

# **Three Essays on Information and Asset Pricing**

by

Zhengyifan (Victor) Chen

A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of

The University of Manitoba

In partial fulfillment of the requirements of the degree of

**DOCTOR OF PHILOSOPHY**

Department of Accounting and Finance

I.H. Asper School of Business

University of Manitoba

Winnipeg, Manitoba, Canada

Copyright © 2026 by Zhengyifan (Victor) Chen

## Abstract

This thesis includes three essays. In the first essay, we study how racial residential integration (RRI) influences stock price efficiency. We find that communities with higher RRI exhibit more upward intergenerational mobility and a smaller White–Black income gap. Consistent with the idea that more integrated communities facilitate better information diffusion, firms headquartered in these regions demonstrate higher stock price informativeness and lower information asymmetry. Using staggered adoptions of source-of-income laws as exogenous shocks to RRI, we establish a causal relationship. Finally, we show that the superior local stock-picking ability of actively managed mutual funds is weakened for firms located in high RRI communities.

In the second essay, we extend the theoretical model of Han and Yang (2013) to study how racial integration influences asset prices with investors from two racial groups (majority and minority). Greater racial integration across the two groups increases price informativeness. When the number of informed investors is exogenously given, racial integration strengthens cross-group transmission of private signals, thereby incorporating more private information into prices and raising informativeness. When the information acquisition decision is endogenous, integration lowers acquisition costs, strengthens incentives to become informed, mitigates free riding on others' signals, and increases price informativeness. In both settings, higher racial integration lowers the cost of capital, deepens liquidity, and narrows the welfare gap between groups.

The third essay examines how mutual fund stock holdings contain predictive information about important common factors in stock returns. Specifically, we use differences between the holdings attributes of highly-skilled and low-skilled mutual funds to forecast characteristic-related factor returns. For instance, we find that a one-decile increase in our measure forecasts a 30-bp increase in the value factor (HML) return in the following month. The predictive signal is concentrated among funds with greater discretion, such as growth funds. Our measure can be

interpreted as an information variable that aggregates informed portfolio tilts towards forecasting factor premia associated with fundamentals, including value, investment, and profitability.

## **Contribution of Authors**

This thesis is the original work of the author, Zhengyifan (Victor) Chen. The research presented herein was conducted under the supervision of Dr. Lei Lu and co-supervision of Dr. Gady Jacoby. Chapters 2, 3, and 4 are based on collaborative research that has been submitted or accepted for presentation at peer-reviewed venues. The specific contributions of all authors for each of these works are detailed below.

### **Chapter 2: Racial Integration and Price Efficiency**

This chapter is based on an unpublished working paper co-authored with Gady Jacoby, Hao Jiang, Chi Liao, and Lei Lu, accepted for presentation at the 2024 American Finance Association Annual Meeting (all authors are listed in alphabetical order following convention in finance). As the primary author, my contributions included conceiving the research idea, collecting and processing the data, conducting all empirical analyses, and writing the manuscript. My advisor, Lei Lu, offered insightful guidance that helped refine the research direction and strengthen the paper's arguments. Co-advisor Gady Jacoby provided substantive feedback on the methodology and contributed to improving the overall framing of the study. Hao Jiang contributed through constructive discussions and offered valuable comments that sharpened the empirical design. Chi Liao supplied detailed writing instructions and thorough editorial feedback that enhanced the clarity and presentation of the manuscript.

### **Chapter 3: Racial Integration in Social Networks and Asset Pricing Efficiency**

This chapter is based on an unpublished working paper co-authored with Gady Jacoby. As the first author, I was responsible for developing the theoretical model, deriving the equilibrium conditions, conducting numerical simulations, and drafting the manuscript. Gady Jacoby contributed by exploring the initial model setup, engaging in extensive discussions on the theoretical results throughout the project, and providing critical input that shaped the direction of the analysis.

### **Chapter 4: Factor Timing: Evidence from Mutual Fund Skill**

This chapter is based on an unpublished working paper co-authored with Lei Lu, Gady Jacoby, and Yu Xia, accepted for presentation at the 2026 China International Conference in Finance (scheduled). Yu Xia and I jointly originated the core research idea during collaborative research discussions, and in that capacity we share equal intellectual leadership of this project; all authors are listed in alphabetical order following convention in finance. Yu Xia also developed the theoretical model underpinning the central argument of the paper. My responsibilities included designing the empirical framework to test the model's predictions, cleaning and processing the raw data, performing all empirical analyses, and authoring the remaining chapters of the paper. Together, Yu Xia and I wrote the literature review and introduction in an iterative manner. Lei Lu and Gady Jacoby offered valuable comments and suggestions that refined the analysis throughout the project.

## Acknowledgements

I am deeply grateful to my advisor, Dr. Lei Lu, whose mentorship has shaped me into the researcher I am today. His guidance on research methodologies, academic writing, and scholarly collaboration has been invaluable. His detailed feedback during our research meetings pushed me to think more critically, and his thoughtful career counsel and organizational discipline helped me develop the habits essential for rigorous academic work.

I owe profound gratitude to my co-advisor, Dr. Gady Jacoby, whose dedication to both scholarship and students has been truly inspirational. His genuine passion for research, meticulous attention to detail, and unwavering integrity set a standard I continually aspire to reach. Over the past five years, his patient mentorship has been transformative. He believed in my potential when I was uncertain and provided the support that made pursuing this academic path possible.

My thanks extend to Dr. Chi Liao, who provided rigorous training in literature review and academic writing. Her thorough feedback challenged me to engage more deeply with existing research. I am grateful to Dr. Fang Wan for her insightful feedback on each chapter and for her warmth when I first joined the program, which made me feel welcomed during a challenging transition. I would also like to express my appreciation to Yu Xia for his collaboration on the third essay. His constructive feedback and thoughtful insights significantly strengthened the work.

This journey would have been lonelier without my fellow doctoral students: Xingyu Chen, Ruiqi Guan, Linlin Jin, Zhihao Huang, Yang Liu, and Dr. Qi Zhang. Your camaraderie and mutual support sustained me through the challenges of doctoral work.

Finally, I acknowledge myself for the perseverance to continue when progress seemed impossible and for choosing to remain on this path despite countless obstacles. Completing this dissertation required resilience I did not know I possessed.

# Table of Contents

Abstract.....	i
Contribution of Authors.....	iii
Acknowledgements .....	v
Table of Contents.....	vi
List of Tables .....	viii
List of Figures.....	ix
Chapter 1: General Introduction .....	1
Chapter 2: Racial Integration and Price Efficiency .....	4
2.1 Introduction.....	4
2.1.1 Motivation and findings .....	4
2.1.2 Contributions .....	10
2.2 Data and Variables .....	12
2.2.1 Measure of racial residential integration.....	12
2.2.2 Upward social mobility and the income gap .....	14
2.2.3 Firm-level data and variables .....	16
2.2.4 Mutual fund-level variables.....	18
2.2.5 Geographic control variables .....	20
2.3 RRI and Economic Integration .....	21
2.4 RRI, Price Informativeness, and Information Asymmetry .....	24
2.4.1 Stock price informativeness .....	24
2.4.2 Firm-level information asymmetry .....	26
2.4.3 Heterogeneity analysis .....	27
2.4.4 Endogeneity.....	28
2.4.5 Robustness.....	34
2.5 RRI and Mutual Funds' Local Investment.....	35
2.5.1 RRI, local bias, and local stock-picking ability .....	35
2.5.2 Heterogeneity analysis .....	37
2.5.3 Robustness.....	39
2.6 Conclusion .....	40
Appendix to Chapter 2 .....	54
Appendix 2A.1: Variable definitions .....	54
Appendix 2A.2: Supplementary information .....	58
Appendix 2A.3: Full tables and extra tests .....	65
Appendix 2A.4: Extra robustness tests .....	85
Chapter 3: Racial Integration in Social Networks and Asset Pricing Efficiency .....	103
3.1 Introduction.....	103
3.2 Model .....	107
3.2.1 Economy and agents.....	108
3.2.2 Information structure and heterogeneous diffusion .....	109

3.2.3 Price-implied signal extraction.....	111
3.3 Financial Market Equilibrium under Exogenous Private Signal Acquiring Design...	112
3.3.1 Information structure and heterogeneous diffusion .....	115
3.3.2 Cost of equity and racial integration .....	117
3.3.3 Market liquidity and racial integration.....	118
3.3.4 Trading volume, welfare, and racial integration .....	120
3.4 Financial Market Equilibrium under Endogenous Information Acquisition .....	124
3.4.1 Asymmetric information acquisition cost and acquisition decision.....	125
3.4.2 Price informativeness, cost of equity, and liquidity .....	128
3.4.3 Trading volume and welfare analysis .....	130
3.5 Conclusion .....	132
Appendix to Chapter 3 .....	136
Appendix 3A.1: Derivation for Proposition 1.....	136
Appendix 3A.2: Derivation for trading volumes .....	139
Chapter 4: Factor Timing: Evidence from Mutual Fund Skill .....	144
4.1 Introduction.....	144
4.2 Data .....	149
4.2.1 Mutual fund sample.....	149
4.2.2 Mutual fund skill measures .....	151
4.2.3 Factor returns.....	153
4.3 Fund Skill Gap .....	153
4.3.1 Construct FSG using factor exposures and characteristics .....	153
4.3.2 Skill Aggregation .....	155
4.4 Predictability of FSG on Factor Returns.....	157
4.4.1 Empirical framework.....	157
4.4.2 Long-short characteristics portfolio predictive regressions .....	159
4.4.3 Mechanism: predictability within different fund mandates .....	165
4.4.4 Robustness issues .....	167
4.5 Conclusion .....	171
Appendix to Chapter 4 .....	187
Appendix 4A.1: Time–series comovement between factor returns and FSG.....	187
Appendix 4A.2: Detailed regression results of baseline results.....	188
Appendix 4A.3: Alternative FSG constructions .....	189
Chapter 5: General Conclusion.....	193
Reference List.....	194

## List of Tables

Table 2.1: Summary Statistics of Community-, Stock-, and Fund-Level Variables.....	43
Table 2.2: RRI, Social Mobility, and the Mean White–Black Income Gap.....	46
Table 2.3: RRI, Stock Price Informativeness, and Information Asymmetry .....	47
Table 2.4: Heterogeneity in RRI’s Effect on Information Environment.....	48
Table 2.5: Instrumental Variable Regression .....	50
Table 2.6: Difference-In-Differences Regression .....	51
Table 2.7: RRI, Local Bias, and Local Stock-Picking Ability .....	52
Table 2.8: RRI and Local Stock-Picking: Economic Conditions and Manager Race.....	53
Table 2A.2.1 Classification for Race and Ethnicity from the US Census Bureau .....	58
Table 2A.3.1: RRI and the Social Connectedness Index.....	67
Table 2A.3.2: Complete Version of Table 2.3 .....	68
Table 2A.3.3: Complete Version of Table 2.4 .....	72
Table 2A.3.4: Extended Version of Table 2.5 and Extra Tests.....	77
Table 2A.3.5: Extended Version of Table 2.6.....	81
Table 2A.3.6: Complete Versions of Table 2.7 and Table 2.8.....	83
Table 2A.4.1: Alternative Community Definitions and Annual ACS Results.....	88
Table 2A.4.2: Results of Alternative Evenness Measure .....	96
Table 2A.4.3: Placebo Test and Other Robustness Checks at the Fund Level .....	101
Table 4.1: Number of Funds by Style (1991 – 2023).....	175
Table 4.2: Summary of Mutual Fund Skill Measures .....	176
Table 4.3: Summary Statistics.....	177
Table 4.4: FSG and Future Factor Return Correlations (Unconditional).....	178
Table 4.5: FSG Predictive Regressions with Controls .....	179
Table 4.6: PCA/PLS Components of FSG Signals and Factor Return Predictability .....	180
Table 4.7: Results for Style Funds with Factor-Exposure-Based FSG .....	181
Table 4.8: FSG Predictability Among Growth Funds .....	182
Table 4.9: FSG Predictability Among Cap-Targeted Funds .....	183
Table 4.10: Average Skill Ranks by Fund-Flow Terciles .....	184
Table 4.11: FSG Predictability Excluding Extreme Flow Funds.....	185
Table 4.12: Results of the Exposure Turnover-Filtered Subsample .....	186
Table 4A.1: Factor-Exposure-Based FSG Regression .....	188
Table 4A.2: Factor Return Predictability Using Daily-Return Beta-Based FSG.....	191
Table 4A.3: Factor Return Predictability Using Monthly-Return Beta-Based FSG .....	192

## List of Figures

Figure 2.1: Distribution of Mutual Funds in the United States .....	42
Figure 2A.1: Correlation Analysis for Demographic Variables .....	63
Figure 3.1: Volume, Welfare Effects of Racial Integration .....	134
Figure 3.2: Racial Integration and Price Efficiency, Liquidity, and Cost of Equity .....	135
Figure 4.1: Prediction Timeline.....	173
Figure 4.2: HML Factor Returns and Lagged Average-Skill FSGs.....	174
Figure 4A.1: Comovement of Factor Returns with Lagged Average-Skill FSG .....	187

## Chapter 1: General Introduction

This dissertation comprises five chapters: a general introduction, three essays, and a general conclusion. The dissertation is organized around a central question in asset pricing: how does information produced and transmitted by people and institutions embedded in real communities and markets become reflected in prices? I address this question through two complementary perspectives. The first examines how social context shapes the production and diffusion of firm-relevant information, thereby affecting price efficiency. The second investigates how institutional investors' portfolio choices transmit information into expected returns through factor exposures. Together, these perspectives provide a coherent framework for understanding the interplay between social structure, institutional behavior, and market outcomes.

