To assert causality in this context, we need exogenous shocks to a firm's diversified stock ownership. To this end, I resort to the granular instrumental variable approach (Gabaix, 2011; Gabaix and Koijen, 2021) applied to portfolios of institutional owners. The key idea of granular instrumental variables in the institutional ownership context is to aggregate idiosyncratic shocks of flows to asset managers into a firm-level instrument that is orthogonal to common and firm-level factors (Zhang, 2024). As before, I posit that institutional ownership is a good proxy for diversified ownership and therefore claim that any exogenous shock to institutional ownership is still suitable in this context.

I provide a thought experiment to build intuition for the identification: consider the sudden closure of a mutual fund. This closure unexpectedly reduces aggregate demand for stocks within this fund by the asset manager. The outflow of institutional ownership in these stocks is completely exogenous to firm fundamentals because all firms are equally affected. Further, the outflow is exogenous to macroeconomic observables because the closure is idiosyncratic to this specific asset manager. The key idea of the granular instrumental variable method is to aggregate these idiosyncratic shocks at the firm-time level.

As with any instrumental variable the two key assumptions are relevance and exogeneity. The relevance condition requires that sizeable shocks to large investors exist such that aggregate idiosyncratic shocks have power. I take this condition as given for a few reasons. First, a growing literature shows that the U.S. asset management industry is highly concentrated and dominated by a few large players (e.g., Amel-Zadeh et al. (2022)). Second, portfolio constraints such as fixed equity shares and capital requirements are common and have real effects on investment flows (Harvey et al., 2025). I will later show empirically that the instrument I create for this context is indeed a strong predictor of institutional ownership.

The second assumption is exogeneity. The granular approach in Gabaix and Koijen (2024)

ensures that the estimated instrument is orthogonal to firm fundamentals as well as observed and latent systematic factors which could predict a firm's ownership makeup. I find ample evidence that my instrument is indeed exogenous. First, the instrument is close to normally distributed and has a mean near zero (see Appendix Figure A.4). Second, my results are close to unchanged when controlling for additional firm fundamentals, when changing the number of latent factors accounted for, or when including industry, time and industry-time fixed effects. Third, I find no evidence of pre-trends in the outcome variable: in untabulated tests I find that my estimates remain unchanged when including lagged outcome variables in the regressions. Finally, I rely on tests in Zhang (2024) which show that the instrument constructed with this method strongly predicts other instances of exogenous demand shocks on the stock level such as large fund in- and outflows (Edmans et al., 2012) as well as reinvestment of dividends (Schmickler and Tremacoldi-Rossi, 2022).

Below I outline the construction of the granular instrument, which closely follows Zhang (2024).

1.3.3.1 Instrument construction

Step 1:

I aggregate all 13F holdings across investor types obtained from Brian Bushee's personal website[13]. For investor type i of firm n at time t I then calculate the change in holdings each quarter $\Delta q_{i,t}(n) = \frac{Q_{i,t}(n) - Q_{i,t-1}(n)}{Q_{i,t-1}(n)}$ where $Q_{i,t}(n)$ is the value of owned shares for any given quarter. I then calculate the time-invariant regression weight $E_i(n) = \frac{\sigma_i^{-2}(n)}{\sum_{i=1}^{I} \sigma_i^{-2}(n)}$ where $\sigma_i^2(n) = \max(\sigma_i^2(n), \text{median}(\sigma_i^2(n)))$. This allows for heteroskedasticity of errors (Gabaix and Koijen, 2024).

Step 2:

For each firm n, I estimate the following panel regression across investors and time with regres-

[13] https://accounting-faculty.wharton.upenn.edu/bushee/

sion weights $E_i(n)$:

$$\Delta q_{i,t}(n) = \alpha_i(n) + \beta_t(n) + \gamma_i(n)\eta_t^o(n) + u_{i,t}(n) \tag{1.8}$$

where $\alpha_i(n)$ are investor fixed effects, $\beta_t(n)$ are time fixed effects and $\eta_t^o(n)$ is an observable factor which predicts changes in stock holding. I follow the convention and use quarterly GDP growth obtained from the St. Louis Fed. Then, the residual $u_{i,t}(n)$ captures unexplained variation in investor demand.

Step 3:

To account for unobserved common shocks to investor demand, I apply Principal Component Analysis (PCA) to the weighted residual panel $\sqrt{E_i(n)}u_{i,t}(n)$. Let the first K latent factors be denoted by $\eta_{k,t}^l(n)$ with corresponding investor loadings $\gamma_{k,i}(n)$. I then estimate:

$$\sqrt{E_i(n)}u_{i,t}(n) = \sum_{k=1}^{K} \gamma_{k,i}(n)\, \eta_{k,t}^l(n) + \varepsilon_{i,t}(n), \tag{1.9}$$

where $\varepsilon_{i,t}(n)$ represents the investor-level idiosyncratic demand shock after removing both observed and latent common components. For the empirical analysis I set $K = 2$ as is convention. The exogeneity assumption requires that this error term cannot be explained by macroeconomic or firm-level factors.

Step 4:

I then aggregate the idiosyncratic, firm-investor pair demand shocks to the firm level by taking the sum weighted by each investor i's position in stock n, $S_{i,t}(n) = \frac{Q_{i,t-1}(n)}{\sum_j Q_{j,t-1}(n)}$. This yields the firm-quarter level granular instrument:

$$z_t(n) = \sum_i [S_{i,t}(n) - E_i(n)]\varepsilon_{i,t}(n). \tag{1.10}$$

Step 5:

I winsorize $z_t(n)$ at 2.5 and 97.5% to eliminate the impact of severe outliers. I then sum the

instrument across 5 years (20 quarters) at the firm-level to $GIV(5y)_t(n) = \sum_{t=-20}^{0} z_t(n)$. This allows the instrument to be measured at the same time frame as the primary outcome variable $Skew$. Finally, I normalize the resulting instrument to be mean zero and have unit standard deviation.

1.3.3.2 Causal relationship

To estimate the causal relationship between diversified ownership and idiosyncratic return skewness I run two-stage least squared regressions in which the lagged instrument $GIV(5y)_{t-5y}$ predicts fitted values of lagged institutional ownership $\widehat{InstOwn}_{t-5y}$ in the first stage.[14] I then use the fitted values to predict skewness over the next 5-year interval for the second stage. Figure 1.4 shows the measurement timelines for key variables in the empirical setup. By construction, the instrument and outcome window do not overlap, ensuring that the estimated effects cannot arise mechanically from shared observations. This setup allows me to measure causally how exogenous changes in firm-level institutional, and therefore diversified, ownership affects subsequent return skewness.

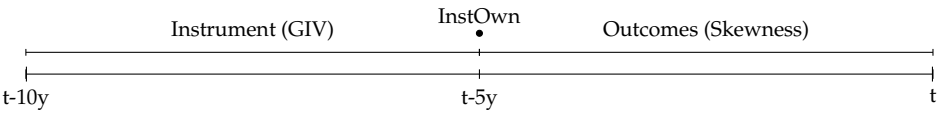


Figure 1.4. Timeline of measurement windows for causal estimation.
The granular instrument (GIV) is constructed from shocks occurring strictly in the pre-period $[t-10y,\ t-5y]$. Institutional ownership is measured at $t-5y$, and firm outcomes (idiosyncratic skewness and later, innovation) are computed over the subsequent window $[t-5y,\ t]$.

The results are shown in Table 1.3. Both regressions include industry and time fixed effects, while in the second column I also include the same time-varying firm-level variables as in the cross-sectional panel regressions. Panel A in Table 1.3 shows that the instrument is highly

[14] It is not strictly necessary to use a two-stage setup in this setting. I choose this setup because (1) the causal chain is easily understood, (2) the first stage serves to show the relevance of the instrument and (3) because the regression coefficients are more easily interpretable and comparable to the evidence presented earlier. The results are robust to predicting skewness directly with the lagged instrument.

significant and the first-stage regressions have F-statistics well above the critical value of 10. In Panel B the results show that an exogenous 10 percentage point increase in institutional ownership linearly predicts a reduction in idiosyncratic return skewness over the following five years of 0.23 units, which corresponds to 33% of the mean and 25% of the standard deviation. The inclusion of firm-level control variables does not change the point estimate significantly. In Appendix Table A.3 I show that the results are robust to including the orthogonal firm-level latent components as controls, increasing the number of latent factors to three, as well as including lagged skewness to account for a potential pre-trend.

Table 1.3. Causal relationship between institutional ownership and skewness
This table shows the results of a two-stage least squared regression predicting idiosyncratic return skewness from a 5-year window estimated with the CAPM. The instrument for lagged institutional ownership is $GIV(5y)_{t-5y}$. In column (1) the model has no control variables besides industry and time fixed effects. In column (2) the model also includes the time-varying firm-level controls from Table 1.2. Standard errors are clustered at the firm level.

	Idiosyncratic Return Skewness	
	(1)	(2)
Panel A: First Stage		
$GIV(5y)_{t-5y}$	0.022***	0.011***
	(0.003)	(0.002)
Num. Obs.	26231	20831
F-statistic	193.97	68.29
R^2	0.29	0.55
Controls		X
FE: Quarter	X	X
FE: Industry	X	X
Panel B: Second Stage		
$\widehat{InstOwn}_{t-5y}$	-2.394***	-2.324***
	(0.311)	(0.605)
Num. Obs.	26231	20831
R^2	0.05	0.20
Controls		X
FE: Quarter	X	X
FE: Industry	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

To summarise, the empirical results in this chapter show a causal link between diversi-

fied institutional ownership and realised idiosyncratic skewness. However, it is still unclear through which channel diversified ownership reduces idiosyncratic skewness. Therefore, in the following section I propose a simple equilibrium model where a firm's idiosyncratic return skewness is endogenised via a project choice problem. The model can reconcile the downwards trend in return skewness over the past decades and link it to the rise in diversified ownership.

1.4 Model

The model loosely builds on Schmalz and Zame (2024) but innovates in key areas. Most notably, I model firms as value-maximisers by endogenising their choice of skewness of idiosyncratic shocks. The second large deviation lies in the distribution of firm profits, where I assume continuous distributions to separate the effects of risk and skewness in the firm's idiosyncratic shock. It is essential to note that the purpose of the model is not to make quantitative predictions about firms, investors or the economy. The model is highly stylised and serves simply to make qualitative predictions, which I find are in line with empirical evidence. Below I first describe the model setup before discussing the solution method and parametrisation.

1.4.1 Environment

M investors indexed by $i \in \mathcal{I}$ choose portfolios among three assets: a stock, a diversified fund, and a riskless bond. The economy contains N identical firms who endogenously choose the skewness of their idiosyncratic shock. Each firm has one share outstanding, in line with contemporary research on stock market inelasticity (Gabaix and Koijen, 2021).[15] The model reflects a two-period economy.

The stock delivers a stochastic cash payoff per share Π with mean and higher moments that depend on a continuous choice of skewness $s \geq 0$. I denote the expected per-share payoff

[15]This is, of course, a gross simplification. While in the short term equity supply might well be fixed, in the long term this is not the case (Sammon and Shim, 2024).

(before the skewness punishment cost) by π_0. The realized payoff can be decomposed as

$$\Pi \;=\; \pi(s) \;+\; \delta \;+\; \varepsilon(s), \tag{1.11}$$

where δ is a symmetric systematic mean-zero shock with standard deviation σ_δ,[16] and $\varepsilon(s)$ is an asymmetric idiosyncratic mean-zero shock with standard deviation σ_ε and skewness $\gamma_1(s)$. The shock is drawn from a skew normal distribution (Azzalini, 2005) with skewness parameter s determined endogenously by the firms.[17] The shocks are not correlated.

The firm can increase its idiosyncratic shock's skewness $\gamma_1(s)$ by adjusting the skewness parameter $s \in [0, S_{\max}]$. In line with U.S. empirics (see Appendix Table A.1), increasing skewness comes at a linear cost to expected payoff. This cost can be understood as investment in areas outside of the firm's core competency, by the fact that firms with moonshot innovation projects are more likely to fail, or by the fact that moonshot innovation projects take longer to pay off.

$$\pi(s) \;=\; \pi_0\Big(1 - \alpha s\Big), \qquad \alpha \geq 0 \tag{1.12}$$

The stock trades at price p. The fund is a simple diversified instrument that grants exposure to firm returns without idiosyncratic shock at a cost $k \geq 0$. This exogenous parameter can be understood as the fund's fee. The bond pays a deterministic payoff 1 per unit and trades at price q.

Investors are heterogenous in terms of wealth and risk aversion. Investor i has wealth w_i, risk-aversion parameter $A_i \geq 0$, and skewness preference coefficient $B_i \geq 0$. Wealth is exponentially distributed and I assume risk aversion is negatively related to wealth (Bali et al.,

[16] In unreported tests I allow the systematic shock to have a negatively skewed distribution. While this change makes the fund less attractive to investors, the results yield the same predictions.

[17] Keeping the mean and variance of the shock fixed implies that idiosyncratic skewness $\gamma_1(s) = \frac{\left(\frac{4-\pi}{2}\right)\left(\omega\sqrt{\frac{2}{\pi}}\right)^3}{\left(1-\frac{2\delta^2}{\pi}\right)^{3/2}}$, where $\omega = \frac{s}{\sqrt{1+s^2}}$.

2023), while skewness preferences are assumed independent of risk aversion (Bougherara et al., 2021) and constant across investors such that $B_i = B$.[18]

1.4.2 Investor preferences and choices

I follow the literature on higher-order preferences and use the Kraus and Litzenberger (1976) third-order Taylor approximation of logarithmic utility to define investor i's mean-variance-skewness utility function (equation 1.13). The advantage of using this type of utility is that I can isolate the effect of skewness preferences on equilibrium outcomes. Furthermore, the function collapses to mean-variance utility when $B = 0$, enabling comparisons to the traditional Markowitz framework (Levy and Markowtiz, 1979).

$$U_i \;=\; \mathbb{E}[C_i] \;-\; \frac{A_i}{2}\,\mathrm{Var}(C_i) \;+\; \frac{B_i}{6}\,\mathrm{Skew}(C_i). \tag{1.13}$$

Let $(x_{S,i}, x_{F,i}, x_{B,i})$ denote investor i's nominal holdings of stock, fund and bond respectively. The budget constraint at date 0 is

$$p(x_{S,i} + x_{F,i}) \;+\; q\,x_{B,i} \;=\; w_i. \tag{1.14}$$

Given price p, diversification cost k, and the payoff parametrization, investor i chooses $(x_{S,i}, x_{F,i}, x_{B,i})$ to maximize (1.13) subject to (1.14) and nonnegativity constraints (no shorting).

The three cumulants of consumption from investor i's portfolio can be written as:

$$\mathbb{E}[C_i] = (x_{S,i} + \kappa\, x_{F,i})\,\pi(s) \;+\; x_{B,i}, \tag{1.15}$$

$$\mathrm{Var}(C_i) = (x_{S,i} + \kappa\, x_{F,i})^2\sigma_\delta^2 \;+\; x_{S,i}^2\sigma_\varepsilon^2, \tag{1.16}$$

$$\mathrm{Skew}(C_i) = x_{S,i}^3\;\gamma_1(s)\;\sigma_\varepsilon^3, \tag{1.17}$$

[18] In unreported tests I relax this assumption. The main conclusions still hold.

where $\kappa = \frac{1}{1+k}$. The coefficient $\gamma_1(s)$ is the third standardized moment of the distribution of the firm's idiosyncratic shock and represents its skewness. The key modelling choice is that idiosyncratic risk enters solely through direct stock exposure. Note that in the model the fund is non-redundant, making asset markets incomplete.

1.4.3 Market equilibrium and firm choice

Given a candidate firm skewness s and a diversification cost k, equilibrium price p clears the stock market:

$$\sum_{i \in \mathcal{I}} \left(x_{S,i}(p; k, s) + x_{F,i}(p; k, s) \right) \;=\; N. \tag{1.18}$$

The firm chooses skewness s anticipating the equilibrium price $p(p \dashv s, k)$ and the aggregate demand response. The firm chooses the skewness parameter s to maximize the equilibrium price p:

$$s^*(k) \;=\; \arg \max_{s \in [0, S_{\max}]} \; p(k, s). \tag{1.19}$$

The firm decision incorporates a trade-off between higher skewness demand and the mean payoff reduction in (1.12).

1.4.4 Parameters

Table 1.4 lists all of the key parameters for the model. I loosely follow Schmalz and Zame (2024) in matching key distributional parameters to the U.S. economy around 1980. I assume there are 100 mn. investors with total wealth of 200mn. Wealth and prices are measured in units of 10,000. Wealth is exponentially distributed and negatively correlated with risk aversion. This represents a Gini coefficient of 0.5. Risk aversion is uniformly distributed with values between 0.1 and 1.5. There are 5,000 listed firms with base profit of 50,000 each. Both the systemic and idiosyncratic shock to firm profit are mean-zero with a standard deviation of 25,000. The skewness punishment parameters of firm profit are $\alpha = 0.01$. I set the exogenous bond price to

$2/3$ and each firm issues a single share which can be held in fractions.

Table 1.4. Model parameters.
This table presents the parameter choices made to numerically solve the model.

Investors	
Number of investors	100,000,000
Mean wealth	2
Risk aversion	$A_i \sim \mathcal{U}[0.1,\ 1.5]$
Wealth distribution	$w_i = \left(\frac{10}{1-e^{-5}}\right) e^{-A_i}$
Firms	
Number of firms N	5,000
π_0	50,000
δ	$\delta \sim \mathcal{N}(0, 25000^2)$
ε	$\varepsilon \sim \mathcal{SN}(0, 25000^2, s)$
α	0.01
Market	
Bond price q	0.667
Shares per firm	1

1.4.5 Solution strategy

I look for numerical solutions to this model because analytical solutions are impractical. For example, the cubic term $x_{S_i}^3$ entering investor utility because of skewness preferences leads to quadratic FOCs. This creates multivalued candidate solutions and corner cases alongside the nonnegativity constraints. The nonlinearity creates further complexities when considering that investors are heterogenous and their demand must clear the market. The numerical approach is in line with other theoretical research on skewness (Mitton and Vorkink, 2007; Barberis and Huang, 2008). For each pair of (B, k) in the parameter space I follow these steps to obtain solutions:

1. **Obtain optimal portfolios:** for fixed diversification cost k, skewness s, skewness prefer-

ence B and price p, I solve each investor type i's portfolio choice problem:

$$\max_{x_{S_i}, x_{F_i}, x_{B_i} \geq 0} U_i(x_{S_i}, x_{F_i}, x_{B_i}) \quad \text{s.t.} \quad p(x_{S_i} + x_{F_i}) + q x_{B_i} = w_i \tag{1.20}$$

For interior solutions I solve the quadratic FOCs while I derive optimal portfolios for exterior solutions numerically. Among all candidate portfolios, the optimal portfolio is the one that generates the highest investor utility U_i.

2. **Solve for equilibrium price:** given (k, s) and optimal portfolios $\{x^*_{S_i}(p), x^*_{F_i}(p)\}_i$ for each price p, aggregate demand for shares is:

$$D(p; k, s) = \sum_{i=1}^{T} \mu_i \left(x^*_{S_i}(p) + x^*_{F_i}(p) \right) \tag{1.21}$$

where μ_i is the number of investors of type i and T is the number of discrete types. Market clearing requires $D(p) = N$. I solve this using bisection on the price interval $[p_{\min}, p_{\max}] = [\$1{,}000, \$10^{10}]$.

3. **Solve for optimal skewness:** for fixed diversification cost k, the firm chooses skewness to maximize total revenue:

$$s^*(k) = \arg \max_{s \in [0, S_{\max}]} p^*(k, s) \cdot N \tag{1.22}$$

I use a golden section search algorithm, which can account for a non-concave function, and account for corner solutions, to find potential candidates for s. For each candidate s, I then compute equilibrium price $p^*(k, s)$ and select the s^* that maximizes p^*.

1.4.6 Discussion

Before analysing the solutions to the model, it is worth addressing some concerns. A first concern is that my model provides no governance mechanism. How can owner preferences then affect firm policy? My model predicts there need not be any communication between owners and managers. The model provides an equilibrium framework where the mechanism arises simply from firms that maximize their valuations. This necessitates that they take into account their owners' preferences. Such 'catering' mechanisms are not novel: Baker and Wurgler (2004) show that firms change their dividend policy to "cater to implied investor demand" to capture a dividend premium. In a similar vein, Polk and Sapienza (2008) show that firms might undertake negative NPV projects if investors put a premium on investment.[19]

A second concern is related to the modelling choice for the fund return. In the model the fund eliminates any idiosyncratic firm shock, which represents full diversification typically associated with index funds. Because index funds only became widely available in the late 1990s, one might argue that my model does not reflect the investor choice set at the onset of the decline in skewness in the 1980s. It is important to consider that the benefits of diversification can be realised even in small portfolios. Simkowitz and Beedles (1978) simulate portfolio characteristics using real stock data and show that 80% of idiosyncratic volatility and 92% of idiosyncratic skewness disappear in portfolio returns after including just 5 stocks.[20] Hence, a diversified portfolio need not be an indexed portfolio.

A third concern is that in my model every investor has the same inherent skewness preferences, despite the previous evidence that institutional owners have lower revealed tastes for stocks with positive idiosyncratic skewness. This deliberate choice reflects a key feature of the

[19] In section 1.6 I empirically test whether my results might be driven by a governance channel but no find no evidence in support of this alternative hypothesis.

[20] I replicate these results using Monte Carlo simulations with the sample from this study. I find that all positive skewness disappears after including only 20 stocks in a portfolio. Active mutual funds typically hold between 50 and 120 unique stocks (Pollet and Wilson, 2008).

model: *revealed* preferences are estimated from asset allocations and are distinct from *inherent* preferences defined in the utility function. The revealed preferences reflect the trade-off between investing in the stock and fund. If the opportunity cost of investing in the single stock is too large, asset allocations might reflect low revealed preferences for skewness.

1.4.7 Results

The model makes key predictions about the relationship between skewness preferences, idiosyncratic return skewness of firms, stock prices, and diversification costs. The purpose of this subsection is to discuss the theoretical results and extract testable predictions.

The key innovation of the model is to endogenise the skewness of idiosyncratic firm returns. How does the firm's choice in optimal skewness s^* vary in fund fees and skewness preferences? Figure 1.5 shows that for low values of B and k the firm elects projects with symmetric returns. In the model, the firm only picks $s^* > 0$ if there is any allocation to the stock *and* when skewness preferences are sufficiently high. In the edge case where the fund is free and skewness preferences are zero the fund dominates the stock and the firm caters only to index fund owners who prefer higher mean return $\pi(s)$ by setting $s^* = 0$. On average, the optimal skewness is increasing in both fund fees and skewness preferences – albeit not monotonically. This implies that as diversification costs decrease, firms shy away from skewed projects to appeal to index investors' interests. In practice, cheap diversification leads to an increase in investments through financial intermediaries. Hence, one would expect institutional ownership to negatively predict idiosyncratic return skewness. Finally, idiosyncratic return skewness is expected to positively correlate with skewness preferences.

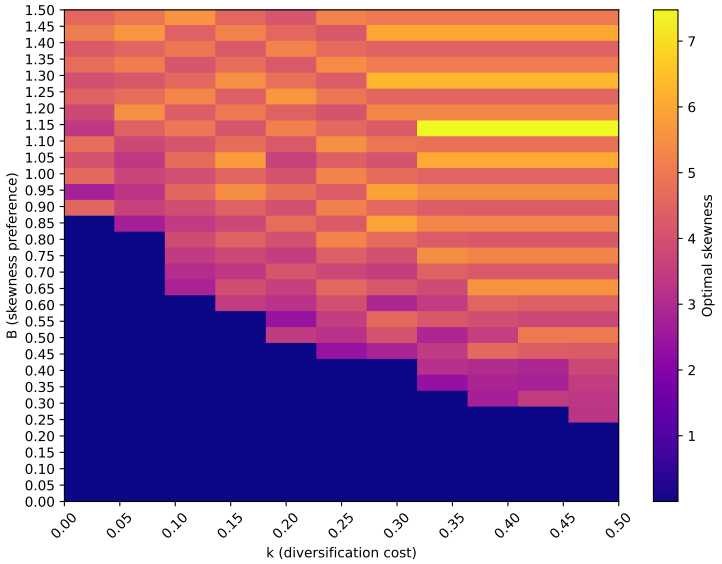


Figure 1.5. Optimal skewness for (k, B) pairs.
This figure plots the optimal choice of firm skewness s^* for different pairs of fund fee k and skewness preference B.

Hence the model's first three predictions are:

P_1 : Idiosyncratic return skewness is increasing in diversification costs.

P_2 : Idiosyncratic return skewness is decreasing in institutional ownership.

P_3 : Idiosyncratic return skewness is increasing in skewness preferences.

An additional observation is that when skewness preferences are very low there is less variation in optimal skewness as k changes. This is intuitive: when investors care only little about skewness, their investment choice is more sensitive to fund fees and expected mean returns. The firms cater to these preferences:

P_4 : Idiosyncratic return skewness is more sensitive to diversification costs when skewness preferences are high.

Beyond the predictions on diversification and ownership which are new to the academic literature, the results from the model also match those of previous work on skewness (e.g., Mitton and Vorkink (2007); Simkowitz and Beedles (1978)). The model matches previous works' predictions about returns and prices under skewness preferences. Panel A of Figure 1.6 shows that for any fund fee k, the stock price is increasing in skewness preferences B, with the slope being highest for large B. This implies that skewness preferences can lead to reductions in returns due to stock overpricing. In panel B of the figure I plot the percentage price difference between the equilibrium price $p(s^*)$ and a counterfactual price $p(s = 0)$ for different parameter pairs of (k, B). Even though the model does not feature heterogenous firms, this can be interpreted as a skewness premium. The pattern reveals that investors are willing to pay higher premia when diversification is expensive. Less surprising is that this figure again confirms that higher skewness preferences are associated with larger premiums. Note that because the equilibrium price at $s = 0$ does not depend on B, the two panels are analytically equivalent. The premium investors pay for stocks when they have higher skewness preferences arises entirely through the firm's endogenous skewness choice, not through a direct effect of preferences on prices independent of firm behaviour.

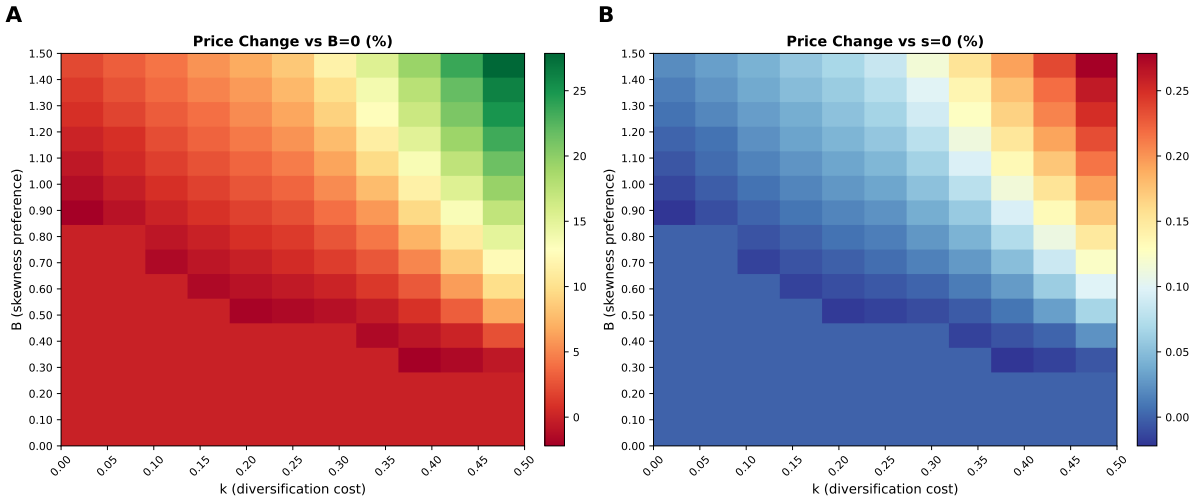


Figure 1.6. Equilibrium outcomes for (k, B) pairs.
Panel A plots the percentage stock price change for fixed k's but varying skewness preferences B: $\frac{P(k,B)-P(k,0)}{P(k,0)}$ and Panel B plots the percentage stock price change compared to allocations where $s = 0$: $\frac{P(k,B,s=s*)-P(k,B,s=0)}{P(k,B,s=0)}$ for exogenous pairs of (k, B).

Thus, the specific predictions are:

P_5 : Equities with higher owners' skewness preferences trade at a premium.

P_6 : The skewness premium is increasing in diversification costs.

Second, the model also matches results on under-diversification versus mean-variance frameworks (Figure 1.7). In this model, under-diversification can be interpreted as each investor's share of risky investment in the single stock $\left(\frac{X_{S,i}}{X_{S,i}+X_{F,i}}\right)$. When skewness preferences are not prevalent ($B = 0$) the stock allocation and under-diversification are zero for small k because the fund dominates the stock in the mean-variance plane. Total stock allocation and under-diversification increase with B and k.

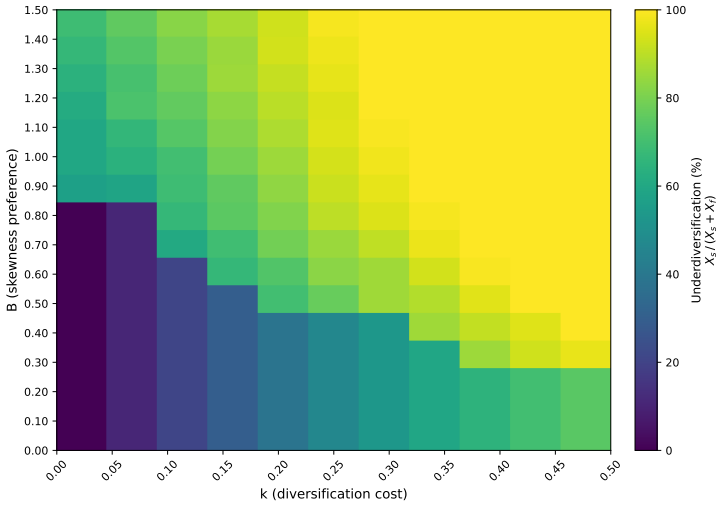


Figure 1.7. Under-diversification for (k, B) pairs.
This figure plots the percentage of total risky wealth in the economy allocated to the stock $\left(\frac{X_{S,i}}{X_{S,i}+X_{F,i}}\right)$ for exogenous pairs of (k, B).

Hence, the final prediction is:

P_7 : Skewness preferences are associated with portfolio under-diversification.

Finally, the model also provides some guidance on a concern raised in Gomez et al. (2020). The authors state that "the significant downward trend in skewness (...) could have dramatic consequences for asset prices" as "the disappearance of firm-level skewness makes all stocks less attractive to investors." In real terms, the disappearance of skewness could reduce prices while increasing the cost of capital. Figure 1.8 shows the equilibrium stock price for pairs of (k, B). Under the conditions in the model prices increase when diversification becomes cheaper (Schmalz and Zame, 2024). This is true for any parameter of skewness preference in the specified range. In the model, the disappearance of skewness is solely the result of cheap diversification opportunities. Therefore, any price change must be understood to originate from the fund fee k. If the model is to be believed, the disappearance of skewness is therefore associated

with a price *increase*, as opposed to a price decrease as hypothesized in Gomez et al. (2020). The results imply that the concerns raised by the authors do not materialize in this framework.

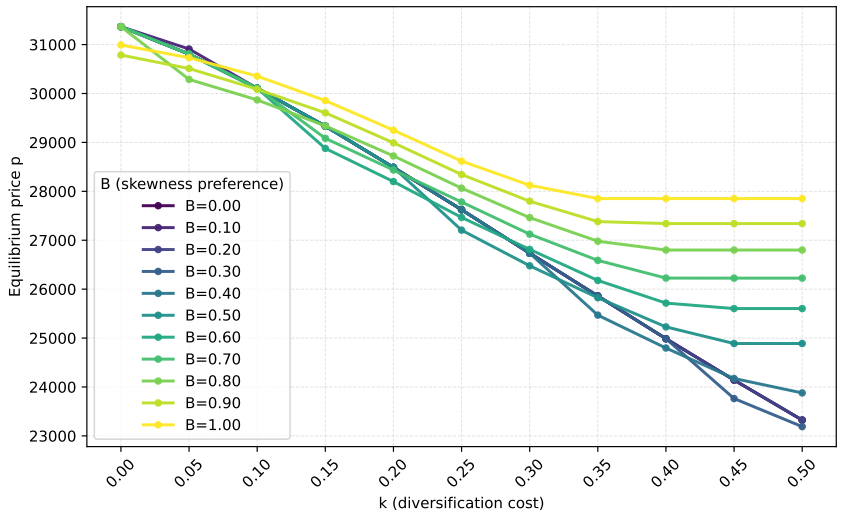


Figure 1.8. Equilibrium price for (k, B) pairs.
This figure plots the equilibrium stock price for firms in the model, for different pairs of exogenous variables k and B.

1.5 Empirical model validation

This section asks whether the predictions from the model reflect real relationships present in the U.S. stock market. To this end, I provide empirical evidence in support of the six predictions laid out in the previous section.

The key predictions of the model link ownership and diversification to the secular decline in idiosyncratic return skewness since the 1980s:

P_1 : Idiosyncratic return skewness is increasing in diversification costs.

P_2 : Idiosyncratic return skewness is decreasing in institutional ownership.

P_3 : Idiosyncratic return skewness is increasing in skewness preferences.

The motivating evidence in section 1.3 provides strong support for these predictions. Figures 1.1 and 1.3 show that aggregate skewness preferences and realised idiosyncratic return skewness have decrease in the last decades; trends contemporary with the increase in institutional ownership and cheap diversification opportunities. Furthermore, the results from Tables 1.2 and 1.3 showed that these relationships are causal and present in the cross-section. Firms with higher (lower) diversified institutional ownership (owners' skewness preferences) have lower future idiosyncratic return skewness, in line with firms catering to their investor base.

The model's 4th prediction is that when skewness preferences are very low there is less variation in optimal skewness as k changes.

P_4 : Idiosyncratic return skewness is more sensitive to diversification costs when skewness preferences are high.

I find this matches cross-sectional evidence from U.S. data. Figure 1.9 plots the 3-year rolling median skewness (and its linear trend) over time for two types of stocks selected in the year 2000: those in the top 10% of owners' skewness preferences $SkewPref$ and those in the bottom 10%. The plot reveals that as fund fees shrank, stocks with owners who have high skewness preferences have seen a larger relative decline in skewness.

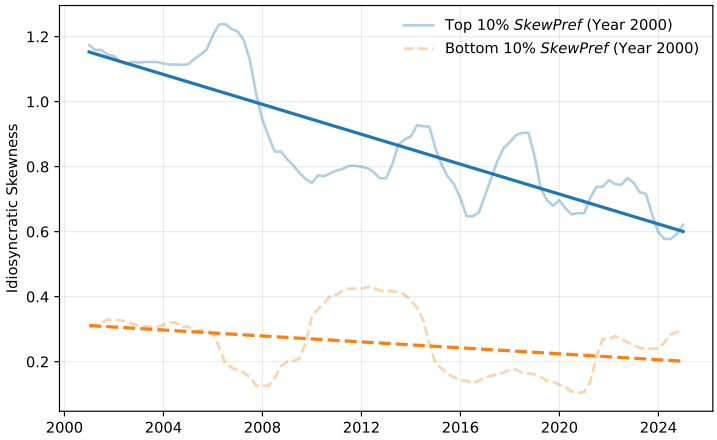


Figure 1.9. Median skewness for $SkewPref$-sorted portfolios.
This figure plots the 3-year rolling median idiosyncratic return skewness, and its linear time trend, for two groups of stocks: stocks in the top and bottom decile of $SkewPref$ at the start of the estimation sample.

Prediction 5 concerns stock valuations:

P_5 : Equities with higher owners' skewness preferences trade at a premium.

To assess whether this prediction is true I use the quarterly stock-level $SkewPref$ variable and conduct portfolio tests. For each quarter between 2000 and 2025 I build equal- and value-weighted portfolios going long all stocks in the quintile with the highest $SkewPref$ and short the quintile with the lowest $SkewPref$. I then test whether these portfolios have significant differences in returns as measured from raw returns, the CAPM, Fama-French three-factor and Carhart four-factor alphas. The results are reported in Table 1.5. My results show that long-short portfolios of stocks selected based on the owners' skewness preferences generate significant negative alphas. The results validate prediction 5.

Table 1.5. Portfolio tests.
This table reports alphas from portfolio tests using stocks sorted quarterly on $SkewPref$. The alphas are calculated from the quarterly return of a portfolio going long the quintile of stocks with the highest $SkewPref$ and going short the stocks in the lowest quintile of $SkewPref$. t-statistics are reported in parentheses. I use Newey-West standard errors with one year lag.

	Raw Return	CAPM (%)	FF3 (%)	Carhart (%)
Equal-Weighted	-3.032***	-2.886***	-2.644***	-2.601***
	(-3.63)	(-3.32)	(-3.53)	(-3.53)
Value-Weighted	-1.763***	-1.687**	-1.671**	-1.614**
	(-2.59)	(-2.23)	(-2.41)	(-2.35)

* p < 0.1, ** p < 0.05, *** p < 0.01

Prediction 6 relates the skewness premium to diversification costs:

P_6 : The skewness premium is increasing in diversification costs.

It is challenging to empirically test this prediction cross-sectionally because diversification costs for individual stocks are closely tied to their liquidity and size. I therefore use the time-series variation of the skewness premium to track whether it is broadly in line with the decrease in diversification costs since the 1980. I calculate quarterly returns of long-short portfolios using the top and bottom quintile of stocks sorted on idiosyncratic returns skewness. Figure 1.10 plots the 15-year rolling average of alphas for three different factor models using value-weighted portfolio construction. The figure reveals a upward trend in the alpha over time, corresponding to a decrease in the skewness premium.[21] The overall trend is in line with Boyer et al. (2010) whose sample ends in 2005. The drop in skewness premium is firmly in line with the decreasing costs of diversification over the last decades.

[21] In Appendix Figure A.5 I plot the results using equal-weight portfolio construction.