In Chapter 2, I investigate how racial residential integration (RRI) influences stock price efficiency. I propose that more racially integrated communities promote broader information diffusion, diminishing the informational edge of particular groups and yielding more informationally efficient prices. Using commuting zones as the unit of community, I find that firms in high-RRI areas have stock prices that better anticipate future earnings and display lower information asymmetry, with stronger effects for firms facing greater informational frictions. I mitigate endogeneity through an instrumental variable leveraging staggered adoptions of source-of-income non-discrimination laws and a difference-in-differences design exploiting headquarters relocations. I further demonstrate that higher RRI corresponds to diminished alpha from local investments by mutual funds, driven by White fund managers, consistent with an information-leveling channel.

Chapter 3 constructs a theoretical framework analyzing how racial integration shapes asset pricing efficiency and investor welfare. I incorporate a racially stratified islands topology into a noisy rational expectations equilibrium setting, where the extent of racial integration determines

signal transmission quality and information acquisition costs. The model reveals that integration improves price informativeness by releasing diverse signals previously trapped within demographic clusters, reduces the cost of soft information acquisition for minority investors (boosting participation and deepening liquidity), and closes the structural welfare gap by weakening informational privilege among majority investors. Chapters 2 and 3 jointly offer empirical and theoretical foundations for understanding how social structure influences information production in financial markets.

Chapter 4 turns to the transmission of systematic risk factor information through institutional investors' portfolio decisions. I propose the fund skill gap (FSG), measured as the difference in stock holding characteristics between high-skill and low-skill mutual funds, and test whether it predicts common factor returns. I show that FSG anticipates future factor returns, most notably for the value factor, with meaningful but smaller effects for investment and profitability factors. This predictability is driven by growth-oriented funds whose mandates allow greater benchmark deviations, supporting a channel rooted in managerial discretion. A battery of robustness tests confirms that skilled managers' portfolio tilts capture time-varying views on conditional expected factor premia.

The narrative arc across these three chapters is clear and cumulative. Chapters 2 and 3 focus on how social context—specifically, racial integration—shapes the production and diffusion of firm-specific soft information, which is transmitted through interpersonal networks and reflected in stock price efficiency. Chapter 4 examines how institutional investors' portfolio choices transmit information about systematic risk factors through their holdings, revealing skilled managers' forward-looking assessments of conditional expected returns. Together, these studies provide a comprehensive understanding of the mechanisms through which information, whether firm-specific or factor-related, becomes embedded in asset prices. They highlight the importance of social structure in facilitating information flow, the role of institutional investors as information

intermediaries, and the distributional consequences of informational frictions in financial markets.

# Chapter 2: Racial Integration and Price Efficiency

## 2.1 Introduction

### 2.1.1 Motivation and findings

To what extent is a firm’s stock price efficiency shaped by the community in which it is located? While prior research has documented that diverse social interactions among residents influence individual and community-level economic outcomes (e.g., Alesina, Baqir, and Easterly, 1999; Alesina and La Ferrara, 2000; Chetty et al., 2022a,b), much less is known about how such interactions affect the informational efficiency of firms located in these communities. This study fills this gap by focusing on a key feature of American communities: the level of racial residential integration (RRI), defined as the extent to which individuals of different racial groups are evenly distributed across geographic space (Trounstine, 2016).<sup>1</sup> We examine whether and how this integration affects the stock price efficiency of firms located in these communities.

We first document that communities with higher RRI show greater economic integration along two key dimensions: higher upward social mobility and narrower White–Black income disparities. Using data from the Opportunity Atlas (Chetty et al., 2018, 2024), we find that in more racially integrated commuting zones, children from low-income households are more likely to move up to higher income ranks, and the average income rank gap between White and Black children is significantly smaller.<sup>2</sup>

---

<sup>1</sup> Consistent with U.S. census guidelines, we categorize White, Black, Asian, and American Indian as racial groups, and recognize Hispanic as an ethnicity (U.S. Census Bureau, 2024). This classification follows the conventional distinction between “race,” which is associated with physical characteristics, and “ethnicity,” which includes aspects such as national origin, culture, language, and religious beliefs. For simplicity, we use the term “race” throughout the remainder of this dissertation to refer to both race and ethnicity. The detailed classification and racial compositions can be found in Appendix 2B.1.

<sup>2</sup> Commuting zones, developed by the Economic Research Service at the U.S. Department of Agriculture, are geographic units defined by commuting patterns that reflect where individuals live and work (Tolbert and Sizer, 1996). Designed to capture economic interdependence across county boundaries, commuting zones often span areas where residents routinely commute for employment or business. Their boundaries are shaped by economic linkages, which may extend across multiple counties or even include non-contiguous areas. Thus, a commuting zone can be understood

Consistent with the idea that more economically and socially integrated communities facilitate better information diffusion, we ask: How is a community's level of RRI related to the stock price efficiency of firms located within it? To address this question, we draw on two insights from the literature on information diffusion and learning in social networks. First, greater social interaction leads to denser social networks within a community, which facilitate the transmission of information among individuals (e.g., Jackson and Rogers, 2007; Rajkumar, Saint-Jacques, Bojinov, Brynjolfsson, and Aral, 2022). Second, in more segregated communities, individuals are more likely to interact with others who share similar attributes such as race and ethnicity (e.g., McPherson, Smith-Lovin, and Cook, 2001). This tendency toward homophily leads to modular social networks, where information circulates within groups but often remains confined to them (e.g., Hinz and Spann, 2008; Golub and Jackson, 2012; Halberstam and Knight, 2016). In contrast, more racially integrated communities exhibit lower social network modularity, allowing information to diffuse more efficiently across the entire social network (Reagans and McEvily, 2003; Hinz and Spann, 2008; Golub and Jackson, 2012; Furutani, Shibahara, Akiyama, and Aida, 2023).

In financial markets, broader diffusion of information across a wider population can reduce the informational advantages of select groups, such as corporate insiders or individuals with privileged social ties, by allowing a larger set of market participants to access and act on value-relevant signals. As a result, stock prices may be more efficient for firms located in communities with higher RRI. Hence, we hypothesize that firms located in communities with higher RRI tend to have greater stock price efficiency. To empirically test this hypothesis, we use commuting zones, geographically and economically integrated areas that span multiple counties, as our primary unit of community.

---

as a cluster of economically connected counties.

To test this prediction, we examine the relationship between RRI and stock price informativeness, measured by the correlation between stock prices and future firm earnings (e.g., Grossman and Stiglitz, 1980; Kacperczyk, Sundaresan, and Wang, 2021). Since corporate headquarters are central to information exchange (Davis and Henderson, 2008), we assign firms to communities based on the location of their headquarters. Our empirical analysis shows that firms headquartered in high-RRI communities have stock prices that more strongly predict future earnings, consistent with our hypothesis.

We conjecture that information about individual firms transmitted through social networks is more likely to take the form of qualitative, subjective, and relatively difficult-to-quantify soft information (Liberti and Petersen, 2019; Dyer, Köchling, and Limbach, 2024; Nguyen and Nguyen, 2025). Such information is relevant for firm valuation and complements public information disseminated through corporate disclosures and analyst coverage (Liberti and Petersen, 2019). Therefore, we hypothesize that the positive effect of RRI on price informativeness is more pronounced for smaller firms (with lower total assets), less visible firms (with lower analyst coverage), and younger firms (with lower firm age). Our empirical results support these conjectures.

Next, we examine whether firms headquartered in communities with higher RRI exhibit lower information asymmetry among stock market participants. While price informativeness and information asymmetry are conceptually distinct, they are closely related. In general, the more accurately a security's price reflects its fundamental value, the lower the expected loss for an uninformed investor trading that security. Thus, price informativeness and information asymmetry tend to be inversely related, with both greater price informativeness and reduced information asymmetry reflecting improvements in a firm's information environment.<sup>3</sup> In our setting, broader information diffusion resulting from greater racial integration lowers the cost of information

---

<sup>3</sup> For intuition and formal analysis, see, e.g., Grossman and Stiglitz (1976, 1980).

acquisition, reducing the advantages of otherwise informationally privileged traders. As a result, information asymmetry in the stock market also declines.

We use two proxies to measure information asymmetry: the adverse selection component of the bid–ask spread and the volume coefficient of variation (Lof and van Bommel, 2023). Consistent with the notion that higher RRI promotes stronger community integration and more efficient information diffusion, we find that both measures of information asymmetry decline significantly for firms located in communities with higher RRI. Similar to our findings on price informativeness, this effect is more pronounced for smaller, less visible, and younger firms. Taken together, these results show that higher RRI improves the informational efficiency of firms located in the community.

One potential concern is that a community's racial integration may be correlated with its future growth, confounding the relationships between RRI and both price informativeness and information asymmetry. To address this, we construct an instrumental variable for RRI using the staggered adoptions of source-of-income (SOI) non-discrimination laws across cities. These laws promote integration by prohibiting landlords from rejecting tenants based on their use of housing vouchers, thereby expanding access to opportunity-rich neighborhoods for low-income and minority households. Using a newly compiled dataset from the Urban Institute, we identify 58 city-level adoptions of SOI laws between 1971 and 2021 and construct the instrument as the share of cities within a commuting zone that have adopted such laws. Weak instrument tests confirm its relevance, and we control for key drivers of SOI adoption to mitigate omitted variable bias. Using this instrument, we find that higher RRI is associated with more informative stock prices and lower firm-level information asymmetry. Details of this identification strategy are provided in Section 2.4.4.1.

Second, we also examine the impact of corporate headquarters relocations on stock price efficiency. Using a difference-in-differences framework restricted to firms that relocated once

during the sample period, and a matching approach based on pretreatment characteristics, we find that firms moving from higher- to lower-RRI commuting zones experience a decline in stock price informativeness and an increase in the volume coefficient of variation, a proxy for information asymmetry. The opposite holds for relocations from lower- to higher-RRI zones. While the sample sizes are modest, the results across both outcomes provide additional support for the hypothesis that local RRI improves stock price efficiency.

To explore the mechanism through which RRI affects stock price efficiency, we examine its influence on the local investment behavior and performance of actively managed mutual funds, a potential channel through which RRI improves firms' information environments. Prior literature has documented a strong link between geographic proximity to corporate headquarters and mutual fund performance. For instance, Coval and Moskowitz (1999) find that mutual fund managers exhibit a strong preference for investing in locally headquartered firms, likely driven by their attempts to exploit information asymmetry between local and nonlocal investors. In a follow-up work, Coval and Moskowitz (2001) show that fund managers earn superior returns on local investments. They attribute this outperformance to improved monitoring abilities or privileged access to private information.

Our analysis shows that higher RRI is significantly associated with a decline in the alpha generated from mutual funds' local investments, which is consistent with a diminished informational advantage in communities with higher RRI. Turning to local bias, we find that higher RRI is insignificantly related to mutual funds' tendency to overweight local stocks relative to the market portfolio. This result shows that the diminished informational advantage does not discourage mutual fund managers from over-weighting local stocks in their portfolios, suggesting that informational advantages are not the sole driver of fund managers' revealed preference for local stocks.

To further examine how RRI weakens local informational advantages, we test for

heterogeneity along two dimensions. First, we examine the performance of local mutual fund investments across the business cycle. Previous literature shows that stock picking is more valuable during economic downturns, which translates to a larger fund alpha (Moskowitz, 2000). Consistent with this evidence, we find that the negative effect of RRI on the performance of local investments by fund managers is particularly pronounced during recessions. Second, we examine whether the effect of RRI on local performance varies by fund manager race. We find that the negative effect of higher RRI on local stock-picking performance is concentrated among funds managed by White managers and is statistically insignificant for funds managed by minority managers. This asymmetry aligns with sociology literature showing that White male networks provide greater access to local information and social capital (McDonald, Lin, and Ao, 2009; McDonald and Day, 2010; McDonald, 2011).

To ensure the robustness of our findings, we perform several additional tests. First, we replicate the analyses using alternative definitions of local communities, including the county in which the firm or fund headquarters is located, the surrounding counties within a 40-mile radius of the headquarters (i.e., local regions), and the headquarters' home state. Second, to address concerns about the use of interpolated RRI measures from decennial census data, we construct an annual, state-level RRI measure using the American Community Survey (ACS), which is conducted annually.<sup>4</sup> Overall, the results support the relationships between RRI and local stock-picking performance, as well as the associations between RRI, stock price informativeness, and firm-level information asymmetry.<sup>5</sup>

---

<sup>4</sup> We conduct the test at the state level because the ACS annual data lack census-tract detail needed to construct commuting zone RRI measures (see Section 2.4.5).

<sup>5</sup> For fund-level analysis, our baseline results remain robust when we (i) exclude funds headquartered in cluster states (California, Massachusetts, New York), (ii) control for past local and nonlocal stock-picking performance, and (iii) include fund fixed effects. Taken together, these findings support the robustness of our main conclusions at the fund level as detailed in the appendix of this chapter.

### 2.1.2 Contributions

Our main contribution is to establish a causal link between community structure and the stock price efficiency of local firms, thereby connecting two previously separate literatures. Community structure, understood as patterns of connections across salient social boundaries, has been shown to shape individual and collective outcomes (Alesina, Baqir, and Easterly, 1999; Alesina and La Ferrara, 2000; Chetty et al., 2022a). RRI changes this structure by increasing cross-racial proximity and routine contact, thereby strengthening cross-group connections (Sigelman, Bledsoe, Welch, and Combs, 1996; Mouw and Entwisle, 2006) and shifting community structure toward greater between-group connectivity. Prior research shows that RRI is associated with higher intergenerational mobility (Chetty, Hendren, Kline, and Saez, 2014; Chyn, Haggag, and Stuart, 2022) and improved outcomes in education (Ananat, 2011), job access (Kain, 2004), public safety (Ananat, 2011), and public goods provision (Trounstine, 2016). Yet, the implications of racial integration for financial markets, and in particular for the price efficiency of local firms, remain underexplored.

The existing literature shows that price efficiency improves when informed investors and analysts generate and disseminate information (Grossman and Stiglitz, 1980; Kelly and Ljungqvist, 2007; Boehmer and Kelley, 2009; Saffi and Sigurdsson, 2011; Kacperczyk, Sundaresan, and Wang, 2021) and when disclosure, monitoring, and fair market rules mitigate information asymmetry (Diamond and Verrecchia, 1991; Leuz and Verrecchia, 2000; Bhattacharya and Daouk, 2002; Boehmer, Jones, and Zhang, 2013; Boone and White, 2015). More recent studies highlight the importance of a firm's structural position within national social networks. Greater network centrality accelerates price adjustments to news (Hirshleifer, Peng, and Wang, 2025), whereas broader connectedness across counties can amplify collective biases and attention cascades that generate mispricing (Bali, Hirshleifer, Peng, Tang, and Wang, 2021). We provide causal evidence that firms located in more racially integrated communities exhibit lower information asymmetry

and greater price informativeness. These findings indicate that racial integration improves the diffusion of value-relevant information while mitigating collective biases and mispricing, thereby connecting community sociology to fundamental questions in financial economics.

Second, we contribute to the growing literature on how social networks shape institutional investor outcomes. Prior studies have examined the impacts of educational ties (e.g., Cohen, Frazzini, and Malloy, 2008), neighborhood interactions (e.g., Hong, Kubik, and Stein, 2005; Pool, Stoffman, and Yonker, 2015), and social networks (e.g., Ahern, 2017; Kuchler, Li, Peng, Stroebel, and Zhou, 2022) on institutional portfolio decisions and performance. Research also shows that geographic proximity to corporate headquarters provides mutual funds with a local informational advantage, resulting in higher abnormal returns from nearby investments (Coval and Moskowitz, 1999, 2001; Baik, Kang, and Kim, 2010; Bernile, Korniotis, Kumar, and Wang, 2015). Building on these findings, we examine how racial integration within networks influences the diffusion of information and, in turn, affects institutional investors' local informational advantage. Prior studies have shown that social network resources (i.e., social capital) and information flows differ across racial and ethnic lines (Bertrand, Luttmer, and Mullainathan, 2000; Fernandez and Fernandez-Mateo, 2006; Agarwal, Choi, He, and Sing, 2019; Pedulla and Pager, 2019). Our findings extend this literature by showing that race-based residential patterns influence institutional investors' local informational advantage, and that this effect varies depending on the fund manager's race.

Lastly, we contribute to the literature on the role of race in the mutual fund and hedge fund industries. Prior studies reveal the disadvantages faced by fund managers from racial and ethnic minority groups, often driven by name-based stereotypes or in-group bias (e.g., Kumar, Niessen-Ruenzi, and Spalt, 2015; Han, Huang, Kadan, and Wu, 2021; Lu, Naik, and Teo, 2022). More recently, Agarwal, Jiang, Luo, and Zou (2025) show that rising anti-East Asian sentiment in 2020 and 2021 negatively affected mutual fund performance and stock-picking ability for East Asian female managers. We extend this literature by showing that racial integration, a sociological

dimension of race, also influences mutual fund stock-picking performance. Specifically, we find that minority fund managers do not experience the same local informational advantage as their White counterparts in low-RRI communities, an additional disadvantage faced by minority managers in the mutual fund industry.

The remainder of this chapter is organized as follows. Section 2.2 describes the data and key measures. Section 2.3 examines the relationship between RRI and economic integration. Section 2.4 presents firm-level evidence on the relationship between RRI and the firm's information environments. Section 2.5 presents mutual fund-level results on RRI and local stock selection. Section 2.6 concludes. The Appendix provides detailed variable definitions, supplementary tests, and robustness checks.

## **2.2 Data and Variables**

### **2.2.1 Measure of racial residential integration**

Racial residential integration (RRI) refers to the extent to which racial groups are evenly distributed across geographic space (Trounstine, 2016).<sup>6</sup> The opposite of integration, segregation, can be measured along several dimensions, including evenness, exposure, clustering, concentration, and centralization (Massey and Denton, 1988). Among these, evenness is considered the most important dimension of integration, based on principal component analysis (Massey and Denton, 1988). Our analysis focuses on this dimension.

---

<sup>6</sup> The sociology and economics literature have identified four main factors that influence patterns of RRI (e.g., Charles, 2003; Lichter, Parisi, and Taquino, 2015; Ambrose, Conklin, and Lopez, 2021; Deroncourt, 2022). First, labor market discrimination, often reflected by income inequality, leads to lower earnings for minority groups, thus constraining their residential choice. Second, race-based in-group preferences may lead individuals to select neighborhoods inhabited by members of their own racial group. Third, institutionalized discrimination, embedded in exclusionary zoning practices and racially biased pricing and recommendations by real estate agents, further prevents integration. Finally, exogenous shocks, such as the Great Migration, industrialization, and urbanization, have historically shaped and continue to influence patterns of RRI.

We measure evenness using the entropy index of Theil (1972), which captures both the distribution of population groups across neighborhoods and their relative sizes. This nuance allows the index to better reflect residential patterns that support diverse and meaningful social interactions (Trounstone, 2016).<sup>7</sup> Entropy, often used to measure uncertainty, randomness, or diversity, here reflects racial diversity within a geographic unit.

Considering a racial minority group  $r$  and White residents in census tract  $i$ , we measure racial diversity using entropy, defined as:  $E_{r,i} = -P_i^r \times \log\left(\frac{1}{P_i^r}\right) + P_i^w \times \log\left(\frac{1}{P_i^w}\right)$ , where  $P_i^r$  and  $P_i^w$  denote the proportions of racial minority  $r$  and White individuals in census tract  $i$ , respectively, at time  $t$ , such that  $P_i^w = 1 - P_i^r$ .<sup>8</sup> Based on tract-level entropy, we construct a community-level racial integration index as the negative of the Theil index (Theil, 1972), so that higher values reflect greater integration.

$$RRINegTheil_{r,j} = -1 \times \sum_{i \in j} \frac{N_{(r+w),i}}{N_{(r+w),j}} \times \frac{(E_{r,i} - E_{r,j})}{E_{r,j}}, \quad (1)$$

where  $N_{(r+w),i}$  and  $N_{(r+w),j}$  denote the total populations of racial minority  $r$  and White individuals in census tract  $i$  and commuting zone  $j$ , respectively.<sup>9</sup> To incorporate the three major racial

---

<sup>7</sup> The dissimilarity index is a widely used measure of evenness, indicating the proportion of a racial group that would need to relocate between neighborhoods to match the overall racial composition of the broader community (Trounstone, 2016). However, it overlooks important aspects of social network formation. Notably, it does not account for subunit population size, which can affect the density and connectivity of social networks. For robustness, we show that our findings are unchanged when using the dissimilarity index.

<sup>8</sup>  $E_{r,i}$  reaches its maximum of 0.69 when the two groups are equally represented ( $P_i^r = P_i^w = 0.5$ ), indicating the highest level of two-group diversity, and falls to zero when only one group is present.

<sup>9</sup> Following established guidance in the literature, we exclude commuting zones where the combined Black, Asian, and Hispanic populations account for less than 5% of the total residents at any point in time, as very small minority shares can distort measures of racial residential segregation. This filter criterion enhances the reliability of the RRI measure. Our results remain robust when using a higher threshold of 10% or when including all commuting zones without exclusions; these alternative specifications are available upon request. The 5% threshold exclusion affects fewer than 2.5% of observations in both the firm-level and fund-level analyses (see Appendix 2B.2 for details). When the population of either the minority group or White individuals is zero in a census tract, we follow Lichter, Parisi, and Taquino (2015) and Trounstone (2016) in setting  $\log(P_i^{r/w})$  to zero. We exclude commuting zones that contain only a single census tract.

minority groups (i.e., Asian, African American, and Hispanic), we calculate a population-weighted negative Theil index for each commuting zone:

$$RRINegTheil_j = \sum_r \frac{N_{r,j}}{N_{m,j}} \times RRINegTheil_{r,j}, \quad (2)$$

where  $N_{R,j}$  denotes the total population of the three major minority groups of commuting zone  $j$ , and  $\frac{N_{r,j}}{N_{m,j}}$  represents the share of minority group  $r$  in the commuting zone, relative to the combined three major minority populations.<sup>10</sup> This weighted index ranges from  $-1$  to  $0$ , with higher values indicating greater racial integration.

We construct the commuting zone RRI index using data from the 2000, 2010, and 2020 U.S. Decennial Census of Population and Housing.<sup>11</sup> Annual values for 2000–2020 are estimated through linear interpolation. Panel A of Table 2.1 reports summary statistics.

[Insert Table 2.1 Here]

## 2.2.2 Upward social mobility and the income gap

We examine the relationship between RRI and economic outcomes from two perspectives: upward social mobility and the mean racial income gap, both measured at the commuting zone level. Following Chetty et al. (2014, 2024), we define upward mobility as the mean income rank in adulthood of children whose parents were at the 25th percentile of the national income distribution.<sup>12</sup> This metric captures the average economic advancement of children from low-

---

<sup>10</sup> As a robustness check, we also examine alternative versions of the negative Theil index. Specifically, we compute entropy by dividing the population into non-White and White groups, and by using the full racial composition rather than a population-weighted average of three two-group measures, following the approach of Trounstein (2016). Our main results remain robust for these alternative measures of racial integration.

<sup>11</sup> The 2000 and 2010 census data are publicly accessed and can be processed using the US Census Bureau's API and the R package "totalcensus." Data from the 2020 census are available at: <https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.html#P1>.

<sup>12</sup> The commuting zone-level measure is calculated using the following formula:  $UpwardMobility_j = E(y_{j,i} | x_{j,i} = 25th)$ , where  $y_{j,i}$  is the adult income rank of child  $i$  (based on mean earnings at age 27) and  $x_{j,i}$  denotes their parental income rank, both from commuting zone  $j$ . In our analysis, we use published individual-level mean income rank

income families, with higher values indicating greater upward mobility. To capture racial disparities, we define the mean income gap as the difference in predicted adult income ranks between Black and White children from low-income households within a commuting zone.

To construct our sample, we merge our 2000 RRI index with data from the Opportunity Atlas, developed by Chetty et al. (2024). Based on administrative records for 57 million children born between 1978 and 1992, the Opportunity Atlas documents intergenerational mobility over time and across geographic areas, with significant variation by race and socioeconomic status. To protect privacy, the dataset reports mean income ranks at age 27, calculated relative to the national income distribution of each birth cohort and measured at both the commuting zone and county levels.<sup>13</sup> These estimates are disaggregated by race, gender, and parental income percentile (i.e., 1st, 25th, 50th, 75th, and 100th). Notably, percentiles are based on the national (not race-specific) income distribution, meaning that families of different races at the same percentile have the same income level (Chyn, Haggag, and Stuart, 2022).

We restrict our sample to children born in 1983–1992 to ensure that each individual spent at least part of their childhood (under age 18) in a commuting zone after the year 2000. For example, children born in 1983 turned 18 in 2001, ensuring exposure to local environments consistent with the 2000 RRI measure. To improve the reliability of our estimates, we limit the sample to commuting zones with at least 100 Black and 100 White children. The final sample includes 6,070 commuting zone cohort observations for income ranks and 3,035 observations for the average

---

predictions reported separately for White and Black children from the same commuting zone. The explanation of this mean predicted income rank can be found from Chyn, Haggag, and Stuart (2022), in which they use the metropolitan-level data from the Opportunity Atlas.

<sup>13</sup> The construction of commuting zone-level income ranks for children is detailed in Chetty et al. (2024); a brief summary is provided here for interpretative clarity. Children are grouped by birth cohort (1978–1992), childhood commuting zone (based on tax return addresses), race, and parental income percentile. Parental income is defined as the average household income when children in the household are ages 13 to 17 and is ranked within the national income distribution for the corresponding cohort. Children’s adult income rank at age 27 is estimated using locally weighted regressions (LOWESS) of child income rank on parental income rank within each county-race-cohort group. Each child is weighted by the proportion of childhood spent in that county. These county-level estimates are then aggregated to the commuting zone-level using population weights.

White–Black income gap, calculated as the difference in predicted adult income ranks between the two groups, across 320 commuting zones. Panel A of Table 2.1 presents summary statistics for average income ranks across different racial groups and income gaps.

### **2.2.3 Firm-level data and variables**

#### *2.2.3.1 Stock price informativeness*

To examine how RRI affects stock price informativeness, we follow the approach of Kacperczyk et al. (2021), who measure the strength of the relationship between market valuation and future firm fundamentals. Specifically, we measure informativeness as the conditional correlation between a firm's market valuation in year  $t$ , proxied by the log of the market capitalization-to-asset ratio, and its future cash flow in year  $t + 1$ , proxied by EBIT scaled by total assets. We then evaluate the impact of RRI by estimating its effect on this conditional correlation.

We link firm-level data to RRI and other geographical controls using historical headquarters' zip codes, integrating the CRSP/Compustat Merged Database with zip codes extracted from SEC filings.<sup>14</sup> We then map firms to commuting zones using the ZIPCODE dataset from the SAS software. Our sample includes all Compustat firms with fiscal year-ends between 2000 and 2018. We restrict the sample period to ensure that future cash flow data for year  $t + 1$  are observed before the onset of the COVID-19 pandemic in early 2020. We exclude financial firms (those with SIC codes 60–69), firms with market capitalization below \$1 million, and observations with missing data. The final sample consists of 23,972 firm-year observations from 2,479 unique firms across 165 commuting zones. To mitigate the influence of outliers, all variables are winsorized at the 1st and 99th percentiles. Summary statistics for this sample are presented in panel B of Table 2.1.

---

<sup>14</sup> The detailed procedure can be found in Appendix 2B.3.