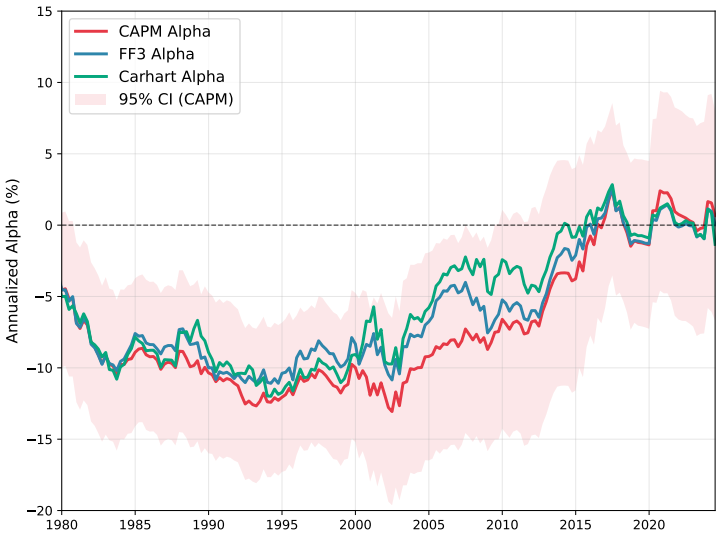


Figure 1.10. Skewness premium over time.
This figure plots the 15-year rolling annualized alpha of value-weighted long-short portfolios of skewness-sorted stocks. The alphas are calculated using the CAPM, Fama-French three factor and Carhart four factor model.

Finally, prediction 7 relates under-diversification with skewness preferences:

P_7 : Skewness preferences are associated with portfolio under-diversification.

In the model under-diversification is defined as the share of risky wealth invested in the stock instead of the fund. Empirically, a proxy for under-diversification at the aggregate level can therefore be understood as the share of stock retail ownership. I have shown ample evidence that retail ownership (1 - institutional ownership) is positively associated with skewness preferences. Mitton and Vorkink (2007) confirm this logic using data from individual investor's portfolios. The authors finds that portfolios of under-diversified investors have higher return skewness because such investors overweight stocks which individually have higher idiosyncratic return skewness – at the cost of mean-variance optimality.

The model's key assertions can all be validated using data from U.S. listed equities. The results reveal interesting dynamics in how cheap diversification affects firms, prices and in-

vestment decisions. The next part of this paper asks whether my findings can be explained by an alternative governance channel.

1.6 The governance channel

A natural alternative explanation for my findings is that institutional ownership reduces skewness through enhanced monitoring and governance rather than through the catering mechanism described above. It seems plausible that institutional owners are better equipped to enforce and understand governance issues than direct household owners. Indeed, much research examines how institutional investors may address agency frictions by imposing tighter oversight (Burns et al., 2010), altering managerial pay incentives (Hartzell and Starks, 2003), constrain empire-building behaviour and entrenchment (Berger et al., 1997), or influence capital structure decisions (Jensen, 1986). If such governance interventions by institutional investors change firm project selection, they could lead to lower idiosyncratic skewness.

Because governance operates through many channels, only some of which are observable, it is impossible to definitively rule out this alternative in favour of a catering channel. However, the two mechanisms generate distinguishable predictions, which I test in three ways. First, if governance is the operative channel, exogenous increases in institutional ownership should produce measurable changes in specific governance instruments. Second, the effect on skewness should be concentrated among firms where the marginal increase in monitoring is largest. Third, the composition, and not just the level of diversified ownership should matter. The catering channel predicts the opposite on all three counts: it operates through equilibrium prices and requires no governance intervention, it implies a uniform effect regardless of the pre-existing governance environment, and it depends only on the aggregate level of diversified ownership, not on who the diversified owners are. I test each prediction in turn.

1.6.1 Direct governance mechanisms

I focus on two measurable channels through which owners align managerial incentives with their own: financial policy and CEO pay. Institutional owners could use financial policy to constrain managerial entrenchment and empire building. For example, debt servicing and dividend payments can discipline managers to reduce discretionary spending of free cash flow (Jensen, 1986). This could reduce the number of positively skewed projects at the firm level if such projects are believed to be pet projects. Similarly, if institutional owners systematically reduce managerial pay convexity, managers might be less likely to undertake positively skewed projects. A large body of work documents the link between convex compensation structures and managerial project selection (Coles et al., 2006; Low, 2009). In particular high pay convexity, the sensitivity of CEO wealth to stock return volatility, called vega, encourages managers to pursue riskier and potentially more positively skewed projects. If institutional ownership reduces vega, this could mechanically reduce skewness.

To assess the validity of these channels I use the granular instrument in a two-stage setup as before. I use the second stage to check whether exogenous changes in diversified ownership cause material changes in the relevant governance mechanisms. I specifically look at the effects on firm leverage, dividend payout as well as the delta and convexity (vega) of CEO pay. The CEO compensation data is obtained from Coles et al. (2013), who rely on S&P 1500 firm data from Execucomp.[22]

The results in Table 1.6 show no evidence suggesting that the same exogenous shocks to diversified ownership which decrease idiosyncratic skewness materially affect the relevant governance variables. As before, the first-stage F-statistics are well above conventional thresholds in all columns, confirming the instrument's relevance. In the second stage, the coefficient on instrumented diversified ownership is statistically insignificant for leverage, log(CEO delta) and

[22]Due to the limited coverage of Execucomp, the sample shrinks significantly for tests related to CEO pay.

log(CEO vega) (columns 2 to 4). The point estimate for dividends (column 1) is marginally significant at the 10% level, but its magnitude is modest: a 10 percentage point increase in institutional ownership is associated with a 0.7 percentage point increase in the dividend-to-book equity ratio which equates to only 16% of the standard deviation in the sample. This could reflect general payout preferences of institutional investors rather than a disciplining mechanism aimed at curbing skewed project selection.

Crucially, the absence of a significant effect on CEO vega is informative. If institutional owners were reducing skewness by flattening managerial compensation, one would expect a strong negative coefficient on vega. Instead, the point estimate is close to zero and statistically indistinguishable from it. Likewise, the null result on leverage suggests that the exogenous variation in institutional ownership that drives the skewness decline does not operate through changes in capital structure discipline. These findings are consistent with the catering mechanism proposed in this paper: firms adjust their project choices without requiring explicit governance interventions.

1.6.2 Monitoring intensity

Some governance mechanisms do not require direct intervention. Increased monitoring alone might suffice to constrain managers and alter firm behaviour (Shleifer and Vishny, 1986; Admati and Pfleiderer, 2009). For instance, institutional investors might discipline managers simply through their presence or threat of exit. If the exogenous increase in institutional ownership raises monitoring intensity for certain firms, the resulting managerial discipline could affect project selection and hence skewness – even without observable changes in compensation or financial policy.

To account for the possibility that increased monitoring from institutional owners could be driving my results, I adjust my identification method to allow for heterogeneous effects along the monitoring dimension. Specifically, I interact the instrumented diversified ownership

Table 1.6. Causal relationship between diversified ownership and governance mechanisms.
This table shows the results of a two-stage least squared regression predicting governance variables. The instrument for lagged institutional ownership is $GIV(5y)_{t-5y}$. In column (1) the dependent variable is dividends divided by book equity. In column (2) it is leverage, defined as the book value of total debt divided by total assets. In columns (3) and (4) the dependent variables are the five-year rolling averages of the logarithm of the delta and vega of CEO compensation from Coles et al. (2013). All models include time and industry fixed effects as well as time-varying firm level control variables from Table 1.2. Standard errors are clustered at the firm level.

	Dividend	Leverage	log(Delta)	log(Vega)
	(1)	(2)	(3)	(4)
Panel A: First Stage				
$GIV(5y)_{t-5y}$	0.011***	0.012***	0.012***	0.012***
	(0.002)	(0.002)	(0.003)	(0.003)
Num. Obs.	20738	20836	9807	9807
F-statistic	67.32	70.96	39.28	39.28
R^2	0.55	0.55	0.35	0.35
Controls	X	X	X	X
FE: Time	X	X	X	X
FE: Industry	X	X	X	X
Panel B: Second Stage				
$\widehat{InstOwn}_{t-5y}$	0.073*	0.029	1.822	-0.428
	(0.037)	(0.101)	(1.422)	(1.418)
Num. Obs.	20738	20836	9807	9807
R^2	0.24	0.71	0.45	0.37
Controls	X	X	X	X
FE: Time	X	X	X	X
FE: Industry	X	X	X	X

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

variable with a proxy for ex-ante monitoring intensity. If monitoring is the operative channel, the causal effect of institutional ownership on skewness should be concentrated among firms where the marginal increase in monitoring is largest. That is, among firms with low pre-existing monitoring. In such firms, additional institutional ownership plausibly generates the greatest marginal increase in oversight. By contrast, if the catering mechanism dominates, the interaction term should be insignificant because the effect of diversified ownership on skewness operates through shifts in aggregate investor preferences, irrespective of monitoring intensity.

I measure monitoring using two distinct proxies. The first is the concentration of institu-

tional investors at each quarter. Studies show that firms with more concentrated institutional ownership have more engaged monitors because large shareholders face stronger incentives to bear the costs of oversight (Shleifer and Vishny, 1986; Hartzell and Starks, 2003). As in Hartzell and Starks (2003), I measure concentration as the Herfindahl index of shares outstanding owned by each unique 13F filer. Higher values indicate that a small number of institutional owners hold large positions, consistent with more intensive monitoring. The second measure is the number of analysts covering a firm each quarter, as measured from EPS estimates in the I/B/E/S database. Previous research shows that analyst coverage serves as an external monitoring mechanism. Analysts produce information that constrains managerial discretion and reduces information asymmetry (Yu, 2008; Chang et al., 2006).

To implement this test I estimate a model with two endogenous regressors: diversified institutional ownership and its interaction with the monitoring proxy. Both are instrumented using the granular instrument $GIV(5y)_{t-5y}$ and its interaction with the same monitoring variable, following standard practice for instrumenting interaction terms. The two first stage regressions coefficients for the level of the instrument and their F statistics are reported in Panels A and B of Table 1.7, confirming that both instruments have adequate power.[23] The second-stage results are reported in Panel C. All regressions include the standard firm-level control variables and fixed effects, alongside the monitoring proxy as additional control. The coefficient on instrumented institutional ownership remains large, negative, and highly significant in both specifications. These estimates are economically indistinguishable from the baseline estimate in Table 1.3. Critically, neither interaction term is statistically significant. These results imply that the causal effect of diversified ownership on skewness does not vary with the pre-existing level of monitoring at the firm. The uniformity of the causal effect of diversified ownership on skewness across monitoring regimes is difficult to reconcile with a monitoring-based explanation

[23]For conciseness I omit reporting the first stage coefficient for the interaction term between $GIV(5y)_{t-5y}$ and the monitoring variable.

but is entirely consistent with a catering mechanism.

Table 1.7. Causal relationship between diversified ownership, monitoring and skewness.
This table shows the results of a two-stage least squared regression predicting idiosyncratic return skewness from a 5-year window estimated with the CAPM, allowing for heterogeneous effects by monitoring intensity. The instruments for lagged institutional ownership and its interaction with each monitoring proxy are $GIV(5y)_{t-5y}$ and $GIV(5y)_{t-5y}x[Monitoring]_{t-5y}$. In column (1) monitoring is proxied by $IOConc_{t-5y}$, the Herfindahl index of institutional ownership shares, and in column (2) by $NumAnalysts_{t-5y}$, the number of analysts covering the firm. All models include the time-varying firm-level controls from Table 1.2, time and industry fixed effects, as well as the relevant monitoring proxy as a main effect control. Standard errors are clustered at the firm level.

	Idiosyncratic Return Skewness	
	(1)	(2)
Panel A: First Stage Level		
$GIV(5y)_{t-5y}$	0.010***	0.013***
	(0.003)	(0.002)
Num. Obs.	20831	20831
F-stat	30.99	33.82
R^2	0.61	0.56
Controls	X	X
FE: Time	X	X
FE: Industry	X	X
Panel B: First Stage Interaction		
$GIV(5y)_{t-5y}$	0.003***	0.017*
	(0.001)	(0.009)
Num. Obs.	20831	20831
F-stat	10.25	14.01
R^2	0.08	0.91
Controls	X	X
FE: Time	X	X
FE: Industry	X	X
Panel C: Second Stage		
$\widehat{InstOwn}_{t-5y}$	-2.326***	-2.542***
	(0.673)	(0.665)
$\widehat{InstOwn_{t-5y}xIOConc_{t-5y}}$	0.419	
	(1.242)	
$\widehat{InstOwn_{t-5y}xNumAnalysts_{t-5y}}$		-0.015
		(0.074)
Num. Obs.	20831	20831
R^2	0.21	0.20
Controls	X	X
FE: Time	X	X
FE: Industry	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

1.6.3 Passive ownership

In this section I ask the question whether the composition of active and passive institutional ownership matters for skewness, holding total diversification levels constant. The catering and governance channels make opposing predictions.

Under the governance channel, composition should matter. Passive and active institutional owners differ substantially in their monitoring capacity, engagement strategies, and voting behaviour (e.g., Gormley and Kim (2025); Heath et al. (2022)). Under the catering channel, composition should not matter. What determines firm project selection is the aggregate diversification level of the investor base, regardless of whether those diversified portfolios are managed passively or actively. A reallocation of ownership from active to passive institutional investors leaves total diversified ownership unchanged and should therefore have no effect on skewness.

I exploit Russell 1000/2000 index reconstitutions to test between these predictions. Firms which leave the Russell 1000 index and drop into the Russell 2000 index experience increases in passive ownership because the Russell 2000 index is more popular with investors (and vice versa). Because the index classification is entirely determined by the relative market capitalization of a firm compared to other similarly valued firms (see Appel et al. (2024) for exact details), the index assignment is random for firms near the threshold (Limburg, 2024). Crucially for this experiment, Gormley and Kim (2025) show that the index switch leaves aggregate institutional ownership unchanged because it is not households but active institutional owners which take the other side to index funds' trades. I closely follow the empirical specifications in Appel et al. (2024) (AGK) and Gormley and Kim (2025) (GK). The results are tabulated in Table 1.8.

In columns (1) to (3) I show the results of the first stage of the AGK instrumental variable specification for the 500 firms just above and below the index inclusion thresholds. The results show that Russell 2000 inclusion does not shock institutional diversified ownership or owners' skewness preferences. It only exogenously shocks passive, index fund ownership – by one

percentage point, which is almost 10% in relative terms. In column (4) I predict a firm's idiosyncratic return skewness over the following years using the estimate for passive ownership obtained in the first stage (column (3)). The coefficient for $\widehat{PassiveOwn}$ is not distinguishable from zero. To provide further robustness to this result I estimate the alternative GK model in column (5) and find again that Russell 2000 inclusion does not affect idiosyncratic return skewness. This result implies that the exogenous change in passive ownership which has a net-zero impact on total diversified ownership does not affect idiosyncratic return skewness. This is in line with a skewness disappearance channel that does not depend on characteristics and governance preferences of owners.

Table 1.8. Russell inclusion study.
This table presents the results from index inclusion studies based on the methodology of Appel et al. (2024) (AGK) and Gormley and Kim (2025) (GK). Columns (1) to (4) plot the first stage in the AGK specification for institutional ownership, owners' skewness preferences ($SkewPref$) and passive ownership and the second stage for the passive ownership estimation. In column (5) I follow the OLS specification from GK. $R2000Dummy$ is a dummy variable equal to one for firms which are members of the Russell 2000 Index after reconstitution day. $\widehat{PassiveOwn}$ is the fitted passive ownership estimate from the first stage regression. All specifications include the AGK control variables. Standard errors are clustered at the firm level.

	AGK 2019 2SLS Specification				GK 2025 OLS Specification
	First Stage			Second Stage	
	Inst.Own	SkewPref	Passive Own.	Skewness (t+5)	
	(1)	(2)	(3)	(4)	(5)
$R2000Dummy$	−0.001	−0.001	0.003***		0.030
	(0.009)	(0.001)	(0.001)		(0.018)
$\widehat{PassiveOwn}$				8.246	
				(7.647)	
Num. Obs.	24 489	23 475	24 511	24 511	23 846
R^2	0.51	0.57	0.89	0.03	0.40
AGK Controls	X	X	X	X	X
FE: Year	X	X	X	X	X
FE: Firm					X

* p < 0.1, ** p < 0.05, *** p < 0.01

Taken together, the results in this section show that it is unlikely that the causal link between diversified ownership and skewness is related to the specific governance preferences of institutional or passive owners. The exogenous shocks to institutional ownership which causally reduce skewness do not operate through changes in CEO compensation convexity, leverage, or dividend policy. The effect does not vary with the pre-existing intensity of monitoring, whether measured by ownership concentration or analyst coverage. Further, an exogenous shock to passive ownership that leaves total diversified ownership unchanged has no effect on skewness. These findings collectively support the interpretation that the decline in idiosyncratic return skewness is driven by the catering mechanism formalised in the model. In the next and final part of this paper I examine whether the choice to take on fewer positively skewed projects has real effects at the firm level and unveil an unintended consequence of cheap diversification.

1.7 Real effects of diversified ownership

My model suggests that an increase in diversified ownership makes managers select projects with lower return skewness. In this section I make the claim that one particular type of project is especially affected by such choices: breakthrough innovation.

To test whether breakthrough innovation projects could be affected by firms' skewness choices, I use three distinct measures of breakthrough innovation from the patent and innovation literature. First, I use the 5-year sum of each firm's patent values ($PatentValue(5y)$) as estimated from stock returns in Kogan et al. (2017), divided by the firm's market capitalisation. Second, $MaxPatentCitations(5y)$ captures the logarithm of number of citations from the firm's most cited patent in a five year window. Third, I use the 5-year median patent importance for each firm obtained from Kelly et al. (2021). $PatentImportance(5y)$ is highest for patents which are both textually distinct from previous patents and are heavily cited by future patents. The authors show that the measure is a good identifier of technological breakthroughs. To minimize

the impact of outliers and increase interpretability, I winsorize the innovation variables at 2.5% and 97.5% and normalize them subsequently. The summary statistics for the non-normalized variables are reported in Table 1.1. Over half of all observations are zero and their distributions are highly skewed, akin to Marsili and Salter (2005).

I first ask whether breakthrough innovation efforts indeed correlate with skewness. The results in Table 1.9 show that breakthrough innovation is positively associated with skewness even after controlling for a host of time-variant firm characteristics as well as industry and time fixed effects. The coefficients for all three measures of breakthrough innovation are significantly positive.[24] A one standard deviation increase in patent value is associated with an increase in idiosyncratic return skewness of approximately 28% compared to the mean. For the values of maximum patent citation and importance the same increase coincides with skewness increases of 2 and 3%.[25]

Taking together that breakthrough innovation efforts are associated with skewed idiosyncratic returns, and that firm managers are less likely to select projects with idiosyncratic return skewness when owners are diversified, diversified ownership should causally reduce breakthrough innovation. In Table 1.10 I test this claim using the granular instrumental variable approach at the firm-level. In the odd columns I use two latent factors and in the even columns three factors in the creation of the instrument as robustness. Panel A of the table shows that the instrument is highly statistically significant with first stage F-statistics of greater than 60. The difference in first-stage coefficient magnitudes reflects differences in residual instrument variation. Because 2SLS is scale-invariant, this does not affect second-stage estimates. The second

[24] It is worth noting that at least some of the relationship between $PatentValue(5y)$ and skewness is mechanical, since Kogan et al. (2017) use stock returns to estimate patent value. However, I assume that the mechanical link is rather small because Kogan et al. (2017) use three day event windows after patent announcement to estimate the economic value. This time window is in stark contrast to the five year skewness measure constructed from monthly returns.

[25] In unreported tests I run these regressions limited to the subset of firms which issue patents. The regression coefficients remain similar, although they lose some significance. This is likely due to shrinkage in sample size.

Table 1.9. Skewness and breakthrough innovation.
The dependent variable is the idiosyncratic return skewness from a 5-year window estimated with the CAPM. $PatentValue(5y)$ is the 5-year sum of a firms patent values from Kogan et al. (2017), divided by the firm market capitalisation. $MaxPatentCitations(5y)$ is the logarithm of number of citations from the firms most cited patent in a five year window. $PatentImportance(5y)$ is the 5-year median patent importance for each firm obtained from Kelly et al. (2021). $Log(Mcap)$ is the log of firm market cap. $Log(Assets)$ is log of total assets. $Leverage$ is the book value of total debt divided by the book value of total assets. $Market-to-Book$ is the market value of equity divided by the book value of equity. $Log(Age)$ is the log of firm age in years since IPO. $CapitalExpenditure$ is capital expenditures scaled by total assets. $Investment$ is the annual change in total assets scaled by total assets. $Spread$ is the average daily bid-ask spread scaled by price. Standard errors are clustered at the firm-level.

	Skewness		
	(1)	(2)	(3)
$PatentValue(5y)$	0.251***		
	(0.030)		
$MaxPatentCitations(5y)$		0.022***	
		(0.006)	
$PatentImportance(5y)$			0.016***
			(0.006)
$Log(MCap)$	−0.157***	−0.161***	−0.160***
	(0.022)	(0.022)	(0.022)
$Log(Assets)$	−0.031	−0.034	−0.034
	(0.022)	(0.022)	(0.022)
$Leverage$	−0.072	−0.077	−0.083
	(0.061)	(0.061)	(0.061)
$Log(Age)$	0.118***	0.115***	0.116***
	(0.008)	(0.008)	(0.008)
$Investment$	0.232***	0.239***	0.237***
	(0.022)	(0.022)	(0.022)
$Market-to-Book$	0.172***	0.175***	0.177***
	(0.022)	(0.022)	(0.022)
$CapEx$	−1.280***	−1.282***	−1.308***
	(0.327)	(0.328)	(0.328)
$Spread$	1.168***	1.123***	1.137***
	(0.293)	(0.291)	(0.291)
Num. Obs.	29 708	29 708	29 708
R^2	0.18	0.18	0.18
FE: Quarter	X	X	X
FE: Industry	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

stage results in Panel B reveal that a 10 percentage point increase in institutional ownership causally reduces the economic value of patents in the next five years by 0.04 standard deviations (column 1-2), the maximum citations on a patent in the subsequent five years by 0.26

standard deviations (column 3-4) and the importance of patents by 0.24 standard deviations (column 5-6).

Furthermore, the causal effect of diversified ownership on skewness extends beyond the initial five-year window. In Appendix Table A.4 I repeat this test predicting skewness and innovation between 5 and 10 years after I measure diversified ownership. This is consistent with the longer horizons over which innovation projects affect firm outcomes. From Table 1.1 we see that breakthrough innovation variables are highly skewed, in part because most firms do not issue patents. To account for the possibility that my results are affected for this reason, in Appendix Table A.5 I replicate this test including only firms which have issued at least one patent in the previous five years. The results are qualitatively similar.

Table 1.10. Causal link between institutional ownership and breakthrough innovation.
This table shows the results of a two-stage least squared regression predicting normalized breakthrough innovation variables. The instrument for lagged institutional ownership is $GIV(5y)_{t-5y}$. In columns (1) and (2) I predict $PatentValue(5y)$, the 5-year sum of a firms patent values from Kogan et al. (2017), divided by the firm market capitalisation. In columns (3) and (4) I predict $MaxPatentCitations(5y)$, the logarithm of number of citations from the firms most cited patent in a five year window. In columns (5) and (6) I predict $PatentImportance(5y)$, the 5-year median patent importance for each firm obtained from Kelly et al. (2021). All dependent variables are normalized and all models include the time-varying firm-level controls from Table 1.2 as well as time and industry fixed effects. In odd (even) columns the instrument is calculated using two (three) latent factors. Standard errors are clustered at the firm level.

	Pat.Value		MaxPat.Citations		Pat.Importance	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First Stage						
$GIV(5y)_{t-5y}$	0.011***	0.012***	0.011***	0.012***	0.011***	0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Num. Obs.	20836	20679	20836	20679	20836	20679
R^2	0.55	0.55	0.55	0.55	0.55	0.55
F-statistic	68.33	68.02	68.33	68.02	68.33	68.02
#Latent Factors	2	3	2	3	2	3
Firm Controls	X	X	X	X	X	X
FE: Quarter	X	X	X	X	X	X
FE: Industry	X	X	X	X	X	X
Panel B: Second Stage						
$\widehat{InstOwn}_{t-5y}$	-0.371**	-0.341**	-2.695***	-2.584***	-2.486***	-2.409***
	(0.164)	(0.162)	(0.905)	(0.896)	(0.856)	(0.852)
Num. Obs.	20836	20679	20836	20679	20836	20679
R^2	0.16	0.16	0.43	0.43	0.36	0.36
#Latent Factors	2	3	2	3	2	3
Firm Controls	X	X	X	X	X	X
FE: Quarter	X	X	X	X	X	X
FE: Industry	X	X	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Finally, it is possible to highlight the corporate strategies which lead to reductions in breakthrough innovation. In Appendix Table A.6 I show that diversified ownership causes significant reductions in R&D expenses and the number of patents issued by firms as well as an increased

focus of patenting activity on fewer unique scientific sub-areas.

Taken together, my results suggest that breakthrough innovation arises from projects with positively skewed returns, and that these are less likely to be undertaken when ownership is diversified. Thus, a negative consequence of cheap diversification is the reduction of breakthrough innovation. Finally, this paper's results and intuition can also be used to think about innovation efforts in private firms. Such firms have less diversified ownership so external validation of my claims requires that private firms should engage in more breakthrough innovation. Gao et al. (2018) and Bernstein (2015) confirm this assertion: patent novelty and breakthrough innovation measures are higher for private firms and drop when firms go public.

1.8 Conclusion

This paper shows that the secular decline in idiosyncratic return skewness is linked to the rise of cheap diversification opportunities since the 1980s. I document that investors' aggregate demand for skewness has declined in the last decades, coinciding with the rise in institutional ownership. Firms with more diversified owners subsequently exhibit lower future idiosyncratic return skewness, consistent with project choices responding to investor composition. A stylised equilibrium model illustrates how cheap diversification increases investments in funds which in turn reduces firms' incentives to adopt positively skewed projects.

One type of project which generates idiosyncratic skewness is breakthrough innovation. In line with the predictions, I reveal a negative causal link between diversified ownership and breakthrough innovation. By changing firm behaviour, the widespread availability of cheap diversification may have unintended social externalities.

Chapter 2

Index Disruption: The Promise and Pitfalls of Self-Index ETFs[†]

with **Sida Li** and **Bige Kahraman**

Abstract Self-indexed ETFs, which track indices created and maintained by ETF issuers themselves, rather than relying on external index providers, are disrupting the index provision market. By avoiding licensing fees to independent index providers, self-indexed ETFs are expected to offer competitive fees for investors. Surprisingly, we find that self-indexed ETFs often charge *higher* fees than their public-index counterparts. This disparity cannot be explained by product differentiation but rather by conflicts of interest. ETF issuers which offer both self-indexed ETFs and wealth management services are incentivized to promote their own high-cost self-indexed products to clients. Increased competition is not always a panacea for end-investors.

[†]We thank the discussants and participants at the SEC Conference on Financial Market Regulation (CFMR), the Midwest Finance Association (MFA) Annual Meeting, the Rising Scholar in Finance Conference (UZH), the FMA Europe Conference, EDHEC Business School, VU Amsterdam University, the Tinbergen Institute Amsterdam, and the University of Oxford Saïd Business School for their useful comments. Anthony Limburg acknowledges support from the Clarendon Fund, the Templeton Education and Charity Trust (TECT) and the Saïd Business School Foundation. Sida Li acknowledges support from ACCESS for supercomputing allocations.

2.1 Introduction

The rise of exchange-traded funds (ETFs) has fundamentally reshaped investment markets and has made ETFs an increasingly important investment vehicle for many U.S. households.[1] Traditionally, most ETFs are "public-indexed," meaning they rely on indices constructed by established industry leaders, such as S&P Dow Jones, FTSE Russell, and MSCI. These providers wield substantial market power and charge licensing fees to ETF issuers, which are ultimately passed on to end investors. Approximately one-third of total ETF fees are estimated to be paid to index providers as licensing fees, with roughly 60% of these fees representing markups.[2] The market dominance and pricing practices of external index providers have drawn scrutiny from U.S. regulators who are considering potential remedies.[3]

Against this backdrop, the emergence of "self-indexed" ETFs marks an important shift away from the traditional model of index provision dominated by external providers. Self-indexing refers to the practice whereby ETF issuers, or their affiliated entities, construct and manage their own indices, thereby avoiding the costs associated with licensing third-party benchmarks. A key catalyst for the expansion of self-indexed ETFs was the SEC's 2013 decision to relax prior restrictions, making it easier for issuers to rely on proprietary indices.[4]

Although relatively uncommon just a decade ago, self-indexed ETFs now account for nearly 20% of ETFs within popular investment styles such as broad equity and large cap. Their cumulative growth in assets under management has doubled that of externally indexed ETFs during our sample period, and several established asset managers (e.g., Goldman Sachs, Fidelity, Wisdom Tree) now offer self-indexed ETFs. Motivated by these developments, this paper studies

[1] According to ICI reports, U.S.-listed ETF assets have surged from approximately 1.75 trillion in 2013 to over $8.1 trillion by year-end 2023, leading to increased ownership by households. For instance, over the last decade, the number of households directly owning ETFs has increased by over 163%.

[2] See An et al. (2023) for a structural model.

[3] See SEC Release No. IA-6050 (June 15, 2022).

[4] See, SEC IM-INFO-2013-09.

how self-indexing shapes competition and outcomes in ETF investing.

While a large literature studies exchange-traded funds, comparatively little attention has been paid to the index provision market and its implications for investor welfare. Although ETF fees are often low on an annual basis, even small fee differences can have substantial effects on long-term investment outcomes.[5] Given the growing adoption of ETFs among households, it is therefore important to examine how competition in index provision affects investor outcomes. Our paper contributes to this literature by focusing on the emerging phenomenon of self-indexing.

To examine how self-indexing shapes ETF investing, we consider three hypotheses. First, the removal of licensing costs may empower issuers to offer more competitive fees, an argument consistent with established theories of competitive markets (e.g., Tirole (1988)) and adopted by many industry commentators.[6] This assertion rests on the assumption that competitive pressures will manifest as price competition. However, fund issuers may struggle to compete down prices if they face an inelastic demand for well-established indices (An et al., 2023) or lack sufficient cost-effectiveness to sustain a price war. According to our second hypothesis, in this environment, fund issuers can resort to product differentiation, consistent with basic models of monopolistic competition (e.g., Dixit and Stiglitz (1977)). Our third hypothesis posits a self-preferencing mechanism within a search costs framework building on Hortaçsu and Syverson (2004) and Roussanov et al. (2021). Index proliferation heightens search frictions as it becomes more difficult for investors to easily compare and choose among them. This enables issuers to sell nearly homogeneous portfolios at higher fees.[7] These sales are facilitated because some self-indexed ETF issuers are also private wealth managers, which allows them to promote their self-indexed products in their advisory businesses.

[5] See, for instance, French (2008), Sharpe (2013) and this SEC Bulletin.
[6] See, for instance, in the Financial Times, Barron's and WatersTechnology.
[7] For S&P 500 index funds, Hortaçsu and Syverson (2004) estimate search costs at 11-20bps and Roussanov et al. (2021) estimate at 39bps for actively managed funds.

To empirically examine these contrasting perspectives, we start by comparing the fees of self-indexed and public-indexed ETFs. Within our sample of passive U.S. equity ETFs, we document that self-indexed ETFs charge significantly *higher* net expense ratios compared to their public-indexed counterparts. Specifically, self-indexed ETFs have fees that are about 10-13 percent higher than those of public-indexed funds.

We then test the notion that self-indexing provides issuers with greater latitude in crafting distinct investment strategies which may fit clients' particular preferences and potentially provide financial value. Using both portfolio holdings similarity as well as return correlation measures, we find that the investment strategies of self-indexed ETFs are actually more *similar* to their peers within the same style, rather than being more distinct. This lack of distinction is also reflected in their performance outcomes: across all our tests, we find no significant evidence that self-indexed ETFs outperform their public-indexed counterparts. This conclusion is established through multiple analyses, using different factor models, daily and monthly data, and regression analyses at both the time-series portfolio and the cross-sectional fund levels.

Our findings on fees, portfolio differentiation and performance are different from related studies as our robust testing controls for: (i) unique investment strategies (e.g., Kostovetsky and Warner (2025)), (ii) "closet active management" by passive ETFs (e.g., Akey et al. (2021); Easley et al. (2021); Cheng et al. (2019)), (iii) ETF secondary market liquidity (e.g., Brown et al. (2024); Khomyn et al. (2024)), (iv) thematic ETFs (e.g., Ben-David et al. (2022)), and (v) smart-beta investment styles (e.g., Huang et al. (2023)). Furthermore, our results are obtained after controlling for a host of fund characteristics (such as size and age) as well as time, style and issuer fixed effects.

Our full-sample findings do not support the first two hypotheses that self-indexing is associated with either price competition or portfolio differentiation. We next examine the third hypothesis, which proposes a self-preferencing mechanism among certain issuers: those that both issue self-indexed ETFs and provide wealth management advisory services may prefer-

entially promote their own funds, particularly to clients facing high search costs. This analysis also allows us to identify cross-sectional variation across issuer types. Overall, the findings align with the predictions of this hypothesis.

To assess this idea empirically, we classify issuers into two groups: specialized fund managers (focused solely on fund management) and investment advisors (offering fund management alongside wealth management advisory services). Consistent with the third hypothesis, we first document that the results thus far are predominantly driven by investment advisors. While these advisors generally offer competitive fees across their other ETFs, their self-indexed ETFs are a notable exception, as these charge 20 percent higher compared with their public-index offerings. Moreover, their funds tend to be more similar to their peers. In contrast, among specialized fund managers, there are no significant fee differences between self- versus publicly-indexed ETFs. Also, within this group, self-indexed funds tend to be more differentiated.

The self-preferencing hypothesis argues that investment advisors favor their self-indexed ETFs in the advisory part of their businesses. This would appear as disproportionately high *Self-Ownership* (fraction of ETF shares reported by its issuer in its 13F filings) to the extent that some of these assets are held in discretionary advisory accounts.[8] Consistent with this, we find that self-indexed ETFs average 14% self-ownership, significantly higher than the 2% observed for public-indexed peers. Crucially, our regression analysis reveals 80 percent of this difference is driven by self-indexed ETFs issued by investment advisors.

As a mechanism test, we evaluate whether model portfolio recommendation is a channel through which issuers promote their self-indexed ETFs. Model portfolios are recommendation portfolios designed by asset management companies, some of which also issue ETFs, create indices and provide advisory services. About 54 percent of advised assets are reportedly allo-

[8]Institutional investment managers in the United States that exercise investment discretion over $100 million in Section 13(f) securities are required to file Form 13F including the assets that they own or the client assets that manage. See Section 13(f)(1)(b) of the Securities Exchange Act, and 15 U.S.C. 78c(a)(35).

cated in model portfolios, which therefore makes them a useful proxy for otherwise confidential client assets (Brogaard et al., 2021). We find that, on average, ETF issuers which are also model providers tend to recommend their self-indexed ETFs more frequently compared with public-indexed ETFs, and this is exclusively driven by the ones which are also investment advisors.

Last but not least, we conduct an additional mechanism test to examine whether self-indexed ETFs attract investors with higher search costs. The rationale is that index proliferation increases search frictions, allowing issuers to charge higher fees for similar portfolios, particularly to clients facing higher search costs. If these clients can be proxied by limited financial sophistication (Roussanov et al., 2021), we would expect flows to self-indexed ETFs to be more sensitive to market sentiment. Consistent with this prediction, we find that flows to self-indexed ETFs rise during periods of elevated sentiment, whereas flows to publicly indexed ETFs show no significant relationship. These results are consistent with a self-preferencing mechanism operating within a search-cost framework.

Related Literature

An et al. (2023) is the first study to quantify the impact of competitive dynamics between external index providers and ETF issuers on investor costs. Xiao and Xiong (2024) later developed a general equilibrium model studying the agency issues stemming from self-indexing and they conclude that a concentrated index provision market can be more beneficial to retail investors. Legal scholars are increasingly interested in the index market due to the significant expansion of passive investment products. While some work recognizes that self-indexed funds come at higher cost, their primary focus remains on legal and regulatory issues (Robertson, 2019). Similarly, Kostovetsky and Warner (2025) acknowledge self-indexed funds but focus mainly on externally indexed funds.

Our study goes beyond documenting fee differences. We provide a detailed economic analysis of self-indexed ETFs, examining whether these funds offer meaningful portfolio differenti-

ation or performance advantages to investors. Our findings instead point to a self-preferencing mechanism associated with a subset of self-indexing issuers that also operate as investment advisers. More broadly, the results show that entry into index provision through self-indexing is not necessarily a panacea for investors.

To the best of our knowledge, our findings are novel both in the academic literature and in the regulatory debate. Earlier regulatory discussions primarily emphasized concerns that self-indexing could facilitate malpractices such as NAV manipulation or internal front-running within fund families.[9] The SEC's 2013 exemptive orders sought to mitigate these risks through safeguards including daily portfolio disclosure and enhanced board oversight. By contrast, our paper highlights a previously overlooked agency concern that arises when the issuer, index provider, and investment adviser are the same entity, and provides timely insights for ongoing policy discussions surrounding index provision.

Our findings also add to the academic literature investigating the incentives and behavior of financial advisors (Mullainathan et al., 2012; Foerster et al., 2017). Prior research documents misconduct and poor client outcomes in the U.S. financial advice industry, in part attributing these issues to principal-agent problems due to brokers' incentives for mutual fund sales (Bergstresser et al., 2009; Guercio and Reuter, 2014; Kahraman, 2021).[10] The incentives analyzed in most of the earlier literature stem from a now largely outdated commission-based compensation model where brokers require no direct payment from clients and instead draw from product-based commissions, some of which are included in fund fees (Christoffersen et al., 2013; Edelen et al., 2012).

The market's shift toward Exchange-Traded Funds (ETFs) is largely accompanied by a change to the client-paid fee model, with the goal of better aligning incentives. In this model, advisors

[9] Section 2.4 empirically examines these concerns and finds no supporting evidence.

[10] Egan et al. (2019); Dimmock et al. (2018); Clifford and Gerken (2021); Gurun et al. (2021) study misconduct in the U.S. financial advice industry. Focusing on Canadian financial advisors, Linnainmaa et al. (2021) find advisors' personal investments align with how they advise their clients.

are compensated through client-paid fees which are based on clients' total account size and separate from fund expense ratios. Different from most earlier papers, we examine advisor incentives within this client-paid fee environment, which is relatively understudied. Recently, Edelen et al. (2025) studied the client-paid fee model in Australia, where regulators have introduced a novel auto-drop requirement. Focusing on the U.S market, we study the emerging practice of self-indexing, and provide evidence that is suggestive of a self-preferencing mechanism among investment advisors, thereby offering insights into agency concerns in the contemporary financial advice landscape.

Last but not least, this study contributes to a growing body of literature documenting dominated products within the ETF market (Brown et al., 2024). For instance, previous work documents significant underperformance and high fees for thematic ETFs (Ben-David et al., 2022), poor performance linked to index providers' data mining in smart-beta funds (Huang et al., 2023), evidence of closet active management in passive ETFs (Akey et al., 2021; Easley et al., 2021; Cheng et al., 2019), and the ability of liquid ETFs to extract rents from liquidity-sensitive investors (Khomyn et al., 2024). As noted previously, we demonstrate the distinctiveness of our results from these studies and highlight the issues arising from self-indexing practices.