### 2.2.3.2 Information asymmetry

To evaluate information asymmetry at the firm level, we use two proxies. The first is the adverse selection component of the bid-ask spread, which compensates market makers for the risk of trading with informed traders (George, Kaul, and Nimalendran, 1991). This measure isolates the portion of the spread attributable to private information by excluding order-processing costs (George et al., 1991; Jiang and Sun, 2014).<sup>15</sup> Higher values indicate greater information asymmetry. The second measure is the volume coefficient of variation (VCV) proposed by Lof and van Bommel (2023), which is the standard deviation of trading turnover divided by its mean. The VCV increases with the proportion of informed trades and offers a microstructure-based indicator of asymmetric information.<sup>16</sup>

As in the stock price informativeness analysis, we link firm-level data to commuting zone RRI and other commuting zone-level controls using historical headquarters zip codes. Our sample consists of all publicly listed U.S. firms available in CRSP from 2000 Q1 to 2020 Q1. We exclude financial and utility firms, as well as firms with missing covariates (Amiram, Owens, and Rozenbaum, 2016; Cui, Jo, and Na, 2018). Firms with fewer than four quarterly observations (i.e., less than one year of data) are also excluded. To mitigate the influence of penny stocks, we drop firm-quarters where the average daily stock price in the current or previous quarter is below \$5 (Amihud, 2002). After applying these filters, the final sample for the VCV measure consists of 147,191 firm-quarter observations from 5,498 firms across 185 commuting zones between 2000 Q1 and 2020 Q1. Due to the limited availability of transaction-level data, the adverse selection

---

<sup>15</sup> We construct the adverse selection measure using the following equation:  $Adverseselection = S - 2 \times \sqrt{-cov(RD_t, RD_{t-1})}$ , where  $S$  denotes the proportional quoted spread, and  $RD_t$  is the difference between returns computed using last-trade transaction prices and the midpoint of bid-ask prices at close on day  $t$ . Further details on the construction of this variable are available in Jiang and Sun (2014).

<sup>16</sup> We compute the volume coefficient of variation (VCV) using the equation:  $VCV = \sigma_t^{turnover} / \mu_t^{turnover}$ , where  $\sigma_t^{turnover}$  is the standard deviation and  $\mu_t^{turnover}$  is the mean of daily trading turnover (defined as the daily shares traded divided by shares outstanding of that day) in quarter  $t$ . Additional details can be found in Lof and van Bommel (2023), who show through a microstructure model that the VCV increases with a higher proportion of informed trading.

spread is available for a subsample of 29,088 firm-quarters for 4,550 firms across 179 zones, a limitation commonly noted in the literature (Jiang and Sun, 2014). All variables are winsorized at the 1st and 99th percentiles to mitigate the impact of outliers. Panel C of Table 2.1 presents summary statistics for the two information asymmetry measures, which are comparable to those reported in prior studies (Cui et al., 2018).

#### 2.2.4 Mutual fund-level variables

Our primary fund-level dependent variable is local stock-picking ability, which captures a fund's skill in identifying and overweighting nearby stocks with superior expected returns. This concept builds on the idea that geographic proximity offers informational advantages to investors (Coval and Moskowitz, 2001). We construct this measure using the approach of Kacperczyk, Nieuwerburgh, and Veldkamp (2014), defined as:

$$LSP_t^k = \sum_{i=1}^{N_{local}^k} (w_{i,t}^k - w_{i,t}^m) (R_{t+1}^i - \beta_{i,t} R_{t+1}^m), \quad (3)$$

where  $LSP_t^k$  denotes the local stock-picking ability of mutual fund  $k$  in quarter  $t$ , and  $N_{local}^k$  is the number of local stocks held by fund  $k$ . We define local holdings as stocks headquartered in the same commuting zone as the fund's headquarters.  $w_{i,t}^k$  and  $w_{i,t}^m$  represent the portfolio weights of stock  $i$  in the fund and the market, respectively, in quarter  $t$ ;  $R_{t+1}^i$  and  $R_{t+1}^m$  are the excess returns of stock  $i$  and the market in quarter  $t + 1$ , respectively; and  $\beta_{i,t}$  is the CAPM beta of stock  $i$  estimated using returns from quarters  $t - 20$  to  $t - 1$ .<sup>17</sup> This measure captures whether a fund systematically overweightes local stocks that outperform the market, reflecting its ability to extract

---

<sup>17</sup> Unlike Kacperczyk, Nieuwerburgh, and Veldkamp (2014), who estimate CAPM betas using returns from the past 12 months, we use 20 quarters of historical data to ensure reliability when using the Fama-French three-factor model, which includes additional regressors beyond the CAPM. This requirement retains 96% of the 12,659 stocks in our holdings sample.

value from geographically proximate information.

As a complementary outcome, we also examine local bias, which measures the extent to which a fund overweights local firms regardless of their future performance. Higher values indicate a stronger preference for local stocks relative to their market portfolio weight. Following Wei and Zhang (2020), we define local bias as:

$$LB_t^k = \sum_{i=1}^{N_{local}^k} (w_{i,t}^k - w_{i,t}^m). \quad (4)$$

To assign mutual funds to commuting zones and link them with RRI values, we use fund headquarter zip codes available in the CRSP Mutual Fund Database starting in 2000; thus, our sample period begins in Q1 2000. Each fund is mapped to a commuting zone using SAS's zip-to-commuting zone conversion, and the corresponding RRI value is assigned based on that location.

To construct the mutual fund sample, we merge fund attributes from the CRSP mutual fund database with portfolio holdings data from the Thomson Reuters Institutional (13F) Holdings database (Jiang and Zheng, 2018).<sup>18</sup> We include only actively managed U.S. domestic mutual funds, as identified by their investment objectives in CRSP, with at least 80% of assets invested in common stocks. Following Kacperczyk et al. (2014), we exclude funds with annual turnover ratios below 5%, total net assets under \$5 million, fewer than 10 stock holdings, or any missing variables.<sup>19</sup> Funds with fewer than four quarterly observations over the sample period are also excluded. Some firms in fund portfolios are dropped due to missing geographic information or

---

<sup>18</sup> Data from the Thomson Reuters Holding database are reported at the fund-portfolio level, while fund characteristics from the CRSP mutual fund database are reported at the share-class level. As a result, we aggregate CRSP fund characteristics to the fund-portfolio level for portfolios with multiple share classes, using the asset-weighted average of the share classes, following the method of Kacperczyk, Nieuwerburgh, and Veldkamp (2014). To construct quarterly excess returns, we compound monthly returns from CRSP and match them with monthly risk factor data from the Fama-French portfolios and Factors database.

<sup>19</sup> This approach also helps mitigate incubation bias, where fund families may seed multiple small funds and report only those with favorable performance (Kacperczyk, Nieuwerburgh, and Veldkamp, 2014).

stock price data, which can reduce the representativeness of our portfolio-level stock-picking ability measure. To address this, we retain only fund-quarter observations where the remaining portfolio value accounts for at least 75% of the original portfolio value prior to these exclusions, which results in a 10% reduction in the sample, with the average retained portfolio value close to 90% of the original.<sup>20</sup> The final sample includes 49,350 fund-quarter observations from 1,845 funds across 64 commuting zones.

[Insert Figure 2.1 Here]

Figure 1 shows the geographic distribution of mutual funds, with darker states indicating higher fund densities. Summary statistics for these funds are presented in panel D of Table 2.1 and are broadly consistent with prior studies (e.g., Wardlaw, 2020). The average local stock-picking ability is close to zero (−1.8 basis points). Although mutual funds' local advantages have been well documented by the literature (Coval and Moskowitz, 2001), Bernile, Kumar, Sulaeman, and Wang (2019) find that these advantages largely disappeared for institutional investors after 2000, which coincides with the start of our sample period.

### **2.2.5 Geographic control variables**

In all regressions, we include time-varying commuting zone (CZ) geographic controls: the natural log of population, median housing value, median income, population density (scaled by 10,000 persons/km<sup>2</sup>), education attainment, unemployment rate, share of residents aged 65 or older, female share (Christoffersen and Sarkissian, 2009; Shu, Sulaeman, and Yeung, 2012; Wei and Zhang, 2020), CZ-level racial diversity, shares of major racial groups (White, African American, Asian, Hispanic) (Trounstine, 2016), a regional social-trust index (Wei and Zhang, 2020), and a proxy for institutional discrimination—zoning redundancy (Lichter, Parisi, and Taquino, 2015).

---

<sup>20</sup> Our results are robust to a more lenient 50% threshold, which excludes only 588 records (0.70% of the initial sample).

For firm and fund-level tests, we additionally control for local stock-market conditions, including total market capitalization, number of listed firms, dividend-payer share, value-weighted return of locally headquartered firms, Amihud illiquidity (Bernile et al., 2015), and institutional-investor clustering, proxied by CZ-level institutional ownership (Kacperczyk et al., 2021; Kim, Wang, and Wang, 2022).

Data are obtained from the U.S. Census Bureau, the U.S. Department of Agriculture, the Bureau of Labor Statistics, the Federal Housing Finance Agency, the World Values Survey, CRSP, and COMPSUTAT. Summary statistics are reported in panel A of Table 2.1. Detailed variable construction and inclusion rationales are provided in Appendix 2A.2.4. Appendix 2A.2.5 reports bivariate correlations between RRI and key demographic variables, underscoring the importance of these controls.

### **2.3 RRI and Economic Integration**

To assess whether RRI fosters broader economic integration, we examine its relationship with the two key dimensions of economic integration: upward social mobility and the mean racial income gap. Following Chetty et al. (2014, 2024), we define upward mobility as the mean adult income rank of children whose parents were at the 25th percentile of the national income distribution in each commuting zone, capturing how much children from low-income families advance economically, with higher ranks indicating greater mobility and local opportunity. We define the mean racial income gap as the difference in mean adult income ranks between Black and White children from low-income households. Together, these two measures capture distinct aspects of economic integration. High upward mobility signals broader access to opportunities for low-income families, while a narrower mean racial income gap reflects more equal access to opportunity across racial groups. Both outcomes point to a more inclusive and efficient economy

that allocates opportunity based on potential rather than background.

We first examine the effect of RRI on intergenerational mobility among children from low-income families using the following regression:

$$\begin{aligned}
 IR_{j,r,s} = & \alpha + \beta_1 RRI_j + \beta_2 \frac{s - 1983}{9} + \beta_3 W_{j,s} + \beta_4 RRI_j \cdot \frac{s - 1983}{9} + \beta_5 W_{j,s} \cdot \frac{s - 1983}{9} \\
 & + \sum_{b=1983}^{1992} \delta_b \mathbb{1}(s = b) \cdot \mathbf{X}_j + \epsilon_{j,s},
 \end{aligned} \tag{5}$$

where  $IR_{j,r,s}$  denotes the mean adult income rank of children from low-income families in commuting zone  $j$ , racial group  $r$  (i.e., White or Black), and birth cohort  $s \in [1983, 1992]$ .

$RRI_j$  measures RRI for commuting zone  $j$  in the year 2000, and  $W_{j,s}$  is a binary indicator equal to 1 if the racial group is White for birth cohort  $s$  in commuting zone  $j$ , and 0 otherwise. Following Chetty et al. (2024), we normalize cohort trends by dividing  $(s-1983)$  by 9.<sup>21</sup> Vector  $\mathbf{X}_j$  includes commuting zone controls described in Section 2.2.5 (e.g., education levels, unemployment ratio), and  $\mathbb{1}(s = b)$  denotes birth cohort fixed effects.<sup>22</sup> Standard errors are two-way clustered by commuting zone and birth cohort to account for intra-group correlation, and all variables are winsorized at the 1st and 99th percentiles. The coefficient  $\beta_1$  captures the average relationship between RRI and community-level income rank (i.e., a proxy for upward social mobility) across all cohorts, while  $\beta_4$  reflects how this relationship varies across birth cohorts. Together,  $\beta_1 + \beta_4 \frac{s-1983}{9}$  allows us to trace the cohort-specific effect of RRI over time. The inclusion of  $W_{j,s}$  and its interaction with cohort trends ( $\beta_3, \beta_5$ ) is intended to control for the effect of racial identity on

---

<sup>21</sup> This approach captures cohort fixed effects while allowing the coefficients of interaction terms,  $RRI_j \cdot ((s - 1983)/9)$  or  $W_{j,s} \cdot ((s - 1983)/9)$ , to be interpreted as the incremental effect across different cohorts. As a result, the coefficients on  $RRI_j$  and  $W_{j,s}$  denote the average effects of these variables on outcome variables over a nine-year period (1983-1992).

<sup>22</sup> Since demographic variables usually change gradually, we assume that commuting zone characteristics from the year 2000 can proxy for earlier periods. Given that each cohort is exposed to these characteristics differently, we adopt this structure to capture cohort-specific effects while maintaining a consistent baseline for comparison.

mean income rank.<sup>23</sup>

Estimation results are presented in Table 2.2. Column (1) shows that a higher RRI is significantly associated with improved upward mobility: the coefficient on  $RRI_j$  is 0.034 and statistically significant at the 5% level. The interaction term  $RRI_j \cdot \frac{s-1983}{9}$  is not statistically significant, suggesting that the effect is stable across cohorts. Economically, a one-standard-deviation increase in RRI (0.10) is associated with a 0.34% increase in mean income rank. Second, we confirm the presence of a persistent mean White–Black income gap. Controlling for commuting zone characteristics and cohort trends, the average gap remains at 0.06, with an average decline of 0.03 across cohorts. These patterns are consistent with Chetty et al. (2024) and motivate further analysis of how RRI affects racial disparities in mean predicted income rank. To this end, we estimate the following regression:

$$IG_{j,s} = \alpha + \beta_1 RRI_j + \beta_2 \frac{s - 1983}{9} + \beta_3 RRI_j \cdot \frac{s - 1983}{9} + \sum_{b=1983}^{1992} \delta_b \mathbb{1}(s = b) \cdot \mathbf{X}_j + \epsilon_{j,s}, \quad (6)$$

where  $IG_{i,s}$  denotes the mean income rank gap between White and Black children from low-income families in commuting zone  $j$  and cohort  $s$ . All other variables and model specifications are the same as in Equation (5). The coefficient  $\beta_1$  captures the average association between RRI and the mean White–Black income gap across all birth cohorts in the sample. The interaction term  $\beta_3$  reflects how this relationship evolves across birth cohorts. A negative  $\beta_1$  indicates that a higher commuting zone RRI is associated with a smaller mean racial income gap.