2.2 Data and sample construction

Our main source of data is ETF Global which covers all ETFs listed in the U.S. and Canada with no survivorship biases. ETF Global provides information on a wide range of monthly ETF characteristics (e.g., fund name, issuer, inception date, benchmark index, AUM, leverage ratio, listing exchange, sector exposures, investment region, fund focus, asset class, active management dummy, currency and sector exposure, put and call options volume, short interest, management fee, and total/net expenses) as well as daily (monthly) holdings data for opaque (nonopaque) ETFs.

ETF Global has three key benefits over the CRSP Mutual Fund database. First and most importantly, it provides information on each ETF's benchmark index (variable labeled 'primary_benchmark'). This allows us to identify self-indexers as we will explain below. Second, the database provides daily ETF holdings, based on which we assess the similarity of investment strategies among ETFs. Thirdly, ETF Global updates fund characteristics monthly, enabling more accurate analyses. One downside of ETF Global is that the data are only available from 2012 onwards. However, this does not have an important impact on our study since self-indexing is a recent phenomenon. To illustrate, at the start of our sample in January 2012, we find only 13 self-indexing ETFs.

Our sample includes all passive unlevered U.S. equity ETFs from 2012 to 2020.[11] To identify the active (passive) status of the fund, we use the active management identifier provided by ETF Global. We remove all data points with missing information on fund primary benchmark and net expense ratios. We also eliminate cases where total expenses minus fee waivers are not equal to the ETFs net expenses.

We complement our data with supplementary information from Morningstar Direct, 13F institutional filings from Thomson/Refinitiv, and the CRSP stock and Mutual Fund databases. From Morningstar Direct, we extract a smart-beta identifier and model portfolio recommendations. We match approximately 90% of our sample with Morningstar using tickers and ETF names. We use 13F institutional filings and data on shares outstanding from the CRSP stock database to calculate the share of institutional ownership in our sample of ETFs. We use the CRSP Mutual Fund database to obtain daily, ETF share-split adjusted, NAV (net asset value) returns as well as to construct characteristics pertinent to ETF issuing investment management companies (e.g., age and total assets under management). Finally, we obtain data on factor returns from Kenneth R. French's website.

[11] It is ambiguous whether levered ETFs can be considered "passive". Nevertheless, we find that the underlying indices (before leverage) are all public indices in our data.

A self-indexed ETF is an exchange-traded fund that tracks an index created and maintained by the ETF issuer or one of its affiliated companies, rather than tracking an index provided by an independent third-party index provider, such as S&P Dow Jones Indices, MSCI, or FTSE Russell. A public-indexed ETF, on the other hand, is an ETF that licenses its index from an external third-party index provider.[12]

In the early years of the ETF market, issuers seeking to launch self-indexed ETFs were required to obtain exemptive relief from the SEC and were subject to several requirements.[13] A turning point came on July 10, 2013, when the SEC issued guidance (IM-INFO-2013-09) and new exemptive orders that significantly relaxed prior conditions on self-indexed ETFs. In particular, the Commission eased the requirements, relying instead mostly on daily portfolio transparency and governance.[14] These changes brought self-indexed ETFs closer to regulatory parity with ETFs tracking third-party indices, making the model more attractive to issuers and arguably contributing to the growth of self-indexing over the subsequent decade. The 2019 "ETF Rule" by the SEC removed "exemptive order" regulations, instead modernizing the regulation of ETFs by establishing a clear and consistent framework for the majority of ETFs operating today. The 2019 rule did not change the status quo rules and requirements for self-indexed ETFs.[15]

We use the variable 'primary_benchmark' from ETF Global to identify ETFs that are self-

[12]Our definition of self-indexing is based on the formal SEC definition, which states: "A "Self-Indexing Fund" is a Fund for which an Affiliated Person, or a Second-Tier Affiliate, of the Trust or a Fund, of the Adviser, of any Sub-Adviser to or promoter of a Fund, or of the Distributor (each, an "Affiliated Index Provider") will serve as the Index Provider" (SEC Release No. 30560)

[13]These requirements mainly stemmed from concerns of conflicts of interest, such as the manipulation of constituent pricing and internal front-running. The requirements generally included: (i) requirements to promote transparency of the index methodology (e.g., making index rules freely available and announcing methodology changes with at least 60 days' notice); (ii) use of an unaffiliated third party to calculate the index; and (iii) creation of rigid firewall arrangements to separate the fund's adviser and the index provider.

[14]The SEC states its objectives in broad terms, as to protect investors, promote informed investment decisions and facilitate innovation in investment products. Some industry commentators interpreted it as an effort to promote competition in the index provision market. See, for instance, SEC Issues New Relief for Self-Indexing ETFs | Morgan Lewis - JDSupra.

[15]https://www.sec.gov/news/press-release/2019-190

and public-indexing. Our procedure involves multiple steps. Self-indexed ETFs are defined as ETFs which are using a benchmark index that is offered by the same or an affiliated company, offering the ETF. Our first step involves parsing the text reported in 'primary_benchmark', taking the last data point available for a given fund (that is, 2020's cross section snapshot for alive funds and last available one for dead funds). If the text includes one of the key words corresponding to commonly used third-party indices such as "S&P", "FTSE", "Russell", "CRSP", "MSCI", "Morningstar", "Dow Jones", "NYSE", "NASDAQ", "Nasdaq", "StrataQuant", we mark these observations as public-indexed ETFs. If the text in the benchmark does not include these keywords but instead includes the name of the issuer (or its affiliates), we then mark these as self-indexed ETFs. For instance, the JP Morgan US momentum ETF (ticker: JMOM) is marked as a self-indexed ETF since its benchmark is "JP Morgan US Momentum Factor Index" and the issuer name is "JP Morgan". Second, for the cases that we were unable to identify through text parsing, we manually collect information on funds' benchmark indices using online sources and check whether the index is provided by the same (including its subsidiaries) or an affiliated company offering the ETF.[16] Third, to capture the changes in benchmark index over time for a given fund, we look at changes in the text reported in the 'primary_benchmark' variable and evaluate if it is a genuine change or a change that is simply capturing a name change or change in name abbreviation. We then manually update the self-indexed (public-indexed) identifiers accordingly. Through this procedure, we can identify the self-indexed (public-indexed) status of more than 99% of funds for which ETF Global provides a 'primary_benchmark'. This classification method allows us to define self-indexed ETFs in line with the formal definition offered by the SEC (see above). Our final sample includes 786 unique ETFs.

ETF Global also provides a 'focus' classification for ETFs which broadly classifies the investment style of a fund into one of 26 categories (e.g., 'Large Cap', 'High Dividend Yield', 'Broad

[16]Examples of ETF issuers being affiliated with index providers include FlexShares ETFs tracking Northern Trust indices and Lattice ETFs tracking Hartford indices.

Equity'), including sector-specific ETFs (e.g., 'Financials') as well as information on the issuing management company (issuer). For both variables we correct spelling errors (e.g., 'JP Morgan' and 'JPMorgan' and 'High Dividend Yielld' and 'High Dividend Yield').

2.2.1 Descriptive statistics

While self-indexing is a recent phenomenon in the ETF industry, it has been gaining traction over the past decade (Figure 2.1). Although there were only a few self-indexed funds a decade ago, by the end of our sample period, this has increased to 96 ETFs in total (Panel A). Moreover, the total cumulative AUM growth of self-indexed ETFs has doubled the growth of public-indexed ETFs during this time (Panel B).

Table 2.1 shows the distribution of self- and public-indexed ETFs across investment styles.[17] Self-indexed funds appear to be most prevalent in investment styles which are also popular among public-indexed funds (e.g., broad equity and large cap). Within these investment styles, self-indexed funds constitute about 20% of all funds by the end of our sample period. Throughout our analysis, we control for style fixed effects.

Table 2.1. Number of ETFs by style and indexing group.
Self-Indexed ETFs are ETFs which track indices provided by the ETF issuer or an affiliated company. Public-Indexed ETFs track indices provided by independent third parties. Styles are obtained from the focus grouping within ETF Global. For simplicity, in this table we bunch all sector-related styles into one group.

Style	Public-Indexer (N)	Public-Indexer (%)	Self-Indexer (N)	Self-Indexer (%)	Total (N)
Alpha-Seeking	21	84.0	4	16.0	25
Broad Equity	77	79.4	20	20.6	97
Buywrite	2	100.0	0	0.0	2
High Dividend Yield	27	69.2	12	30.8	39
Large Cap	165	79.3	43	20.7	208
Long/Short	12	85.7	2	14.3	14
Micro Cap	4	100.0	0	0.0	4
Mid Cap	43	87.8	6	12.2	49

[17] For expositional purposes, in this table, we group all sector-specific funds as 'sector' funds. This reduces the total number of investment styles to fifteen.

Preferred Stock	9	100.0	0	0.0	9
Real Estate	1	50.0	1	50.0	2
Sector	198	92.1	17	7.9	215
Size and Style	1	100.0	0	0.0	1
Small Cap	74	79.6	19	20.4	93
Target Outcome	1	100.0	0	0.0	1
Theme	41	70.7	17	29.3	58

We classify ETF issuers into three groups based on their mix of self- and public-indexed offerings: ones which offer only self-indexed funds, ones which offer only public-indexed funds, and ones which offer a mix of self- and public-indexed funds. Figure 2.2 shows the share of issuers in each of these groups over time. While most issuers only offered public-indexed ETFs a decade ago, this has been changing over time. The count of issuers which explicitly focus on self-indexed ETFs, and issuers offering both types of funds has been steadily increasing since 2012. By the end of our sample period, 51% of issuers in our sample are pure public-indexed issuers, 28% are pure self-indexed and 20% are issuers which offer a mix of public and self-indexed ETFs.[18]

Appendix Table B.1 lists the 10 largest ETF issuers within each group. Total issuer AUMs reported in this table reflect the total AUMs of ETFs issued by each of these issuers within our sample. Several large investment management companies such as Goldman Sachs, JP Morgan, State Street Global Advisors, and Fidelity stand out as issuers offering a mix of self- and public-indexed ETFs. For instance, within our sample of U.S equity passive ETFs, Fidelity offers 6 self- and 11 public-indexed ETFs; JP Morgan offers 5 versus 4; Goldman Sachs 3 versus 2; and State Street offers 3 versus 59. Other issuers with a mix offering include Invesco (2 versus 105), Victory Capital Management (7 versus 2) and Pacer Financial with (8 versus 4), and Northern Trust (5 versus 2), and Charles Schwab (1 versus 10).

When we look at the group of issuers which thus far offer only self-indexed funds, we see

[18]When companies start offering self-indexed ETFs, they generally do this by starting new funds instead of converting the existing ones, hence we oberver very few index-switching events in the data.

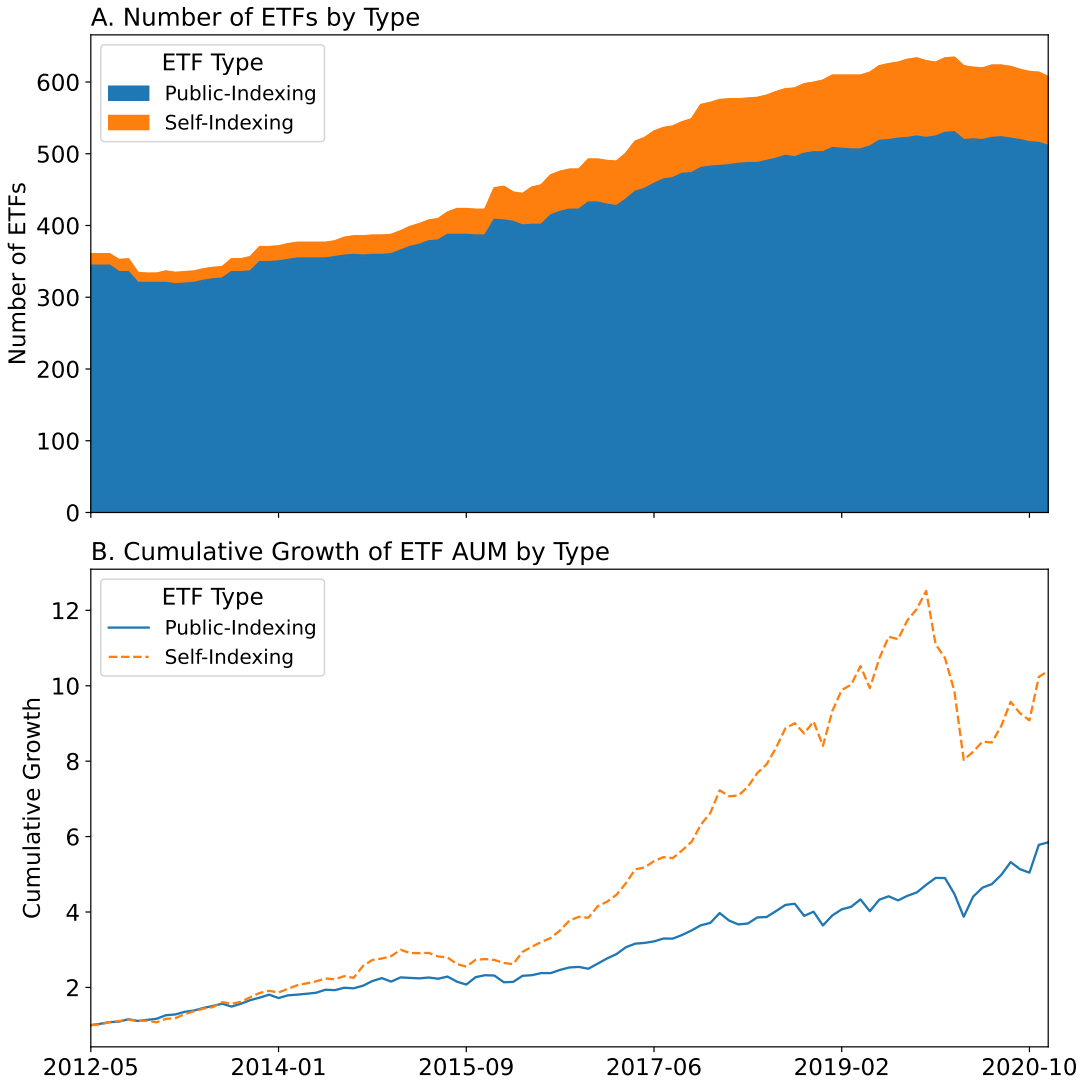


Figure 2.1. Number of self- and public-indexed ETFs over time.
Panel A presents the monthly number of self-indexed and public-indexed ETFs in our sample from May 2012 until December 2020. Panel B shows the cumulative growth of total ETF assets under management (AUM) for both types of ETFs over the same time frame (AUM is indexed to 1 in May 2012). Self-indexed ETFs are ETFs which track an index created and maintained by the ETF issuer themselves or one of its affiliated companies. Public-indexed ETFs are ETFs which track indices from unaffiliated index providers such as S&P Dow Jones or FTSE Russell.

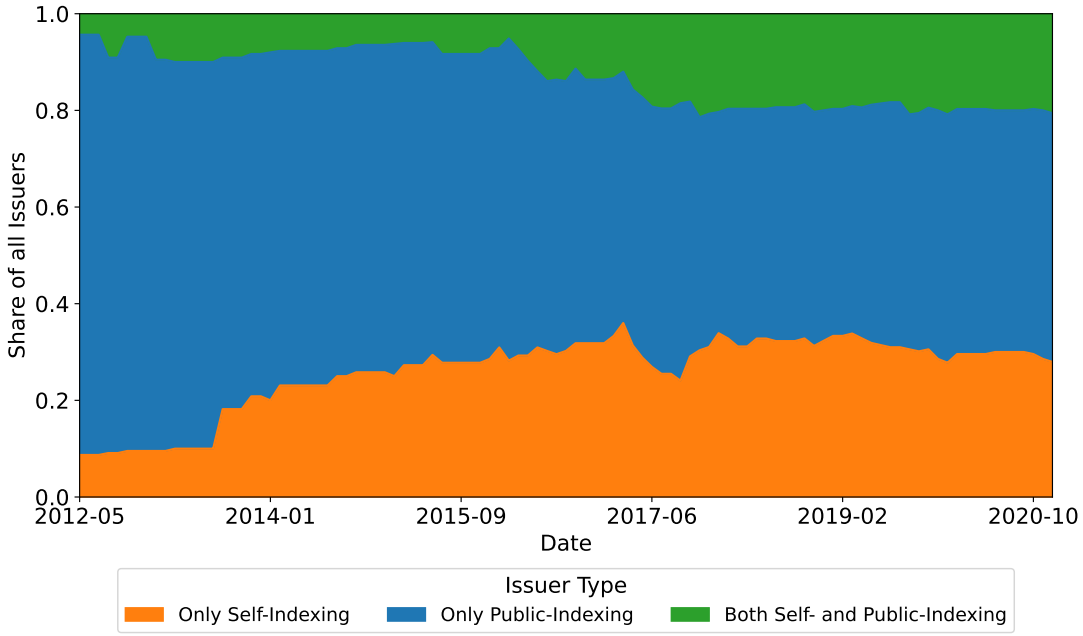


Figure 2.2. Share of issuer types over time.
This figure presents the monthly share of issuers in our sample from May 2012 until December 2020 which provide only self-index, only public-index, or both types of ETFs.

that these issuers tend to be smaller. John Hancock, Renaissance, and American Century are the largest ones within this group. The issuers which thus far offer only public-indexed funds include some of the largest players in the market such as BlackRock and Vanguard. This group however also includes smaller issuers such as Van Eck, DWS, Principal, and Nuveen.

In Table B.2, we consider the full sample of issuers to examine the systematic relation between issuer characteristics and issuers' propensity to offer self-indexed ETFs. For each issuer, we extract the total AUM of all outstanding funds (across all asset-classes), as well as the age relative to the earliest offer date of any fund for each issuer from the CRSP Mutual Fund database. When we look at the role of issuer size (total issuer AUM) and issuer age, and we don't find statistically significant effects. Figure B.1, which shows the distribution of issuer size and issuer age within each issuer group, provides insights into these findings. Self-indexed ETFs are primarily issued by either large and old or small and young issuers and this drives the coefficient

estimates for issuer size and age towards zero in a linear regression model. The only significant effect the regression picks up is the pre-existence of a self-indexed ETF. Issuers which already offer self-indexed ETFs are more likely to offer another one.

Panel A of Table 2.2 shows the descriptive statistics on various fund characteristics for self- versus public-indexed funds. Panel B and C report summary information for our similarity measures, to be defined in the next section. $Net\ Expense\ Ratio\ (\%)$ is the annual net expense ratio after fee waivers, as provided by ETF Global. $Log(Age)$ is the natural logarithm of fund age, where age of an ETF is defined as the difference between the month-end date and the inception date, divided by 365. $Log(AUM)$ is the natural logarithm of fund's total assets under management. If information on AUM for an ETF is missing in ETF Global, we extract it from the CRSP Mutual Fund database. We then remove all remaining observations with missing or zero AUM. All variable descriptions are provided in Appendix Table B.6.

We examine ETFs' gross (before fees) NAV-based returns ($Gross\ Return\ (\%)$). Since ETFs subtract expense ratios from NAV on a daily basis, we first calculate daily gross returns by adding annual net expense ratio/252 (assuming 252 trading days in a year) to daily NAV net returns. Monthly gross returns are then calculated as the cumulative product of one plus each day's return over a full month, minus one. $Gross\ Carhart\ Alpha\ (\%)$ is the fund's excess return estimated over a 36-month rolling window using the 4-factor Carhart model. Following the literature, we define fund flow as:

$$FundFlow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1} \cdot (1 + R_{i,t})}{TNA_{i,t-1}}. \tag{2.1}$$

As expected, public-indexed ETFs are on average older ($Log(Age)$ is 2 versus 1.37) and consequently, they tend to be larger ($Log(AUM)$ is 19.4 versus 18.2). Self-indexed funds, however, seem to be receiving more flows – *Fund Flow* is 0.031 vs 0.019. ETF performance measures seem

Table 2.2. Summary statistics
In Panel A we report summary statistics per ETF type. $NetExpenseRatio(\%)$ is the annual net expense ratio. $GrossReturn(\%)$ is the net (NAV) return adjusted for the net expense ratio. $Volatility$ is the standard deviation of net (NAV) returns in a month. $GrossCarhartAlpha(\%)$ is the excess return over a Carhart four-factor model prediction with a 3-year rolling window. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $FundFlow$ measures the net investment inflow in proportion to the ETFs previous AUM, adjusted for the ETFs returns. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Avg.ReturnCorrtoStyle\text{-}Peers$ is an ETFs average return correlation compared to other ETFs in the same style category. $Inst.SelfOwnership$ is 13F ownership by the same or related investment management companies. $\#AffiliatedModels$ is the number of model portfolios by related investment management companies which report holding the ETF. Ownership variables and model portfolios are available at quarterly frequency only. Panel B reports the average cosine similarity of ETFs holdings compared to other ETFs in the same style-group, grouped by ETF type. Panel C reports the average return correlation with other ETFs in the same style group for the full sample and the five largest style groups.

	Public-Indexing			Self-Indexing		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Panel A: Summary Statistics of Key Variables						
Net Expense Ratio (%)	44261	0.419	0.541	6173	0.428	0.265
Gross Return (%)	44111	1.05	5.44	6091	1.02	5.41
Volatility	44104	0.169	0.13	6086	0.164	0.136
Gross Carhart Alpha (%)	24942	-0.166	2.69	2314	-0.184	2.07
Log(AUM)	43876	19.4	2.42	6081	18.2	1.9
Log(Age)	44256	2	0.748	6173	1.37	0.693
Log(Holdings)	43414	4.88	1.4	6009	5.07	1.46
Fund Flow	43454	0.0198	0.119	5963	0.0319	0.124
Smart Beta?	38888	0.533	0.499	5381	0.775	0.418
Avg. Return Corr to Style-Peers	42797	0.702	0.218	5833	0.737	0.218
Inst. Self Ownership	15036	0.0183	0.0937	2083	0.14	0.283
# Affiliated Models	15050	0.695	4.01	2104	0.874	4
Panel B: Average Cosine Similarity of Holdings Compared to Style-Peers						
Full Sample	551	0.263	0.147	106	0.267	0.146
Large Cap	119	0.318	0.137	34	0.328	0.133
Mid Cap	40	0.267	0.109	5	0.288	0.037
Small Cap	58	0.257	0.112	15	0.231	0.0815
Broad Equity	61	0.249	0.14	11	0.246	0.107
Theme	40	0.201	0.13	16	0.139	0.0968
Panel C: Average Return Correlation with Style-Peers						
Full Sample	590	0.874	0.131	113	0.876	0.199
Large Cap	141	0.911	0.0497	35	0.919	0.0562
Mid Cap	39	0.937	0.024	6	0.876	0.189
Small Cap	62	0.891	0.166	14	0.925	0.039
Broad Equity	62	0.914	0.0489	15	0.902	0.062
Theme	36	0.881	0.0668	15	0.852	0.109

comparable. *Gross Return* for public- vs self-indexed ETFs is 1.05 vs 1.02, and *Cahart Alpha* is -0.166 and -0.184. In the following sections, we will formally examine the systematic differences in fees, performances and flows between the two types of ETFs in regression models where we control for various fund characteristics, investment styles, issuer and time fixed effects.

This panel also reports summary statistics on our ownership variable (available at quarterly frequency). $Institutional\ Self\ Ownership$ is the fraction of total ETF shares reported in 13F filings of the ETF issuing company or one of its affiliates. Self-indexed ETFs have 14% self-ownership, while public-indexed ETFs have only 2% on average. Later in the paper, we formally examine also the ownership differences between the two groups of ETFs.

2.3 Do self-indexed ETFs provide financial benefits to investors?

2.3.1 Differences in annual fees

We start by analyzing the extent to which fees charged by self-indexing ETFs differ from those charged by public-indexing ETFs. Table 2.3 reports the results. Throughout all regressions, standard errors are two-way clustered by month-issuer interaction, and we find that self-indexed ETFs have fees that are statistically higher.

In column 1, we first regress $Net\ Expense\ Ratio\ (\%)$ on $SelfIndexer?$, which is an indicator variable that equals one if the fund is identified as a self-indexed ETF, including month-style fixed effects. The coefficient estimate for $SelfIndexer?$, is 0.055 (5.5 basis points) and statistically significant. In the second column, we extend the model by including several control variables $Log(AUM)$, $Log(Age)$, $Log(Holdings)$, $SmartBeta?$ (an indicator variable taken from Morningstar which equals one for smart beta ETFs) as well as issuer fixed effects. The coefficient estimate for $SelfIndexer?$ stands similarly significant at 0.052.

As ETFs fees are in basis points on average, these estimates are economically meaningful: they imply that self-indexed ETFs on average charge 13% more their public-indexed counter-

Table 2.3. Explaining net expense ratios.
The dependent variable is an ETFs annual Net Expense Ratio in percentage terms. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Uniqueness$ is the 12-month rolling average absolute difference of ETF gross returns versus the value-weighted average ETF gross return in the same style-category, as in Kostovetsky and Warner (2025). $Activeness$ is defined as $(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model. $TurnRatio$ is an ETFs yearly portfolio turnover ratio. $MAvgSpread$ is the monthly average ETF secondary market price spread scaled by price. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Net Expense Ratio (%)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$SelfIndexer?$	0.055***	0.052***	0.057***	0.056***	0.045***	0.052***	0.035***
	(0.006)	(0.008)	(0.011)	(0.009)	(0.008)	(0.008)	(0.013)
$Log(AUM)$		−0.023***	−0.024***	−0.025***	−0.022***	−0.023***	−0.027***
		(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$Log(Holdings)$		−0.012***	0.000	0.005**	−0.010***	−0.012***	0.026***
		(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.004)
$Log(Age)$		0.070***	0.075***	0.078***	0.066***	0.070***	0.083***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
$SmartBeta?$		0.001	0.001	0.007	0.000	0.001	0.004
		(0.010)	(0.012)	(0.010)	(0.010)	(0.010)	(0.012)
$Uniqueness$			0.056***				0.046***
			(0.004)				(0.004)
$Activeness$				0.712***			0.796***
				(0.059)			(0.073)
$TurnRatio$					0.017***		0.038***
					(0.003)		(0.009)
$MAvgSpread$						−0.003	−0.005
						(0.003)	(0.005)
Num. Obs.	50 434	43 214	35 644	42 938	42 096	43 173	34 969
R^2	0.23	0.47	0.50	0.49	0.48	0.47	0.55
FE: Issuer		X	X	X	X	X	
FE: Month-Style	X	X	X	X	X	X	X
FE: Month-Issuer							X

* p < 0.1, ** p < 0.05, *** p < 0.01

parts. Although ETF fees are seemingly low on an annual basis, even small differences in fees can have an important impact on long-term investments.[19] These results are interesting in that they represent contrary evidence to the popular view expressed by numerous industry commentators that self-indexing can be a way for ETF issuers to offer more competitive fees.[20]

We next assess our main findings against related research and demonstrate how they differ from previous work. It is useful to first highlight two points. We incorporate style fixed effects throughout our analysis, to ensure that our findings are not driven by specific style categories, such as thematic ETFs (as discussed in Ben-David et al. (2022)). We also include issuer fixed effects to confirm that these results are not attributable to potential differences in issuer-level factors, such as differences in brand recognition, customer services or other business models that might influence their fees or strategies (Hortaçsu and Syverson, 2004; Kostovetsky and Warner, 2025).

Kostovetsky and Warner (2025) document that certain ETFs employ distinct investment strategies compared to their peers, often accompanied by higher fees. To ensure this does not account for our results, we control for their measure of *Uniqueness* (the 12-month rolling absolute difference in a fund's gross returns versus its style category average). Consistent with the literature, *Uniqueness* is associated with higher fees. Crucially, however, our core finding persists after controlling for this measure (column 3). We will formally examine portfolio differentiation in the next section to evaluate related hypotheses around it.

In columns 4 and 5, we evaluate the potential role of the "closet active management" of passive ETFs (Akey et al., 2021; Easley et al., 2021; Cheng et al., 2019). It is useful to note that we control for smart-beta in our baseline regressions. Moreover, we address this point by two additional proxies of activeness, following the related literature. In column 4, we include

[19] See, for instance, French (2008), Sharpe (2013) and this SEC Bulletin.

[20] For instance, Todd Rosenbluth, director of ETF and Mutual Fund Research at CFRA has recently said "Self-indexing is a reaction to investor preference for low-cost strategies (...) licensing a third-party index can be expensive." For more, also see in the Financial Times, Barron's and WatersTechnology.

$Activeness$, defined as $(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model (Amihud and Goyenko, 2013), in our regression. In column 5, we introduce *TurnRatio*, which is ETF's yearly portfolio turnover ratio (Pástor et al., 2017). With both measures, our finding remain stable, indicating self-indexing is empirically distinct from proprietary indexing practices identified by Akey et al. (2021) and more broadly, closet active management.

In column 6, we control for the influence of ETF secondary market liquidity, which is a dimension of ETF competition (Khomyn et al., 2024) and thus a potential confounder for fund fees. As a proxy for share liquidity, we use an ETFs bid-ask spread, *MAvgSpread*, which is defined as the monthly average difference between the ask and bid as reported in CRSP, scaled by the price. Our result is virtually unchanged. Finally, in column 7, we consider all of the additional control variables together alongside month-style and month-issuer fixed effects, and confirm that our results are robust to additional extensions of model specification. To summarize, this section demonstrates that self-indexed ETFs charge significantly higher fees, contradicting the predicted outcome of fee competition with public-indexed funds.

2.3.2 Differentiated products?

In this section, we build on our earlier discussion and delve deeper into whether self-indexing provides issuers the flexibility to create distinct investment strategies that may provide performance gains to customers. Our analysis reveals that self-indexed ETFs are, in fact, more similar to their peers within the same style, rather than being distinct. They also do not provide significantly better performance. Overall, our results contradict the differentiated products hypothesis.

2.3.2.1 Differences in holdings and return correlations

We start by introducing a cosine similarity measure between each pair of ETFs within our sample, to examine the differences in holdings (Sias et al., 2016). We pull the end-of-month

portfolios of all ETFs in the sample from ETF Global and use all equity holdings with a valid ticker. For each pair of ETFs within our sample, we identify months where both ETFs have holding data (index for overlapped months is m) and calculate the cosine similarity between the two ETF's portfolios. Specifically, i is the index of stocks held by either ETF, ETF A's holding for stock i is $A_{i,m}$ dollars and ETF B's holding for stock i is $B_{i,m}$, where $A_{i,m}$ and $B_{i,m}$ equal zero if the ETF does not hold stock i. The cosine similarity is calculated as:

$$Cos_{A,B,m} = \frac{\sum_{i=1}^{N} A_{i,m} B_{i,m}}{\sqrt{\sum_{i=1}^{N} A_{i,m}^2} \cdot \sqrt{\sum_{i=1}^{N} B_{i,m}^2}} \tag{2.2}$$

N is the total number of stocks that either ETF A or B holds. By definition $Cos_{A,B,m} = Cos_{B,A,m}$. We take the numerical average of $Cos_{A,B,m}$ across all months m, and obtain our pairwise cosine similarity measure, $Cos_{A,B}$.[21] Fund-level cosine similarity is constructed by averaging the cosine similarity of each fund against all other "peer" funds within the same style category and same month. As an alternative, for each ETF-month, we also calculate average pairwise daily return correlations (against its peer funds within the same style). Panel B and C of Table 2.2 report summary statistics for fund-level cosine similarity and return correlations, respectively. Return correlations tend to be quite high (around 0.90) while cosine similarity appears relatively lower (around 0.3), arguably due to synthetic replication, value-weighting (with value weighting, a handful of large companies drive a significant portion of index returns) and exponential changes in cosine similarity measure due to its definition.

Our focus is to evaluate the differences in similarity measures between self- and public-indexed ETFs. We start by conducting network analyses to visualize the potential differentiation. Based on $Cos_{A,B}$, we calculate a minimum spanning tree representing the network of

[21] To confirm our calculations we checked the cosine similarity between well-established S&P 500 ETFs such as SPY, IVV, and VOO and find that their measures of cosine similarity are greater than 0.998.

all ETFs in our sample. Intuitively, this method calculates a parsimonious network which connects all funds, minimizing the necessary connections, where funds with more similar holdings become bunched together, forming "clusters". In Figure 2.3, we plot this network using the Fruchterman Reingold layout, which minimizes overlapping connections and spreads the network evenly on a canvas. The network diagrams clearly illustrate that self-indexed ETFs are generally situated at the center of various clusters, which demonstrates a strong similarity with peer funds in the same style. We reproduce the same diagrams using return correlations in B.2. Our conclusions carry over.

Table 2.4 presents a formal regression analysis. In columns 1 and 2, the dependent variable is the average monthly cosine similarity of a fund to its peers. We regress this variable on various ETF characteristics including $SelfIndexer?$, $Log(AUM)$, $Log(Age)$, $Log(Holdings)$ and $SmartBeta?$ as well as issuer, style and month fixed effects. Our findings consistently show that, when compared to their public-indexed peers, self-indexed ETFs have more *similar* (i.e., less differentiated) portfolio holdings. In columns 3 and 4, we replace cosine similarity with average monthly return correlations (as before, measured against peers within the same style-category). The results are similar.[22]

In summary, we find no evidence of portfolio differentiation. If anything, we find self-indexed ETFs are actually more similar to their peer funds within the same style categories, compared with their public-indexed counterparts. Although our analysis shows no differentiation, investors may still perceive these products as distinct due to marketing and branding. This perception aligns with our third hypothesis, which will be examined in a later section.

[22]Table B.3 shows the robustness of these findings when we also control for additional explanatory variables (Uniqueness, Activeness, Turnover and Spread) as in Table 2.3.

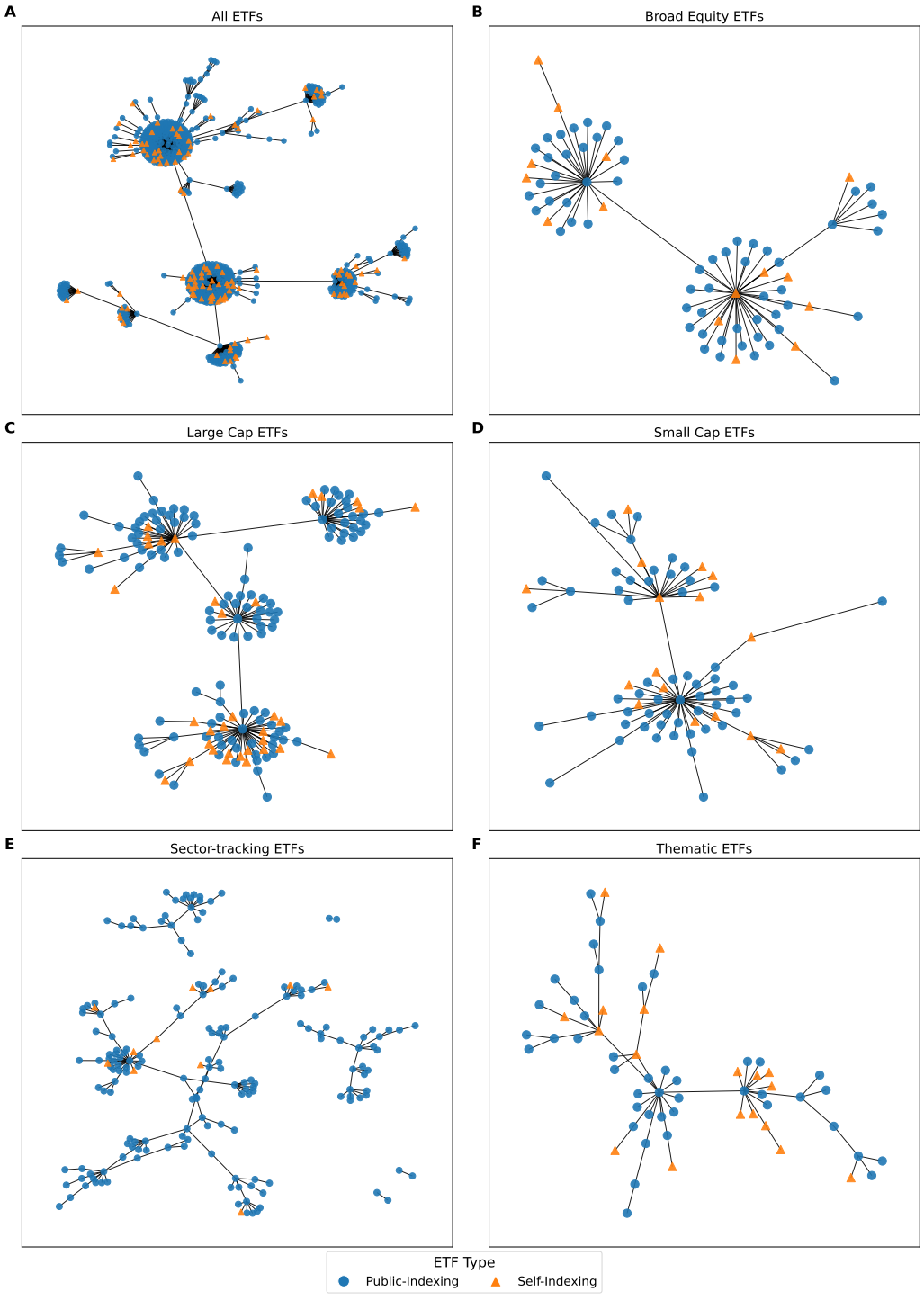


Figure 2.3. ETF networks based on cosine similarity of stock holdings.
This figure presents minimum spanning tree network plots of ETFs based on their monthly average pairwise cosine similarity of stock holdings. The networks are plotted using the Fruchterman Reingold Layout. Panel A presents the network of all ETFs whereas panels B to F plot networks within ETF subgroups.