Column (2) in Table 2.2 shows that a higher RRI is significantly associated with a narrower mean White–Black income gap: the coefficient on  $RRI_j$  is  $-0.045$  and statistically significant. As

---

<sup>23</sup> As a side note, the inclusion of  $W_{j,s}$  confirms the existence of a persistent mean White-Black income gap and how this gap evolves across birth cohorts. Since our data are a subsample of the Opportunity Atlas dataset used by Chetty et al. (2024), these findings confirm that the well-documented racial disparity is also evident in our commuting zone level analysis for cohorts born in 1983-1992. Establishing this disparity lays the foundations for our subsequent analysis of how RRI affects racial disparities in mobility.

in the previous model, the interaction term is not significant, indicating that the effect is stable across cohorts. Economically, a one-standard-deviation increase in RRI (0.10) corresponds to a 9.38% reduction in the mean income gap, relative to its mean of 0.05.

Taken together, our findings suggest that RRI is positively related to economic integration. RRI is associated with greater upward mobility among low-income children and a narrower mean racial income gap, both of which point to a more equitable distribution of opportunity and access to information.

[Insert Table 2.2 Here]

As an additional test, we examine whether higher RRI is associated with greater local social connectedness, measured by a within commuting zone Social Connectedness Index (SCI).<sup>24</sup> Although the SCI (Bailey, Cao, Kuchler, Stroebel, and Wong, 2018) does not identify interracial ties, increases in cross-racial friendships should raise overall connectedness by relaxing homophily constraints. The positive RRI SCI relation thus provides suggestive evidence that RRI improves the connectivity of the local social network. Details on variable construction and results are reported in Appendix 2A.3.1.

## **2.4 RRI, Price Informativeness, and Information Asymmetry**

### **2.4.1 Stock price informativeness**

We have hypothesized that RRI is associated with lower information asymmetry among investors. When fewer investors have privileged access to information, price formation tends to be more efficient. Thus, if RRI improves the transparency of a firm's information environment, we should observe not only higher price informativeness but also reduced information asymmetry.

---

<sup>24</sup> As defined by Bailey et al. (2018), the SCI captures inter-county social connections. In our study, we apply this measure to within-commuting zone (CZ) connectedness by focusing on linkages among counties in the same CZ.

This section tests these two predictions, explores cross-sectional heterogeneity in the effects, addresses endogeneity concerns, and presents a series of robustness checks.

We begin by testing whether RRI is positively associated with stock price informativeness. Following Kacperczyk et al. (2021), we measure stock price informativeness by evaluating how strongly a firm's market valuation in year  $t$  (proxied by the log of the market capitalization-to-asset ratio) predicts its future cash flow in year  $t + 1$  (proxied by EBIT scaled by total assets). RRI enters the model as a moderator of this relationship. Specifically, we estimate the following regression model:

$$\frac{E_{i,t+1}}{A_{i,t}} = \alpha + \beta_1 \log(M/A)_{i,t} + \beta_2 RRI_{i,t} + \beta_3 \log(M/A)_{i,t} \times RRI_{i,t} + \beta_4 \mathbf{X}_{i,t} + \beta_5 \mathbf{X}_{j,t} + FE_{firm} + FE_{year} + \epsilon_{i,t+1}, \quad (7)$$

where  $E_{i,t+1}/A_{i,t}$  denotes the ratio of firm  $i$ 's EBIT in year  $t + 1$  to total assets in year  $t$ , and  $\log(M/A)_{i,t}$  is the log of firm  $i$ 's market valuation in year  $t$ .  $RRI_{i,t}$  is the level of racial integration in firm  $i$ 's commuting zone in quarter  $t$ . To isolate the effect of RRI, the firm-level control vector  $\mathbf{X}_{i,t}$  includes firm characteristics known to predict expected cash flows following Kacperczyk et al. (2021) and for board racial diversity following Bernile, Bhagwat, and Yonker (2018). The details can be found in Appendix 2A.2.3. The commuting zone-level control vector  $\mathbf{X}_{j,t}$ , described in Section 2.2.5, includes local demographic and socioeconomic characteristics, as well as stock market conditions. All variables are winsorized at the 1st and 99th percentiles.  $FE_{firm}$  and  $FE_{year}$  denote firm and year fixed effects, respectively, controlling for unobserved heterogeneity across firms and over time. Standard errors are clustered at the commuting zone level.

Our primary variable of interest is the interaction term  $\beta_3$ , which captures whether the effect of market valuation on future fundamentals varies with RRI. A positive and statistically significant  $\beta_3$  indicates that higher RRI strengthens the relationship, reflecting improved price efficiency. The expression  $\beta_1 + \beta_3 RRI$  captures how the link between valuation and future cash flows varies with

the level of RRI in a firm's local community. In columns (1)–(3) of Table 2.3, the coefficients on the interaction between valuation and RRI ( $\beta_3$ ) are positive and statistically significant across all model specifications. To evaluate whether this effect is confounded by other forms of diversity, column (3) includes interactions with board and community racial diversity, and the results are reported in Appendix 2A.3; the RRI interaction ( $\beta_3$ ) remains positive and robust. These findings suggest that firms headquartered in more racially integrated communities show a stronger alignment between market valuations and future performance, consistent with greater stock price informativeness, supporting our hypothesis that RRI improves stock price informativeness by enhancing the local information environment.

[Insert Table 2.3 Here]

#### 2.4.2 Firm-level information asymmetry

To test whether RRI is negatively associated with firm-level information asymmetry, based on the idea that broader information access reduces the advantage of informed investors and improves price formation, we estimate the following model:

$$IA_{i,t} = \beta_0 + \beta_1 RRI_{i,t} + \beta_2 X_{i,t} + \beta_3 X_{j,t} + FE_{\text{year} \times \text{quarter}} + FE_{\text{industry}} + FE_{\text{cz}} + \varepsilon_{i,t}, \quad (8)$$

where  $IA_{i,t}$  denotes information asymmetry for firm  $i$  in quarter  $t$ , measured using either the adverse selection component of the bid-ask spread or the volume coefficient of variation (VCV).  $RRI_{i,t}$  is the level of racial integration in the commuting zone of firm  $i$ 's headquarters in quarter  $t$ . Firm-level control variables,  $\mathbf{X}_{i,t}$ , including size, liquidity, and trading activity (Full list can be found in Appendix 2A.2.6), are included alongside commuting zone characteristics  $\mathbf{X}_{j,t}$  (Section 2.2.5). To address concerns about extreme VCV values in low-volume stocks, we also control for the inverse of total quarterly trading volume. All variables are winsorized at the 1st and 99th percentiles. Year-quarter fixed effects control for macroeconomic shocks, while industry and commuting zone fixed effects account for cross-sectional variation in information asymmetry

across sectors and communities. Standard errors are clustered by commuting zone.

Columns (4)–(9) of Table 2.3 show that RRI is negatively associated with both measures of information asymmetry. Economically, after including all controls, a one-standard-deviation increase in RRI (0.10) is associated with a 51.22% decrease in adverse selection (in column (6)) relative to its mean of 0.41, and a 9.73% decline in the VCV (column (9)) relative to its mean of 0.85. These results suggest that firms headquartered in more racially integrated communities face lower information asymmetry, reinforcing the view that racial integration improves the transparency of the local information environment.

In summary, higher levels of RRI are associated with more informative stock prices and lower firm-level information asymmetry. Firms headquartered in high-RRI communities show a stronger relationship between market valuation and future earnings. RRI is also negatively related to adverse selection and, to a lesser extent, the volume coefficient of variation, suggesting that improved local social networks enhance the quality of information reflected in stock prices. Overall, these findings indicate that higher RRI contributes to the improvement of firms' information environments.

### **2.4.3 Heterogeneity analysis**

When formal disclosure is limited, investors place greater emphasis on alternative, informal channels of information, including those facilitated by social networks (Liberti and Petersen, 2019; Hirshleifer et al., 2025).<sup>25</sup> Smaller, younger, and less visible firms often lack an extensive earnings history or analyst scrutiny and thus are less represented in hard information sources. Therefore, the diffusion of soft information enabled by higher RRI should disproportionately improve the information quality of these firms by mitigating information frictions.

---

<sup>25</sup> RRI enhances information environments by facilitating broader diffusion of firm-specific insights through social networks. This diffusion is particularly important for soft information—qualitative, subjective, or difficult-to-verify knowledge—because such information typically spreads via interpersonal interactions (Liberti and Petersen, 2019). In contrast, hard information, such as firm fundamentals and research generated by sell-side analysts, is more easily accessible through formal channels.

Based on this reasoning, we hypothesize that the positive effects of RRI on stock price informativeness and the reduction in information asymmetry should be stronger for smaller firms (lower total assets), less visible firms (lower analyst coverage), and younger firms (lower firm age). To test this, we sort firms annually (for price informativeness tests) and quarterly (for information asymmetry tests) into high and low groups based on the median value of each characteristic. We then replicate the baseline regressions on each subsample using the same model specifications as in Sections 2.4.1 and 2.4.2.

Table 2.4 presents the results. Consistent with our hypothesis, we find that the impact of RRI on both stock price informativeness and information asymmetry is more pronounced for smaller, less visible, and younger firms. These patterns align with prior evidence showing that social networks are especially valuable for information diffusion in settings where hard information is scarce, and firms are more opaque (Hirshleifer et al., 2025; Kuchler et al., 2022). Overall, the findings support our conjecture that RRI enhances access to value-relevant soft information, thereby contributing to more efficient pricing, particularly for firms with limited formal visibility.

[Insert Table 2.4 Here]

#### **2.4.4 Endogeneity**

To establish a causal link between RRI and firm-level price efficiency, we address endogeneity concerns arising from omitted variables and reverse causality. RRI may be correlated with a community's future economic development or unobserved local factors that also shape firms' information environments. Reverse causality is also a concern if firms with more informative stock prices or lower information asymmetry actively support local integration through targeted community investments. To strengthen causal identification, we adopt two complementary strategies. First, we use the staggered adoption of source of income (SOI) non-discrimination laws as an instrumental variable to capture exogenous variation in RRI. Second, we implement a

difference in differences design based on firm relocations across commuting zones with differing levels of RRI. These approaches help isolate variation in RRI and allow us to assess its causal impact on stock price informativeness and information asymmetry.

#### *2.4.4.1 Instrumental variable*

We instrument RRI with the staggered adoption of SOI non-discrimination laws, which expand housing access for low-income, often minority, households by preventing landlords from rejecting tenants on the basis of their use of housing vouchers. This facilitates residential integration and relocation to higher opportunity neighborhoods. This instrument is grounded in the structure of the Housing Choice Voucher (HCV) program, the largest rental assistance initiative in the United States, which was established in 1974 under Section 8 of the Housing and Community Development Act and currently serves over two million households. By subsidizing a portion of rent for low-income families, disproportionately from minority groups, the program improves residential mobility and reduces exposure to concentrated poverty (Turner, 2003; Sard, 2008). In the absence of SOI protections, however, landlords can legally refuse voucher holders, limiting integration. Prior studies show that prohibiting such discrimination improves access to higher-quality neighborhoods, better schools, and safer environments (Galvez, 2010; McClure, 2010). We instrument RRI using the percentage of cities in the commuting zone that have adopted SOI laws explicitly protecting HCV holders.<sup>26</sup>

To be valid, an instrumental variable (IV) must satisfy two key conditions: it must be relevant (i.e., strongly correlated with the endogenous variable—RRI in our case) and exogenous (i.e.,

---

<sup>26</sup> SOI laws can be enacted at the city, county, or state level. As documented by Teles and Su (2022), at least 20 states and over 100 local jurisdictions had adopted some form of SOI protection by 2021, often with overlapping coverage. While state and federal laws set broad policy frameworks, cities are better positioned to design targeted interventions that reflect local conditions and can be implemented through their own administrative systems to address specific housing challenges (Cho and Lucio, 2025). However, the scope and enforceability of these laws vary widely. Many statutes provide only general protection, leaving it to courts or administrative agencies to determine whether voucher holders are covered. To address these concerns and capture concrete policy shifts relevant to residential integration, we focus on 58 city level SOI law adoptions that explicitly name HCV holders as protected.

uncorrelated with the error term in the outcome equation conditionally on the other covariates—in other words, it should affect the outcome only through its effect on RRI) (Wooldridge, 2010). The first condition is supported by our weak instrument tests, which show that the share of cities in a commuting zone that adopted the law strongly predicts RRI. To assess exogeneity, we discuss whether the instrument could influence the firm-level information environment through channels other than RRI. Theoretically, the SOI non-discrimination laws specifically target landlord acceptance of vouchers and do not directly regulate or financially affect firms. And there is no clear theoretical pathway linking SOI adoption to firm-level metrics. Therefore, the share of adopting cities in a commuting zone should not affect firm outcomes except through changes in residential integration.

However, concerns remain that an unobserved driver of SOI adoption may also influence the firm's information environment, introducing endogeneity. To address this, we draw on recent evidence from Cho and Lucio (2025), who identify three factors that predict SOI adoption but are not already captured by our baseline controls: state-level policy diffusion, local voucher demand, and election timing.<sup>27</sup> Building on their work, we address each of the three drivers' impact on our IV to rule out potential endogeneity.<sup>28</sup> First, to account for policy diffusion, defined as the tendency of cities to adopt policies after others in the same state, we restrict our sample to commuting zones entirely within a single state and use time by state fixed effects to absorb diffusion-driven shocks as well as broader state-level policy environments.<sup>29</sup> Second, we control for variation in local

---

<sup>27</sup> Other factors significantly associated with SOI adoptions are education and median rent, which we control for through education attainment and median housing value. In contrast, factors such as city size, demographic composition, and political partisanship (proxied by the mayor's party affiliation) do not significantly correlate with SOI adoption when the primary drivers are controlled.

<sup>28</sup> The concerns for each factor are: (1) regional policy diffusion may correlate with unobserved regional shocks influencing both firm level information environments and the share of cities with a commuting zone that adopted the law; (2) local demand for housing assistance may reflect broader regional economic conditions affecting firm level information environments; (3) election timing creates political uncertainty, potentially leading to temporary shifts in firm disclosure practices and affecting measures of information asymmetry (Bird, Karolyi, and Ruchti, 2023).

<sup>29</sup> The restriction also helps create a cleaner setting, given the historical importance of state level institutions, such as Jim Crow housing policies, in shaping residential segregation.

housing assistance needs using the annual number of voucher-assisted households in each commuting zone, based on HUD's Picture of Subsidized Households. This variable enters both stages of the IV estimation. Third, we lag the SOI adoption variable by one year to limit election year bias, which aligns with lease renewal timing and accounts for the 8% of cities that adopted the law prior to its official effective date. Finally, we exclude firms that moved across commuting zones during the sample period to avoid confounding from firm relocation induced changes in RRI.<sup>30</sup> Together, these identification strategies—combined with strong results from weak instrument tests—support the validity of our instrument-based estimation.

After applying the above empirical strategies, our final instrumental variable sample includes 15,171 firm year observations (63.28% of the original sample) for 1,630 firms across 113 commuting zones for the stock price informativeness test; 17,503 firm quarter observations (60.17% of the original sample) for 2,947 firms across 137 commuting zones for the adverse selection test; and 91,052 firm quarter observations (61.86% of the original sample) for 3,708 firms across 137 commuting zones for the VCV test.

We estimate all specifications using panel IV regressions with the same control variables and fixed effects as our baseline models, except that we replace time fixed effects with time by state fixed effects. As shown in Table 2.5, the Kleibergen–Paap Wald statistic rejects the null of weak instruments, and the Durbin-Wu-Hausman test suggests the endogeneity of RRI, justifying the use of IV estimation. Table 2.5 confirms our baseline findings: Column (1) shows that higher RRI strengthens the relationship between market valuation and future firm cash flows, indicating more informative stock prices in high RRI commuting zones. Columns (2) and (3) demonstrate that greater RRI also reduces firm-level information asymmetry, as reflected by lower adverse selection and VCV measures.

---

<sup>30</sup> Our results are robust without this filter.

[Insert Table 2.5 Here]

#### 2.4.4.2 Firm relocations

We examine firm relocations as a supplementary setting to explore how changes in RRI are associated with stock price informativeness and information asymmetry. While relocation decisions are likely endogenous to broader firm strategies and local economic conditions, they nonetheless generate discrete shifts in a firm's geographic exposure that can be informative. Following Ellis, Smith, and White (2020), we focus on firms that relocated only once during the sample period to ensure a clean identification window. Commuting zones are ranked into RRI quintiles in the year of relocation, and a firm is classified as treated if it moves from a higher to lower RRI quintile (or vice versa for inverse treatment).<sup>31</sup> We construct a control group using a 1:2 nearest neighbor match based on firm-level characteristics and pre-relocation RRI, using Mahalanobis distance for matching to address small sample limitations that hinder convergence in propensity score models.<sup>32</sup> To ensure sufficient post-treatment observations, we exclude firms with only one observation following relocation. As a result, the stock price informativeness analysis focuses on relocations before 2018, while the information asymmetry test, using the VCV due to sample size constraints for the adverse selection measure, relies on relocations before 2019.

First, we estimate the following regression for the price informativeness test:

---

<sup>31</sup> Alternative grouping choices (e.g., using 2, 3, or 4 groups) result in very few treated firms (currently, we have 11 firms for a high to low RRI move and 13 firms with a low to high move). In contrast, decile-based classifications fail to capture meaningful RRI shifts.

<sup>32</sup> For the stock informativeness test, we include the following covariates: RRI, market valuation, leverage, investment, tangible assets, total assets, cash holdings, cash flow, sales to asset ratio, price volatility, illiquidity, institutional block ownership, number of analysts covering, and board racial diversity, as detailed in Section 2.2.3.1. For the information asymmetry test, the covariates are age, R&D disclosure dummy, advertisement disclosure dummy, leverage, tangible assets, sales to assets ratio, book to market ratio, cash holdings, payout ratio, proportion of institutional blockholder ownership, and number of covering analysts, lagged returns, lagged price level, lagged price standard deviation, and lagged market capitalization.

$$\begin{aligned} \frac{E_{i,t+1}}{A_{i,t}} = & \alpha + \beta_1 \log(M/A)_{i,t} + \beta_2 RRI_{i,t} + \beta_3 \log(M/A)_{i,t} \times Post \\ & + \beta_4 \log(M/A)_{i,t} \times Treatment + \\ & \beta_5 \times Post + \beta_6 \times Treatment + \beta_7 X_{i,t} + \beta_8 X_{j,t} + FE_{\text{firm}} + FE_{\text{year}} + \epsilon_{i,t+1}, \end{aligned} \quad (9)$$

where *Treatment* equals 1 for treated firms and 0 otherwise; *Post* equals 1 for periods following relocation for treated firms. The coefficient  $\beta_3$  captures the change in stock price informativeness attributable to relocation; we expect a negative  $\beta_5$  for relocations from high to low RRI zones, indicating a decline in informativeness, and a positive  $\beta_3$  for relocations from low to high RRI zones. Second, for the VCV test, we estimate:

$$IA_{i,t} = \beta_0 + \beta_1 post + \beta_2 Treatment + \beta_2 X_{i,t} + \beta_3 X_{j,t} + FE_{\text{year} \times \text{quarter}} + FE_{\text{firm}} + \epsilon_{i,t}, \quad (10)$$

where the coefficient  $\beta_1$  captures the average post-relocation effect on the VCV. We expect  $\beta_1$  to be positive for relocations from high to low RRI zones, indicating increased information asymmetry, and negative for moves in the opposite direction. For both the price informativeness and VCV regression, all fixed effects and controls follow the baseline specifications in Section 2.4.1 and Section 2.4.2, except that commuting zone fixed effects are replaced with firm fixed effects. Standard errors are clustered at the firm level.

Columns (1) and (3) in Table 2.6 show that relocations from high to low RRI commuting zones lead to declines in stock price informativeness (  $-0.080$ ) and an increase in VCV (0.119), while the opposite holds for relocations from low to high RRI areas. Despite modest sample sizes, the consistent direction and significance of the effects support the causal impact of RRI on the efficiency with which firm-specific information is incorporated into stock prices.

[Insert Table 2.6 Here]

### 2.4.5 Robustness

To test the robustness of our firm-level findings, we use (1) alternative definitions of local communities beyond commuting zones (CZs), (2) annual American Community Survey (ACS) data, and (3) alternative evenness-based measures of racial integration. First, while CZs are designed to capture commuting flows, they may overlook two key aspects. CZs often span multiple counties that differ in politically relevant characteristics such as governance and policy orientation. As a result, counties may reveal localized variation that CZs obscure. Additionally, increased cross-zone commuting in recent decades may weaken the ability of CZ boundaries to reflect functional local areas. To address these concerns, we replicate our baseline analysis using three alternative geographic units: counties (smaller than CZs), firm-defined local regions based on a 40-mile radius around each headquarters (larger than CZs), and states.

Second, our baseline RRI measure relies on decennial Census data, which requires interpolation to construct annual estimates. This may introduce forward-looking bias if future values affect current estimates. While this approach is common in work on residential patterns and cultural variables (e.g., Trounstein, 2016), we address potential concerns by re-estimating our results using annual data from the ACS at the state level, available from 2005 onward.<sup>33</sup>

Finally, to test whether our findings are sensitive to the specific construction of the RRI measure, we re-estimate our models using three alternative evenness-based indices, including the dissimilarity index (Jahn, Schmid, and Schrag, 1947), each of which captures racial composition differently. In sum, these tests confirm that the relationship between racial integration and firms' information environments is not sensitive to specific data choices or geographic definitions. Full methodological details and estimations are provided in Appendix 2A.4.1–4.2.

---

<sup>33</sup> Because the ACS only includes data for geographic areas with populations over 65,000, many smaller counties and census tracts are excluded. As a result, we construct state level RRI using data aggregated by Public Use Microdata Areas (PUMAs), the smallest consistently available geographic units in the ACS, each with at least 100,000 residents.

## 2.5 RRI and Mutual Funds' Local Investment

To further understand the mechanism behind RRI's impact on price efficiency, we examine how RRI affects the ability of local institutional investors to exploit informational advantages. We expect mutual funds in higher RRI commuting zones to exhibit weaker local stock picking performance than those in lower RRI areas, as higher RRI diminishes local funds' advantages by improving local firms' information environment.

### 2.5.1 RRI, local bias, and local stock-picking ability

To test this prediction, we construct a quarterly measure of mutual fund local stock picking performance based on each fund's ability to forecast the alphas of firms headquartered in the same commuting zone. We examine the relationship between RRI and both local stock picking performance and local bias using the following specification:

$$DV_{k,t} = \beta_0 + \beta_1 RRI_{k,t} + \beta X_{k,t} + FE_{\text{year} \times \text{quarter} \times \text{style}} + FE_{\text{CZ}} + \varepsilon_{k,t}, \quad (11)$$

where  $DV_{k,t}$  represents either the local alpha or the local bias of fund  $k$  in quarter  $t$ , and  $RRI_{k,t}$  denotes the RRI in fund  $k$ 's commuting zone in quarter  $t$ . Following the literature (Coval and Moskowitz, 2001; Cohen et al., 2008; Kacperczyk et al., 2014; Chuprinin and Sosyura, 2018),  $X_{k,t}$  includes standard predictors of mutual fund performance, such as trading style, net flow, and expense ratio (the full list can be found in Appendix 2A.2.7), as well as commuting zone characteristics from Section 2.2.5. We include year by quarter by style fixed effects ( $YQS_{s,t}$ ) to account for unobserved time-varying factors across trading styles, such as the seasonality of active fund performance (Brown, Sotes-Paladino, Wang, and Yao, 2017).<sup>34</sup> We also include commuting

---

<sup>34</sup> We classify a mutual fund's trading style by sorting funds into four categories according to their portfolio-level size and book-to-market factor loadings, estimated using the Fama and French (1993) three-factor model. The construction of all other variables is described in detail in Appendix 2A.1. To mitigate the influence of outliers, we winsorize both net flow and turnover ratio at the 1st and 99th percentiles, following Kacperczyk et al (2014).

zone fixed effects to control for unobserved, time-invariant community-level differences in stock picking ability, such as proximity to major economic centers or large cities, which can influence fund manager performance through knowledge spillovers and learning opportunities (Christoffersen and Sarkissian, 2009). Standard errors are clustered at the commuting zone level.

Table 2.7 shows that RRI, measured using the negative Theil index (RRI NegTheil Index), does not significantly affect mutual funds' local bias (columns (1) and (2)), but is significantly negatively associated with local stock picking performance (columns (3) and (4)). Economically, the coefficient in column (4) ( $-0.045$ ) indicates that a one standard deviation increase in RRI (0.10) corresponds to a decrease of 45 basis points in local stock picking performance, equivalent to 80.36% of its standard deviation (56 basis points). This effect is economically significant given the large net asset size of mutual funds (Chuprinin and Sosyura, 2018).

While local performance declines, local bias remains unchanged. This divergence suggests that mutual funds continue to invest in nearby firms even as their informational advantage erodes. If fund managers adjusted their portfolios in response to diminished informational advantages, we would expect to see reduced investment in local firms in high RRI communities. However, the absence of a significant relationship between RRI and local bias suggests that fund managers may continue to invest locally due to behavioral tendencies such as familiarity bias, or because stronger overall information environments in integrated communities lower perceived risks and transaction costs.

[Insert Table 2.7 Here]

## 2.5.2 Heterogeneity analysis

This section investigates whether the effect of RRI on mutual fund performance varies with economic conditions and manager characteristics. Understanding these sources of heterogeneity helps determine whether the relationship between RRI and local information advantage is structural or context-dependent. By examining variation across business cycles and fund manager race, we can better understand how broader economic and racial dynamics affect access to and use of local information.

### 2.5.2.1 Economic conditions

Economic conditions shape how mutual funds generate alpha through stock selection and market timing (Kacperczyk et al., 2014). Prior research shows that stock picking skill is particularly valuable in economic downturns, when investors place the highest premium on a manager's ability (Moskowitz, 2000). During expansions, abundant opportunities and clearer signals allow managers to rely more evenly on stock selection, while in recessions the payoff to skill is amplified: funds that identify resilient firms can deliver meaningful alpha, whereas unskilled managers may incur severe losses. This widening gap makes the erosion of local informational advantages in high RRI communities even more damaging during recessions, though also less precise to estimate due to greater heterogeneity across funds.

We classify economic conditions using quarterly data from the NBER Business Cycle Dating Committee, which defines expansions as the period between a trough and the next peak, and recessions as the period between a peak and the next trough.<sup>35</sup> Column (1) of Table 2.8 shows that

---

<sup>35</sup> The determination of peak and trough months is based on a range of monthly indicators of aggregate real economic activity published by federal statistical agencies. Separately, the NBER Business Cycle Dating Committee identifies calendar quarters of peaks and troughs using quarterly measures of aggregate economic activity. Key metrics used to determine quarterly turning points—where monthly data are unavailable—include both expenditure side and income side estimates of real gross domestic product (GDP and GDI). The committee also considers quarterly averages of monthly indicators, such as payroll employment. According to the NBER's definition, recession periods between 2000 and 2020 include: 2001Q1–Q4 (associated with the dot com crash and the September 11th attacks), 2007Q4–2009Q2

during expansions, RRI is significantly negatively related to local stock picking ability, consistent with the baseline results in Table 2.7. Column (2) indicates that this negative relationship becomes even stronger during recessions, although it is estimated with lower statistical precision. The weaker detectability likely reflects both higher volatility and greater dispersion in outcomes: while skilled funds tend to generate alpha when it matters most, unskilled funds suffer disproportionately larger losses. Overall, the findings suggest that the erosion of local informational advantage in high RRI communities is structural and persists throughout the business cycle and becomes particularly pronounced—though more difficult to measure—during recessions.