Table 2.4. Explaining ETF differentiation.
In columns (1) - (2) we predict the monthly average portfolio cosine similarity of an ETF to its style-peers and in columns (3) - (4) we predict the average return correlation between an ETF and its style-peers within a month. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $NetExp.Ratio(\%)$ is the ETFs annual expense ratio in percentage terms. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Cosine Similarity		Return Correlation	
	(1)	(2)	(3)	(4)
$SelfIndexer?$	0.019***	0.021***	0.015***	0.018***
	(0.002)	(0.002)	(0.004)	(0.004)
$NetExp.Ratio(\%)$	−0.007***	−0.006***	−0.013***	−0.011***
	(0.001)	(0.001)	(0.002)	(0.002)
$Log(AUM)$	0.001***	0.001***	−0.003***	−0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
$Log(Holdings)$	0.028***	0.035***	0.020***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
$Log(Age)$	−0.002	−0.001	0.015***	0.012***
	(0.001)	(0.002)	(0.002)	(0.002)
$SmartBeta?$	−0.008***	−0.006***	0.010***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
Num. Obs.	29 902	29 902	41 887	41 887
R^2	0.76	0.78	0.83	0.86
FE: Issuer	X		X	
FE: Month-Style	X	X	X	X
FE: Month-Issuer		X		X

* p < 0.1, ** p < 0.05, *** p < 0.01

2.3.2.2 Differences in performance

We next move to evaluating the potential differences in fund performances. We conduct multiple analyses, using different factor models, daily and monthly data, and tests at both the time-series portfolio and cross-sectional fund levels. Across all these tests, we never find significant evidence that self-indexed ETFs outperform their public-indexed counterparts.

We start by presenting results of various portfolio tests in which we estimate alphas to the 'spread' portfolio – the portfolio that goes long on self-indexed ETFs and short on public-indexed ETFs. Columns 1 and 2 of Table 2.5 report the monthly alphas, while columns 3 and 4 report the daily alphas. The daily analysis not only provides a higher statistical power for our tests but also serves as an independent check on the results derived from the monthly alphas. For both monthly and daily alphas, we report results for value- and equal-weighted portfolio returns, and use three different factor models: the CAPM, the Fama-French three-factor and Carhart four-factor model. In total, we estimate alphas under 12 different scenarios. All returns are before fees and we require at least 20 observations in each portfolio of funds.[23] We use Newey-West standard errors allowing correlation up to 6 months. We find that all 12 alphas for the spread portfolios are statistically indistinguishable from zero, indicating no significant performance difference between the two types of ETFs.

Next, we move to cross-sectional fund level regression analyses where we control for the influence of fund characteristics on fund performance (Table 2.6). The dependent variable is the gross risk-adjusted return where returns are adjusted using a Carhart four-factor model with loadings to each factor estimated from 36-months rolling windows. Columns 1 and 2 use monthly, and columns 3 and 4 use daily alphas, significantly increasing the number of observations. The main variable of interest is $SelfIndexer?$ and other explanatory variables include

[23] Effectively, this moves the start date for this analysis to July 2015, marginally shortening the sample period. Cross-sectional regression tests (presented next) provide an alternative empirical approach and show the robustness of our findings.

Table 2.5. Alphas from portfolio tests with monthly and daily returns.
Self-Indexers (-) Public-Indexers represents alphas from returns of the portfolio of self-indexed ETFs minus the returns of the portfolio of public-indexed ETFs. We report alphas from predictions of gross returns in excess of the risk-free rate. Gross returns are calculated by backing out net expense ratios from net (NAV) returns. To predict excess gross returns we use the CAPM (Panel A), the Fama-French three-factor model (Panel B) and the Carhart four-factor model (Panel C). Only observations with at least 20 ETFs per portfolio are included. We use Newey-West Standard Errors with six lags and report them in parentheses.

	Monthly Returns (%)		Daily Returns (%)	
	Gross, VW	Gross, EW	Gross, VW	Gross, EW
Panel A: CAPM				
Self-Indexers (-) Public-Indexers	-0.135	0.061	0.006	0.001
	(0.088)	(0.049)	(0.007)	(0.002)
Panel B: FF3				
Self-Indexers (-) Public-Indexers	-0.011	0.033	0.004	0.000
	(0.069)	(0.045)	(0.006)	(0.002)
Panel C: FF3 + Mom				
Self-Indexers (-) Public-Indexers	-0.012	0.033	0.004	0.001
	(0.069)	(0.045)	(0.006)	(0.002)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

$Log(AUM)$, $Log(Age)$, $Log(Holdings)$, $SmartBeta?$, $Volatility$ as well as a host of month, style, and issuer fixed effects. The results are consistent with the portfolio tests. There are no significant performance differences between self- and public-indexed ETFs.[24]

In Table 2.7, we consider the degree to which self- and public-indexed funds vary on volatility and Sharpe ratio. In panel A, the dependent variable is $Volatility$, which is defined as the standard deviation of daily net fund returns within a month, and in panel B, it is $Sharpe\ Ratio$, which is the monthly net fund return divided by the fund volatility. Self-indexed ETFs appear to have somewhat lower volatility, however estimates are economically modest (around a 2-4% difference). In a similar vein, self-indexed ETFs seem to have either similar or weakly lower Sharpe Ratios. Taken altogether, our findings highlight that self-indexing ETFs do not deliver superior performance outcomes compared to their public-indexed counterparts.

[24] Appendix Table B.3 shows the results when we control for additional explanatory variables as in Table 2.3. There is no evidence that self-indexed ETFs perform better. If anything, in the analysis with daily alphas, their performance seems weakly negative.

Table 2.6. Performance regressions.
The dependent variable is the monthly (daily) gross Carhart four-factor excess return, defined as the gross returns minus predicted returns using factor loadings from 36 month (126 day) rolling windows. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. Standard errors are two-way clustered by time-issuer interaction and reported in parentheses.

	Gross Carhart Four-Factor Excess Returns (%)			
	Monthly Returns		Daily Returns	
	(1)	(2)	(3)	(4)
$SelfIndexer?$	−0.045	−0.042	−0.003	−0.003
	(0.088)	(0.099)	(0.003)	(0.003)
$Log(AUM)$	−0.012	−0.013	0.000	0.000
	(0.010)	(0.011)	(0.000)	(0.000)
$Log(Holdings)$	−0.020	−0.022	0.000	0.000
	(0.020)	(0.021)	(0.001)	(0.001)
$Log(Age)$	0.002	0.000	−0.001	−0.001
	(0.054)	(0.057)	(0.001)	(0.001)
$SmartBeta?$	−0.088**	−0.096***	−0.003**	−0.003**
	(0.034)	(0.035)	(0.001)	(0.001)
$Volatility$	−0.063	−0.544	−0.010	−0.025
	(0.653)	(0.675)	(0.029)	(0.029)
Num. Obs.	23 929	23 929	843 273	843 273
R^2	0.54	0.60	0.53	0.60
FE: Issuer	X		X	
FE: Month-Style	X	X		
FE: Month-Issuer		X		
FE: Day-Style			X	X
FE: Day-Issuer				X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7. Explaining other ETF characteristics.
In Panel A we predict ETF volatility, defined as the standard deviation of daily net (NAV) returns within a month. In Panel B we predict Sharpe Ratios, defined as the monthly net (NAV) returns divided by the volatility. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	(1)	(2)	(3)
Panel A: Volatility			
$SelfIndexer?$	-0.005***	-0.007***	-0.008***
	(0.001)	(0.001)	(0.001)
$Log(AUM)$		-0.003***	-0.003***
		(0.000)	(0.000)
$Log(Holdings)$		-0.002***	-0.003***
		(0.000)	(0.000)
$Log(Age)$		0.007***	0.006***
		(0.001)	(0.001)
$SmartBeta?$		-0.006***	-0.006***
		(0.001)	(0.001)
Num. Obs.	50190	43074	43074
R^2	0.90	0.92	0.93
Panel B: Sharpe Ratio			
$SelfIndexer?$	-3.266	-7.239*	-9.197*
	(2.386)	(3.989)	(5.490)
$Log(AUM)$		-0.455	-0.589
		(0.480)	(0.511)
$Log(Holdings)$		2.180	3.031
		(2.040)	(2.724)
$Log(Age)$		0.068	0.420
		(1.136)	(1.126)
$SmartBeta?$		1.166	1.416
		(1.526)	(1.707)
Num. Obs.	50189	43073	43073
R^2	0.14	0.13	0.18
FE: Issuer		X	
FE: Month-Style	X	X	X
FE: Month-Issuer			X

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

2.4 Self-preferencing by issuer-advisors?

Our analyses thus far reveal that, on average, self-indexed funds neither perform better nor provide portfolio differentiation. In this section, we examine the "self-preferencing hypothesis": index proliferation increases search frictions and ETF issuers which are also in the business of providing advisory services promote their own high-fee self-indexed ETFs to arguably high search-cost clients. The results of this section are consistent with this hypothesis.

2.4.1 Investment advisors and self-indexed ETFs

We start by analyzing whether self-indexed ETFs issued by issuers which are likely to be more prone to self-preferencing charge higher fees. To test this idea, we classify issuers into two groups: 'specialized fund managers' and 'investment advisors' (or 'issuer-advisors'). While the former group explicitly focuses on fund management (e.g., Global X, Defiance ETFs, Van Eck), the latter group offers a range of financial services including fund management as well as wealth management advisory services (e.g., Fidelity, State Street, Goldman Sachs) and is more prone to self-preferencing. There are 105 unique issuers in our sample, with 34 classified as investment advisor and 71 as specialized fund managers. A full list of these issuers is provided in Appendix Table B.5.

Column 1 of Table 2.8 tests the differences in fees between the two groups of issuers. We take the specification of column 2 of Table 2.3 and introduce the interaction term $SelfIndexer?$ x $InvAdvisor?$, where $InvAdvisor?$ is an indicator variable equal to one for ETFs issued by issuers classified as 'investment advisors'. The regression includes month-style fixed effects alongside the control variables from Table 3. The results are striking. Investment advisors, on average, charge lower fees, which is consistent with these companies being able to offer more competitive prices for their products. However, their self-indexed ETFs systematically charge higher fees. The $SelfIndexer?$ x $InvAdvisor?$ interaction term is significantly positive.

The coefficient for $SelfIndexer?$ alone, however, is indistinguishable from zero: among ETFs issued by specialized fund managers, there is no significant difference in fees between self- and public-indexed ETFs.

Table 2.8. Exploring the role of investment advisors.
In column (1) we predict annual Net Expense Ratios in percentage terms. In column (2) we predict Gross Returns, defined as net (NAV) returns adjusted for net expense ratios. In column (3) we predict the average cosine similarity of holdings of an ETF compared to all other ETFs in the same style category. In column (4) we predict next-period Fund Flow, defined as the net investment inflow in proportion to the ETFs previous AUM, adjusted for returns. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. $AvgNetRet(Yr)(\%)$ is the 12-month rolling average net (NAV) ETF return. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Net Expense Ratio (%)	Gross Return (%)	Cosine Similarity	Fund Flow (t+1)
	(1)	(2)	(3)	(4)
$SelfIndexer?$	0.002	0.136	−0.020***	0.011**
	(0.013)	(0.086)	(0.002)	(0.005)
$InvAdvisor?$	−0.320***	−0.006	0.015***	0.005**
	(0.011)	(0.042)	(0.002)	(0.002)
$Log(AUM)$	−0.024***	0.018*	0.004***	−0.002***
	(0.001)	(0.009)	(0.000)	(0.000)
$Log(Holdings)$	−0.045***	0.001	0.032***	0.001
	(0.003)	(0.014)	(0.001)	(0.001)
$Log(Age)$	0.061***	−0.022	−0.010***	−0.016***
	(0.004)	(0.025)	(0.001)	(0.002)
$SmartBeta?$	0.049***	−0.050**	−0.011***	0.000
	(0.008)	(0.025)	(0.001)	(0.002)
$SelfIndexer?$:$InvAdvisor?$	0.090***	−0.120	0.013***	−0.019***
	(0.016)	(0.093)	(0.003)	(0.006)
$Volatility$		1.141		−0.003
		(0.735)		(0.017)
$AvgNetRet(Yr)(\%)$				0.017***
				(0.001)
Num. Obs.	43 214	43 074	29 902	35 955
R^2	0.21	0.88	0.70	0.17
FE: Month-Style	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

In columns 2 and 3, respectively, we check whether self-indexed ETFs from investment advisors perform differently or have more differentiated portfolios. We find no such evidence. There is no significant difference in performance, and if anything, self-indexed ETFs by invest-

ment advisors are more similar to other ETFs in the same style category. Self-indexed ETFs issued by specialized fund managers, on the other hand, seem to offer more differentiated portfolios.

In column 4, we analyze the differences in investor flows. We first observe that $SelfIndexer?$ is significantly positive, consistent with our previous discussion that self-indexed funds have been gaining traction over the past decade attracting more flows than public-indeed funds. Moreover, the coefficient for $InvAdvisor?$ is also significantly positive (albeit smaller in magnitude), demonstrating that ETFs offered by such issuers on average attract more capital. However, $SelfIndexer? : InvAdvisor?$ is negative. Overall estimates indicate that, on average, these self-indexed ETFs have flows comparable with those of a typical public-indexed ETF with similar characteristics.

In Table 2.9, we focus on testing how flows to self-indexed ETFs vary with aggregate market sentiment. Under the self-preferencing hypothesis, index proliferation amplifies search frictions, allowing issuers to market similar portfolios at higher fees, arguably to high search-cost clients. These clients can be characterized by limited financial sophistication (Roussanov et al., 2021). If self-indexed ETFs are popular with such clients, we can expect flows of self-indexed funds to be more sensitive to market sentiment. Using the University of Michigan Consumer Sentiment Index as a proxy for aggregate market sentiment (e.g., Giannetti and Kahraman (2017)), we confirm that this is indeed the case. During periods of heightened market sentiment, self-indexed funds experience significantly higher flows, while publicly indexed funds display no comparable sensitivity to sentiment.

2.4.2 Institutional self-ownership of self-indexed ETFs

Under the self-preferencing hypothesis, investment advisors promote their self-indexed ETFs through the advisory part of their businesses. These assets would appear in the 13F filings of the investment advisor to the extent that some of these assets are purchased via (and held in)

Table 2.9. Sentiment-flow sensitivity.
In all specifications we predict next period fund flow, defined as the net investment inflow in proportion to the ETFs previous AUM, adjusted for returns. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. $Log(MichSent)$ is the natural logarithm of the University of Michigan Consumer Sentiment Index, $MichSentNorm$ is its normalized value and $MichSentTophalf$ a dummy variable that equals to one when the index is above its median value. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Fund Flow (t+1)		
	(1)	(2)	(3)
$SelfIndexer?$	-0.335^{***}	-0.010^{**}	-0.015^{***}
	(0.087)	(0.004)	(0.005)
$AvgNetRet(Yr)(\%)$	0.011^{***}	0.011^{***}	0.011^{***}
	(0.001)	(0.001)	(0.001)
$Log(AUM)$	-0.003^{***}	-0.003^{***}	-0.003^{***}
	(0.001)	(0.001)	(0.001)
$Log(Holdings)$	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
$Log(Age)$	-0.020^{***}	-0.020^{***}	-0.020^{***}
	(0.002)	(0.002)	(0.002)
$SmartBeta?$	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)
$Volatility$	0.039^{***}	0.038^{***}	0.037^{***}
	(0.007)	(0.007)	(0.006)
$Log(MichSent)$	0.009		
	(0.012)		
$SelfIndexer?$:$Log(MichSent)$	0.073^{***}		
	(0.019)		
$MichSentNorm$		0.001	
		(0.002)	
$SelfIndexer?$:$MichSentNorm$		0.011^{***}	
		(0.003)	
$MichSentTophalf$			0.002
			(0.003)
$SelfIndexer?$:$MichSentTophalf$			0.014^{***}
			(0.005)
Num. Obs.	35 955	35 955	35 955
R^2	0.04	0.04	0.04
FE: Issuer	X	X	X
FE: Style	X	X	X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

discretionary advisory accounts.[25] If this is the case, we would expect to find Institutional Self-ownership (fraction of ETF's total shares outstanding which are reported by the ETF issuer, or an affiliated company, in 13F filings) to be disproportionately high for self-indexed ETFs issued by investment advisors.

The results reported in Table 2.10 confirm this prediction. As noted previously, when we look at the summary statistics (Table 2.2), we observe that self-indexed ETFs have 14% self-ownership, while public-indexed ETFs have only 2% on average. Regression analyses formally establishes the significance of this difference. In column 1, the estimate for the self-index coefficient is 11%, revealing a significant difference in Institutional Self-ownership on average. In column 2, we include the $SelfIndexer?$ x $InvAdvisor?$ interaction and find that the difference in self-ownership is largely driven by investment advisors. In column 2, the self-indexed dummy has a positively significant but economically small coefficient (around 3.5%). However, the estimate for $SelfIndexer?$ x $InvAdvisor?$ stands at 9.5%, which is both statistically significant and economically important. The investment advisor estimate alone is statistically not different from zero. Overall, our findings provide supporting evidence for the self-preferencing hypothesis.

One might wonder whether self-indexed ETFs are bundled with advisory services and their fees, in part, reflect compensation for such services. This is highly unlikely for two reasons. First, just like public-indexed ETFs, self-indexed ETFs are exchange traded and also widely available for direct purchase by individual investors without an advisor. Second, investment advisors are compensated for selling or recommending ETFs through "client-paid" fees, which are *separate* from ETF expense ratios.[26] Client-paid fees are recurring annual fixed or percentage charges based on the total value of a client's account. This is different from the older commis-

[25] Institutional investment managers in the United States that exercise investment discretion over $100 million in Section 13(f) securities are required to file Form 13F including the assets that they own or the client assets that manage. See Section 13(f)(1)(b) of the Securities Exchange Act, and 15 U.S.C. 78c(a)(35).

[26] For instance, see 2025 SEC Bulletin.

Table 2.10. Explaining institutional self ownership.
The dependent variable is Institutional Self Ownership, defined as 13f ownership by the same or related investment management companies. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $AvgNetRet(Yr)(\%)$ is the 12-month rolling average net (NAV) ETF return. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. Standard errors are two-way clustered by quarter-issuer interaction and reported in parentheses.

	Institutional Self Ownership	
	(1)	(2)
$SelfIndexer?$	0.108***	0.035**
	(0.017)	(0.015)
$Log(AUM)$	0.008***	0.008***
	(0.001)	(0.001)
$Log(Holdings)$	0.011***	0.009***
	(0.003)	(0.002)
$Log(Age)$	−0.052***	−0.053***
	(0.006)	(0.006)
$SmartBeta?$	0.033***	0.029***
	(0.005)	(0.005)
$AvgNetRet(Yr)(\%)$	−0.001	−0.001
	(0.002)	(0.002)
$InvAdvisor?$		0.000
		(0.004)
$SelfIndexer?$:$InvAdvisor?$		0.095***
		(0.025)
Num. Obs.	12 307	12 307
R^2	0.18	0.19
FE: Quarter-Style	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

sions model used by broker-sold mutual funds where part of the advisory compensation could be paid via a fund's expense ratio (specifically, 12b-1 fees).

2.4.3 Model portfolio recommendations and self-indexed ETFs

In this section, we provide a mechanism test by evaluating whether model portfolio recommendation is one channel through which issuers promote their self-indexed ETFs. Model portfolios are recommendation portfolios designed by asset managers and strategists for finan-

cial advisors. Some asset management companies play multiple roles such as managing and issuing funds, creating indices as well as providing investment advice and model portfolio recommendations. Brogaard et al. (2021) show that asset manager model providers recommend their affiliated ETFs more frequently and therefore enjoy significant flows. Building on these findings, we are interested in examining whether ETF issuers which are model providers are more likely to include their self-indexed ETFs in their recommended model portfolios compared to their public-indexed ETFs. Reportedly, 54% of advised assets are invested in model portfolios.[27] Therefore, model portfolios serve as a useful proxy for client portfolios which are otherwise confidential.

For each global and US-focused model portfolio in the Morningstar Direct database, we extract quarterly portfolio holdings between 2012 and 2020. 133 unique model providers report 1125 model portfolios with at least one ETF. Within our sample, 322 out of 785 unique ETFs are at least part of one model portfolio, and 29 out of 105 unique issuers are identified as model portfolio providers. Of these 29 ETF issuers, 20 are investment advisors. This means that a much higher percentage of investment advisors provide model portfolios. Only 12% of the specialized fund managers offer model portfolios (9 out of 71). The same figure is 58% for investment advisors (20 out of 34).

We classify an ETF as affiliated to a model provider if the ETF issuing company is affiliated with the model provider (as in Brogaard et al. (2021)). For example, an iShares S&P 500 ETF in the "BlackRock 50/50 Long-Horizon Allc" model portfolio counts as an affiliated ETF. Out of 29 model providing ETF issuers within our sample, 17 include their ETFs in their model portfolios.[28] Out of these 17 issuers, 11 are classified as investment advisors. In sum, 32% of investment advisors within our sample (11 out of 34) offer model portfolios which include at

[27] https://www.broadridge.com/_assets/pdf/broadridge-fa-model-portfolio-may-2019.pdf

[28] The 17 issuers which include their affiliated ETFs in their model portfolios are Vanguard, First Trust, SSgA, WisdomTree, Goldman Sachs, Northern Trust, John Hancock, BlackRock, Charles Schwab, Fidelity, Franklin Templeton Investments, Inspire Investing, Nuveen, Global X, Invesco, Columbia, and JPMorgan.

least one of their own or affiliated ETFs.

For each ETF quarter, we count the number of affiliated model portfolios an ETF is part of. If we cannot match an ETF to any model portfolio, then the ETF is taken to have zero affiliated model portfolios. This is the main dependent variable for the regressions in Table 2.11. Explanatory variables include $Log(AUM)$, $Log(Age)$, $Net\ Expense\ Ratio\ (\%)$, $AvgNetRet(Yr)\ (\%)$ and a smart-beta flag. Standard errors are clustered by quarter-issuer interaction.

Table 2.11. Explaining affiliated model portfolio membership.
The dependent variable is the number of affiliated model portfolios an ETF is a part of. $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $NetExp.Ratio(\%)$ is the annual net expense ratio. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $AvgNetRet(Yr)(\%)$ is the 12-month rolling average net (NAV) ETF return. $InvAdvisor?$ is a dummy variable which is one if the ETF issuer is classified as investment advisor as opposed to a fund specialist. Standard errors are two-way clustered by quarter-issuer interaction and reported in parentheses.

	# Affiliated Model Portfolios			
	(1)	(2)	(3)	(4)
$SelfIndexer?$	−0.083	0.788*	1.171**	−1.190***
	(0.184)	(0.454)	(0.513)	(0.164)
$NetExp.Ratio(\%)$	−0.018	−0.021	−0.029	0.088***
	(0.032)	(0.042)	(0.041)	(0.029)
$Log(AUM)$	0.540***	0.737***	0.769***	0.530***
	(0.063)	(0.091)	(0.094)	(0.062)
$Log(Age)$	−0.788***	−2.036***	−2.156***	−0.853***
	(0.175)	(0.315)	(0.330)	(0.186)
$SmartBeta?$	0.367***	−0.442***	−0.425***	0.336***
	(0.080)	(0.089)	(0.090)	(0.082)
$AvgNetRet(Yr)(\%)$	−0.091	−0.084	−0.043	−0.088
	(0.082)	(0.080)	(0.085)	(0.082)
$InvAdvisor?$				0.339***
				(0.130)
$SelfIndexer?$:$InvAdvisor?$				1.351***
				(0.258)
Num. Obs.	12 385	12 385	12 385	12 385
R^2	0.10	0.16	0.19	0.11
FE: Issuer		X		
FE: Quarter-Style	X	X	X	X
FE: Quarter-Issuer			X	

* p < 0.1, ** p < 0.05, *** p < 0.01

Issuers which are model providers tend to recommend their self-indexed ETFs more often than their public-indexed ETFs. Note that for issuers that are not model providers the dependent variable equals zero for both self- and public-indexed ETFs. This effect is economically large, given that the average count of affiliated model portfolio membership is around 0.6 (the estimates for $SelfIndexer?$ are 0.788 and 1.171 in columns 2-3). In the last column, we introduce the $SelfIndexer?$ x $InvAdvisor?$ interaction to examine the behavior of investment advisors. The estimate for $InvAdvisor?$ is significantly positive – ETFs issued by investment advisors are part of more affiliated model portfolios, consistent with most model providers being classified as investment advisors. Importantly, $SelfIndexer?$ x $InvAdvisor?$ is significantly positive. That is, investment advisors tend to recommend their self-indexed ETFs more frequently than their public-indexed ETFs in model portfolios, indicating that model portfolio recommendations are indeed a channel through which issuer-advisors promote their self-indexed ETFs to investors.

2.4.4 Policy discussion

Unlike previous regulatory discussions, this paper highlights a self-preferencing mechanism among issuer-advisors, drawing attention to a previously overlooked agency concern within the indexing market. Earlier commentary focused on potential malpractices, such as manipulation of constituent pricing to inflate NAV and internal front-running within fund families.[29] While these concerns had been widely discussed, they are not the primary focus of our study, as our analysis finds limited evidence supporting their relevance in our sample.

First of all, the risk of price manipulation is limited given our focus on liquid U.S. equities. Second, we believe that the risk of front-running is arguably more acute for public-indexed ETFs. Their rebalancing and reconstitution dates are pre-announced, making them highly sus-

[29]See, for instance, https://www.etfstream.com/articles/self-indexing-etf-issuers-must-answer-conflict-of-interest-question. For more, see Invesco's SEC comment letter on Rule 6c-11.

ceptible to large-scale front-running by the broader market given the predictable nature of such trades and the vast aggregate capital tracking these indices (Chinco and Sammon, 2024; Li, 2021). In contrast, self-indexed funds are unlikely to generate large price impact in comparison, since their issuers can avoid these issues (e.g., by not pre-announcing or shifting rebalancing dates). However, for completeness, we empirically tested for possible internal front-running and found no supporting evidence (Table B.4). For each self-indexed ETF, we examine the trading activities of all other ETFs offered by the same issuer around the self-indexed funds' rebalancing dates. We find no trace of strategic trading that would suggest that other funds are front-running self-indexed ETFs (panel A). We also examine whether the presence of self-indexed ETFs in a fund family is associated with superior performance among the issuer's actively managed mutual funds. This would suggest that active funds are capitalizing on informational advantages. However, we find no such relationship (Panel B).

2.5 Conclusion

This paper examines self-indexing as a potential competitive mechanism and evaluates how this practice shapes the landscape of ETF investing. Contrary to predictions of fee competition and increased portfolio differentiation, self-indexed ETFs sponsored by issuers who are also investment advisors feature higher fees, exhibit more portfolio similarity, and deliver no performance gains compared to peers. Further analyses indicate a self-preferencing interpretation by these issuers. Self-indexed ETFs from other issuers, however, are associated with greater portfolio differentiation at comparable fees.

Chapter 3

Do Firms Self-Select Into Russell Indices?[†]

Abstract I investigate whether firms change their behaviour to increase the odds of joining stock indices, in the context of the yearly Russell 1000/2000 reconstitutions. I show that firms close to Russell 2000 inclusion do not manage real earnings, do not adjust earnings guidances or dividend payments, and do not time equity issuance and repurchase announcements in order to manipulate firm valuation. These findings hold even for firms with managers which would benefit most from index inclusion – those with high performance-pay, low inside ownership and short decision horizons. I provide novel evidence that Russell 1000/2000 Index assignment is random near the threshold.

[†]I thank Amir Amel-Zadeh, Bige Kahraman, Martin Schmalz and Mungo Wilson for their helpful comments. I acknowledge support from the Clarendon Fund, the Templeton Education and Charity Trust (TECT) and the Saïd Business School Foundation.

3.1 Introduction

Stock indices matter for public firms because they are strongly affected from being included in or excluded from an index. Mechanically, joining a heavily tracked index brings large increases in institutional ownership which leads to short-term price hikes (Shleifer, 1986), increased analyst coverage (Yu, 2008) and weaker governance (Heath et al., 2022). Further, joining a prestigious index grants access to cheaper capital and improved credit ratings (Brisker et al., 2013). Given these effects, firms and their managers might strongly value an index inclusion. Co-CEO of AppolloMed, a small-cap which in 2021 joined the S&P 600 small cap index, stated that "Our inclusion is an absolute honor that we expect will increase our visibility within the investment community and enhance shareholder value."[1] Anticipating these positive effects, it is natural to ask: do firms take actions specifically to join stock indices?

This question matters because the research design of much influential research on institutional ownership and governance assumes that firms do not take action to join stock indices (e.g., Appel et al. (2016); Crane et al. (2016); Heath et al. (2022)). These, and many other papers using index-inclusions as exogenous shock to institutional ownership, assume that index membership for size-based indices follows random assignment at the margin. In other words, it is assumed that firms who just manage to reach a market capitalisation threshold to join an index are comparable to firms that barely miss to reach the threshold. However, given the incentives for firms and their managers to join certain indices, it is ex-ante not obvious that the assignment is random. Rather, it seems plausible that some firms self-select into indices.

Typically, researchers look for signs of bunching to check whether assignment is random near a discontinuity threshold. This is not possible in the setting of index inclusions because a firm's distance to the threshold is usually determined relative to other firms in the sample.

[1] https://www.apollomed.net/news-media/press-releases/detail/169/apollo-medical-holdings-inc-to-join-sp-smallcap-600

To overcome this issue, this paper carefully examines actions firms could take to self-select into stock indices.

Each index provider separately determines the rules which dictate the criteria for joining and leaving its indices. For example, while firm inclusion in the S&P 500 is at the discretion of an appointed Index Committee (S&P Dow Jones Indices, 2023), the Russell 1000 and 2000 indices follow a mechanical rule to roughly include the largest 1000 and the following 2000 U.S. firms by market cap, determined at a pre-specified ranking date in May of each year (FTSE Russell, 2023).

These differences in inclusion criteria are largely what dictate the priors regarding corporate behaviour around index thresholds. For example, firms eligible for the S&P 500 index could benefit most from business ties to S&P Global companies (Li et al., 2026), news which might increase salience for the stock in the index committee, or they could try and signal quality as suggested by Hong et al. (2021). The levers are different for firms around the mechanically determined Russell 1000 Index cutoff. Those could manipulate equity valuations, which might suggest they manage real earnings, earnings guidance, or take other corporate actions to manipulate firm value.

There are few reasons why the Russell 1000/2000 index assignment is relatively more popular as an identification tool compared to other index inclusions. First, it comprises stocks listed in the U.S., which is by far the largest ETF market globally (Blackrock, 2023), suggesting that inclusion effects on institutional ownership are relatively strong. Second, as opposed to many other indices such as the S&P 500, Russell Index reconstitutions are purely mechanical and based on a pre-defined ranking date. This allows researchers to set up RDD-like experiments. Third, the Russell 1000 Index has a relatively high index turnover, increasing sample size for statistical inference. Since 2006, there have been on average 46 new firms that joined the Index each year, while only around 20 new firms join the S&P 500 per year.

At first, it may seem obvious that firms around the Russell 1000/2000 Index threshold

would prefer increasing equity valuations to attain or keep Russell 1000 membership. The Russell 1000 Index is more heavily tracked by asset managers and firms usually release press statements which celebrate the Russell 1000 inclusion as an achievement. For example, Bill.com, which joined the Russell 1000 in 2020 announced that "Inclusion in the Russell 1000 Index is an important milestone and will increase the overall awareness and exposure we are seeing within the investment community."[2] However, recent research suggests that managers of firms near the Russell 1000/2000 threshold could prefer Russell 2000 membership. Heath et al. (2022) find that firms that drop into the Russell 2000 pay higher total executive compensation and lower the pay-performance sensitivity for executives. Further, they invest less and perform worse, which hints towards managerial entrenchment. This effect is attributable to the increased ownership by passive index funds, which bring an overall reduction in stringent governance.[3]

To answer whether firms do indeed engage opportunistically to move into the Russell 2000 Index, I assign each Russell eligible firm its market capitalisation rank and use the Russell methodology to determine the proximity of the firm to the Russell 1000/2000 Index cutoff. I then analyse whether firms close to this cutoff engage opportunistically, and whether firms time those actions in line with the Russell reconstitution calendar. In particular, I analyse whether firms engage in real earnings management, earnings guidance and dividend payout management and whether firms time their equity issue and repurchase announcements. These tools of opportunistic action are readily available to managers, frequently employed in other settings and have been shown to affect firm valuation. It is therefore reasonable to assume they would also be used in this setting.

I first test whether on average, managers engage opportunistically to drop from the Russell 1000 into the Russell 2000 Index. The evidence is conclusive and suggests that managers do not take actions intended to change index membership. I then employ a battery of tests with the

[2] https://www.bill.com/press-release/billcom-set-join-russell-1000-index

[3] These claims are still an area of academic debate. See Heath et al. (2022); Gormley and Kim (2025) for discussions.

purpose of revealing cross-sectional differences among firms. Due to the potentially weaker governance regime for large Russell 2000 firms, managers with lower total compensation, high performance-pay and low inside ownership could be especially inclined to move from the Russell 1000 into the Russell 2000. I find no evidence for this claim. Alternatively, one might argue that managers believe that the value manipulations are not worth their associated costs. Since Russell Indices are reconstituted yearly, managers might worry that the benefits of joining the Russell 2000 will be short-lived. To account for this, I show that firms controlled by CEOs with short decision horizons are not more likely to engage in opportunistic behaviour for the sake of index inclusion. It is also plausible that managers believe that index inclusion is random at the margin. Indeed, 30% of firms close to the index threshold move more than 100 market capitalisation ranks each quarter. Managers of firms with volatile rankings could be deterred from taking costly action. I therefore test whether firms which are consistently close to the index threshold are more likely to engage in opportunistic behaviour, but find no conclusive evidence. I also find no evidence that firms engage opportunistically to move the opposite way, from the Russell 2000 into the Russell 1000 Index. Taking together, the results strongly suggest that Russell Index assignment is random at the threshold.

The findings of this study contribute to several strands of literature. First, they are closely related to studies which examine whether firms behave strategically to attain index membership. Hong et al. (2021), for example, claim that S&P 500 eligible firms strategically reduce leverage in order to signal stability to the S&P Index Committee, which could improve the odds of joining the index. Li et al. (2026) claim that amongst S&P 500 eligible firms, those that purchase S&P provided ratings have better chances of joining. Further, they argue that eligible firms purchase more S&P ratings when there are openings in the index. Chattopadhyay et al. (2020) find that Japanese firms who have a chance to join the Nikkei 400 Index cut discretionary spending so as to increase return on equity, a key variable that determines index membership. Most closely related to this study are a paper by Cho et al. (2017) and a study by Guo (2019). Cho et al. (2017)

find that firms that are just above the Russell 1000 Index threshold engage in negative earnings management via discretionary accruals so as to drop into the Russell 2000 Index. By accounting for more confounding variables, Guo (2019) find that the opposite is true: compared to the 100 firms just above the Russell 1000 Index threshold, firms just below the threshold have higher abnormal accruals and revenues, lower discretionary expenses and higher abnormal production costs. These findings suggest that managers prefer Russell 1000 over Russell 2000 membership, which goes against the logic of Heath et al. (2022). I improve on this research in several ways. First, I look at corporate behaviour that goes beyond earnings management. Firms might manipulate their market capitalisation using financing and payout policy and earnings guidance. Second, I analyse the timing of corporate behaviour. Given the costs associated with earnings and real activities management (e.g., Zang (2012)) managers might time their actions to periods just before the index ranking date. Third, I exploit the Russell reconstitution mechanism to employ an improved identification strategy.

My paper also adds to the literature on index inclusion effects, which has thoroughly analysed how index inclusion might effect, among other things, stock prices (Shleifer, 1986), information production (Boone and White, 2015), governance (Heath et al., 2022) and financing (Brisker et al., 2013). Most studies in this area use the Russell 1000/2000 Index reconstitution as exogenous shock to index membership and institutional ownership. For this empirical strategy to work, the shock must be exogenous such that treatment is truly random. Given the evidence on the effects of index inclusion, it is ex-ante unclear whether firms would in fact not self-select into an index. So far, no studies have thoroughly analysed this possibility. For example, a contribution by Heath et al. (2022) assumes that "firms cannot control small variations in ranking, [making] index assignment near the threshold (...) as good as random." My paper is the first to provide robust evidence that the assignment is indeed likely random, which validates a key assumption in the identification of said studies.

The study is also related to the literature on entrenchment and the "quiet life" of managers

which broadly asks whether managers in weaker governance regimes have a stronger tendency towards pursuing personal goals instead of those of their firm's shareholders (Shleifer, 1986; Bertrand and Mullainathan, 2003). Assuming that managers have tools to nudge their firm's valuation in the short-term, the Russell 1000/2000 Index threshold provides a setting which might reveal managerial preferences for specific governance regimes. Admittedly, the setting is not perfect. First, managers cannot freely choose index membership – they can merely change the odds. Second, the fact that the index is reconstituted on a yearly basis likely decreases the value of changing the governance regime for managers. With these caveats in mind, I find that managers either have no preference for one governance regime over the other, or do not act on their preference, even if they might have particularly strong incentives (based on compensation, inside ownership and share of performance-pay).

3.2 Institutional background: Russell 1000 Index reconstitution

To understand the empirical nuances of this study, it is essential to discuss the Russell 1000 Index reconstitution mechanism. Unless otherwise stated, the following information stems from the official FTSE Russell methodology (FTSE Russell, 2023).

Russell constructs all of its indices from its universe of U.S. companies, which as of 2023 covers 99% of the U.S. equity market. Some of the key criteria for inclusion in the universe are that the securities must be U.S. based,[4] traded on eligible U.S. exchanges,[5] have a closing price of at least $1 and a total market capitalisation of $30 million. Further, at least 5% of shares must be available in the marketplace, companies cannot produce unrelated business taxable income,[6] or be of a certain company structure including ETFs, mutual funds, limited liability or special-purpose acquisition companies (SPACs). REITs are generally eligible.

[4] ADRs are not included in the Russell stock universe.

[5] As of 2023, those are the CBOE, NYSE, NYSE American, NASDAQ and ARCA.

[6] Companies with unrelated business taxable income have restricted ownership from U.S. tax-exempt investors. Russell excludes those to ensure its indices are as investable as possible.

The most famous Russell Indices are the Russell 1000, 2000 and 3000 Indices. The Russell 1000 roughly tracks the largest 1000 U.S. firms and the Russell 3000 tracks the largest 3000 U.S. firms. The Russell 2000 Index tracks Russell 3000 members that are not in the Russell 1000 Index. From its inception in 1984, the Russell 1000 Index was reconstituted once per year (in June) to include the 1000 largest U.S. firms in the Russell stock universe by end-of-may market capitalisation (rank day).[7] In 2007, Russell announced the introduction of its 'banding' policy. Essentially, this new rule gives firms inside the Russell 1000 a market cap buffer before they lose index membership. The goal of this was to reduce unnecessary index turnover. Under the banding policy, Russell 1000 membership is determined using the following procedure on rank day:

1. Sort all Russell Index eligible securities by descending market capitalisation.
2. Calculate the cumulative market cap percentile of each firm.
3. Calculate a 5 percentile range around the cumulative market cap percentile of the 1000th firm.
4. Russell 1000 members that have a cumulative market cap percentile above the range are replaced by Russell 2000 firms with cumulative market percentile below the range.

Besides the yearly reconstitution, firms can only enter or exit the Russell 1000 Index via IPO or delistings, which are assessed once per quarter. Table 3.1 is taken from the Russell Index methodology handbook and provides a clear example of the yearly reconstitution mechanism.

The result of this banding procedure is that since 2007, Russell 1000 members must become much smaller in size than the 1000th ranked firm to drop into the Russell 2000 Index. Figure 3.1 shows which size rank a Russell 1000 member would have to reach in order to drop into the the Russell 2000 Index (exclusion rank), if it was reconstituted on a quarterly basis. After

[7] In 2015, the rank day was moved forward to a date in the middle of May.

Table 3.1. Fictional example of the Russell 1000 Index reconstitution methodology since 2007. Firms in-between the horizontal lines on rank day are in the banding range of ($89.99 \pm 2.5\%$). In this example, ABC company replaces RYT Inc. in the Russell 1000 Index. Further, Foods Inc. drops below the 1000th rank but since its cumulative percentile is not outside the banding range, the firm remains in the index.

Rank	Company	Current index	Market cap	Cumulative market cap	Cumulative percentile	New index
995	XYZ Company	Russell 1000	$2,115	$154,000	84.38%	Russell 1000
996	ABC Company	Russell 2000	$2,105	$156,105	85.54%	Russell 1000
997	Drugstore Inc.	Russell 1000	$2,100	$158,205	86.69%	Russell 1000
998	PYK Shipping	Russell 2000	$2,011	$160,216	87.79%	Russell 2000
999	Z Technology	Russell 2000	$2,010	$162,226	88.89%	Russell 2000
1,000	RE Trust	Russell 2000	$2,000	$164,226	89.99%	Russell 2000
1,001	Foods Inc.	Russell 1000	$1,995	$166,221	91.08%	Russell 1000
1,002	PETs & More	Russell 2000	$1,950	$168,171	92.15%	Russell 2000
1,003	RYT Inc.	Russell 1000	$1,923	$170,094	93.20%	Russell 2000

the banding policy was introduced, the exclusion rank was approximately 1250. The upward trend of this rank since 2007 mechanically follows from the increasing concentration of market cap amongst the largest U.S. companies in the past decade.

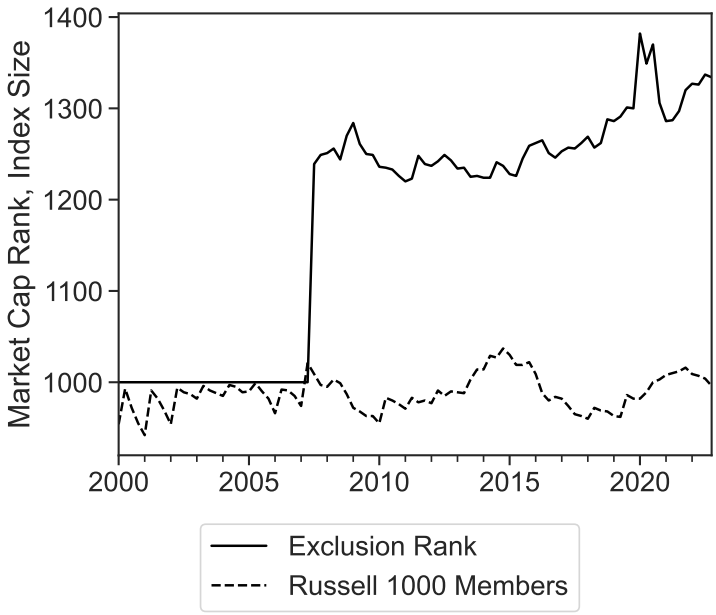


Figure 3.1. Russell 2000 exclusion rank and index size.
This figure plots the market capitalisation rank needed for a Russell 1000 company to enter the Russell 2000 Index if index reconstitution occurred quarterly, and the Russell 1000 index member count. The rank is calculated following the Russell methodology using quarterly CRSP and Russell data.

3.3 Data and variable construction

I obtain firm-level balance sheet, and P&L data from Compustat. Stock data, including merger events, are from the CRSP database and institutional ownership is from the Thomson S34 file, which covers 13F holdings. Firm issue and repurchase announcements are from the SDC Platinum database. Analyst estimates and management guidance data are from the I/B/E/S databases. As is common practise, if a firm has no I/B/E/S coverage it is assumed to have no analyst coverage. Manager compensation is extracted from Execucomp. Finally, index membership data is gathered from FTSE Russell. All raw continuous variables are winsorized at 0.5% and 99.5%.

In line with Appel et al. (2024), the time frame of this study is from 2007 until 2022. 2007 marks the start of Russell's banding policy and roughly coincides with the start of a rapid

growth phase of index investing.

Since the market cap size rank of U.S. firms determines index membership, it is one of the key variables in this study. In theory, it should be straightforward to calculate using firm market capitalisation. Even though Russell publicly provides an exact methodology for market cap calculation, their market cap unfortunately differs from the market cap researchers usually compute using the CRSP and Compustat databases. As Ben-David et al. (2019) show, the differences stem from conflicting measures of common shares outstanding, and disagreements in issue- and company-level aggregation of shares between the databases and Russell. Therefore, researchers generally fail to perfectly predict Russell 1000 Index inclusion using rank day variables. This matters especially when using Russell 1000 Index inclusion as exogenous shock in an identification strategy. Because I use end-of-quarter data at dates where index membership is already determined, this concern does not apply.

However, since CRSP or Compustat do not provide perfect rankings, one should consider other options. One potential method to obtain correct Russell ranking data could be to use end-of-quarter Russell Index weights as determinants of size ranks. However, Russell Index *weights* are based on float-adjusted market cap, while Russell Index *membership* is based on total market cap, including shares outstanding that are unavailable to the public. This means that to extrapolate the official market cap ranking which determines index membership, I would require Russell's proprietary measure of free-float and market cap, only the latter of which is available, only from 2020 onwards. I therefore opt to use the CRSP database and aggregate end-of-quarter market cap at the firm-level using the $permco$ identifier. Admittedly, my market cap rank slightly deviates from the proprietary ranking calculated by Russell. However, slight differences in ranks are not crucial in the design of this study. I use the ranking variable to determine a proxy for the feasibility of managerial action to join an index. Thus, the ranking variable should also be readily available to managers. It seems unsound to expect firm managers apply the same scrutiny to Russell rankings as researchers do.

On an end-of-quarter basis, I then define the $Distance$ of a Russell 1000 firm to exclusion as the exclusion rank minus the firm's market cap rank. The $Distance$ of a Russell 2000 firm to inclusion is its market cap rank minus the inclusion rank. The exclusion (inclusion) rank is determined using the upper (lower) bound of Russell's banding range at the end of a quarter. Figure 3.2 plots a yearly snapshot of the distance for each firm in the Russell 3000, split into Russell 1000 and 2000 members. This single plot captures the most important institutional details relevant to the research question. First, it is apparent that the minimum market cap for inclusion has risen over the years. Second, the fact that index reconstitution occurs only yearly allows some Russell 1000 members to temporarily be below the index threshold but retain index membership, and vice versa for Russell 2000 firms being above the threshold. Third, the sudden increase of distance in 2007 that is very apparent in the left hand panel is the result of the banding policy. After 2007, all firms in the Russell 1000 have gained additional distance to the index cutoff, hence the sudden shift. This detail also explains why the market cap of a firm with distance 0 is not the same for Russell 1000 and 2000 members after 2007.

Earnings management is not straightforward to detect and define. I therefore adopt well established methodologies from the accounting literature to define the different levers of earnings management. To estimate discretionary accruals, I use the modified Jones (1991) model. Accruals are defined as net income minus net cashflow from operations.[8] Then, I follow Bartov et al. (2000). First, for each quarter and all firms in a sector, regression equation 3.1 is estimated using OLS. The firm subscript i is omitted for simplicity. For all of the following estimations, I require at least eight observations and use the first two digits of the SIC code as industry classification.

$$\frac{TA_t}{A_{t-1}} = \alpha_1 \cdot \frac{1}{A_{t-1}} + \alpha_2 \cdot \frac{\Delta REV_t}{A_{t-1}} + \alpha_3 \cdot \frac{PPE_t}{A_{t-1}} + \epsilon_t \tag{3.1}$$

[8]For robustness, in unreported tests, I also use the balance sheet estimation of discretionary accruals. The results are qualitatively the same.

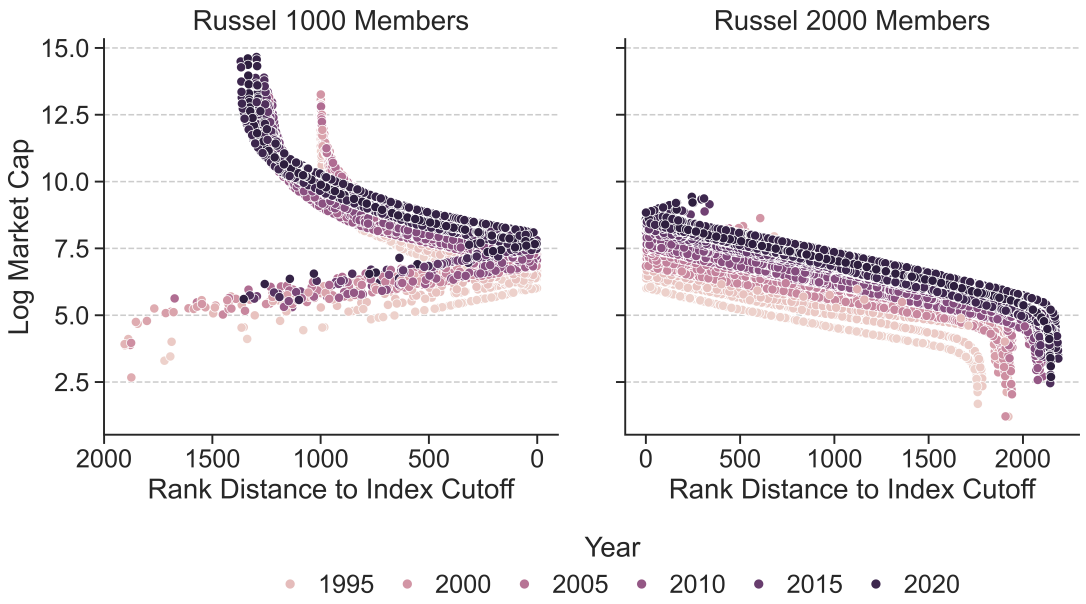


Figure 3.2. Distance to index switch for Russell 3000 members.
This figure plots the rank distance of Russell 3000 firms to Russell 2000 in- and exclusion in the first calendar year quarter for all years between 1992 and 2022. The left panel shows the rank distance of Russell 1000 members to the exclusion rank. The right panel shows the rank distance of Russell 2000 members to the inclusion rank.

TA_t are total accruals, A_t firm assets, REV_t revenue and PPE_t gross property plant and equipment. Then, I use the estimated parameters to estimate firm-level nondiscretionary accruals scaled by assets (NDA_t) for each firm quarter as in equation 3.2, where REC_t are net receivables. Nondiscretionary accruals are then simply total accruals minus discretionary accruals.

$$NDA_t = \hat{\alpha_1} \cdot \frac{1}{A_{t-1}} + \hat{\alpha_2} \cdot \frac{\Delta REV_t - \Delta REC_t}{A_{t-1}} + \hat{\alpha_3} \cdot \frac{PPE_t}{A_{t-1}} \tag{3.2}$$

For quarterly unexpected cashflows from operations and discretionary expenses, I follow Roychowdhury (2006). Using the same requirements as for discretionary accruals, I estimate regression equations 3.3 and 3.4, the residuals of which are the scaled firm measures for unexpected cashflows from operations and unexpected discretionary expenses. CFO are the

cashflows from operations and $DISEXP$ is the sum of research and development (R&D) and selling, general and administrative expenditure (SG&A). Other variables are defined in the appendix.

$$\frac{CFO_t}{A_{t-1}} = \alpha_0 + \alpha_1 \cdot \frac{1}{A_{t-1}} + \alpha_2 \cdot \frac{REV_t}{A_{t-1}} + \alpha_3 \cdot \frac{\Delta REV_t}{A_{t-1}} + \epsilon_t \tag{3.3}$$

$$\frac{DISEXP_t}{A_{t-1}} = \alpha_0 + \alpha_1 \cdot \frac{1}{A_{t-1}} + \alpha_2 \cdot \frac{\Delta REV_t}{A_{t-1}} + \epsilon_t \tag{3.4}$$

For all real earnings management variables I remove heavy outliers that lie beyond the 1st and 99th percentile.

In line with Chen et al. (2023), I define the managerial forecast error as the actual earnings per share minus the managerial guidance, in percentage terms of the stock price. For simplicity, I only include quarterly earnings per share guidances. I then remove pre-announcements. Those are guidances issued for a quarter which has already ended but for which earnings have not been announced. I then restrict the sample to guidances which are in the same quarter as the guided quarter. Amongst the remaining guidances I select only the last guidance per quarter. Finally, I remove data points outside of the 1st and 99th percentile to omit significant outliers.

Summary statistics of key variables are provided in Table 3.2. Data for quarterly real earnings management is relatively scarce for the sample. The main reasons for this lie in the requirements for the estimations. First, they require quarterly cashflow data, which is not available for all firms. Second, for each industry-quarter data must be available for at least eight firms in the sample. Coverage for managerial earnings guidance is scarce primarily because of the limitations in I/B/E/S.

Table 3.2. Summary statistics.
This table presents the firm-level summary statistics for all members of the Russell 3000 Index between 2007 and 2022. The variables are measured at quarterly frequency.

Variable	N	Mean	Std. Dev.	25th Perc.	Median	75th Perc.
DiscrAccr	64,432	-0.002	0.037	-0.016	-0.002	0.012
UnexpDiscrExp	57,739	0.023	0.071	-0.013	0.005	0.048
UnexpCFO	159,449	-0.002	0.041	-0.012	0.002	0.015
IssueAnn?	187,067	0.030	0.169	0.000	0.000	0.000
RepAnn?	187,067	0.036	0.187	0.000	0.000	0.000
Dividend?	183,996	0.501	0.500	0.000	1.000	1.000
MFE	19,468	0.096	0.814	0.008	0.084	0.226
DaysUntilAnn	20,449	88	17	84	91	95
MCap (USD Mn.)	187,067	817,2	41,058	431	1,275	4,132
InstOwn	168,443	0.696	0.285	0.551	0.775	0.908
NumAnalysts	187,067	5	5	2	4	7

3.4 Managerial actions for index inclusion

It is ex ante unclear how close a firm must be to the index inclusion threshold for the firm manager to believe that inclusion is realistic if action was taken. It is also not obvious whether the relationship between distance to inclusion and managerial incentive is linear. Therefore, the following tests use rank distance bins to classify firms based on their distance to the index threshold. I deliberately limit the sample of firms for which I expect corporate action to Russell 1000 firms that have a distance to the index threshold of less than 150. I find that 70% of firms close to the index cutoff move fewer than 100 ranks per quarter and only 15% of firms move more than 150 ranks per quarter. I therefore argue that it is unlikely for a manager of a firm with a distance of more than 150 to believe that their actions will significantly improve the chances of index inclusion. I limit the overall sample of Russell 1000 firms in the study to those less than 300 ranks above the exclusion threshold. This ensures that the firms are not too dissimilar and that I do not capture managerial actions intended for joining other indices. In particular, this excludes most firms close to the S&P 500 index. The results are qualitatively similar for other

sample sizes and bin-specifications.

3.4.1 Earnings management

If managers act on their incentives to join the Russell 2000 Index, it is reasonable to expect that they engage in earnings management (Roychowdhury, 2006; Bergstresser and Philippon, 2006). In particular, earnings management should be most prevalent in the first quarter of the calendar year. The first reason for this is that amongst all quarter-ends, the distance to the index threshold is the most accurate predictor of inclusion chance at the end of Q1, since the Russell rank day is in May. The second reason is that earnings management is costly (Zang, 2012) and effects on firm valuation are potentially short-lived. Since Q1 results are usually reported in April, managers might hope that the effects of earnings surprises distort valuations until the rank day in May. I therefore conduct the tests for both all quarters and quarter one only. I estimate regression equation 3.5, where $Y_{i,t}$ is an earnings management measure for firm i in quarter t.

$$Y_{i,t} = \alpha + \sum_{Q} \beta_q \cdot RankBinDummy_{q,i,t} + \sum_{K} \gamma_k \cdot Ctrl_{k,i,t} + \epsilon_{i,t} \tag{3.5}$$

Accurately measuring the information set of managers is paramount for this study, which is why the timing of variables should be carefully examined. It is assumed that discretionary accruals can be adjusted after the calendar quarter ends such that managers can adjust accruals between the quarter-end date and the earnings announcement date. This implies that managers are aware of end-of-quarter t market capitalisation rankings. For changes in cashflow from operations and discretionary expenses, the same assumption does not hold. Hence, I lag all independent variables by one quarter.

Tables 3.3, 3.4, and 3.5 report the estimated regression coefficients and standard errors for the dummy variables which capture a firm's distance to inclusion, and key control variables.

Besides controlling for size, tobin's Q and return on assets, I include the number of analysts and institutional ownership percentage to capture monitoring functions that these parties provide (Yu, 2008; Chung et al., 2002). I also add a REIT dummy and year, fiscal calendar, industry and quarter fixed effects.

Table 3.3. Predicting unexpected discretionary expenses.
The dependent variable is unexpected discretionary expenses as calculated in Roychowdhury (2006). Columns (1) to (3) use the full sample while columns (4) to (6) include each calendar year's first quarter only. Included firms are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. Timeframe: 2007 to 2022. Standard errors are clustered at the firm-level and shown in brackets.

	All Q			Q1		
	(1)	(2)	(3)	(4)	(5)	(6)
$0 \leq Distance_{t-1} \leq 50$	0.0078	0.0024	0.0021	0.0209**	−0.0011	−0.0051
	(0.0054)	(0.0047)	(0.0050)	(0.0105)	(0.0165)	(0.0156)
$51 \leq Distance_{t-1} \leq 100$	0.0044	0.0002	−0.0005	0.0077	−0.0073	−0.0081
	(0.0042)	(0.0041)	(0.0040)	(0.0066)	(0.0126)	(0.0125)
$101 \leq Distance_{t-1} \leq 150$	0.0024	−0.0007	−0.0019	0.0109	0.0003	0.0000
	(0.0033)	(0.0035)	(0.0033)	(0.0066)	(0.0102)	(0.0096)
$NumAnalysts_{t-1}$		0.0014***	0.0019***		0.0015**	0.0021***
		(0.0005)	(0.0005)		(0.0007)	(0.0007)
$TobinQ_{t-1}$		0.0129***	0.0134***		0.0133***	0.0133***
		(0.0026)	(0.0026)		(0.0025)	(0.0026)
ROA_{t-1}		−0.9000***	−0.9541***		−0.9492***	−1.0038***
		(0.0982)	(0.1004)		(0.1042)	(0.1137)
$LogMcap_{t-1}$		−0.0221**	−0.0181**		−0.0621	−0.0598
		(0.0091)	(0.0088)		(0.0447)	(0.0423)
$InstOwn_{t-1}$		0.0188*	0.0230**		0.0275	0.0299*
		(0.0101)	(0.0101)		(0.0168)	(0.0167)
Num. Obs.	1667	1356	1356	379	305	305
R^2	0.05	0.42	0.48	0.09	0.42	0.49
FE: Year	X	X	X	X	X	X
FE: Fiscal Calendar	X	X	X	X	X	X
FE: Industry			X			X
FE: Quarter	X	X	X			
FE: REIT Flag	X	X	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

I conclude that managers closer to index inclusion do not engage in different earnings management than managers of firms that are farther away from inclusion. This holds for the full

Table 3.4. Predicting unexpected cashflow from operations.
The dependent variable is a firm's unexpected cash flow from operations as calculated in Roychowdhury (2006). Columns (1) to (3) use the full sample while columns (4) to (6) include each calendar year's first quarter only. Included firms are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. Timeframe: 2007 to 2022. Standard errors are clustered at the firm-level and shown in brackets.

	All Q			Q1		
	(1)	(2)	(3)	(4)	(5)	(6)
$0 \leq Distance_{t-1} \leq 50$	−0.0021	−0.0001	−0.0003	−0.0051	−0.0049	−0.0021
	(0.0017)	(0.0023)	(0.0023)	(0.0038)	(0.0065)	(0.0064)
$51 \leq Distance_{t-1} \leq 100$	−0.0018	−0.0006	−0.0005	−0.0038*	−0.0033	−0.0030
	(0.0013)	(0.0017)	(0.0016)	(0.0022)	(0.0045)	(0.0043)
$101 \leq Distance_{t-1} \leq 150$	−0.0012	−0.0001	−0.0003	−0.0020	−0.0023	−0.0012
	(0.0012)	(0.0014)	(0.0014)	(0.0022)	(0.0035)	(0.0033)
$NumAnalysts_{t-1}$		0.0003**	0.0001		0.0002	−0.0001
		(0.0001)	(0.0002)		(0.0002)	(0.0002)
$TobinQ_{t-1}$		−0.0014	−0.0004		0.0000	0.0019
		(0.0014)	(0.0011)		(0.0021)	(0.0018)
ROA_{t-1}		0.1948***	0.2137***		0.1305**	0.1572***
		(0.0424)	(0.0383)		(0.0555)	(0.0500)
$LogMcap_{t-1}$		0.0033	0.0019		−0.0007	0.0051
		(0.0040)	(0.0039)		(0.0153)	(0.0146)
$InstOwn_{t-1}$		0.0015	0.0021		0.0049	0.0082
		(0.0029)	(0.0027)		(0.0054)	(0.0050)
Num. Obs.	4960	4179	4179	1241	1059	1059
R^2	0.02	0.08	0.15	0.04	0.08	0.24
FE: Year	X	X	X	X	X	X
FE: Fiscal Calendar	X	X	X	X	X	X
FE: Industry			X			X
FE: Quarter	X	X	X			
FE: REIT Flag	X	X	X	X	X	X
Std. Errors	by: Firm	by: Firm	by: Firm	by: Firm	by: Firm	by: Firm

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 3.5. Predicting discretionary accruals.
The dependent variable is a firm's calculated discretionary accruals based on the modified version of the Jones model based on Dechow et al. (1995). Columns (1) to (3) use the full sample while columns (4) to (6) include each calendar year's first quarter only. Included firms are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. Timeframe: 2007 to 2022. Standard errors are clustered at the firm-level and shown in brackets.

	All Q			Q1		
	(1)	(2)	(3)	(4)	(5)	(6)
$0 \leq Distance_t \leq 50$	−0.0001	0.0002	0.0021	0.0034	0.0088	0.0106
	(0.0030)	(0.0034)	(0.0031)	(0.0040)	(0.0098)	(0.0099)
$51 \leq Distance_t \leq 100$	−0.0012	−0.0014	−0.0015	−0.0004	0.0034	0.0044
	(0.0021)	(0.0025)	(0.0024)	(0.0038)	(0.0078)	(0.0080)
$101 \leq Distance_t \leq 150$	0.0012	0.0017	0.0019	0.0000	0.0031	0.0021
	(0.0021)	(0.0023)	(0.0023)	(0.0039)	(0.0062)	(0.0060)
$NumAnalyst_t$		−0.0003*	−0.0003		−0.0001	0.0001
		(0.0001)	(0.0002)		(0.0005)	(0.0005)
$TobinQ_t$		−0.0004	−0.0001		0.0017**	0.0017*
		(0.0005)	(0.0007)		(0.0008)	(0.0010)
ROA_t		0.0051	0.0063		0.0268	0.0165
		(0.0201)	(0.0202)		(0.0379)	(0.0438)
$LogMcap_t$		0.0010	0.0025		0.0168	0.0210
		(0.0050)	(0.0050)		(0.0272)	(0.0280)
$InstOwn_t$		−0.0034	−0.0042		−0.0058	−0.0052
		(0.0038)	(0.0038)		(0.0082)	(0.0083)
Num. Obs.	1965	1849	1849	469	442	442
R^2	0.02	0.03	0.10	0.07	0.08	0.17
FE: Year	X	X	X	X	X	X
FE: Fiscal Calendar	X	X	X	X	X	X
FE: Industry			X			X
FE: Quarter	X	X	X			
FE: REIT Flag	X	X	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

sample as well as the Q1 subsample.

3.4.2 Earnings guidance

Besides managing earnings, managers might also try to manage investor expectations in such a way that increases the odds of index inclusion. The most common form of earnings guidance management is to understate earnings forecasts such that on the earnings announcement date, the firm experiences a positive shock to its valuation. This practise is called 'lowballing' and evidence suggests that lowballing does indeed have positive short term effects on valuations so long as firms are not known for being serial-lowballers (Bartov et al., 2002; Chen et al., 2023). In the context of this study, Russell 1000 managers might try to explicitly overstate Q1 earnings forecasts to entice a negative valuation shock just before Russell rank day in May (highballing). At the same time, managers could understate their earnings expectations for firm-quarters which have earnings announcements after the Russell rank day (lowballing).

To test these predictions, I use the managerial forecast error (MFE) which is defined by the actual earnings per share minus the latest managerial guidance, in percentage terms of the stock price. A positive MFE indicates lowballing. I split firm-quarter observations in two subsets. The 'resolved' subset includes firm-quarters for which the actual earnings per share is revealed at a date prior to the Russel rank day. The 'unresolved' subset includes observations where the earnings announcement occurs after the rank day. If firms manage their earnings forecasts to maximise their odds of Russell 2000 Index inclusion, one should find a negative effect of the proximity to the index threshold on MFE in the resolved (highballing) and a positive effect in the unresolved sample (lowballing), after controlling for key variables. In addition to the controls previously used, I add a variable that captures the difference in days between the earnings guidance and the actual announcement. That is because previous research has shown that managers frequently 'walk down' earnings guidances as the announcement day approaches (Elton et al., 1984). I also add the quarterly stock price volatility as a measure for

uncertainty of firm valuation and lag the independent variables such that they best represent the manager's information set when announcing the earnings guidance (equation 3.6).

$$MFE_{t,i} = \alpha + \sum_{Q} \beta_q \cdot RankBinDummy_{q,i,t-1} + \sum_{K} \gamma_k \cdot Ctrl_{k,i,t-1} + \epsilon_{i,t} \tag{3.6}$$

From the results in Table 3.6 I infer that managers do not manage their earnings guidance so as to join the Russell 1000.

3.4.3 Issue and repurchase announcements, and dividend payout

Managers might time stock issue and repurchase announcements, or refrain from paying out dividends to improve the odds of joining the Russell 2000 Index. Basic corporate finance theory suggests that due to information asymmetry, stock issue announcements are followed by negative price reactions (Myers and Majluf, 1984). By symmetry, stock repurchase announcements should be followed by positive returns. Much empirical literature supports these claims (e.g., Dann (1981); Korajczyk et al. (1989)). Since open-market repurchase announcements are not binding, and a significant number of firms do not follow through with their repurchase intent (Ikenberry and Vermaelen, 1996), those types of announcements seem especially suited for opportunistic managers. When it comes to dividends, research has shown that omissions lead to strong negative market reactions (Healy and Palepu, 1988). Hence, managers attempting to reduce market valuations could be incentivised to refrain from paying dividends. To test these predictions, I run logit regressions that predict dummies for whether a firm i has announced equity issuance or repurchase, or paid dividends in a given quarter t. In addition to previous control variables, I further add short-term debt, the one year stock return and cash. I exclude financial services firms and utility companies (equation 3.7).

$$Y_{t,i} = \alpha + \sum_{Q} \beta_q \cdot RankBinDummy_{q,i,t-1} + \sum_{K} \gamma_k \cdot Ctrl_{k,i,t-1} + \epsilon_{i,t} \tag{3.7}$$

Table 3.6. Predicting managerial forecast error (MFE).
The dependent variable is the managerial forecast error, defined as the quarterly actual eps minus the managerial eps guidance, in percentage terms of the stock price. In columns (1) to (3) the model predicts resolved forecasts, defined as forecasts for which the guidance and the actual earnings announcement occur before the Russel rank day. In columns (4) to (6) the models predict unresolved forecasts, defined as forecasts for which the guidance is before the rank day and the earnings announcement occurs after the rank day. Included firms are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. Timeframe: 2007 to 2022. Standard errors are clustered at the firm-level and shown in brackets.

	Resolved			Unresolved		
	(1)	(2)	(3)	(4)	(5)	(6)
$0 \leq Distance_{t-1} \leq 50$	0.2067	0.0542	−0.4094	−0.1366	−0.6247**	−0.2664
	(0.2105)	(0.4474)	(0.5270)	(0.1354)	(0.2803)	(0.2598)
$51 \leq Distance_{t-1} \leq 100$	0.0257	0.0281	−0.2923	0.0462	−0.3595*	−0.0055
	(0.0828)	(0.2510)	(0.3656)	(0.0771)	(0.2101)	(0.3249)
$101 \leq Distance_{t-1} \leq 150$	0.0641	−0.0533	−0.3292	−0.1734	−0.5029	−0.3754
	(0.1177)	(0.2378)	(0.3023)	(0.1882)	(0.3107)	(0.2400)
$DaysUntilAnn_t$		0.0070	0.0106		−0.0020	−0.0036**
		(0.0071)	(0.0091)		(0.0012)	(0.0018)
$NumAnalysts_{t-1}$		0.0091	0.0218*		0.0151	0.0061
		(0.0090)	(0.0115)		(0.0123)	(0.0090)
$TobinQ_{t-1}$		−0.0639	−0.0633		−0.0860*	−0.1137*
		(0.0431)	(0.0744)		(0.0472)	(0.0576)
$Leverage_{t-1}$		−0.0439	0.6443		0.1260	0.2937
		(0.1620)	(0.5097)		(0.1805)	(0.3195)
ROA_{t-1}		5.8030	6.3316		4.5754	8.9292
		(4.6032)	(4.6095)		(3.6763)	(5.6362)
$LogMcap_{t-1}$		−0.5788	−1.5488		−1.6092**	−1.0172
		(1.1928)	(1.4123)		(0.7137)	(0.7154)
$InstOwn_{t-1}$		0.1012	−0.0208		−0.3401	−0.3920
		(0.2834)	(0.3734)		(0.3490)	(0.4107)
$VolatilityQ_{t-1}$		0.5150	2.1313		−5.3702	−7.0082
		(6.0468)	(7.4202)		(8.3767)	(7.4276)
Num. Obs.	217	175	172	176	133	133
R^2	0.13	0.14	0.48	0.13	0.27	0.54
FE: Year	X	X	X	X	X	X
FE: Fiscal Calendar	X	X	X	X	X	X
FE: Industry			X			X
FE: Quarter		X	X		X	X
FE: REIT Flag	X	X	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 3.7. Predicting issuance announcements (logit).
The dependent variable is a dummy equal to one if the firm announced a stock issuance in quarter t. Columns (1) to (3) use the full sample while columns (4) to (6) include each calendar year's first quarter only. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. Financial services and utility companies are excluded. Timeframe: 2007 to 2022. Standard errors are clustered at the firm-level and shown in brackets.

	All Q			Q2		
	(1)	(2)	(3)	(4)	(5)	(6)
$0 \leq Distance_{t-1} \leq 50$	0.1137	0.0025	0.0642	0.7648	2.7377	7.3147
	(0.3953)	(0.6590)	(0.7315)	(0.7488)	(2.2837)	(8.6258)
$51 \leq Distance_{t-1} \leq 100$	−0.8004*	−0.7915	−0.5518	−0.0369	0.0885	2.3037
	(0.4786)	(0.6216)	(0.6124)	(0.7852)	(2.4937)	(6.3866)
$101 \leq Distance_{t-1} \leq 150$	−0.4301	−0.3555	−0.3674	−0.3252	0.8253	2.6870
	(0.3986)	(0.5769)	(0.6141)	(0.7802)	(1.3128)	(4.9637)
$TobinQ_{t-1}$		0.1450	0.1752		0.3422	0.3300
		(0.1374)	(0.1581)		(0.3726)	(2.0339)
$LogAssets_{t-1}$		−0.1553	−0.2203		2.3583	2.1313
		(0.6831)	(0.8496)		(1.6693)	(8.9218)
$Leverage_{t-1}$		0.1638	0.7235		−6.7584**	−11.5624
		(1.0065)	(0.9974)		(2.8173)	(13.8211)
$Cash_{t-1}$		0.4774	0.2621		0.9864	−3.0603
		(1.2662)	(1.1868)		(2.4458)	(3.1077)
$ShortDebt_{t-1}$		3.8247	4.1218		13.7612**	21.8387
		(2.7487)	(2.6405)		(6.4804)	(15.0687)
ROA_{t-1}		−3.0846	−2.3825		−14.6035**	−34.6002
		(2.3109)	(2.9843)		(5.8665)	(27.4625)
$LogMcap_{t-1}$		−1.0247	−0.7522		4.3146	12.3487
		(1.4500)	(1.5723)		(7.9950)	(19.0572)
$InstOwn_{t-1}$		1.3044	1.9634*		5.2299	14.2052
		(0.8408)	(1.1283)		(4.5862)	(16.8944)
$VolatilityQ_{t-1}$		26.4479***	22.9418**		6.5686	45.1037
		(9.9477)	(11.1434)		(22.3079)	(38.5397)
$ReturnY_{t-1}$		2.0927***	1.9268***		4.4011*	7.1742
		(0.6751)	(0.7086)		(2.5625)	(6.8159)
Num. Obs.	3325	2032	1264	478	225	127
R^2	0.08	0.12	0.17	0.06	0.25	0.49
FE: Year	X	X	X	X	X	X
FE: Fiscal Calendar	X	X	X	X	X	X
FE: Industry			X			X
FE: Quarter	X	X	X			
FE: REIT Flag	X	X	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

The results provide no evidence that firms close to Russell 1000 exclusion adjust their is-

Table 3.8. Predicting repurchase announcements (logit).
The dependent variable is a dummy equal to one if the firm announced a stock repurchase programme in quarter t. Columns (1) to (3) use the full sample while columns (4) to (6) include each calendar year's first quarter only. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. Included firms are members of the Russell 1000 that are less than 300 ranks away from exclusion. Financial services and utility companies are excluded. Timeframe: 2007 to 2022. Standard errors are clustered at the firm-level and shown in brackets.

	All Q			Q2		
	(1)	(2)	(3)	(4)	(5)	(6)
$0 \leq Distance_{t-1} \leq 50$	−0.1484	0.5388	0.5315	−1.0928	−1.3532	−0.4186
	(0.2911)	(0.4621)	(0.4758)	(1.0748)	(1.5398)	(2.1659)
$51 \leq Distance_{t-1} \leq 100$	−0.3011	−0.0039	0.0000	−0.7906	−0.4857	0.4481
	(0.2568)	(0.4104)	(0.4149)	(0.7750)	(1.1699)	(1.5374)
$101 \leq Distance_{t-1} \leq 150$	−0.0749	0.2758	0.3242	−0.0204	0.0808	1.2082
	(0.2096)	(0.2938)	(0.2996)	(0.5226)	(0.7878)	(0.9794)
$TobinQ_{t-1}$		−0.3864	−0.5036		−0.7391	−1.4339***
		(0.2633)	(0.3381)		(0.4872)	(0.5431)
$LogAssets_{t-1}$		−0.5160	−0.9247		−1.3237	−1.5842
		(0.6130)	(0.7820)		(1.3677)	(1.6161)
$Leverage_{t-1}$		−0.3862	−0.5890		0.6808	0.9413
		(0.6447)	(0.7007)		(1.2278)	(1.7680)
$Cash_{t-1}$		1.1828	1.8072		0.7996	5.4828**
		(0.9539)	(1.2579)		(2.1606)	(2.6727)
$ShortDebt_{t-1}$		2.8154*	3.5782**		−0.9636	2.4871
		(1.4435)	(1.4681)		(3.6350)	(4.7342)
ROA_{t-1}		5.5331	7.8744		8.9813	39.2404**
		(6.9685)	(10.4015)		(5.5410)	(17.3821)
$LogMcap_{t-1}$		2.0947*	2.2454**		0.2256	4.6674
		(1.0878)	(1.1250)		(3.2877)	(5.8454)
$InstOwn_{t-1}$		1.1210**	0.8458*		−0.4490	−1.2236
		(0.4986)	(0.4894)		(0.9288)	(1.2735)
$VolatilityQ_{t-1}$		6.3029	8.5495		19.5627	37.5549***
		(8.7134)	(8.1239)		(12.9543)	(13.2434)
$ReturnY_{t-1}$		0.5644	0.6154		1.9634*	3.7406**
		(0.5110)	(0.5284)		(1.1479)	(1.6704)
Num. Obs.	3928	3052	2790	889	602	342
R^2	0.05	0.07	0.11	0.07	0.11	0.27
FE: Year	X	X	X	X	X	X
FE: Fiscal Calendar	X	X	X	X	X	X
FE: Industry			X			X
FE: Quarter	X	X	X			
FE: REIT Flag	X	X	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 3.9. Predicting a dividend payout dummy (logit).
The dependent variable is a dummy equal to one if the firm paid a dividend in quarter t. Columns (1) to (3) use the full sample while columns (4) to (6) include each calendar year's first quarter only. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. Included firms are members of the Russell 1000 that are less than 300 ranks away from exclusion. Financial services and utility companies are excluded. Timeframe: 2007 to 2022. Standard errors are clustered at the firm-level and shown in brackets.