[Insert Table 2.8 Here]

#### *2.5.2.2 Mutual fund manager race*

Social capital refers to the resources that individuals access through social connections, which are crucial for achieving their goals (Lin, 2002). Prior studies have shown that social networks are often segregated by race and gender, with White male networks offering more resources and higher status connections (McDonald et al., 2009; McDonald and Day, 2010; McDonald, 2011). Building on this insight, we examine whether the relationship between RRI and local stock picking performance varies by the race of the fund manager. Specifically, we hypothesize that White managers benefit more from local social capital, making their stock picking ability more sensitive to changes in RRI. In less integrated communities, where racially homogeneous networks are more prevalent and exclusive, White managers may benefit from preferential access to firm-specific information or local business ties that improve investment outcomes. As RRI increases and social networks become more inclusive and less segmented, this advantage may decline. We therefore expect the negative relationship between RRI and local stock picking performance to be more

---

(corresponding to the global financial crisis and the housing market collapse), and 2019Q4–2020Q2 (the COVID-19 recession). More details are available at <https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions>.

pronounced for White fund managers.

To test this, we restrict the sample to funds managed by a single individual to avoid confounding organizational effects (Csaszar, 2012). We also exclude managers who oversee four or more funds simultaneously, as their superior skills or professional reputations may drive both local performance and location choice, potentially biasing our estimates. Manager race is inferred using NamePrism, following the same method used to classify board director ethnicity, as detailed in Appendix 2A.2.3. A manager is assigned to a racial group if the predicted probability exceeds 50%; otherwise, the observation is excluded. This yields 10,492 fund quarter observations from 817 managers: 759 White (9,715 observations across 54 commuting zones) and 58 non-White (777 observations across 17 commuting zones).

Column (3) in Table 2.8 shows that for White managers, RRI is significantly negatively associated with local stock picking performance, with a coefficient of  $-0.041$ , comparable to the baseline in Table 2.7. In contrast, the coefficient for non-White managers in column (4) is statistically insignificant. While a limited sample size may reduce power for this group, the results suggest that non-White managers may be less reliant on racially segmented local networks for accessing firm-specific information, indicating potential differences in how managers from different racial backgrounds engage with and benefit from local social capital.

### **2.5.3 Robustness**

This section evaluates the robustness of our mutual fund level findings by employing alternative definitions of local community, varying the sample composition, and considering performance persistence as well as unobserved fund-specific characteristics. Consistent with our firm-level analysis, we reexamine the relationship between RRI and local stock picking performance using different community definitions, including the fund's county, surrounding counties within a 40-mile radius (i.e., local region), and the state in which the fund is located. We

further use annual RRI measures constructed from ACS data and alternative evenness-based integration metrics. The results remain robust across all specifications, with the exception of county-level RRI, which appears to capture a geographic unit that is too narrow to reflect the broader social and informational networks fund managers rely upon. All results are presented in Appendix 2A.4.1–4.2.

We also conduct several fund-level robustness checks to rule out confounding effects from sample composition or unobserved fund characteristics. First, we exclude all funds headquartered in New York, California, and Massachusetts, three states that account for nearly half of the sample, and find that the results remain significant despite the reduction in sample size. Second, to address concerns related to autocorrelation and fund level persistence, we control for lagged local stock picking performance and contemporaneous nonlocal performance. The main findings remain unchanged. Third, we include fund fixed effects to account for time-invariant heterogeneity across funds, and the negative association persists. All robust results are consistent with our baseline results and are reported in Table 2A.4.3 of the Appendix.

## **2.6 Conclusion**

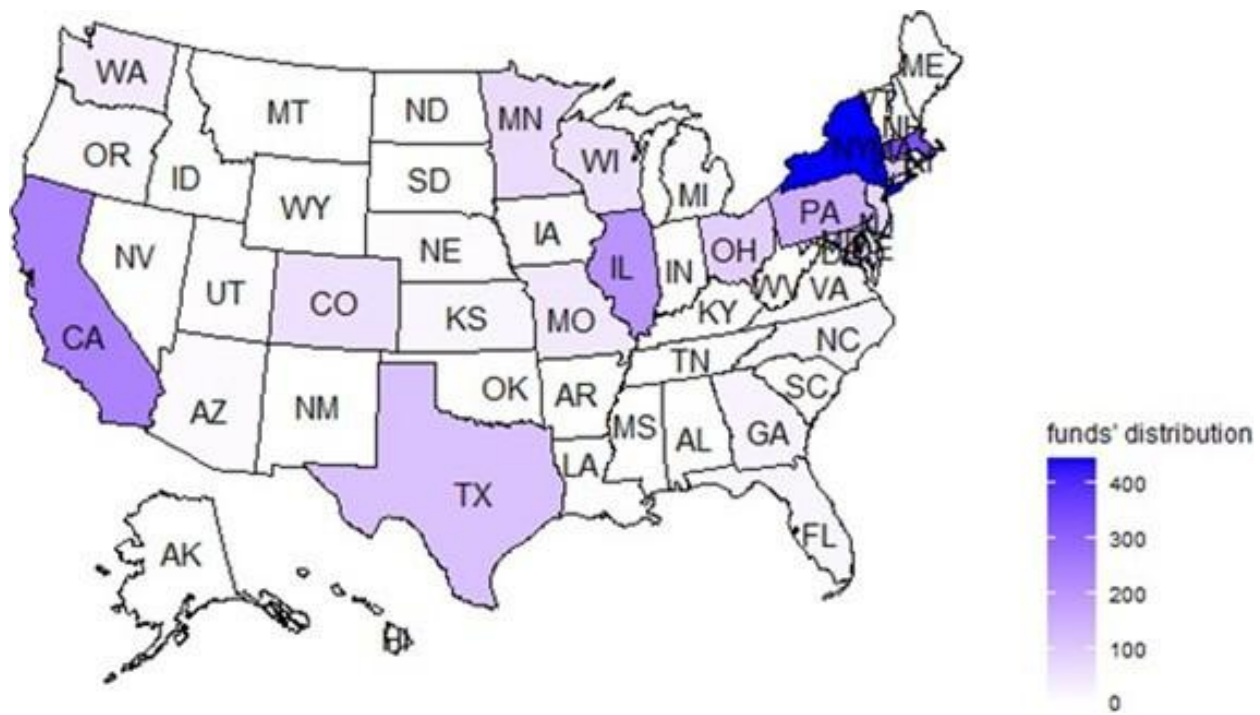
We show that RRI enhances economic integration by promoting upward social mobility and narrowing average income gaps, suggesting improved interpersonal connections and access to opportunities. Building on this foundation, we examine how RRI influences the quality of firm-level information environments and investor behavior in financial markets. Using several identification strategies, including firm relocations and instrumental variable regressions, we find that firms headquartered in more racially integrated areas exhibit significantly more informative stock prices and lower information asymmetry. Specifically, these firms show a stronger relationship between current market valuations and future cash flows, as well as reduced

information asymmetry, reflected in narrower adverse selection components of bid-ask spreads and lower volume coefficient of variation.

These effects appear to be driven by improved local information networks in more integrated areas, which facilitate broader and faster information diffusion. In contrast, less integrated communities allow a relative advantage to active institutional investors who can access private, local information that is less widely shared. This local edge is concentrated in White fund managers, who perform better when investing in local firms in less integrated communities. In contrast, non-White managers do not experience similar benefits, suggesting unequal access to local information networks and highlighting a structural barrier facing minority managers in the asset management industry.

Overall, our findings show that racial integration plays a meaningful role in shaping how information flows through financial markets. It not only improves price efficiency but also affects which investors benefit from local informational advantages. These findings have important implications: fostering more racially integrated communities can support fairer and more efficient markets by improving access to information and reducing disparities in investment outcomes.

Figure 2.1: Distribution of Mutual Funds in the United States



## Table 2.1: Summary Statistics of Community-, Stock-, and Fund-Level Variables

This table provides descriptive statistics of the variables used in this study. **Panel A** reports community level characteristics (i.e., commuting zone-level). *RRI Negative* captures the evenness dimension of racial integration. *IR (Black)* and *IR (White)* represent the average predicted national adult income ranks of children from low-income families, measured at the commuting zone level by race. *Average White-Black Income Gap* is the difference between the two. *Senior Ratio* is the proportion of the population aged 65 or older. *Trust Index* follows Wei and Zhang (2020). *Racial Diversity* is defined as one minus the Herfindahl-Hirschman index based on five racial groups (Section 2.2.5). *Education* is the proportion of adults aged 25 and older with a bachelor's degree or higher. *White Ratio*, *Hispanic Ratio*, *African American Ratio*, and *Asian Ratio* reflect the respective racial group proportions in the local population. All non-ratio local capital market variables are value weighted, except for total market capitalization (*Regional MKTCAP*), which is the sum of individual stocks' market capitalization. *Regional Ret* is the value-weighted return. *Regional Illiq* is Amihud's (2002) illiquidity measure  $\times 10^6$ . *Regional Fpd* is the ratio of firms paying dividends. *Regional IO* is the commuting zone institutional ownership. All variables are winsorized at 1% and 99% levels. **Panel B** presents stock-level variables used in the price informativeness test. *Future Cash Flow* is EBIT in year  $t+1$  scaled by total assets in year  $t$ . *Market Valuation*,  $\log(M/A)$ , is the log of the market-to-asset ratio. *Board Racial Diversity* is defined as one minus the Herfindahl-Hirschman Index using the racial composition of board directors (Section 2.2.3). *Age* is the number of years since the firm first appeared in the Compustat database. Institutional ownership ratio (IO) represents holdings of at least 5% of firm shares by institutional investors. Sales to total assets (*Sale*), tangible assets to total assets (*Tangibility*), book to market ratio (*B/M*), EBIT to assets in year  $t$  (*Cash Flow*), leverage, cash holdings to total assets (*Cash Holdings*), and investment to total assets (*Invest*) are obtained from the same fiscal year data, updated annually. *Volp* is the daily stock price standard deviation over the past 12 months. *Illiquidity* denotes illiquidity. *Analyst Coverage* is defined as the natural log of one plus the number of analysts following the stock. All these variables are winsorized at the 1st and 99th percentiles. **Panel C** includes additional stock level variables used in the information asymmetry analysis. The construction of *Adverse Selection* follows Jang and Sun (2014). *Volume Coefficient of Variation* is the ratio of the standard deviation to the mean of trading turnover. *Trading Volume* is the quarterly trading volume denominated at \$1 million. *Pre( $t - 1$ )* is the lagged mean stock price. *Volp( $t - 1$ )* is the lagged price standard deviation. *Market Cap( $t - 1$ )* is denominated at \$1,000. *Sales*, *Intangibility*, *B/M Ratio*, *Leverage*, *Cash Holdings*, and *Payout* (payout ratio) are measured using data from the most recent fiscal year and are updated annually. *R&D (0/1)* and *Advertising (0/1)* are dummy variables set to 1 if the firm disclosed respective expenditures from the most recent fiscal year. Other firm level variables are quarterly variables. According to the literature, all variables are winsorized at 1% and 99% levels. **Panel D** reports fund level variables. *Turnover*, *expense ratio*, and *total load* are all annualized. *Fund Age* is based on the mutual fund's initial date and measured in years. *Local Investment* is defined as the fraction of the mutual fund's portfolio invested in stocks headquartered in the same commuting zone as the mutual fund's headquarters. *TNA* denotes total net assets. *Team Managed* is a binary variable set to 1 if a fund is managed by at least four managers. *Factor loading* is a portfolio level measure categorizing funds into four groups. *Fund Family Size* is defined as the number of fund share classes under the same family. Full variable definitions are provided in Appendix 2A.1. All fund level variables are winsorized at the 1st and 99th percentiles, except return-related variables and local bias.