	All Q			Q2		
	(1)	(2)	(3)	(4)	(5)	(6)
$0 \leq Distance_{t-1} \leq 50$	−0.1626	−0.1899	−0.2084	−0.1785	−0.3859	−0.7190
	(0.1501)	(0.1891)	(0.2123)	(0.2350)	(0.5531)	(0.6309)
$51 \leq Distance_{t-1} \leq 100$	−0.0558	−0.0754	−0.2217	−0.3225*	−0.3239	−0.6733
	(0.1260)	(0.1508)	(0.1604)	(0.1957)	(0.4393)	(0.4930)
$101 \leq Distance_{t-1} \leq 150$	−0.1434	−0.0912	−0.2026	−0.3361*	−0.4525	−0.6900*
	(0.1114)	(0.1319)	(0.1485)	(0.1881)	(0.3475)	(0.4056)
$TobinQ_{t-1}$		−0.4209**	−0.4975**		−0.3913	−0.4727
		(0.1780)	(0.2198)		(0.2425)	(0.3483)
$LogAssets_{t-1}$		−0.0993	−0.2029		−0.0346	−0.4033
		(0.5343)	(0.6009)		(0.6181)	(0.7430)
$Leverage_{t-1}$		0.2839	0.4759		0.3305	0.7760
		(0.6441)	(0.6990)		(0.7315)	(0.8764)
$Cash_{t-1}$		−2.8728***	−3.5714***		−2.6217**	−3.3278**
		(0.9036)	(1.0562)		(1.0449)	(1.3802)
$ShortDebt_{t-1}$		1.6803	3.3334*		1.3959	4.2127*
		(1.8539)	(2.0170)		(2.1929)	(2.4448)
ROA_{t-1}		8.8851**	6.9308*		13.1470*	8.5273
		(3.6799)	(4.0324)		(7.1214)	(8.9127)
$LogMcap_{t-1}$		−0.1923	−0.2754		−0.5875	−1.0204
		(0.4508)	(0.5218)		(1.5665)	(1.8145)
$InstOwn_{t-1}$		−1.4089***	−1.3495***		−1.5241***	−1.4742**
		(0.4075)	(0.4622)		(0.5164)	(0.6093)
$VolatilityQ_{t-1}$		−21.2816***	−28.7508***		−24.2343*	−27.8399
		(6.8636)	(7.7344)		(13.9980)	(18.7949)
$ReturnY_{t-1}$		0.3980	0.7286**		0.4373	1.0117
		(0.3369)	(0.3584)		(0.5684)	(0.6698)
Num. Obs.	3915	3064	3000	996	768	742
R^2	0.04	0.13	0.30	0.04	0.15	0.29
FE: Year	X	X	X	X	X	X
FE: Fiscal Calendar	X	X	X	X	X	X
FE: Industry			X			X
FE: Quarter	X	X	X			
FE: REIT Flag	X	X	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

suance and repurchase announcements, or dividend payouts to attain Russell 2000 membership. In unreported results I find this also holds for total payouts, including repurchases.

3.5 Concerns and cross-sectionality

This study provides no evidence that the average manager takes opportunistic action to drop from the Russell 1000 Index into the Russell 2000 Index. However, there are some empirical concerns related to the setting which require further investigation. This section addresses these concerns by looking for cross-sectional differences in the sample.[9]

The first concern stems from the fact that Russell rankings fluctuate strongly from quarter to quarter. If managers believe that these fluctuations are either too noisy and/or cannot be reasonably overcome using costly actions, they will not take such actions. However, if a firm has been consistently close to the index threshold, it seems plausible that its manager believes that action towards Russell 2000 inclusion could be more fruitful. To test this, I interact the previously used distance-dummies with the number of months a firm has been within a 200 rank distance to exclusion in the previous 12 months. The results are reported in Tables C.1 to C.3 in the Appendix. I find no conclusive evidence that firms with more consistent distances to the Russell exclusion threshold engage in earnings management to attain Russell 2000 membership.

Another plausible hypothesis is that managers understand that Russell membership determination is a repeated game. If the fundamental value of a firm does not justify index membership, then the membership could be corrected at future index reconstitution dates. This will deter managers from taking opportunistic actions. In that case, managers with a shorter decision horizon, that is, managers which value short-term benefits more than long-term value

[9]For simplicity reasons, I omit the cross-sectional results with respect to repurchase and issuance announcements as well as dividend policy. These omitted results are in line with those of the other managerial tools.

creation, should be more inclined to engage in opportunistic behaviour. I calculate a measure of CEO decision horizon proposed by Antia et al. (2010) and interact the variable with the distance dummies. The results are tabulated in Tables C.4 to C.6 and indicate that CEOs with shorter decision horizons do not engage in more earnings management around the index threshold than other CEOs. However, the results do suggest that some managers who are close to the Russell cutoff and have a low decision horizon, are more likely to highball earnings guidances when those are resolved before the rank day. The result is economically and statistically significant but only holds for firms within 50 ranks of exclusion. Considering the small sample and the fact that this is the only significant effect in this group of regressions, the result is probably spurious.

Another concern is that not all managers would benefit from the change in governance following Russell 2000 inclusion. If the leading force behind managerial incentives of moving from the Russell 1000 to the Russell 2000 Index is a change in governance regime which affects managerial compensation, CEOs who are already compensated better than their peers could see little benefit in taking opportunistic action. Further, CEOs with low performance-pay and high firm-ownership would also be less likely to destroy firm value to drop into the Russell 2000. To see whether these cross-sectional differences drive my results, I rank firms and their industry peers into quintiles based on total CEO compensation and pay-performance sensitivity, and use CEO stock ownership. I interact these variables with the distance-dummies. The results are reported in Tables C.7 to C.15 in the Appendix. Although for some of the governance measures, coefficients are economically and statistically significant, the coefficients are not robust to all of the three specifications based on CEO ownership, performance-pay and total compensation. A potential change in the governance regime stemming from the Russell reconstitution does not seem to entice managers to act opportunistically.

It could also be possible that managers do take actions to join the Russell 2000 Index but that I consider the wrong methods. If there are any effective actions, firms close to the index thresh-

old should be more likely to have negative earnings surprises, and be less likely to marginally beat earnings forecasts. In untabulated tests, I do not find evidence that would suggest this is the case. Further, I analyse the subset of Russell 2000 firms with negative distance to the index threshold (those that would jump into the Russell 1000 Index if indices were reconstituted in the same quarter). These firms might worry that they could join the Russell 1000 Index if they keep their market capitalisation rank. Then, reducing the market capitalisation slightly could be optimal. Again, I find no evidence of opportunistic behaviour in these firms.

Given that the bulk of results indicate no managerial action, it is worth checking whether managers might not simply prefer Russell 1000 over Russell 2000 membership. Joining a new index is often a landmark moment for a company and the associated prestige could also be a factor for firm managers. Indeed, as shown earlier, firms' public statements after Russell 1000 inclusions are usually very positive in tone and mention hard work and dedication. Tables C.16 to C.18 tabulate the results of the main specifications for this study, but inspect firms in the Russell 2000, which are within 300 ranks of promotion to the Russell 1000. The tables provide no evidence that firm managers in the Russell 2000 systematically engage in opportunistic behaviour to jump into the Russell 1000 Index.

3.6 Conclusion

All of the evidence in this study suggests that firms, even if the incentives of their managers should be especially high, do not behave opportunistically in order to obtain specific Russell Index membership. How can these findings be reconciled?

One explanation could be that investors are aware of the inclusion incentives such that the tools at the management's disposal are ineffective. If managers anticipate this, they will not take such actions. However, this explanation fails when taking into account the vast literature on (1) the predictors of real earnings management (e.g., CEO stock participation (Bergstresser and

Philippon, 2006), seasoned equity issuance (Cohen and Zarowin, 2010) and firm performance (Roychowdhury, 2006)) and (2) the stock price effects of real earnings and guidance management (e.g., Sloan (1996); Chen et al. (2023)). The most plausible explanations for the findings is that managers do not have a strong preference for whether they are amongst the largest Russell 2000 firms or smallest Russell 1000 firms, or that if managers do have a preference, the potential benefit of switching the index is not large enough to entice action.

Overall, the results provide novel evidence that Russell 1000/2000 Index membership is likely random at the margin. Even for firms who would be most incentivised to take action to move from one of the Russell Indices into the other, I find no evidence of stock price manipulation before the index reconstitution. No previous study has sufficiently scrutinised the possibility of self-selection into Russell Indices, even though it underlies most of the empirical research studying the effects of Russell Index inclusions.

Conclusion

This PhD thesis uncovered three ways in which cheap diversification has transformed financial markets and the real economy. Across its three chapters, it traced the consequences of the rise of diversified investment strategies. The evidence paints a consistent picture: as diversification becomes nearly costless, it reshapes what investors value, how financial products are organized, and ultimately how firms innovate and allocate resources.

The first chapter showed that access to diversification alters investors' preferences for positively skewed payoffs and that this change has real effects on innovation. As investors no longer reward firm-level lottery-like outcomes, firms optimally reduce exposure to highly skewed, exploratory projects, investing less in transformative innovation – uncovering an important externality of diversified investing. The second chapter demonstrated that the shift in industrial organization of indexing that evolved alongside this demand shift can result in poor outcomes to investors. Rather than reducing costs for investors, self-indexing has led to higher fees and potential conflicts of interest when fund issuers also act as wealth advisors. The third chapter validated that despite these structural changes and the importance of index inclusion for firms, managers do not take opportunistic action to increase their odds of of joining Russell stock indices. It provides novel evidence that Russell Index reconstitutions create a credible empirical setting for studying passive ownership shocks.

The ability to hold the market portfolio at minimal cost has delivered enormous private

gains to investors through lower fees, improved risk sharing, and broad market access. Yet, as this thesis has shown, the diffusion of diversification technologies has implications that extend far beyond portfolio efficiency. The results in this thesis suggest that diversification's general equilibrium consequences are more nuanced than conventional welfare arguments assume. Low-cost indexing has increased allocative efficiency in financial markets, but it may also have dulled the marginal incentive for breakthrough innovation. Modelling and quantifying these externalities of diversification in equilibrium settings mark the logical next step for this research agenda. Doing so requires frameworks that jointly account for investor demand, industry structure, and firm behaviour.

A few papers have recently made important advances in building equilibrium models with index investing. Bond and Garcia (2022) find that as index investing rises, price efficiency of indexed assets fall. Chabakauri and Rytchkov (2021) posit that index investing increases stock return volatilities and cross-correlations. Coles et al. (2022) use a Grossman-Stiglitz model and conclude that price efficiency of assets remains constant when index investing rises because active investors will always trade on private information if it is profitable. Finally, Schmalz and Zame (2024) specifically focus on the welfare effects of index investing by accounting for changes in equilibrium asset prices. In their model, index investing increases asset prices and decreases future returns. In aggregate, this reduces investor welfare. An interesting avenue for future research is to incorporate the innovation externality, and other endogenous firm responses, into an equilibrium framework. This would allow for a more complete welfare analysis of index investing, going beyond its effects on prices and returns. I leave these considerations to future research.

Bibliography

Admati, Anat R and Paul Pfleiderer, "The "Wall Street Walk" and shareholder activism: Exit as a form of voice," *The Review of Financial Studies*, 2009, *22* (7), 2645–2685.

Aghion, Philippe, John Van Reenen, and Luigi Zingales, "Innovation and institutional ownership," *American Economic Review*, 2013, *103* (1), 277–304.

Akey, Pat, Adriana Robertson, and Mikhail Simutin, "Closet active management of passive funds," *Rotman School of Management Working Paper*, 2021.

Amel-Zadeh, Amir, Fiona Kasperk, and Martin C Schmalz, "Mavericks, universal, and common owners-The largest shareholders of US public firms," *European Corporate Governance Institute Working Paper*, 2022.

Amihud, Yakov and Ruslan Goyenko, "Mutual fund's R^2 as predictor of performance," *The Review of Financial Studies*, 2013, *26* (3), 667–694.

An, Yu, Matteo Benetton, and Yang Song, "Index providers: Whales behind the scenes of ETFs," *Journal of Financial Economics*, 2023, *149* (3), 407–433.

Anghel, Dan, Petre Caraiani, Alina Rosu, and Ioanid Rosu, "Asset pricing with systematic skewness: Two decades later," *HEC Paris Research Paper No. FIN-2021-1432*, 2022.

Antia, Murad, Christos Pantzalis, and Jung Chul Park, "CEO decision horizon and firm performance: An empirical investigation," *Journal of Corporate Finance*, 2010, *16* (3), 288–301.

Antón, Miguel, Florian Ederer, Mireia Giné, and Martin Schmalz, "Common ownership, competition, and top management incentives," *Journal of Political Economy*, 2023, *131* (5), 1294–1355.

Appel, Ian R, Todd A Gormley, and Donald B Keim, "Passive investors, not passive owners," *Journal of Financial Economics*, 2016, *121* (1), 111–141.

_ , _ , and _ , "Standing on the Shoulders of Giants: The Effect of Passive Investors on Activism," *The Review of Financial Studies*, 2018, *32* (7), 2720–2774.

_ , _ , and _ , "Identification using Russell 1000/2000 index assignments: A discussion of

methodologies," *Critical Finance Review*, 2024, *13* (1-2), 151–224.

Azar, José, Martin C Schmalz, and Isabel Tecu, "Anticompetitive effects of common ownership," *The Journal of Finance*, 2018, *73* (4), 1513–1565.

Azzalini, Adelchi, "The skew-normal distribution and related multivariate families," *Scandinavian Journal of Statistics*, 2005, *32* (2), 159–188.

Baker, Malcolm and Jeffrey Wurgler, "Appearing and disappearing dividends: The link to catering incentives," *Journal of Financial Economics*, 2004, *73* (2), 271–288.

Bali, Turan G, A Doruk Gunaydin, Thomas Jansson, and Yigitcan Karabulut, "Do the rich gamble in the stock market? Low risk anomalies and wealthy households," *Journal of Financial Economics*, 2023, *150* (2), 103715.

Balloch, Cynthia and Julian Richers, "Asset Allocation and Returns in the Portfolios of the Wealthy," *LSE Financial Markets Group Working Paper*, 2023.

Barber, Brad M and Terrance Odean, "Trading is hazardous to your wealth: The common stock investment performance of individual investors," *The Journal of Finance*, 2000, *55* (2), 773–806.

Barberis, Nicholas and Ming Huang, "Stocks as lotteries: The implications of probability weighting for security prices," *American Economic Review*, 2008, *98* (5), 2066–2100.

Bartov, Eli, Dan Givoly, and Carla Hayn, "The rewards to meeting or beating earnings expectations," *Journal of Accounting and Economics*, 2002, *33* (2), 173–204.

_ , Ferdinand A. Gul, and Judy S.L. Tsui, "Discretionary-accruals models and audit qualifications," *Journal of Accounting and Economics*, 2000, *30* (3), 421–452.

Baruch, Shmuel and Xiaodi Zhang, "The distortion in prices due to passive investing," *Management Science*, 2022, *68* (8), 6219–6234.

Bednarek, Pawel, "Rise of Passive Investing - Effects on Price Level, Market Volatility, and Price Informativeness," *Available at SSRN 4551509*, 2023.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, "Do ETFs increase volatility?," *The Journal of Finance*, 2018, *73* (6), 2471–2535.

_ , _ , and _ , "An improved method to predict assignment of stocks into Russell indexes," 2019. NBER Working Paper.

_ , _ , Byungwook Kim, and Rabih Moussawi, "Competition for Attention in the ETF Space," *The Review of Financial Studies*, 2022, *36* (3), 987–1042.

Berger, Philip G., Eli Ofek, and David L. Yermack, "Managerial Entrenchment and Capital Structure Decisions," *The Journal of Finance*, 1997, *52* (4), 1411–1438.

Bergstresser, Daniel and Thomas Philippon, "CEO incentives and earnings management," *Journal of Financial Economics*, 2006, *80* (3), 511–529.

_ , John M. R. Chalmers, and Peter Tufano, "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry," *The Review of Financial Studies*, 2009, *22* (10), 4129–4156.

Bernstein, Shai, "Does Going Public Affect Innovation?," *The Journal of Finance*, 2015, *70* (4), 1365–1403.

Bertrand, Marianne and Sendhil Mullainathan, "Enjoying the Quiet Life? Corporate Governance and Managerial Preferences," *Journal of Political Economy*, 2003, *111* (5), 1043–1075.

Blackrock, "Global ETF Market Facts: three things to know from Q2 2023," https://www.ishares.com/us/insights/global-etf-facts 2023. Accessed: 12-01-2025.

Bond, Philip and Diego Garcia, "The equilibrium consequences of indexing," *The Review of Financial Studies*, 2022, *35* (7), 3175–3230.

Boone, Audra L. and Joshua T. White, "The effect of institutional ownership on firm transparency and information production," *Journal of Financial Economics*, 2015, *117* (3), 508–533.

Bougherara, Douadia, Lana Friesen, and Céline Nauges, "Risk taking with left-and right-skewed lotteries," *Journal of Risk and Uncertainty*, 2021, *62* (1), 89–112.

Boyer, Brian, Todd Mitton, and Keith Vorkink, "Expected idiosyncratic skewness," *The Review of Financial Studies*, 2010, *23* (1), 169–202.

Brisker, Eric R, Gönül Çolak, and David R Peterson, "Changes in cash holdings around the S&P 500 additions," *Journal of Banking & Finance*, 2013, *37* (5), 1787–1807.

Brogaard, Jonathan, Matthew C Ringgenberg, and David Sovich, "The Economic Impact of Index Investing," *The Review of Financial Studies*, 2018, *32* (9), 3461–3499.

_ , Nataliya Gerasimova, and Ying Liu, "Advising the Advisors: Evidence from ETFs," *Available at SSRN 3972562*, 2021.

Brown, David C, Scott Cederburg, and Mitch Towner, "Dominated ETFs," *forthcoming at Critical Finance Review*, 2024.

Burns, Natasha, Simi Kedia, and Marc Lipson, "Institutional ownership and monitoring: Evidence from financial misreporting," *Journal of Corporate Finance*, 2010, *16* (4), 443–455.

Chabakauri, Georgy and Oleg Rytchkov, "Asset pricing with index investing," *Journal of Financial Economics*, 2021, *141* (1), 195–216.

Chang, Xin, Sudipto Dasgupta, and Gilles Hilary, "Analyst coverage and financing decisions," *The Journal of Finance*, 2006, *61* (6), 3009–3048.

Chattopadhyay, Akash, Matthew D. Shaffer, and Charles C.Y. Wang, "Governance through shame and aspiration: Index creation and corporate behavior," *Journal of Financial Economics*, 2020, *135* (3), 704–724.

Chen, Jing, Michael Jung, and Michael Tang, "Does lowball guidance work? An analysis of firms that consistently beat their guidance by large margins," *The British Accounting Review*, 2023, p. 101219.

Cheng, Si, Massimo Massa, and Hong Zhang, "The unexpected activeness of passive investors: a worldwide analysis of ETFs," *The Review of Asset Pricing Studies*, 2019, *9* (2), 296–355.

Chinco, Alex and Marco Sammon, "The passive ownership share is double what you think it is," *Journal of Financial Economics*, 2024, *157*, 103860.

Cho, Jang Hyung, Dima Leshchinskii, and Janis K Zaima, "Earnings Management and the Reconstitution of the Russell Indexes," *Journal of Accounting and Finance*, 2017, *17* (4), 10–31.

Christoffersen, Susan EK, Richard Evans, and David K Musto, "What do consumers' fund flows maximize? Evidence from their brokers' incentives," *The Journal of Finance*, 2013, *68* (1), 201–235.

Chung, Richard, Michael Firth, and Jeong-Bon Kim, "Institutional monitoring and opportunistic earnings management," *Journal of Corporate Finance*, 2002, *8* (1), 29–48.

Clifford, Christopher P and William C Gerken, "Property rights to client relationships and financial advisor incentives," *The Journal of Finance*, 2021, *76* (5), 2409–2445.

Cohen, Daniel A. and Paul Zarowin, "Accrual-based and real earnings management activities around seasoned equity offerings," *Journal of Accounting and Economics*, 2010, *50* (1), 2–19.

Coles, Jeffrey L, Davidson Heath, and Matthew C Ringgenberg, "On index investing," *Journal of Financial Economics*, 2022, *145* (3), 665–683.

_ , Naveen D Daniel, and Lalitha Naveen, "Managerial incentives and risk-taking," *Journal of Financial Economics*, 2006, *79* (2), 431–468.

_ , _ , and _ , "Calculation of compensation incentives and firm-related wealth using Execucomp: Data, program, and explanation," *Available at SSRN 2296381*, 2013.

Conrad, Jennifer, Robert F Dittmar, and Eric Ghysels, "Ex ante skewness and expected stock returns," *The Journal of Finance*, 2013, *68* (1), 85–124.

Crane, Alan D, Sébastien Michenaud, and James P Weston, "The effect of institutional ownership on payout policy: Evidence from index thresholds," *The Review of Financial Studies*, 2016, *29* (6), 1377–1408.

Dann, Larry Y., "Common stock repurchases: An analysis of returns to bondholders and stockholders," *Journal of Financial Economics*, 1981, *9* (2), 113–138.

Dechow, Patricia M., Richard G. Sloan, and Amy P. Sweeney, "Detecting Earnings Management," *The Accounting Review*, 1995, *70* (2), 193–225.

Dimmock, Stephen G, William C Gerken, and Nathaniel P Graham, "Is fraud contagious? Coworker influence on misconduct by financial advisors," *The Journal of Finance*, 2018, *73* (3), 1417–1450.

Dixit, Avinash K and Joseph E Stiglitz, "Monopolistic competition and optimum product diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

Driessen, Joost, Sebastian Ebert, and Joren Koeter, "(Pi)-CAPM: The Classical CAPM with Probability Weighting and Skewed Assets," *The Review of Financial Studies*, 2025, *38* (12), 3497–3541.

Easley, David, David Michayluk, Maureen O'Hara, and Talis J Putnins, "The active world of passive investing," *Review of Finance*, 2021, *25* (5), 1433–1471.

Ebrahimian, Mehran and Alexander Ljungqvist, "Private Innovation and Wealth Inequality," *Available at SSRN 5957735*, 2025.

Edelen, Roger M., Kingsley Y.L. Fong, and Jingyi Han, "Regulating inattention in fee-based financial advice," *Journal of Financial Economics*, 2025, *164*, 103985.

_ , Richard B. Evans, and Gregory B. Kadlec, "Disclosure and agency conflict: Evidence from mutual fund commission bundling," *Journal of Financial Economics*, 2012, *103* (2), 308–326.

Edmans, Alex, Itay Goldstein, and Wei Jiang, "The real effects of financial markets: The impact of prices on takeovers," *The Journal of Finance*, 2012, *67* (3), 933–971.

Egan, Mark, Gregor Matvos, and Amit Seru, "The market for financial adviser misconduct," *Journal of Political Economy*, 2019, *127* (1), 233–295.

Elton, Edwin J., Martin J. Gruber, and Mustafa N. Gultekin, "Professional Expectations: Accuracy and Diagnosis of Errors," *The Journal of Financial and Quantitative Analysis*, 1984, *19* (4), 351–363.

Federal Reserve Bank of St. Louis, "Banks And Investment Funds No Longer Mutually Exclusive," https://www.stlouisfed.org/publications/regional-economist/october-1993/banks-and-investment-funds-no-longer-mutually-exclusive 1993. Accessed: 2026-01-09.

Foerster, Stephen, Juhani T Linnainmaa, Brian T Melzer, and Alessandro Previtero, "Retail financial advice: does one size fit all?," *The Journal of Finance*, 2017, *72* (4), 1441–1482.

France, Clemens J, "The gambling impulse," *The American Journal of Psychology*, 1902, *13* (3),

364–407.

French, Kenneth R, "Presidential address: The cost of active investing," *The Journal of Finance*, 2008, *63* (4), 1537–1573.

FTSE Russell, "Russell U.S. Equity Indices. Construction and Methodology," https://research.ftserussell.com/products/downloads/Russell-US-indexes.pdf 2023. Accessed: 2026-04-13.

Gabaix, Xavier, "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 2011, *79* (3), 733–772.

_ **and Ralph SJ Koijen**, "In search of the origins of financial fluctuations: The inelastic markets hypothesis," *NBER Working Paper*, 2021.

_ **and** _ , "Granular instrumental variables," *Journal of Political Economy*, 2024, *132* (7), 2274–2303.

Gao, Huasheng, Po-Hsuan Hsu, and Kai Li, "Innovation strategy of private firms," *Journal of Financial and Quantitative Analysis*, 2018, *53* (1), 1–32.

Giannetti, Mariassunta and Bige Kahraman, "Open-End Organizational Structures and Limits to Arbitrage," *The Review of Financial Studies*, 2017, *31* (2), 773–810.

Goetzmann, William N. and Alok Kumar, "Equity Portfolio Diversification," *Review of Finance*, 2008, *12* (3), 433–463.

Gomez, Matthieu, Valentin Haddad, and Erik Loualiche, "The Case Of The Disappearing Skewness," *American Finance Association (AFA) 2023 Annual Meeting Paper*, 2020.

Gormley, Todd A and Hwanki Brian Kim, "Do index funds monitor? Revisited," *ECGI Finance Working Paper*, 2025.

Greenwood, Robin and Marco Sammon, "The disappearing index effect," *The Journal of Finance*, 2025, *80* (2), 657–698.

Guercio, Diane Del and Jonathan Reuter, "Mutual fund performance and the incentive to generate alpha," *The Journal of Finance*, 2014, *69* (4), 1673–1704.

Guo, Di, *Do Firms Manage Earnings to Gain Russell 1000 Membership?*, The Pennsylvania State University, 2019.

Gurun, Umit G, Noah Stoffman, and Scott E Yonker, "Unlocking clients: The importance of relationships in the financial advisory industry," *Journal of Financial Economics*, 2021, *141* (3), 1218–1243.

Hartzell, Jay C and Laura T Starks, "Institutional investors and executive compensation," *The Journal of Finance*, 2003, *58* (6), 2351–2374.

Harvey, Campbell R and Akhtar Siddique, "Conditional skewness in asset pricing tests," *The Journal of Finance*, 2000, *55* (3), 1263–1295.

_ , Michele G Mazzoleni, and Alessandro Melone, "The unintended consequences of rebalancing," *NBER Working Paper*, 2025.

Healy, Paul M. and Krishna G. Palepu, "Earnings information conveyed by dividend initiations and omissions," *Journal of Financial Economics*, 1988, *21* (2), 149–175.

Heath, Davidson, Daniele Macciocchi, Roni Michaely, and Matthew C Ringgenberg, "Do index funds monitor?," *The Review of Financial Studies*, 2022, *35* (1), 91–131.

Hong, Eunpyo, Min Hwang, and Jiyoon Lee, "Strategic adjustment of capital structure: Evidence from firms newly added to the S&P 500 index," *Available at SSRN 3263539*, 2021.

Hortaçsu, Ali and Chad Syverson, "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds," *The Quarterly Journal of Economics*, 2004, *119* (2), 403–456.

Hosking, J. R. M., "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1990, *52* (1), 105–124.

Huang, Shiyang, Yang Song, and Hong Xiang, "The Smart Beta Mirage," *Journal of Financial and Quantitative Analysis*, 2023, p. 1–32.

Ikenberry, David L. and Theo Vermaelen, "The Option to Repurchase Stock," *Financial Management*, 1996, *25* (4), 9–24.

Jensen, Michael C., "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers," *The American Economic Review*, 1986, *76* (2), 323–329.

Jiang, Hao, Dimitri Vayanos, and Lu Zheng, "Passive investing and the rise of mega-firms," *NBER Working Paper*, 2020.

Joanes, D. N. and C. A. Gill, "Comparing measures of sample skewness and kurtosis," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1998, *47* (1), 183–189.

Jones, Jennifer J., "Earnings Management During Import Relief Investigations," *Journal of Accounting Research*, 1991, *29* (2), 193–228.

Kahraman, Bige, "Fee complexity and investor mistakes in retail financial markets," *Quarterly Journal of Finance*, 2021, *11* (01), 2150004.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring technological innovation over the long run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

Kelly, Morgan, "All their eggs in one basket: Portfolio diversification of US households," *Journal of Economic Behavior & Organization*, 1995, *27* (1), 87–96.

Khomyn, Marta, Talis Putnins, and Marius Zoican, "The value of ETF liquidity," *The Review of Financial Studies*, 2024, *37* (10), 3092–3148.

Koch, Andrew, Marios Panayides, and Shawn Thomas, "Common ownership and competition in product markets," *Journal of Financial Economics*, 2021, *139* (1), 109–137.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, "Which investors matter for equity valuations and expected returns?," *Review of Economic Studies*, 2024, *91* (4), 2387–2424.

Korajczyk, Robert, Deborah J Lucas, and Robert L McDonald, "Understanding stock price behavior around the time of equity issues," *NBER Working Paper*, 1989.

Kostovetsky, Leonard and Jerold Warner, "Investor heterogeneity and the market for fund benchmarks: Evidence from passive ETFs," *Journal of Banking & Finance*, 2025, *173*, 107412.

Kraus, Alan and Robert H Litzenberger, "Skewness preference and the valuation of risk assets," *The Journal of Finance*, 1976, *31* (4), 1085–1100.

Kumar, Alok, "Who gambles in the stock market?," *The Journal of Finance*, 2009, *64* (4), 1889–1933.

Langlois, Hugues, "Measuring skewness premia," *Journal of Financial Economics*, 2020, *135* (2), 399–424.

Levy, H and HM Markowtiz, "Approximating Expected Utility by a Function of Mean and Variance," *American Economic Review*, 1979, *69* (3), 308–17.

Li, Kun, Xin Liu, and Shang-Jin Wei, "Credit Rating Purchases and S&P 500 Index Membership Decisions," *Management Science*, 2026.

Li, Sida, "Should passive investors actively manage their trades?," *Available at SSRN 3967799*, 2021.

Limburg, Anthony, "Do Firms Self-Select Into Russell Indices?," *Available at SSRN 4816038*, 2024.

Lin, Tse-Chun and Xin Liu, "Skewness, individual investor preference, and the cross-section of stock returns," *Review of Finance*, 2018, *22* (5), 1841–1876.

Linnainmaa, Juhani T, Brian T Melzer, and Alessandro Previtero, "The misguided beliefs of

financial advisors," *The Journal of Finance*, 2021, *76* (2), 587–621.

Lintner, John, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *The Review of Economics and Statistics*, 1965, *47* (1), 13–37.

Liu, Yang, Yao Shen, Jun Wang, and Qijian Wang, "Passive institutional investors and corporate innovation," *Available at SSRN 3415513*, 2019.

Low, Angie, "Managerial risk-taking behavior and equity-based compensation," *Journal of Financial Economics*, 2009, *92* (3), 470–490.

Luong, Hoang, Fariborz Moshirian, Lily Nguyen, Xuan Tian, and Bohui Zhang, "How do foreign institutional investors enhance firm innovation?," *Journal of Financial and Quantitative Analysis*, 2017, *52* (4), 1449–1490.

Markowitz, Harry, "Portfolio Selection," *The Journal of Finance*, 1952, *7* (1), 77–91.

Marsili, Orietta and Ammon Salter, "'Inequality' of innovation: skewed distributions and the returns to innovation in Dutch manufacturing," *Economics of Innovation and New Technology*, 2005, *14* (1-2), 83–102.

Mitton, Todd and Keith Vorkink, "Equilibrium underdiversification and the preference for skewness," *The Review of Financial Studies*, 2007, *20* (4), 1255–1288.

Mossin, Jan, "Equilibrium in a Capital Asset Market," *Econometrica*, 1966, *34* (4), 768–783.

Mullainathan, Sendhil, Markus Noeth, and Antoinette Schoar, "The Market for Financial Advice: An Audit Study," *NBER Working Paper*, 2012.

Myers, Stewart C. and Nicholas S. Majluf, "Corporate financing and investment decisions when firms have information that investors do not have," *Journal of Financial Economics*, 1984, *13* (2), 187–221.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, "Do funds make more when they trade more?," *The Journal of Finance*, 2017, *72* (4), 1483–1528.

Polk, Christopher and Paola Sapienza, "The Stock Market and Corporate Investment: A Test of Catering Theory," *The Review of Financial Studies*, 2008, *22* (1), 187–217.

Pollet, Joshua M. and Mungo Wilson, "How Does Size Affect Mutual Fund Behavior?," *The Journal of Finance*, 2008, *63* (6), 2941–2969.

Robertson, Adriana Z, "Passive in name only: Delegated management and index investing," *Yale Journal on Regulation*, 2019, *36*, 795.

Roussanov, Nikolai, Hongxun Ruan, and Yanhao Wei, "Marketing mutual funds," *The Review of Financial Studies*, 2021, *34* (6), 3045–3094.

Roychowdhury, Sugata, "Earnings management through real activities manipulation," *Journal of Accounting and Economics*, 2006, *42* (3), 335–370.

Sabban, Adam, "It's Official: Passive Funds Overtake Active Funds," https://www.morningstar.com/funds/recovery-us-fund-flows-was-weak-2023 2024. Accessed: 2026-01-09.

Sammon, Marco, "Passive ownership and price informativeness," *Management Science*, 2025, *71* (6), 4582–4598.

_ **and John J Shim**, "Who Clears the Market When Passive Investors Trade?," *Available at SSRN 4777585*, 2024.

Schmalz, Martin C and William R Zame, "Do Index Funds Benefit Investors?," *Available at SSRN 4246321*, 2024.

Schmickler, Simon and Pedro Tremacoldi-Rossi, "Spillover effects of payouts on asset prices and real investment," *Available at SSRN 4287300*, 2022.

Schmidt, Cornelius and Rüdiger Fahlenbrach, "Do exogenous changes in passive institutional ownership affect corporate governance and firm value?," *Journal of Financial Economics*, 2017, *124* (2), 285–306.

Sharpe, William F, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *The Journal of Finance*, 1964, *19* (3), 425–442.

_ , "The arithmetic of investment expenses," *Financial Analysts Journal*, 2013, *69* (2), 34–41.

Shleifer, Andrei, "Do Demand Curves for Stocks Slope Down?," *The Journal of Finance*, 1986, *41* (3), 579–590.

_ **and Robert W Vishny**, "Large shareholders and corporate control," *Journal of Political Economy*, 1986, *94* (3), 461–488.

Sias, Richard, Harry J Turtle, and Blerina Zykaj, "Hedge fund crowds and mispricing," *Management Science*, 2016, *62* (3), 764–784.

Simeth, Markus and David Wehrheim, "On "Innovation and institutional ownership"," *Journal of Corporate Finance*, 2024, *86*, 102569.

Simkowitz, Michael A and William L Beedles, "Diversification in a three-moment world," *Journal of Financial and Quantitative Analysis*, 1978, *13* (5), 927–941.

Sloan, Richard G., "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?," *The Accounting Review*, 1996, *71* (3), 289–315.

S&P Dow Jones Indices, "S&P U.S. Indices Methodology," https://www.spglobal.com/spdji/

en/documents/methodologies/methodology-sp-us-indices.pdf 2023. Accessed: 2026-04-13.

Tian, Xuan and Yuanchen Yang, "Are Passive Institutional Investors Engaged Monitors or Risk-Averse Owners? Both!," *IMF Working Paper*, 2023.

Tirole, Jean, *The Theory of Industrial Organization*, MIT press, 1988.

Viva, Luca Del, Eero Kasanen, and Lenos Trigeorgis, "Real options, idiosyncratic skewness, and diversification," *Journal of Financial and Quantitative Analysis*, 2017, *52* (1), 215–241.

Xiao, Yizhou and Yan Xiong, "Designing Index Provision," *Available at SSRN 4815898*, 2024.

Yu, Fang, "Analyst coverage and earnings management," *Journal of Financial Economics*, 2008, *88* (2), 245–271.

Zang, Amy Y., "Evidence on the Trade-Off between Real Activities Manipulation and Accrual-Based Earnings Management," *The Accounting Review*, 2012, *87* (2), 675–703.

Zhang, Hulai, "Flow-Driven Corporate Finance: A Supply-Demand Approach," *Available at SSRN 5052065*, 2024.

Appendices

Appendix A

Appendix A: Figures

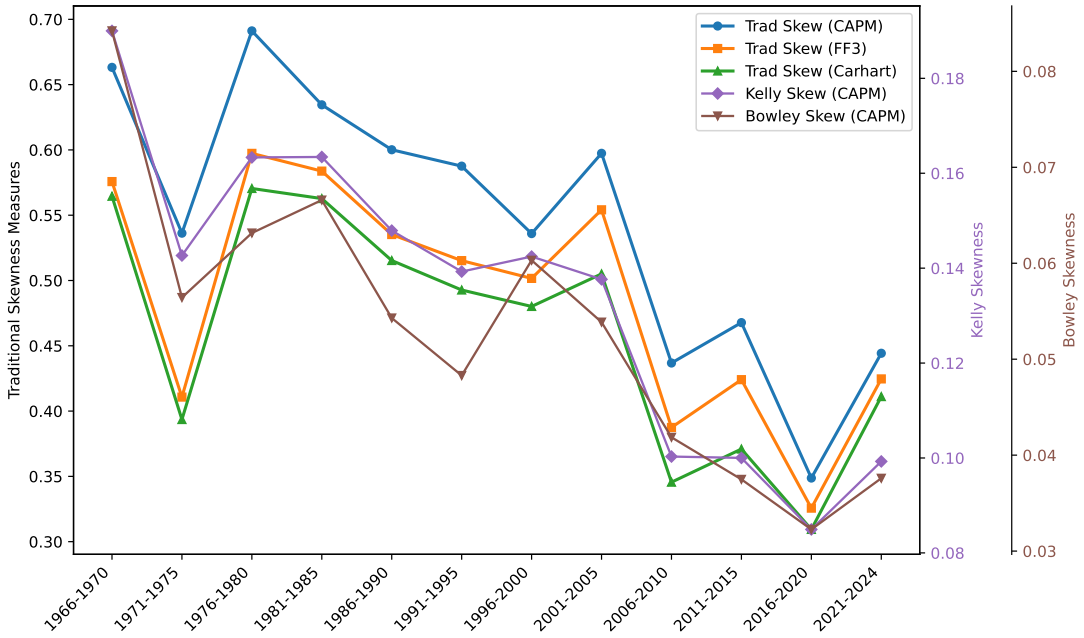


Figure A.1. This figure plots the median idiosyncratic skewness as calculated from residuals from 5-year rolling CAPM, Fama French and Carhart model regressions. The measures on the left axis are calculated using the standard definition of skewness as in equation 1.2. On the right axis I plot skewness as measured by Bowley and Kelly.

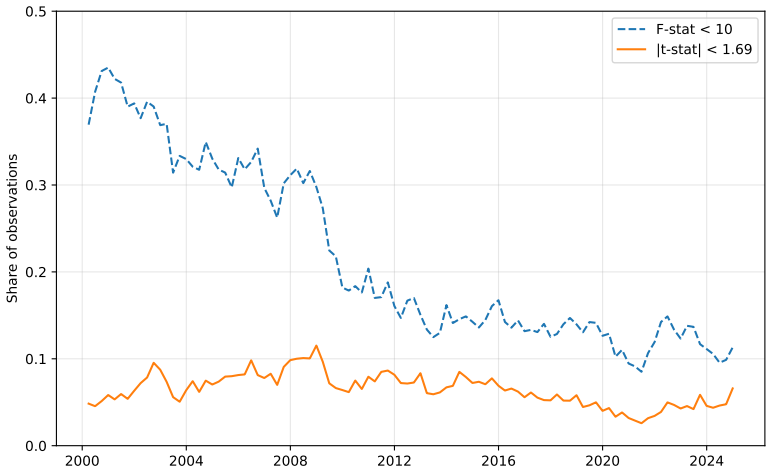


Figure A.2. This figure plots the share of quarterly first-stage regressions (Koijen et al., 2024) with absolute t statistics of the instrument below 1.69 and share of regression F statistics below 10.

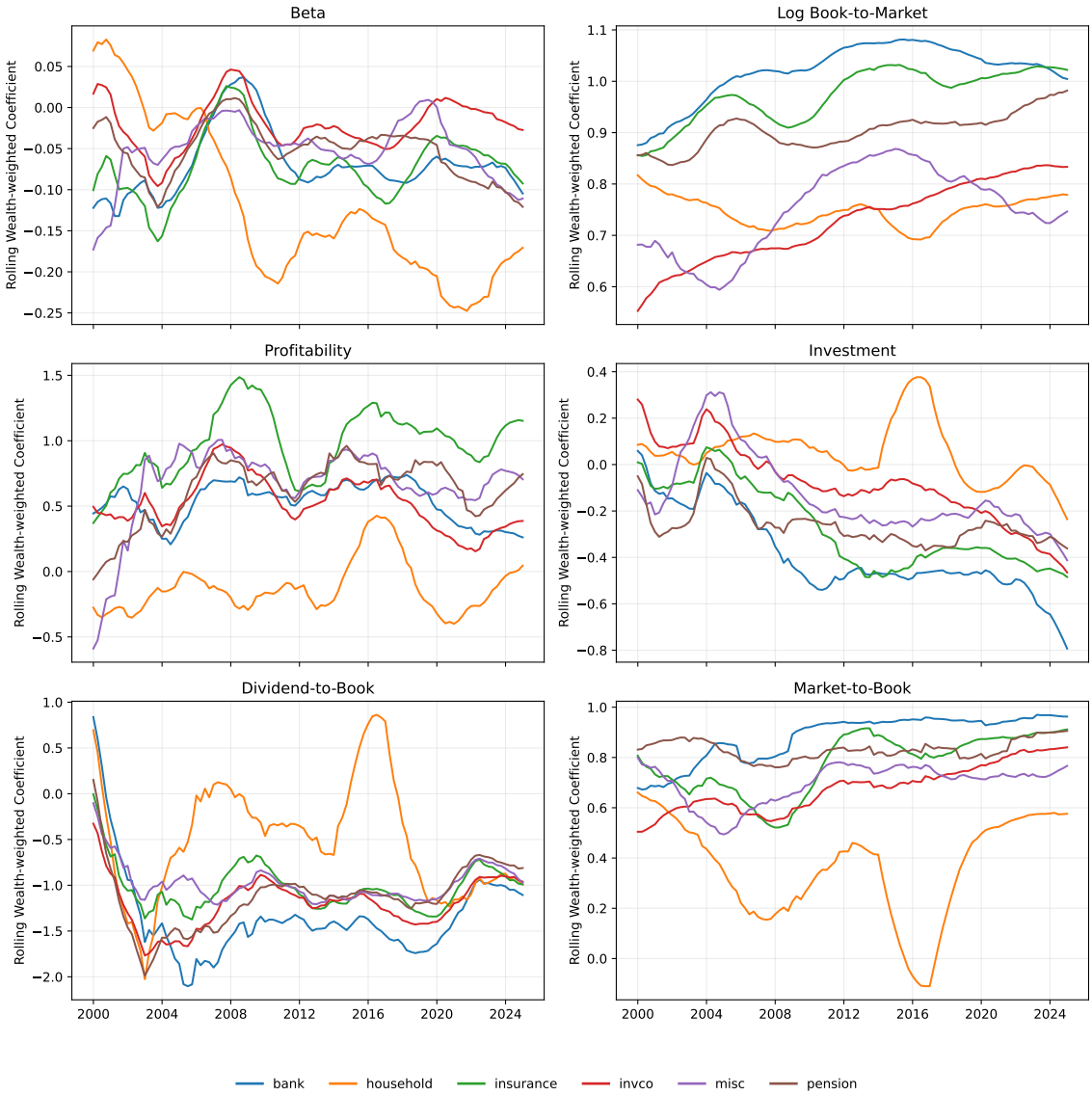


Figure A.3. This figure plots the AUM-weighted taste estimates for different types of investors between 2000 and 2024 estimated from a Koijen et al. (2024) framework. The taste estimate for skewness is omitted.

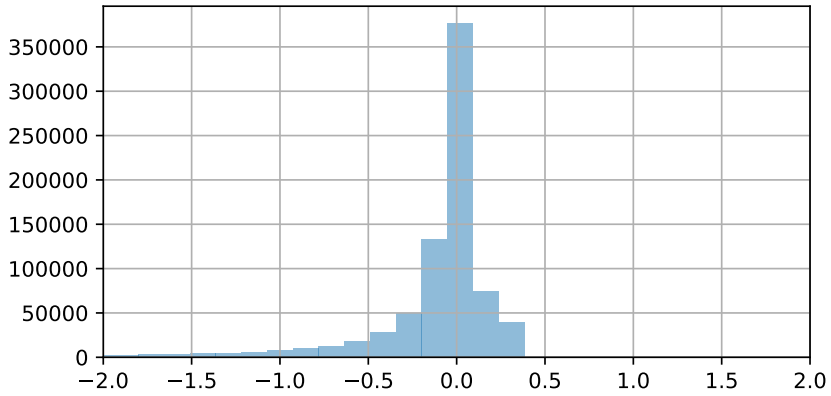


Figure A.4. This figure plots the distribution of $z_t(n)$ when using $K = 2$ latent orthogonal variables. The variable is winsorized at 2.5 and 97.5%.

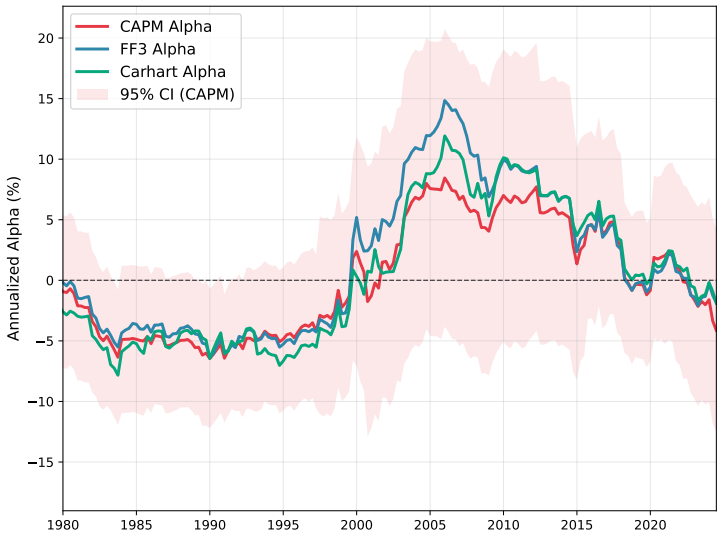


Figure A.5. Skewness premium over time (equal-weight).
This figure plots the 15-year rolling annualized alpha of equal-weighted long-short portfolios of skewness-sorted stocks. The alphas are calculated from the CAPM, Fama-French three factor and Carhart four factor model.

Appendix A: Tables

Table A.1. Skewness and profitability.
The dependent variable is the idiosyncratic return skewness from a 5-year window estimated with the CAPM. $Profitability$ is defined as operating profit divided by book equity. $Log(Mcap)$ is the log of firm market cap. $Log(Assets)$ is log of total assets. $Leverage$ is the book value of total debt divided by the book value of total assets. $Market - to - Book$ is the market value of equity divided by the book value of equity. $Log(Age)$ is the log of firm age in years since IPO. $CapitalExpenditure$ is capital expenditures scaled by total assets. $Investment$ is the annual change in total assets scaled by total assets. $Spread$ is the average daily bid-ask spread scaled by price. Standard errors are clustered at the firm level.

	Idiosyncratic Return Skewness	
	(1)	(2)
$Profitability$	-0.800^{***}	-0.487^{***}
	(0.039)	(0.054)
$Log(MCap)$		-0.161^{***}
		(0.023)
$Log(Assets)$		-0.018
		(0.023)
$Leverage$		-0.075
		(0.063)
$Log(Age)$		0.122^{***}
		(0.008)
$Investment$		0.265^{***}
		(0.023)
$Market - to - Book$		0.181^{***}
		(0.023)
$CapEx$		-1.333^{***}
		(0.340)
$Spread$		1.215^{***}
		(0.323)
Num. Obs.	34 514	28 661
R^2	0.05	0.18
FE: Quarter	X	X
FE: Industry	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table A.2. Skewness and institutional ownership with overlapping windows.
The dependent variable is the idiosyncratic return skewness from a 5-year window estimated with the CAPM. $InstOwn_{t-5y}$ is 5-year lagged institutional ownership and $SkewPref_{t-5y}$ is the 5-year lagged owner skewness preference estimate. $Log(Mcap)$ is the log of firm market cap. $Log(Assets)$ is log of total assets. $Leverage$ is the book value of total debt divided by the book value of total assets. $Market-to-Book$ is the market value of equity divided by the book value of equity. $Log(Age)$ is the log of firm age in years since IPO. $CapitalExpenditure$ is capital expenditures scaled by total assets. $Investment$ is the annual change in total assets scaled by total assets. $Spread$ is the average daily bid-ask spread scaled by price.

	Idiosyncratic Return Skewness			
	(1)	(2)	(3)	(4)
$InstOwn_{t-5y}$	−0.379***		−0.245***	
	(0.027)		(0.029)	
$SkewPref_{t-5y}$		1.087***		0.642***
		(0.118)		(0.124)
$Log(MCap)$			−0.006	−0.012
			(0.019)	(0.020)
$Log(Assets)$			−0.140***	−0.146***
			(0.021)	(0.022)
$Leverage$			0.189***	0.180***
			(0.056)	(0.057)
$Log(Age)$			−0.028	−0.063*
			(0.035)	(0.035)
$Investment$			0.173***	0.184***
			(0.011)	(0.011)
$Market-to-Book$			0.070***	0.075***
			(0.018)	(0.019)
$CapEx$			−1.024***	−0.996***
			(0.204)	(0.213)
$Spread$			−0.168	−0.347
			(0.216)	(0.223)
Num. Obs.	522 792	471 630	399 906	386 402
R^2	0.48	0.48	0.51	0.51
FE: Quarter	X	X	X	X
FE: Firm	X	X	X	X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3. Robustness on causal skewness evidence.
This table shows the results of a two-stage least squared regression predicting idiosyncratic return skewness from a 5-year window estimated with the CAPM. The instrument for lagged institutional ownership is $GIV(5y)_{t-5y}$ as constructed in equation 1.10. The models presented in columns (1) to (4) vary in whether they control for firm-level control variables, latent factors, and in how many latent factors were used to calculate the instrument. Firm-level control variables are the same as in 1.2. Standard errors are clustered at the firm level.

	Idiosyncratic Return Skewness			
	(1)	(2)	(3)	(4)
Panel A: First Stage				
$GIV(5y)_{t-5y}$	0.019***	0.008***	0.002***	0.001***
	(0.003)	(0.002)	(0.000)	(0.000)
Num. Obs.	25651	20602	25930	20674
R^2	0.31	0.56	0.30	0.56
F-statistic	136.35	34.51	163.27	38.31
#Latent Factors	2	2	3	3
Firm Controls		X		X
Latent Controls	X	X		
FE: Quarter	X	X	X	X
FE: Industry	X	X	X	X
Panel B: Second Stage				
$\widehat{InstOwn}_{t-5y}$	-2.864***	-2.529***	-2.732***	-2.550***
	(0.417)	(0.865)	(0.366)	(0.821)
Num. Obs.	25651	20602	25930	20674
R^2	0.05	0.22	0.05	0.22
#Latent Factors	2	2	3	3
Firm Controls		X		X
Latent Controls	X	X		
FE: Quarter	X	X	X	X
FE: Industry	X	X	X	X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4. Causal evidence on longer horizons.
This table shows the results of a two-stage least squared regression predicting idiosyncratic return skewness from a 5-year window estimated with the CAPM (columns 1 and 2) and normalised measures of breakthrough innovation (columns 3 to 5). In column (3) I predict $PatentValue(5y)$, the 5-year sum of a firms patent values from Kogan et al. (2017), divided by the firm market capitalisation. In column (4) I predict $MaxPatentCitations(5y)$, the logarithm of number of citations from the firms most cited patent in a five year window. In column (5) I predict $PatentImportance(5y)$, the 5-year median patent importance for each firm obtained from Kelly et al. (2021). All dependent variables are measured over the t+5 and t+10 years time horizon relative to the measurement of diversified ownership. The instrument for lagged institutional ownership is $GIV(5y)_{t-5y}$ as constructed in equation 1.10. Firm-level control variables and fixed effects are the same as in Table 1.2. Standard errors are clustered at the firm level.

	Skewness (t+5)		Pat.Value (t+5)	MaxPat.Citations (t+5)	Pat.Importance (t+5)
	(1)	(2)	(3)	(4)	(5)
Panel A: First Stage					
$GIV(5y)_{t-5y}$	0.015***	0.009***	0.009***	0.009***	0.009***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Num. Obs.	16597	13180	13317	13317	13317
F-statistic	59.87	28.93	26.96	26.96	26.96
R^2	0.28	0.54	0.55	0.55	0.55
Firm Controls		X	X	X	X
FE: Time	X	X	X	X	X
FE: Industry	X	X	X	X	X
Panel B: Second Stage					
$\widehat{InstOwn}_{t-5y}$	-2.852***	-2.643**	-0.607*	-2.343*	-2.763**
	(0.655)	(1.059)	(0.361)	(1.325)	(1.283)
Num. Obs.	16597	13180	13317	13317	13317
R^2	0.04	0.19	0.15	0.45	0.37
Firm Controls		X	X	X	X
FE: Time	X	X	X	X	X
FE: Industry	X	X	X	X	X

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

Table A.5. Robustness to causal link between institutional ownership and breakthrough innovation. This table shows the results of a two-stage least squared regression predicting normalized breakthrough innovation variables only including firms with patenting activity. The instrument for lagged institutional ownership is $GIV(5y)_{t-5y}$ as constructed in equation 1.10. In column (1) I predict $PatentValue(5y)$, the 5-year sum of a firms patent values from Kogan et al. (2017), divided by the firm market capitalisation. In column (2) I predict $MaxPatentCitations(5y)$, the logarithm of number of citations from the firms most cited patent in a five year window. In column (3) I predict $PatentImportance(5y)$, the 5-year median patent importance for each firm obtained from Kelly et al. (2021). All dependent variables are normalized. All models include the time-varying firm-level controls from Table 1.2 as well as time and industry fixed effects. The instrument is calculated using two latent factors. Standard errors are clustered at the firm level.

	Pat.Value	MaxPat.Citations	Pat.Importance
	(1)	(2)	(3)
Panel A: First Stage			
$GIV(5y)_{t-5y}$	0.014***	0.014***	0.014***
	(0.003)	(0.003)	(0.003)
Num. Obs.	8873	8873	8873
F-statistic	42.63	42.63	42.63
R^2	0.56	0.56	0.56
Firm Controls	X	X	X
FE: Quarter	X	X	X
FE: Industry	X	X	X
Panel B: Second Stage			
$\widehat{InstOwn}_{t-5y}$	-0.255*	-1.443	-1.577**
	(0.146)	(0.905)	(0.725)
Num. Obs.	8873	8873	8873
R^2	0.18	0.56	0.50
Firm Controls	X	X	X
FE: Quarter	X	X	X
FE: Industry	X	X	X

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

Table A.6. Causal evidence on innovation channels.
This table shows the results of a two-stage least squared regression predicting innovation variables. The instrument for lagged institutional ownership is $GIV(5y)_{t-5y}$ as constructed in equation 1.10. In columns (1) and (2) I predict $R\&D$, the 5-year average of a firms R&D intensity, defined as R&D expenditure divided by total assets. In column (3) I predict $PatCount$, the normalized number of patents from in a five year window. In column (4) I predict $HHIPatAreas$, the 5-year average Herfindahl Index of CPC patent classifications. In column (1) I include all firms in the sample while all other columns include only patenting firms. All models include the time-varying firm-level controls from Table 1.2 as well as time and industry fixed effects. The instrument is calculated using two latent factors. Standard errors are clustered at the firm level.

	All Firms	Only Patenting Firms		
	R&D		PatCount	HHIPatAreas
	(1)	(2)	(3)	(4)
Panel A: First Stage				
$GIV(5y)_{t-5y}$	0.011***	0.015***	0.015***	0.015***
	(0.002)	(0.003)	(0.003)	(0.003)
Num. Obs.	20836	8873	8873	8873
F-statistic	68.33	50.06	50.06	50.06
R^2	0.55	0.55	0.55	0.55
FE: Time	X	X	X	X
FE: Industry	X	X	X	X
Panel B: Second Stage				
$\widehat{InstOwn}_{t-5y}$	-0.101***	-0.131***	-2.748*	0.572*
	(0.026)	(0.033)	(1.413)	(0.293)
Num. Obs.	20836	8873	8873	8873
R^2	0.40	0.48	0.48	0.29
FE: Time	X	X	X	X
FE: Industry	X	X	X	X

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01

Appendix B

Appendix B: Figures

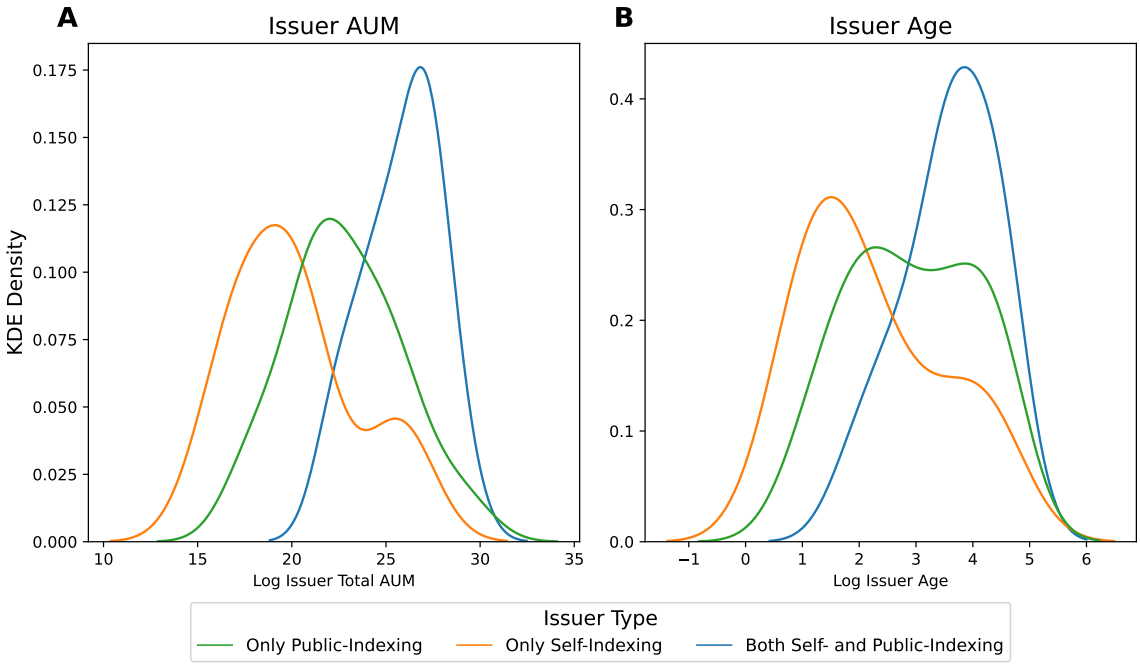


Figure B.1. Density plots of issuer characteristics by issuer type.
This figure presents KDE density plots for key issuer characteristics, grouped by issuer type, as of December 2020. Panel A plots the densities for the logarithm of total issuer AUM, calculated as the total AUM of all outstanding funds by an issuer (across all asset-classes) from the CRSP MF database. Panel B plots the densities for the logarithm of issuer age, defined as the difference in days between December 31st 2020 and the first observation of the issuer in the CRSP MF database, divided by 365.

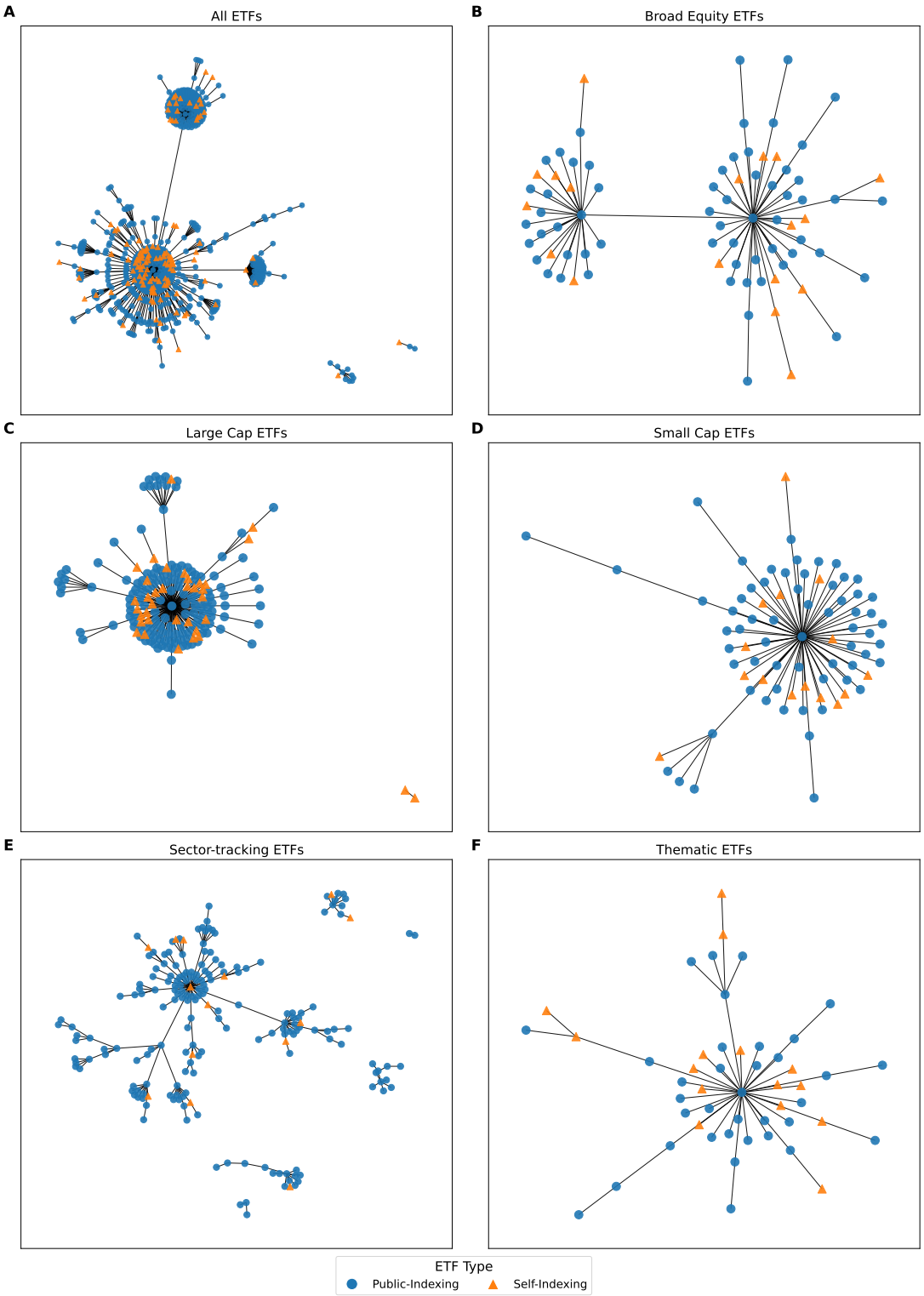


Figure B.2. ETF networks based on pairwise ETF return correlation.
This figure presents minimum spanning tree network plots of ETFs based on their monthly average pairwise return correlations. The networks are plotted using the Fruchterman Reingold Layout. Panel A presents the network of all ETFs whereas panels B to F plot networks within ETF subgroups.

Appendix B: Tables

Table B.1. Largest issuers of ETFs by AUM as of December 2020.
AUM and the number of unique ETFs are from our sample of non-active equity ETFs in ETF Global. Mixing issuers are issuers which offer both self- and public-index ETFs.

	Self-Indexing		Public-Indexing	
Issuer	AUM ($Mn.)	Unique ETFs	AUM ($Mn.)	Unique ETFs
Panel A: Only Self-Indexing Issuers				
John Hancock	854.76	11	-	-
Renaissance	734.30	1	-	-
American Century	384.50	2	-	-
Motley Fool Asset Management	367.76	1	-	-
Tortoise	360.65	1	-	-
Hartford	269.96	2	-	-
Inspire Investing	237.59	2	-	-
Janus Henderson	236.10	2	-	-
Alpha Architect	218.68	2	-	-
Distillate Capital Partners	201.09	1	-	-
Panel B: Mixing Issuers				
WisdomTree	15588.73	13	1006.86	1
Goldman Sachs	11400.19	3	621.38	2
Pacer Financial	2722.28	8	871.15	4
Victory Capital Management	2365.95	7	448.53	2
Northern Trust	2080.11	5	1452.45	2
Charles Schwab	1320.58	1	107147.66	10
Fidelity	980.20	6	17625.00	11
SSgA	976.39	3	633741.90	59
JPMorgan	737.89	5	1280.29	4
ProShares	278.79	2	9032.34	13
Panel C: Only Public-Indexing Issuers				
Blackrock	-	-	961544.19	93
Vanguard	-	-	866575.55	43
First Trust	-	-	55674.09	54
Global X	-	-	6132.46	15
Alps	-	-	5936.66	6
Van Eck	-	-	4300.81	4
DWS	-	-	3444.60	4
Principal	-	-	2858.50	5
Nuveen	-	-	1904.51	6
Franklin Templeton Investments	-	-	1520.72	3

Table B.2. Explaining the issuance of self-indexed ETFs at the issuer-month level.
The dependent variable is a dummy which equals one if the issuer increases the amount of outstanding self-indexed ETFs. In columns (1) - (3) we use an OLS estimation and in columns (4) - (6) we use probit estimation. $IssuerLog(Age)$ is the natural logarithm of the number of days divided by 365 since the first observation of an issuer in the CRSP MF database. $IssuerLog(AUM)$ is the natural logarithm of all assets under management of an issuer across all asset classes in the CRSP MF database. $IssuerNrETFs$ is the number of non-active equity ETFs operated by an issuer and $IssuerNrSelfIndexETFs$ is the number of self-indexed ETFs operated by an issuer in our sample from ETF global. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	OLS			Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	0.042			−1.918		
	(0.051)			(2.457)		
$IssuerLog(Age)$	0.004	0.004	0.004	0.312	0.237	−1.749
	(0.007)	(0.006)	(0.011)	(0.357)	(0.343)	(2.538)
$IssuerLog(AUM)$	−0.002	−0.002	−0.012*	−0.197	−0.157	−1.020
	(0.003)	(0.003)	(0.007)	(0.155)	(0.116)	(1.068)
$IssuerNrETFs$	0.000	0.000	0.001	0.020	0.019	0.047
	(0.000)	(0.000)	(0.001)	(0.014)	(0.014)	(0.042)
$IssuerNrSelfIndexETFs$	0.007***	0.007***	0.010	0.264***	0.266***	0.760
	(0.002)	(0.002)	(0.006)	(0.052)	(0.048)	(0.757)
Num. Obs.	1521	1521	1521	1521	705	286
R^2	0.03	0.05	0.12	0.13	0.17	0.16
FE: Month		X			X	
FE: Issuer			X			X

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

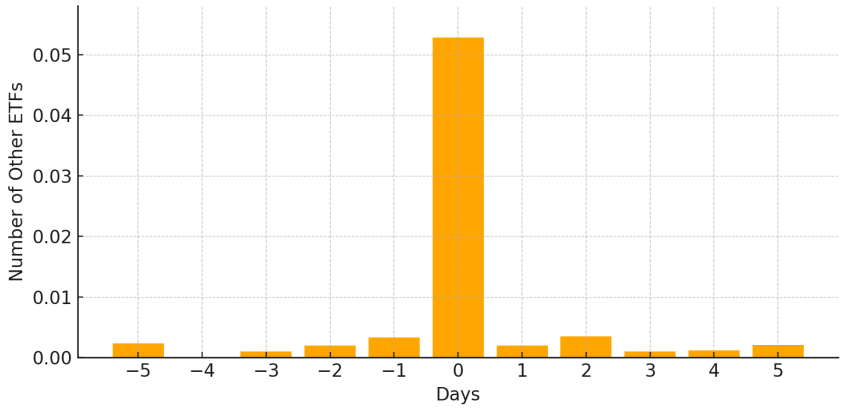
Table B.3. Robustness tests.
In column (1) we predict the monthly average portfolio cosine similarity of an ETF to its style-peers and in column (2) we predict the average return correlation between an ETF and its style-peers within a month. In columns (3) - (4) the dependent variable is the monthly (daily) gross Carhart four-factor excess return, defined as the gross returns minus predicted returns using factor loadings from 36 month (126 day) rolling windows $SelfIndexer?$ is a dummy variable that equals one if an ETF is self-indexed. $NetExp.Ratio(\%)$ is the ETFs annual expense ratio in percentage terms. $Log(AUM)$ is the natural logarithm of an ETFs assets under management. $Log(Holdings)$ is the natural logarithm of an ETFs number of unique portfolio holdings. $Log(Age)$ is the natural logarithm of an ETFs age as measured in days since inception divided by 365. $SmartBeta?$ is a dummy variable that is one if an ETF is classified as smart beta. $Uniqueness$ is the 12-month rolling average absolute difference of ETF gross returns versus the value-weighted average ETF gross return in the same style-category, as in Kostovetsky and Warner (2025). $TurnRatio$ is an ETFs yearly portfolio turnover ratio. $Activeness$ is defined as $(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model. $MAvgSpread$ is the monthly average ETF secondary market price spread scaled by price. $Volatility$ is the standard deviation of daily net (NAV) returns within a month. Standard errors are two-way clustered by month-issuer interaction and reported in parentheses.

	Cosine Similarity	Return Correlation	Monthly Excess Ret	Daily Excess Ret
	(1)	(2)	(3)	(4)
$SelfIndexer?$	0.018***	0.028***	−0.138	−0.010**
	(0.003)	(0.006)	(0.138)	(0.005)
$NetExp.Ratio(\%)$	0.001*	0.001	−0.022	0.001
	(0.001)	(0.002)	(0.076)	(0.003)
$Log(AUM)$	0.001**	−0.004***	0.002	0.000
	(0.000)	(0.000)	(0.014)	(0.000)
$Log(Holdings)$	0.026***	0.013***	−0.048*	−0.001
	(0.001)	(0.001)	(0.028)	(0.001)
$Log(Age)$	−0.009***	0.011***	−0.030	−0.001
	(0.002)	(0.002)	(0.068)	(0.002)
$SmartBeta?$	−0.006***	0.004***	−0.076*	−0.003**
	(0.002)	(0.001)	(0.040)	(0.001)
$Uniqueness$	−0.019***	−0.018***	−0.003	0.000
	(0.001)	(0.001)	(0.050)	(0.002)
$TurnRatio$	−0.001***	0.001	−0.056	−0.001
	(0.000)	(0.001)	(0.041)	(0.001)
$Activeness$	−0.036***	−0.233***	−0.649*	−0.014
	(0.005)	(0.011)	(0.366)	(0.014)
$MAvgSpread$	0.002	0.005*	0.543*	0.004**
	(0.003)	(0.003)	(0.330)	(0.002)
$Volatility$			−0.900	−0.032
			(0.758)	(0.034)
Num. Obs.	22 943	34 181	20 013	693 988
R^2	0.83	0.88	0.62	0.62
FE: Month-Style	X	X	X	
FE: Month-Issuer	X	X	X	
FE: Day-Style				X
FE: Day-Issuer				X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table B.4. Performance of active funds.
Panel A plots the trading behavior of public-indexed ETFs around same-family self-indexed ETF trades. For each self-indexed ETF trade of a stock ticker i on day T, we count the number of same-issuer public-indexed ETFs trading the same ticker in the same direction in [T-5,T+5]. The figure plots the average count. Panel B reports performance regressions for active funds. In Panel B, the dependent variable is monthly net (NAV) return or Net Carhart Alpha using factor loadings from a 36-month rolling window. $IssuerHasSelfIndexer?$ is a dummy variable that equals one if an issuer offers a self-indexed ETF. $Log(AUM)$ is the natural logarithm of a fund's assets under management. $TurnRatio$ is the yearly portfolio turnover ratio. $Log(Age)$ is the natural logarithm of a fund's age (in years). $Volatility$ is the standard deviation of daily net returns within a month. $IssuerLog(AUM)$ is the natural logarithm of the issuer's assets under management. The sample includes all active equity funds in the CRSP MF database from 2012–2020. Standard errors are two-way clustered by month–issuer interaction.

Panel A: Within-Issuer Front-running Test



Panel B: Performance Regressions

	Net Returns (%)			Net Carhart Alpha (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
$IssuerHasSelfIndexer?$	−0.061**	−0.041**	−0.017	−0.072***	−0.029*	−0.035
	(0.027)	(0.017)	(0.021)	(0.023)	(0.016)	(0.023)
$Log(AUM)$	0.016***	0.013***	0.012***	0.010***	0.011***	0.007***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
$TurnRatio$	−0.004*	−0.004**	0.000	−0.004	−0.004	−0.006*
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
$Log(Age)$	−0.045***	−0.035***	−0.034***	−0.024**	−0.027***	−0.034***
	(0.006)	(0.005)	(0.005)	(0.009)	(0.007)	(0.008)
$Volatility$	3.420***	2.501***	2.360***	−0.673***	−0.530**	−0.417*
	(0.260)	(0.242)	(0.250)	(0.230)	(0.264)	(0.245)
$IssuerLog(AUM)$	0.014***	0.014***	−0.002	0.003	0.004**	−0.018**
	(0.002)	(0.001)	(0.006)	(0.002)	(0.002)	(0.007)
Num. Obs.	493 044	492 988	492 986	286 151	286 100	286 096
R^2	0.74	0.87	0.87	0.12	0.57	0.58
FE: Month	X			X		
FE: Issuer			X			X
FE: Style	X			X		
FE: Month-Style		X	X		X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table B.5. Sample of issuers classified as investment advisors and fund specialists.
We classify issuers as 'Investment Advisors' if our manual search shows they offer a range of financial services including fund management and (wealth) advisory services. 'Fund Specialists' are issuers which focus explicitly on fund management.

Investment Advisors	Fund Specialists	
Advisors Asset Management	AGFiQ	LocalShares
Blackrock	Alpha Architect	M-CAM International
Calamos	AlphaClone	Metaurus Advisors
Cboe Vest Financial	Alps	Motley Fool Asset Management
Charles Schwab	American Century	Nuveen
Deutsche Bank	Amplify	O'Shares
Diamond Hill	Aptus Capital Advisors	Oppenheimer
DWS	Beyond Investing	Pacer Financial
Fidelity	BioShares	Pax World
Goldman Sachs	BNY Mellon Investment Management	ProShares
Guggenheim	Cambria	QuantX Funds
Hartford	Change Finance	Reality Shares
Invesco	Columbia	Recon Capital
Janus Henderson	Compass	Renaissance
John Hancock	Cushing	RevenueShares
JPMorgan	Defiance ETFs	Salt Financial
Legg Mason	Direxion	SerenityShares
Nationwide	Distillate Capital Partners	SL Advisors
Northern Trust	Elkhorn	SP Funds
PIMCO	EntrepreneurShares	Sprott
Point Bridge Capital	ETF Managers Group	Syntax
Principal	ETF Securities	Timothy Plan
Redwood	Exchange Traded Concepts	TriLine Index Solutions
Royal Bank of Scotland	Exponential ETFs	TWM Funds
Russell	Falah Capital	US Commodity Funds
Scottrade	FFCM	USAA
SoFi	First Trust	ValueShares
SSgA	Franklin Templeton Investments	Van Eck
Tortoise	Global Beta Advisors	VelocityShares
Vanguard	Global X	Vident Financial
Victory Capital Management	GraniteShares	Virtus
Wahed Invest	Horizons	WBI
WisdomTree	Hoya Capital	-
-	Impact Shares	-
-	IndexIQ	-
-	Innovation Shares	-
-	Innovator Capital Management	-
-	Inspire Investing	-
-	KraneShares	-
-	Lattice	-

Table B.6. Variable definitions

Variable	Definition
$SelfIndexer?$	Dummy variable that equals one if an ETF is self-indexed. Self-indexed ETFs are ETFs which track an index created and maintained by the ETF issuer themselves. ETF indices obtained from 'primary_benchmark' column in ETF Global.
$NetExp.Ratio(\%)$	Yearly net expense ratio after fee waivers from ETF Global.
$Log(AUM)$	Natural logarithm of an ETF's assets under management from ETF Global. If this information is missing in ETF Global, we use data from the CRSP MF database.
$Log(Age)$	Natural logarithm of fund age, where age of an ETF is defined as the difference between the month-end date and the inception date, divided by 365. The inception date is from ETF Global.
$Log(Holdings)$	Natural logarithm of the unique number of held assets by an ETF, obtained from ETF Global.
$FundFlow$	Increase in total net assets of an ETF not attributable to asset performance, as share of previous period's total net assets (see equation 2.1).
$Volatility$	Standard deviation of daily net (NAV) ETF returns within a calendar month. Daily NAV returns are obtained from the CRSP daily stock file.
$GrossReturn(\%)$	We add $NetExp.Ratio(\%)/252$ to daily net (NAV) returns. Monthly returns are then calculated as the cumulative product of one plus each day's return over a full month, minus one. Daily NAVs are from the CRSP daily stock file.
$AvgNetRet(Yr)(\%)$	12-month rolling average net (NAV) ETF return, calculated using NAVs from the CRSP MF database.
$GrossCarhartAlpha(\%)$	ETF's excess $GrossReturn(\%)$ estimated over a 36-month rolling window using the 4-factor Carhart model.
$SharpeRatio$	Monthly net (NAV) ETF return divided by the $Volatility$, using NAVs from the CRSP MF database.
$Avg.RetCorrtoStyle\text{-}Peers$	Average of all pairwise daily net (NAV) return correlations of a given ETF with all other style-peers in each month. Daily NAV returns are obtained from the CRSP daily stock file and style refers to the column 'focus' in ETF Global.
$CosineSimilarity$	Monthly average cosine similarity of ETF holdings compared to all other ETFs in the same style. Data is obtained from ETF Global where style refers to column 'focus'.
$Inst.SelfOwnership$	Fraction of ETF shares held only by the ETF issuing company. Quarterly holdings are from Thomson/Refinitiv's S34 file.
$SmartBeta?$	Dummy variable that is one if an ETF is classified as smart beta. The smart beta classification is taken from Morningstar Direct.
$Uniqueness$	12-month average of an ETFs absolute difference in $GrossReturn(\%)$ versus the value-weighted average $GrossReturn(\%)$ of all ETFs within the same style category. Style categories are obtained from the 'focus' columns in ETF Global.
$Activeness$	$(1 - R^2)$ from regressing daily gross returns on the Carhart four-factor model.
$TurnoverRatio$	Minimum of aggregate sales or purchases of securities by the ETF, divided by the 12-month average TNA. Obtained from the CRSP MF database.
$MavgSpread$	Monthly average difference between the daily ask and bid, scaled by the price. All variables obtained from the CRSP daily stock file.
$Log(MichSent)$, $MichSentNorm$, $MichSentTophalf$	The natural logarithm, normalized value and dummy for when the the University of Michigan Consumer Sentiment Index is above its median value.
$InvestmentAdvisor?$	Dummy variable which is one if the ETF issuer is manually classified as investment advisor as opposed to a fund specialist.
$IssuerLog(AUM)$	Natural logarithm of all assets under management of an issuer across all asset classes in the CRSP MF database.
$IssuerLog(Age)$	Natural logarithm of the number of days divided by 365 since the first observation of an issuer in the CRSP MF database.
$IssuerNrETFs$	Number of passive equity ETFs operated by an issuer in the CRSP MF database.
$IssuerNrSelfIndexETFs$	Number of passive self-indexed ETFs operated by an issuer. Self-indexed ETFs are defined using data from ETF Global (see above).

Appendix C

Table C.1. Predicting real earnings management when firms are consistently close to the index threshold.
Unexpected cash flow from operations and discretionary expenses are calculated based on Roychowdhury (2006). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $MonthsClose_t$ measures for how many of the past 12 months a firm had $Distance \leq 200$. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q		Q1	
	$UnexpDiscrExp_t$	$UnexpCFO_t$	$UnexpDiscrExp_t$	$UnexpCFO_t$
$MonthsClose_{t-1}$	−0.0002	0.0002	−0.0006	0.0005
	(0.0006)	(0.0001)	(0.0010)	(0.0003)
$0 \leq Distance_{t-1} \leq 50$	−0.0120	−0.0067	−0.0291	−0.0106
	(0.0106)	(0.0057)	(0.0286)	(0.0158)
$51 \leq Distance_{t-1} \leq 100$	0.0013	−0.0015	−0.0093	−0.0087
	(0.0082)	(0.0036)	(0.0192)	(0.0077)
$101 \leq Distance_{t-1} \leq 150$	−0.0062	0.0006	0.0096	0.0080
	(0.0060)	(0.0026)	(0.0157)	(0.0055)
$NumAnalysts_{t-1}$	0.0019***	0.0001	0.0021***	0.0000
	(0.0005)	(0.0001)	(0.0007)	(0.0002)
$TobinQ_{t-1}$	0.0135***	−0.0003	0.0136***	0.0021
	(0.0026)	(0.0011)	(0.0029)	(0.0018)
ROA_{t-1}	−0.9546***	0.2135***	−0.9948***	0.1560***
	(0.0998)	(0.0384)	(0.1155)	(0.0499)
$LogMcap_{t-1}$	−0.0197**	0.0028	−0.0750	0.0153
	(0.0089)	(0.0040)	(0.0460)	(0.0154)
$InstOwn_{t-1}$	0.0228**	0.0020	0.0295*	0.0086*
	(0.0102)	(0.0027)	(0.0165)	(0.0050)
$MonthsClose_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	0.0017	0.0006	0.0026	0.0011
	(0.0012)	(0.0006)	(0.0029)	(0.0015)
$MonthsClose_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	−0.0002	0.0001	0.0000	0.0007
	(0.0011)	(0.0004)	(0.0020)	(0.0007)
$MonthsClose_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	0.0007	−0.0002	−0.0016	−0.0013**
	(0.0009)	(0.0003)	(0.0021)	(0.0006)
Num. Obs.	1356	4179	305	1059
R^2	0.48	0.15	0.49	0.25
FE: Year	X	X	X	X
FE: Fiscal Calendar	X	X	X	X
FE: Industry	X	X	X	X
FE: Quarter	X	X		
FE: REIT Flag	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.2. Predicting discretionary accruals when firms are consistently close to the index threshold. Discretionary accruals are calculated using the modified Jones model from Dechow et al. (1995). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $MonthsClose_t$ measures for how many of the past 12 months a firm had $Distance \leq 200$. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q	Q1
	$DiscrAccr_t$	$DiscrAccr_t$
$MonthsClose_t$	0.0006**	−0.0005
	(0.0003)	(0.0005)
$0 \leq Distance_t \leq 50$	−0.0052	−0.0527**
	(0.0121)	(0.0222)
$51 \leq Distance_t \leq 100$	0.0033	−0.0122
	(0.0065)	(0.0143)
$101 \leq Distance_t \leq 150$	0.0029	−0.0047
	(0.0059)	(0.0142)
$NumAnalyst_t$	−0.0007***	−0.0005
	(0.0002)	(0.0004)
$TobinQ_t$	−0.0032***	−0.0024
	(0.0009)	(0.0016)
ROA_t	0.3015***	0.3311***
	(0.0404)	(0.0641)
$LogMcap_t$	0.0098	−0.0271
	(0.0065)	(0.0276)
$InstOwn_t$	−0.0059	−0.0036
	(0.0041)	(0.0075)
$MonthsClose_t$:$0 \leq Distance_t \leq 50$	0.0006	0.0047**
	(0.0012)	(0.0019)
$MonthsClose_t$:$51 \leq Distance_t \leq 100$	−0.0003	0.0015
	(0.0007)	(0.0014)
$MonthsClose_t$:$101 \leq Distance_t \leq 150$	0.0000	0.0001
	(0.0007)	(0.0018)
Num. Obs.	2470	563
R^2	0.15	0.28
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.3. Predicting managerial forecast error (MFE) when firms are consistently close to the index threshold.
MFE is defined as the quarterly actual eps minus the managerial eps guidance, in percentage terms of the stock price. Resolved forecasts are forecasts for which the guidance and the actual earnings announcement occur before the Russel rank day. Unresolved forecasts are forecasts for which the guidance is before the rank day and the earnings announcement occurs after the rank day. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $MonthsClose_t$ measures for how many of the past 12 months a firm had $Distance \leq 200$. Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	Resolved	Unresolved
	MFE_t	MFE_t
$MonthsClose_{t-1}$	0.0154	0.0082
	(0.0179)	(0.0152)
$0 \leq Distance_{t-1} \leq 50$	1.3743*	0.0443
	(0.8226)	(0.5669)
$51 \leq Distance_{t-1} \leq 100$	−0.2049	−0.7701
	(0.4751)	(0.8238)
$101 \leq Distance_{t-1} \leq 150$	−0.0204	−1.1575*
	(0.3632)	(0.6735)
$DaysUntilAnn_t$	0.0107	−0.0038**
	(0.0092)	(0.0017)
$NumAnalysts_{t-1}$	0.0195	0.0095
	(0.0120)	(0.0086)
$TobinQ_{t-1}$	−0.0513	−0.1316**
	(0.0790)	(0.0570)
$Leverage_{t-1}$	0.6910	0.1533
	(0.5434)	(0.2851)
ROA_{t-1}	6.6907	10.1650*
	(4.8406)	(5.1858)
$LogMcap_{t-1}$	−1.1238	−0.6883
	(1.2541)	(0.6966)
$InstOwn_{t-1}$	−0.1420	−0.4613
	(0.4312)	(0.4243)
$VolatilityQ_{t-1}$	0.7151	−4.6012
	(7.7372)	(5.4691)
$MonthsClose_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	−0.1879*	−0.0325
	(0.1019)	(0.0617)
$MonthsClose_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	−0.0056	0.0822
	(0.0446)	(0.0719)
$MonthsClose_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	−0.0380	0.1030
	(0.0437)	(0.0690)
Num. Obs.	172	133
R^2	0.50	0.60
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	X
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.4. Predicting real earnings management when CEO horizon is short.
Unexpected cash flow from operations and discretionary expenses are calculated based on Roychowdhury (2006). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $CEOHorizon_t$ is the CEO decision horizon as calculated by Antia et al. (2010). Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q		Q1	
	$UnexpDiscrExp_t$	$UnexpCFO_t$	$UnexpDiscrExp_t$	$UnexpCFO_t$
$CEOHorizon_{t-1}$	−0.0010	−0.0005	0.0002	−0.0006
	(0.0018)	(0.0005)	(0.0028)	(0.0009)
$0 \leq Distance_{t-1} \leq 50$	0.0011	0.0025	−0.0245	−0.0050
	(0.0114)	(0.0032)	(0.0305)	(0.0088)
$51 \leq Distance_{t-1} \leq 100$	−0.0007	0.0015	−0.0034	−0.0017
	(0.0105)	(0.0025)	(0.0207)	(0.0056)
$101 \leq Distance_{t-1} \leq 150$	−0.0047	−0.0034	0.0122	−0.0102*
	(0.0090)	(0.0022)	(0.0202)	(0.0055)
$NumAnalysts_{t-1}$	0.0020***	0.0000	0.0019**	0.0000
	(0.0006)	(0.0002)	(0.0009)	(0.0003)
$TobinQ_{t-1}$	0.0140**	0.0046***	0.0170***	0.0067***
	(0.0058)	(0.0014)	(0.0059)	(0.0025)
ROA_{t-1}	−0.8608***	0.0749**	−1.0131***	0.0575*
	(0.1506)	(0.0332)	(0.1891)	(0.0319)
$LogMcap_{t-1}$	−0.0116	0.0004	−0.0492	−0.0069
	(0.0105)	(0.0043)	(0.0510)	(0.0158)
$InstOwn_{t-1}$	0.0303**	0.0058*	0.0247	0.0091
	(0.0140)	(0.0033)	(0.0190)	(0.0057)
$CEOHorizon_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	0.0004	−0.0003	0.0100	0.0023
	(0.0038)	(0.0014)	(0.0094)	(0.0032)
$CEOHorizon_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	0.0009	−0.0008	−0.0007	−0.0010
	(0.0036)	(0.0011)	(0.0058)	(0.0017)
$CEOHorizon_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	0.0015	0.0010	−0.0035	0.0035*
	(0.0031)	(0.0009)	(0.0060)	(0.0020)
Num. Obs.	959	3120	230	833
R^2	0.40	0.13	0.49	0.25
FE: Year	X	X	X	X
FE: Fiscal Calendar	X	X	X	X
FE: Industry	X	X	X	X
FE: Quarter	X	X		
FE: REIT Flag	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.5. Predicting discretionary accruals when CEO horizon is short.
Discretionary accruals are calculated using the modified Jones model from Dechow et al. (1995). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $CEOHorizon_t$ is the CEO decision horizon as calculated by Antia et al. (2010). Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q	Q1
	$DiscrAccr_t$	$DiscrAccr_t$
$CEOHorizon_t$	−0.0005	0.0000
	(0.0009)	(0.0016)
$0 \leq Distance_t \leq 50$	0.0031	−0.0078
	(0.0078)	(0.0158)
$51 \leq Distance_t \leq 100$	0.0058	0.0010
	(0.0059)	(0.0092)
$101 \leq Distance_t \leq 150$	0.0089**	−0.0125
	(0.0043)	(0.0103)
$NumAnalyst_t$	−0.0007***	0.0003
	(0.0002)	(0.0004)
$TobinQ_t$	−0.0082***	−0.0104***
	(0.0014)	(0.0030)
ROA_t	0.3395***	0.4146***
	(0.0502)	(0.0574)
$LogMcap_t$	0.0097	0.0128
	(0.0075)	(0.0284)
$InstOwn_t$	−0.0107**	−0.0102
	(0.0045)	(0.0088)
$CEOHorizon_t$:$0 \leq Distance_t \leq 50$	−0.0007	0.0019
	(0.0028)	(0.0040)
$CEOHorizon_t$:$51 \leq Distance_t \leq 100$	−0.0019	0.0012
	(0.0024)	(0.0029)
$CEOHorizon_t$:$101 \leq Distance_t \leq 150$	−0.0021	0.0063
	(0.0022)	(0.0063)
Num. Obs.	1792	428
R^2	0.17	0.33
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.6. Predicting managerial forecast error (MFE) when CEO horizon is short.
MFE is defined as the quarterly actual eps minus the managerial eps guidance, in percentage terms of the stock price. Resolved forecasts are forecasts for which the guidance and the actual earnings announcement occur before the Russel rank day. Unresolved forecasts are forecasts for which the guidance is before the rank day and the earnings announcement occurs after the rank day. Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $CEOHorizon_t$ is the CEO decision horizon as calculated by Antia et al. (2010). Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	Resolved	Unresolved
	MFE_t	MFE_t
$CEOHorizon_{t-1}$	0.0153	0.0038
	(0.0437)	(0.0547)
$0 \leq Distance_{t-1} \leq 50$	−1.7940**	0.4917
	(0.8324)	(0.9089)
$51 \leq Distance_{t-1} \leq 100$	−0.4798	0.2015
	(0.5196)	(0.7063)
$101 \leq Distance_{t-1} \leq 150$	−0.2283	−0.0596
	(0.3907)	(0.3338)
$DaysUntilAnn_t$	0.0102	−0.0011
	(0.0102)	(0.0037)
$NumAnalysts_{t-1}$	0.0197	0.0071
	(0.0170)	(0.0125)
$TobinQ_{t-1}$	−0.0149	−0.1211
	(0.1127)	(0.0847)
$Leverage_{t-1}$	0.6079	0.2131
	(0.6710)	(0.3860)
ROA_{t-1}	5.0289	8.2495
	(7.0608)	(8.6222)
$LogMcap_{t-1}$	−1.9383	−0.5744
	(1.5998)	(0.7650)
$InstOwn_{t-1}$	0.6012	−0.4387
	(0.4111)	(0.4499)
$VolatilityQ_{t-1}$	6.2405	−11.8231
	(10.3444)	(8.9200)
$CEOHorizon_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	0.4548**	−0.2177
	(0.2107)	(0.2824)
$CEOHorizon_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	−0.0222	0.0031
	(0.1216)	(0.2442)
$CEOHorizon_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	−0.1155	−0.0943
	(0.1254)	(0.1395)
Num. Obs.	152	117
R^2	0.54	0.59
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	X
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.7. Predicting real earnings management when CEOs are poorly compensated.
Unexpected cash flow from operations and discretionary expenses are calculated based on Roychowdhury (2006). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $TotalComp_t$ is the quintile of the total compensation of the firm CEO, compared to industry peers in the given quarter. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q		Q1	
	$UnexpDiscrExp_t$	$UnexpCFO_t$	$UnexpDiscrExp_t$	$UnexpCFO_t$
$TotalComp_{t-1}$	0.0005	−0.0005	0.0026	−0.0002
	(0.0018)	(0.0006)	(0.0035)	(0.0013)
$0 \leq Distance_{t-1} \leq 50$	0.0129	0.0075	0.0187	0.0093
	(0.0120)	(0.0048)	(0.0321)	(0.0097)
$51 \leq Distance_{t-1} \leq 100$	0.0066	−0.0019	0.0006	−0.0067
	(0.0118)	(0.0039)	(0.0221)	(0.0062)
$101 \leq Distance_{t-1} \leq 150$	0.0016	0.0010	−0.0191	0.0048
	(0.0087)	(0.0026)	(0.0233)	(0.0058)
$NumAnalysts_{t-1}$	0.0021***	0.0000	0.0019*	0.0001
	(0.0006)	(0.0002)	(0.0010)	(0.0003)
$TobinQ_{t-1}$	0.0141**	0.0044***	0.0174***	0.0066**
	(0.0057)	(0.0015)	(0.0058)	(0.0027)
ROA_{t-1}	−0.8618***	0.0760**	−1.0033***	0.0586*
	(0.1508)	(0.0334)	(0.1840)	(0.0327)
$LogMcap_{t-1}$	−0.0126	0.0002	−0.0517	−0.0075
	(0.0105)	(0.0043)	(0.0503)	(0.0158)
$InstOwn_{t-1}$	0.0302**	0.0062*	0.0268	0.0089
	(0.0140)	(0.0033)	(0.0190)	(0.0058)
$TotalComp_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	−0.0046	−0.0026*	−0.0074	−0.0043
	(0.0046)	(0.0015)	(0.0095)	(0.0030)
$TotalComp_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	−0.0022	0.0007	−0.0029	0.0013
	(0.0043)	(0.0016)	(0.0069)	(0.0022)
$TotalComp_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	−0.0011	−0.0010	0.0085	−0.0034
	(0.0029)	(0.0009)	(0.0064)	(0.0022)
Num. Obs.	959	3120	230	833
R^2	0.40	0.13	0.50	0.25
FE: Year	X	X	X	X
FE: Fiscal Calendar	X	X	X	X
FE: Industry	X	X	X	X
FE: Quarter	X	X		
FE: REIT Flag	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.8. Predicting discretionary accruals when CEOs are poorly compensated.
Discretionary accruals are calculated using the modified Jones model from Dechow et al. (1995). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $TotalComp_t$ is the quintile of the total compensation of the firm CEO, compared to industry peers in the given quarter. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q	Q1
	$DiscrAccr_t$	$DiscrAccr_t$
$TotalComp_t$	−0.0007	0.0005
	(0.0011)	(0.0019)
$0 \leq Distance_t \leq 50$	−0.0071	−0.0074
	(0.0061)	(0.0134)
$51 \leq Distance_t \leq 100$	−0.0031	0.0005
	(0.0055)	(0.0113)
$101 \leq Distance_t \leq 150$	0.0045	0.0120
	(0.0051)	(0.0158)
$NumAnalyst_t$	−0.0007***	0.0003
	(0.0002)	(0.0003)
$TobinQ_t$	−0.0081***	−0.0101***
	(0.0015)	(0.0033)
ROA_t	0.3394***	0.4131***
	(0.0497)	(0.0549)
$LogMcap_t$	0.0096	0.0133
	(0.0074)	(0.0281)
$InstOwn_t$	−0.0109**	−0.0098
	(0.0045)	(0.0088)
$TotalComp_t$:$0 \leq Distance_t \leq 50$	0.0037	0.0018
	(0.0026)	(0.0046)
$TotalComp_t$:$51 \leq Distance_t \leq 100$	0.0021	0.0012
	(0.0019)	(0.0031)
$TotalComp_t$:$101 \leq Distance_t \leq 150$	0.0000	−0.0043
	(0.0019)	(0.0044)
Num. Obs.	1792	428
R^2	0.16	0.32
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.9. Predicting managerial forecast error (MFE) when CEOs are poorly compensated.
MFE is defined as the quarterly actual eps minus the managerial eps guidance, in percentage terms of the stock price. Resolved forecasts are forecasts for which the guidance and the actual earnings announcement occur before the Russel rank day. Unresolved forecasts are forecasts for which the guidance is before the rank day and the earnings announcement occurs after the rank day. Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $TotalComp_t$ is the quintile of the total compensation of the firm CEO, compared to industry peers in the given quarter. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	Resolved	Unresolved
	MFE_t	MFE_t
$TotalComp_{t-1}$	−0.0446	0.1338**
	(0.1006)	(0.0619)
$0 \leq Distance_{t-1} \leq 50$	−0.5314	0.9598
	(0.9583)	(0.8582)
$51 \leq Distance_{t-1} \leq 100$	−0.7452	−0.4739
	(0.6044)	(0.8555)
$101 \leq Distance_{t-1} \leq 150$	−1.4265*	−0.4470
	(0.8346)	(0.4238)
$DaysUntilAnn_t$	0.0113	0.0032
	(0.0098)	(0.0045)
$NumAnalysts_{t-1}$	0.0191	−0.0022
	(0.0142)	(0.0139)
$TobinQ_{t-1}$	−0.0408	−0.0808
	(0.1071)	(0.0700)
$Leverage_{t-1}$	1.0466*	0.5284
	(0.5920)	(0.4103)
ROA_{t-1}	6.5333	11.9066
	(6.8554)	(9.2936)
$LogMcap_{t-1}$	−1.6421	−0.6480
	(1.6053)	(0.8194)
$InstOwn_{t-1}$	0.6177	−0.1076
	(0.5922)	(0.5681)
$VolatilityQ_{t-1}$	−1.1517	−6.6286
	(8.2260)	(8.1138)
$TotalComp_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	−0.0518	−0.3921
	(0.3176)	(0.3004)
$TotalComp_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	0.1271	0.2448
	(0.1530)	(0.3343)
$TotalComp_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	0.4069*	0.0983
	(0.2348)	(0.1316)
Num. Obs.	152	117
R^2	0.54	0.62
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	X
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.10. Predicting real earnings management when CEOs have high performance-pay. Unexpected cash flow from operations and discretionary expenses are calculated based on Roychowdhury (2006). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $PerfPay_t$ is the quintile of the share of bonus and equity compensation of the firm CEO, compared to industry peers in the given quarter. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q		Q1	
	$UnexpDiscrExp_t$	$UnexpCFO_t$	$UnexpDiscrExp_t$	$UnexpCFO_t$
$PerfPay_{t-1}$	0.0031	0.0002	0.0057	0.0006
	(0.0019)	(0.0005)	(0.0036)	(0.0011)
$0 \leq Distance_{t-1} \leq 50$	0.0038	0.0051	−0.0064	0.0132
	(0.0117)	(0.0044)	(0.0286)	(0.0092)
$51 \leq Distance_{t-1} \leq 100$	0.0066	−0.0003	0.0118	−0.0015
	(0.0122)	(0.0031)	(0.0216)	(0.0053)
$101 \leq Distance_{t-1} \leq 150$	−0.0020	−0.0013	−0.0060	0.0046
	(0.0080)	(0.0023)	(0.0216)	(0.0053)
$NumAnalysts_{t-1}$	0.0019***	0.0000	0.0018*	0.0001
	(0.0006)	(0.0002)	(0.0009)	(0.0003)
$TobinQ_{t-1}$	0.0142**	0.0046***	0.0169***	0.0068***
	(0.0057)	(0.0015)	(0.0057)	(0.0026)
ROA_{t-1}	−0.8443***	0.0752**	−0.9699***	0.0594*
	(0.1465)	(0.0333)	(0.1847)	(0.0324)
$LogMcap_{t-1}$	−0.0124	0.0005	−0.0513	−0.0082
	(0.0104)	(0.0043)	(0.0500)	(0.0158)
$InstOwn_{t-1}$	0.0295**	0.0058*	0.0289	0.0083
	(0.0141)	(0.0033)	(0.0195)	(0.0058)
$PerfPay_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	−0.0007	−0.0014	0.0025	−0.0062**
	(0.0043)	(0.0015)	(0.0088)	(0.0028)
$PerfPay_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	−0.0022	0.0001	−0.0079	−0.0012
	(0.0042)	(0.0014)	(0.0070)	(0.0019)
$PerfPay_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	0.0005	0.0000	0.0031	−0.0035*
	(0.0027)	(0.0009)	(0.0065)	(0.0020)
Num. Obs.	959	3117	230	833
R^2	0.41	0.13	0.50	0.25
FE: Year	X	X	X	X
FE: Fiscal Calendar	X	X	X	X
FE: Industry	X	X	X	X
FE: Quarter	X	X		
FE: REIT Flag	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.11. Predicting discretionary accruals when CEOs have high performance-pay.
Discretionary accruals are calculated using the modified Jones model from Dechow et al. (1995). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $PerfPay_t$ is the quintile of the share of bonus and equity compensation of the firm CEO, compared to industry peers in the given quarter. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q	Q1
	$DiscrAccr_t$	$DiscrAccr_t$
$PerfPay_t$	−0.0008	0.0007
	(0.0010)	(0.0015)
$0 \leq Distance_t \leq 50$	−0.0081	−0.0071
	(0.0059)	(0.0118)
$51 \leq Distance_t \leq 100$	0.0027	0.0084
	(0.0052)	(0.0102)
$101 \leq Distance_t \leq 150$	0.0064	0.0130
	(0.0054)	(0.0165)
$NumAnalyst_t$	−0.0007***	0.0003
	(0.0002)	(0.0003)
$TobinQ_t$	−0.0082***	−0.0104***
	(0.0014)	(0.0032)
ROA_t	0.3390***	0.4122***
	(0.0507)	(0.0554)
$LogMcap_t$	0.0094	0.0148
	(0.0074)	(0.0275)
$InstOwn_t$	−0.0108**	−0.0094
	(0.0045)	(0.0087)
$PerfPay_t$:$0 \leq Distance_t \leq 50$	0.0042*	0.0018
	(0.0024)	(0.0039)
$PerfPay_t$:$51 \leq Distance_t \leq 100$	−0.0005	−0.0021
	(0.0019)	(0.0030)
$PerfPay_t$:$101 \leq Distance_t \leq 150$	−0.0009	−0.0052
	(0.0021)	(0.0057)
Num. Obs.	1789	427
R^2	0.17	0.32
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.12. Predicting managerial forecast error (MFE) when CEOs have high performance-pay. MFE is defined as the quarterly actual eps minus the managerial eps guidance, in percentage terms of the stock price. Resolved forecasts are forecasts for which the guidance and the actual earnings announcement occur before the Russel rank day. Unresolved forecasts are forecasts for which the guidance is before the rank day and the earnings announcement occurs after the rank day. Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $PerfPay_t$ is the quintile of the share of bonus and equity compensation of the firm CEO, compared to industry peers in the given quarter. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	Resolved	Unresolved
	MFE_t	MFE_t
$PerfPay_{t-1}$	−0.0193	0.1433*
	(0.0713)	(0.0804)
$0 \leq Distance_{t-1} \leq 50$	−0.6666	0.8459
	(0.9249)	(0.7211)
$51 \leq Distance_{t-1} \leq 100$	−0.5171	−0.8335
	(0.4492)	(0.8187)
$101 \leq Distance_{t-1} \leq 150$	−1.1543*	−0.8481
	(0.6643)	(0.6333)
$DaysUntilAnn_t$	0.0109	0.0035
	(0.0100)	(0.0045)
$NumAnalysts_{t-1}$	0.0226	0.0000
	(0.0149)	(0.0148)
$TobinQ_{t-1}$	−0.0288	−0.1403
	(0.1062)	(0.0886)
$Leverage_{t-1}$	1.1212*	0.3347
	(0.6196)	(0.3991)
ROA_{t-1}	6.3183	16.9350
	(6.9345)	(10.3415)
$LogMcap_{t-1}$	−1.6852	−0.4901
	(1.6353)	(0.7469)
$InstOwn_{t-1}$	0.5459	0.0724
	(0.5430)	(0.5458)
$VolatilityQ_{t-1}$	0.8966	−10.0656
	(8.5687)	(8.6044)
$PerfPay_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	0.0073	−0.3306
	(0.3376)	(0.2381)
$PerfPay_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	0.0451	0.3649
	(0.1532)	(0.3219)
$PerfPay_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	0.3213	0.2509
	(0.1942)	(0.2248)
Num. Obs.	152	116
R^2	0.53	0.69
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	X
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.13. Predicting real earnings management when CEO stock ownership is low.
Unexpected cash flow from operations and discretionary expenses are calculated based on Roychowdhury (2006). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $CEOOwn$ is the share of total stocks outstanding owned by the firm CEO. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q		Q1	
	$UnexpDiscrExp_t$	$UnexpCFO_t$	$UnexpDiscrExp_t$	$UnexpCFO_t$
$CEOOwn_{t-1}$	0.0012	0.0003	0.0008	−0.0001
	(0.0007)	(0.0002)	(0.0012)	(0.0002)
$0 \leq Distance_{t-1} \leq 50$	0.0001	0.0011	−0.0077	0.0005
	(0.0051)	(0.0026)	(0.0192)	(0.0065)
$51 \leq Distance_{t-1} \leq 100$	−0.0021	−0.0007	−0.0150	−0.0035
	(0.0047)	(0.0018)	(0.0154)	(0.0045)
$101 \leq Distance_{t-1} \leq 150$	−0.0008	−0.0022	0.0074	−0.0032
	(0.0040)	(0.0014)	(0.0125)	(0.0036)
$NumAnalysts_{t-1}$	0.0021***	0.0000	0.0019**	0.0000
	(0.0006)	(0.0002)	(0.0009)	(0.0003)
$TobinQ_{t-1}$	0.0139**	0.0044***	0.0161***	0.0068***
	(0.0054)	(0.0015)	(0.0055)	(0.0026)
ROA_{t-1}	−0.8654***	0.0752**	−0.9938***	0.0612*
	(0.1507)	(0.0327)	(0.1876)	(0.0320)
$LogMcap_{t-1}$	−0.0122	0.0008	−0.0569	−0.0065
	(0.0103)	(0.0044)	(0.0500)	(0.0157)
$InstOwn_{t-1}$	0.0326**	0.0068**	0.0295	0.0084
	(0.0140)	(0.0033)	(0.0190)	(0.0058)
$CEOOwn_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	0.0008	0.0006**	0.0065	−0.0003
	(0.0014)	(0.0003)	(0.0090)	(0.0013)
$CEOOwn_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	0.0023**	0.0005	0.0029*	−0.0002
	(0.0011)	(0.0005)	(0.0015)	(0.0007)
$CEOOwn_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	−0.0006	0.0010**	−0.0155	0.0003
	(0.0009)	(0.0004)	(0.0152)	(0.0012)
Num. Obs.	959	3121	230	834
R^2	0.42	0.14	0.50	0.25
FE: Year	X	X	X	X
FE: Fiscal Calendar	X	X	X	X
FE: Industry	X	X	X	X
FE: Quarter	X	X		
FE: REIT Flag	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.14. Predicting discretionary accruals when CEO stock ownership is low.
Discretionary accruals are calculated using the modified Jones model from Dechow et al. (1995). Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $CEOOwn$ is the share of total stocks outstanding owned by the firm CEO. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q	Q1
	$DiscrAccr_t$	$DiscrAccr_t$
$CEOOwn_t$	0.0010***	0.0038***
	(0.0003)	(0.0009)
$0 \leq Distance_t \leq 50$	0.0066*	0.0031
	(0.0036)	(0.0107)
$51 \leq Distance_t \leq 100$	0.0029	0.0047
	(0.0034)	(0.0079)
$101 \leq Distance_t \leq 150$	0.0090***	0.0053
	(0.0031)	(0.0084)
$NumAnalyst_t$	−0.0007***	0.0003
	(0.0002)	(0.0003)
$TobinQ_t$	−0.0086***	−0.0113***
	(0.0014)	(0.0032)
ROA_t	0.3502***	0.4212***
	(0.0495)	(0.0554)
$LogMcap_t$	0.0101	0.0113
	(0.0074)	(0.0281)
$InstOwn_t$	−0.0121***	−0.0100
	(0.0044)	(0.0087)
$CEOOwn_t$:$0 \leq Distance_t \leq 50$	−0.0046***	−0.0079
	(0.0011)	(0.0048)
$CEOOwn_t$:$51 \leq Distance_t \leq 100$	−0.0006	−0.0010
	(0.0008)	(0.0031)
$CEOOwn_t$:$101 \leq Distance_t \leq 150$	−0.0039**	−0.0041***
	(0.0016)	(0.0011)
Num. Obs.	1792	428
R^2	0.19	0.36
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.15. Predicting managerial forecast error (MFE) when CEO stock ownership is low.
MFE is defined as the quarterly actual eps minus the managerial eps guidance, in percentage terms of the stock price. Resolved forecasts are forecasts for which the guidance and the actual earnings announcement occur before the Russel rank day. Unresolved forecasts are forecasts for which the guidance is before the rank day and the earnings announcement occurs after the rank day. Firms included are members of the Russell 1000 that are less than 300 ranks away from exclusion. $Distance_t$ measures how many market cap ranks a firm would need to lose to drop into the Russell 2000, if the index was reconstituted in quarter t. $CEOOwn$ is the share of total stocks outstanding owned by the firm CEO. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	Resolved	Unresolved
	MFE_t	MFE_t
$CEOOwn_{t-1}$	0.0242	−0.0223
	(0.0527)	(0.0327)
$0 \leq Distance_{t-1} \leq 50$	0.0806	−0.0101
	(0.4670)	(0.3582)
$51 \leq Distance_{t-1} \leq 100$	−0.1717	0.1432
	(0.3188)	(0.3546)
$101 \leq Distance_{t-1} \leq 150$	−0.1914	−0.1999
	(0.2806)	(0.2647)
$DaysUntilAnn_t$	0.0130	−0.0007
	(0.0103)	(0.0035)
$NumAnalysts_{t-1}$	0.0045	0.0070
	(0.0137)	(0.0123)
$TobinQ_{t-1}$	−0.0930	−0.0955
	(0.1014)	(0.0755)
$Leverage_{t-1}$	0.4789	0.4881
	(0.5103)	(0.4422)
ROA_{t-1}	8.9360	7.3079
	(7.1111)	(8.1613)
$LogMcap_{t-1}$	−1.6087	−0.6459
	(1.5329)	(0.8772)
$InstOwn_{t-1}$	0.4406	−0.4545
	(0.4360)	(0.6121)
$VolatilityQ_{t-1}$	−1.6718	−11.0274
	(8.3447)	(9.2379)
$CEOOwn_{t-1}$:$0 \leq Distance_{t-1} \leq 50$	−1.3623***	0.0168
	(0.2884)	(0.0689)
$CEOOwn_{t-1}$:$51 \leq Distance_{t-1} \leq 100$	−0.1823	0.0428
	(0.1745)	(0.0492)
$CEOOwn_{t-1}$:$101 \leq Distance_{t-1} \leq 150$	−0.1006	−0.0334
	(0.0607)	(0.0485)
Num. Obs.	152	117
R^2	0.62	0.58
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	X
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

lvi

Table C.16. Predicting real earnings management for firms near Russell 1000 inclusion. Unexpected cash flow from operations and discretionary expenses are calculated based on Roychowdhury (2006). Firms included are members of the Russell 2000 that are less than 300 ranks away from inclusion. $Distance_t$ measures how many market cap ranks a firm would need to gain to join the Russell 1000, if the index was reconstituted in quarter t. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q		Q1	
	$UnexpDiscrExp_t$	$UnexpCFO_t$	$UnexpDiscrExp_t$	$UnexpCFO_t$
$0 \leq Distance_{t-1} \leq 50$	0.0105*	0.0015	0.0013	−0.0074
	(0.0056)	(0.0024)	(0.0118)	(0.0066)
$51 \leq Distance_{t-1} \leq 100$	0.0020	0.0028	−0.0042	−0.0051
	(0.0038)	(0.0020)	(0.0092)	(0.0049)
$101 \leq Distance_{t-1} \leq 150$	0.0046	0.0014	0.0031	−0.0066*
	(0.0030)	(0.0016)	(0.0069)	(0.0037)
$NumAnalysts_{t-1}$	0.0020***	0.0006***	0.0021***	0.0005
	(0.0005)	(0.0002)	(0.0005)	(0.0003)
$TobinQ_{t-1}$	0.0100***	−0.0002	0.0114***	−0.0009
	(0.0009)	(0.0004)	(0.0010)	(0.0007)
ROA_{t-1}	−0.8543***	0.4940***	−0.8567***	0.3902***
	(0.0850)	(0.0315)	(0.0992)	(0.0663)
$LogMcap_{t-1}$	−0.0493***	−0.0036	−0.0304	0.0285
	(0.0110)	(0.0049)	(0.0317)	(0.0185)
$InstOwn_{t-1}$	0.0166**	−0.0003	0.0103	0.0009
	(0.0074)	(0.0028)	(0.0131)	(0.0067)
Num. Obs.	1975	4815	453	1220
R^2	0.63	0.35	0.65	0.31
FE: Year	X	X	X	X
FE: Fiscal Calendar	X	X	X	X
FE: Industry	X	X	X	X
FE: Quarter	X	X		
FE: REIT Flag	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.17. Predicting discretionary accruals for firms near Russell 1000 inclusion.
Discretionary accruals are calculated using the modified Jones model from Dechow et al. (1995). Firms included are members of the Russell 2000 that are less than 300 ranks away from inclusion. $Distance_t$ measures how many market cap ranks a firm would need to gain to join the Russell 1000, if the index was reconstituted in quarter t. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	ALL Q	Q1
	$DiscrAccr_t$	$DiscrAccr_t$
$0 \leq Distance_t \leq 50$	0.0027	0.0047
	(0.0036)	(0.0115)
$51 \leq Distance_t \leq 100$	−0.0012	0.0023
	(0.0030)	(0.0082)
$101 \leq Distance_t \leq 150$	0.0002	−0.0043
	(0.0029)	(0.0070)
$NumAnalyst_t$	−0.0006**	−0.0009
	(0.0003)	(0.0006)
$TobinQ_t$	0.0007	−0.0009
	(0.0006)	(0.0010)
ROA_t	0.0694*	0.0453
	(0.0373)	(0.0743)
$LogMcap_t$	−0.0056	−0.0077
	(0.0078)	(0.0282)
$InstOwn_t$	0.0029	−0.0024
	(0.0044)	(0.0113)
Num. Obs.	2122	455
R^2	0.06	0.20
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

Table C.18. Predicting managerial forecast error (MFE) for firms near Russell 1000 inclusion.
MFE is defined as the quarterly actual eps minus the managerial eps guidance, in percentage terms of the stock price. Resolved forecasts are forecasts for which the guidance and the actual earnings announcement occur before the Russel rank day. Unresolved forecasts are forecasts for which the guidance is before the rank day and the earnings announcement occurs after the rank day. Firms included are members of the Russell 2000 that are less than 300 ranks away from inclusion. $Distance_t$ measures how many market cap ranks a firm would need to gain to join the Russell 1000, if the index was reconstituted in quarter t. Timeframe: 2007 to 2022. Standard errors are clustered at the firm level and shown in brackets.

	Resolved	Unresolved
	MFE_t	MFE_t
$0 \leq Distance_{t-1} \leq 50$	−0.0972	0.1060
	(0.1194)	(0.1717)
$51 \leq Distance_{t-1} \leq 100$	−0.0283	0.0593
	(0.0885)	(0.1371)
$101 \leq Distance_{t-1} \leq 150$	0.0110	0.1193
	(0.0930)	(0.1013)
$DaysUntilAnn_t$	0.0016	−0.0055
	(0.0017)	(0.0043)
$NumAnalysts_{t-1}$	−0.0066	0.0066
	(0.0062)	(0.0047)
$TobinQ_{t-1}$	0.0001	−0.0084
	(0.0162)	(0.0130)
$Leverage_{t-1}$	0.0936	0.0413
	(0.1507)	(0.0974)
ROA_{t-1}	1.7125**	1.4591
	(0.6867)	(1.0638)
$LogMcap_{t-1}$	0.3594	−0.1989
	(0.3964)	(0.4631)
$InstOwn_{t-1}$	0.0403	0.1912
	(0.0904)	(0.1492)
$VolatilityQ_{t-1}$	3.5795	0.8961
	(2.6955)	(2.9005)
Num. Obs.	176	163
R^2	0.39	0.30
FE: Year	X	X
FE: Fiscal Calendar	X	X
FE: Industry	X	X
FE: Quarter	X	X
FE: REIT Flag	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01