Variables	Frequency	N	Mean	Stdev	P25	Median	P75
<b><i>Panel A: Community-level variables</i></b>							
Racial Integration: RRI NegTheil	Year	9,744	- 0.171	-0.096	-0.230	-0.159	-0.097
Average Income Rank: IR (Black)	By cohort	3,035	0.388	0.023	0.374	0.387	0.401
Average Income Rank: IR(White)	By cohort	3,035	0.437	0.026	0.418	0.433	0.452
Average White-Black Income Gap	By cohort	3,035	0.048	0.028	0.030	0.048	0.066
Number of Counties in a CZ	—	452	4.998	2.614	3.000	5.000	6.000
White Ratio	Year	9,744	0.693	0.170	0.588	0.721	0.830
Hispanic Ratio	Year	9,744	0.132	0.153	0.036	0.073	0.161
African American Ratio	Year	9,744	0.116	0.132	0.021	0.060	0.170
Asian Ratio	Year	9,744	0.018	0.025	0.005	0.010	0.020
Racial Diversity	Year	9,744	0.411	0.140	0.296	0.432	0.526
Population (ln)	Year	9,744	12.212	1.491	11.224	12.174	13.258
Population Density ( $10^3/km^2$ )	Year	9,744	0.054	0.085	0.009	0.024	0.057
Education	Year	9,744	0.224	0.079	0.163	0.207	0.273
Unemployment Ratio	Year	9,744	0.059	0.024	0.042	0.054	0.071
Senior Ratio	Year	9,744	0.152	0.035	0.127	0.149	0.172
Female Ratio	Year	9,744	0.502	0.016	0.498	0.506	0.513
Median Housing Value (ln)	Year	9,744	11.106	2.291	11.179	11.495	11.844
Median Income (ln)	Year	9,744	10.691	0.241	10.516	10.686	10.849
Political Fragmentation (ln)	Year	9,744	0.250	0.185	0.123	0.229	0.347
Trust Index	Year	9,744	0.372	0.067	0.317	0.381	0.403
Regional Ret	Quarter	36,281	0.035	0.124	-0.017	0.036	0.090
Regional Fpd	Quarter	36,281	0.537	0.210	0.392	0.517	0.667
Regional Illiq ( $\times 10^6$ )	Quarter	35,474	0.012	0.105	0.001	0.002	0.005
Regional Mktcap (ln)	Quarter	36,281	18.122	2.400	16.861	18.832	20.064
Regional Nstocks (ln)	Quarter	36,281	3.782	1.681	2.565	3.850	5.323
Racial Integration: RRI NegTheil	Year	9,744	0.171	0.096	0.230	0.159	0.097
Regional IO	Quarter	36,141	0.721	0.177	0.676	0.756	0.828
<b><i>Panel B: Firm-level variables for the stock price informativeness test</i></b>							
Future Cash Flow	Year	23,976	0.086	0.130	0.048	0.091	0.146
Market Valuation: log(M/A)	Year	23,976	0.097	0.840	-0.466	0.101	0.660
Board Racial Diversity	Year	23,976	0.272	0.202	0.000	0.292	0.444
Age (ln)	Year	23,976	2.978	0.823	2.398	3.045	3.714
IOR	Year	23,976	0.222	0.140	0.117	0.212	0.312
Sale	Year	23,976	1.018	0.703	0.526	0.853	1.316
Tangibility	Year	23,976	0.269	0.233	0.086	0.189	0.395
Cash Flow	Year	23,976	0.081	0.115	0.049	0.087	0.134
Investment	Year	23,976	0.088	0.081	0.035	0.064	0.113
Asset (ln)	Year	23,976	7.471	1.668	6.258	7.358	8.569
Analyst Cover (ln)	Year	23,976	2.393	0.697	1.946	2.398	2.890
Leverage	Year	23,976	0.215	0.175	0.043	0.207	0.337
Cash Holdings	Year	23,976	0.176	0.199	0.031	0.099	0.250
Illiquidity ( $\times 10^6$ )	Year	23,976	0.236	1.114	0.011	0.038	0.115
Price Volatility: Volp (ln)	Year	23,976	2.268	1.939	1.045	2.136	3.281

Variables	Frequency	N	Mean	Stdev	P25	Median	P75
<b><i>Panel C: Firm-level variables for the information asymmetry test</i></b>							
Adverse Selection (%)	Quarter	29,088	0.412	0.662	0.022	0.134	0.494
Volume Coefficient Variation	Quarter	147,191	0.846	0.537	0.499	0.674	0.997
Institutional Ownership: IO(t-1)	Quarter	147,191	0.212	0.158	0.079	0.198	0.317
Analyst Cover (ln)	Quarter	147,191	1.380	1.056	0.000	1.609	2.197
Trading Volume (\$10 <sup>8</sup> )	Quarter	147,191	24.694	59.500	0.597	3.696	17.812
Return: Ret(t-1)	Quarter	147,191	0.052	0.268	-0.096	0.030	0.164
Price Volatility: Volp(t-1) (ln)	Quarter	147,191	0.269	1.007	-0.332	0.318	0.917
Price: Prc(t-1) (ln)	Quarter	147,191	3.243	1.135	2.493	3.076	3.716
Market Cap(t-1) (ln)	Quarter	147,191	13.507	1.862	12.244	13.437	14.687
Board Racial Diversity	Year	147,191	0.267	0.161	0.198	0.278	0.347
Age (ln)	Year	147,191	2.797	0.811	2.197	2.833	3.434
Sale	Year	147,191	1.028	0.763	0.502	0.867	1.362
Tangibility	Year	147,191	0.231	0.220	0.065	0.156	0.326
B/M	Year	147,191	0.576	0.472	0.256	0.452	0.751
Leverage	Year	147,191	0.191	0.183	0.007	0.159	0.315
Cash Holding	Year	147,191	0.226	0.245	0.039	0.130	0.334
Payout	Year	147,191	0.416	1.244	0.000	0.035	0.625
R&D Disclosure (0/1)	Year	147,191	0.654	0.476	0.000	1.000	1.000
Advertising Disclosure (0/1)	Year	147,191	0.413	0.492	0.000	0.000	1.000
<b><i>Panel D: Fund-level variables</i></b>							
Local Stock Picking (%)	Quarter	49,350	-0.018	0.526	-0.184	-0.006	0.155
Local Bias (%)	Quarter	49,350	0.560	4.629	-1.503	0.390	2.459
Local Investment (%)	Quarter	49,350	5.051	4.799	1.913	3.710	6.418
Expense Ratio (%)	Quarter	49,350	1.130	0.361	0.907	1.102	1.337
Net Flow (%)	Quarter	49,350	-0.183	11.373	-4.823	-1.912	1.709
Turnover (%)	Year	49,350	73.633	54.514	35.000	60.000	97.000
Total Load (%)	Year	49,350	1.579	1.812	0.000	0.985	2.717
Fund Age (ln)	Year	49,350	2.541	0.678	2.079	2.639	2.996
TNA (Billion)	Quarter	49,350	1.716	7.107	0.086	0.332	1.221
Team Managed (0/1)	Quarter	49,350	0.317	0.465	0.000	0.000	1.000
Single Manager (0/1)	Quarter	49,350	0.276	0.447	0.000	0.000	1.000
Factor Loading Size (%)	Quarter	49,350	0.094	2.019	-0.487	0.067	0.691
Factor Loading Value (%)	Quarter	49,350	0.082	1.658	-0.485	0.021	0.614
Number of Stocks in Portfolio	Quarter	49,350	122.387	166.075	50.000	74.000	122.000
Fund Family Asset (Billion)	Quarter	49,350	142.584	411.114	2.823	24.344	81.761
Fund Family Size	Quarter	49,350	63.846	87.183	9.000	35.000	87.000

**Table 2.2: RRI, Social Mobility, and the Mean White–Black Income Gap**

This table shows the result of the effect of racial residential integration (RRI) on social mobility and the mean White–Black income gap. The dependent variable  $IR_{i,r,s}$  is the mean predicted income at age 27 of children in commuting zone  $i$ , for racial group  $r$  (White or Black) and birth cohort  $s$ . The dependent variable  $IG_{i,s}$  is the difference in income ranks between White and Black children from low-income families in commuting zone  $i$  and cohort  $s$ . The key explanatory variable, *RRI NegTheil*, captures the evenness dimension of racial integration, as detailed in Section 2.2.1. The White dummy ( $W$ ) equals 1 for White children. All regressions control for commuting zone demographic and socioeconomic variables described in Section 2.2.5, including racial group proportions (White, African American, Asian, and Hispanic), racial diversity, zoning behavior, median housing value, median income, commuting zone population, population density, education level, unemployment ratio, senior citizen ratio, female ratio, and social trust. Control variables are entered following the specifications outlined in Equations (5) and (6). Standard errors are two-way clustered at the commuting zone and cohort levels.  $*p < .1$ ;  $**p < .05$ ;  $***p < .01$ ;

	Dependent Variables	
	Mean Income Rank	Mean White–Black Income Gap
	$(IR_{i,r,s})$	$(IG_{i,s})$
	(1)	(2)
RRI NegTheil	0.034** (0.012)	−0.045** (0.018)
RRI NegTheil $\times \frac{s-1983}{9}$	−0.008 (0.010)	−0.004 (0.019)
White	0.061*** (0.002)	
White $\times \frac{s-1983}{9}$	−0.028*** (0.003)	
Controls	Yes	Yes
Observations	6,070	3,035
$R^2$	0.635	0.370

**Table 2.3: RRI, Stock Price Informativeness, and Information Asymmetry**

This table shows results for the relationship between RRI, the correlation between firms' future cash flow and their market valuation, and the quality of a firm's information environment. Columns (1)–(3) use future cash flow (EBFF) in year  $t+1$  scaled by total assets in year  $t$  as the dependent variable. Columns (4)–(6) use adverse selection, measured as the bid–ask spread component compensating for informed trading. Columns (7)–(9) use the volume coefficient of variation (VCV), defined as the standard deviation of turnover divided by its average within a quarter. *RRI NegTheil* captures the evenness dimension of racial integration (Section 2.2.1). For stock price informativeness tests, control variables include firm-level characteristics (e.g., age, EBIT-to-asset ratio, leverage, investments, tangible assets, analyst coverage, total assets, cash holdings, sales-to-asset ratio, price volatility, and illiquidity) and commuting zone–level variables (as listed in Table 2.2). For information asymmetry tests, control variables include firm characteristics (e.g., age, R&D and advertising disclosure dummies, leverage, tangible assets, sales-to-assets, book-to-market ratio, cash holdings, payout ratio, institutional ownership, analyst coverage), firm-level market attributes (e.g., lagged returns, lagged price, lagged price volatility, lagged market capitalization), and the commuting zone–level variables, as listed in Table 2.2. Coefficients for controls are detailed in Appendix 2A.1. Fixed effects are indicated in the table. Standard errors are clustered at the commuting zone level and reported in parentheses. \* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ .

	Dependent Variables								
	Price Informativeness			Information Asymmetry					
	Future Cash Flow			Adverse Selection (%)			Volume Coefficient of Variation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\log(M/A) \times \text{RRI NegTheil}$	0.058**	0.031*	0.033**						
	(0.025)	(0.016)	(0.017)						
RRI NegTheil	–0.011	0.004	–0.127***	–0.897***	–0.703***	–2.06***	–0.114	–0.060**	–0.800***
	(0.050)	(0.019)	(0.045)	(0.135)	(0.090)	(0.647)	(0.084)	(0.027)	(0.278)
Firm Controls		Yes	Yes		Yes	Yes		Yes	Yes
Commuting Zone Controls			Yes			Yes			Yes
Year FE and Firm FE	Yes	Yes	Yes						
Year $\times$ Quarter FE						Yes			Yes
Industry (Two Digit SIC) FE					Yes	Yes		Yes	Yes
Commuting Zone FE						Yes			Yes
Observations	23,976	23,976	23,976	29,088	29,088	29,088	147,191	147,191	147,191
$R^2$	0.751	0.821	0.823	0.018	0.168	0.433	0.0004	0.351	0.367

**Table 2.4: Heterogeneity in RRI’s Effect on Information Environment**

This table shows how the effect of RRI on firms’ information environments varies by firm size (panel A), analyst coverage (panel B), and firm age (panel C). Columns (1)–(2) use future cash flow (EBFF in  $t+1$  scaled by total assets in  $t$ ) as the dependent variable; columns (3)–(4) use adverse selection (bid–ask spread component); columns (5)–(6) use volume coefficient of variation (VCV), defined as the standard deviation of turnover over its average within a quarter. *RRI NegTheil* captures the evenness dimension of racial integration (Section 2.2.1). Controls follow Table 2.3. Firms are sorted into high and low groups based on the median value of each characteristic—total assets, analyst coverage, or age—updated annually for price informativeness and quarterly for information asymmetry. Standard errors are clustered at the commuting zone level and shown in parentheses. \* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ .

	Dependent Variables					
	Price Informativeness:		Information Asymmetry:			
	Future Cash Flow		Adverse Selection (%)		VCV	
<i>Panel A: Sorted by Firm Size</i>						
	Smaller	Larger	Smaller	Larger	Smaller	Larger
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(M/A) \times \text{RRI NegTheil}$	0.052*	0.007				
	(0.028)	(0.021)				
RRI NegTheil	-0.173***	-0.088	-2.450**	-1.010*	-1.528**	-0.397
	(0.058)	(0.049)	(1.18)	(0.602)	(0.479)	(0.244)
Controls	Yes		Yes			
	Firm and Year FEs		Year×Quarter, Industry, and CZ FEs			
Observations	11,871	11,871	14,523	14,523	73,578	73,576
<i>Panel B: Sorted by Analyst Coverage</i>						
	Fewer	More	Fewer	More	Fewer	More
	Analysts	Analysts	Analysts	Analysts	Analysts	Analysts
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(M/A) \times \text{RRI NegTheil}$	0.041*	0.015				
	(0.023)	(0.023)				
RRI NegTheil	-0.067	-0.135*	-2.628**	-1.778**	-1.118**	-0.429*
	(0.044)	(0.073)	(1.09)	(0.691)	(0.521)	(0.250)
Controls	Yes		Yes			
	Firm and Year FEs		Year×Quarter, Industry, and CZ FEs			
Observations	11,257	11,356	9,740	12,342	67,105	68,317

*Panel C: Sorted by Firm Age*

	Younger	Older	Younger	Older	Younger	Older
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(M/A) \times \text{RRI NegTheil}$	0.031 (0.029)	0.016 (0.018)				
RRI NegTheil	-0.315*** (0.100)	-0.026 (0.030)	-3.383*** (1.03)	-0.752 (0.811)	-1.398*** (0.431)	-0.378 (0.342)
Controls	Yes		Yes			
	Firm and Year FEs		Year×Quarter, Industry, and CZ FEs			
Observations	11,551	11,576	13,988	14,171	14,113	11,719

**Table 2.5: Instrumental Variable Regression**

This table reports the results of the instrumental variable regressions. In column (1), the dependent variable is future cash flow, proxied by EBFF in year  $t+1$  scaled by total assets in year  $t$ . Column (2) uses adverse selection as the dependent variable, measured by the bid–ask spread component that compensates market makers for the risk of trading with informed investors. Column (3) examines the volume coefficient of variation (VCV), defined as the standard deviation of trading turnover divided by its quarterly average. Market valuation,  $\log(M/A)$ , is calculated as the natural logarithm of the firm's market capitalization divided by total assets in year  $t$ . *RRI NegTheil* captures the evenness dimension of racial integration, as detailed in Section 2.2.1. The instrument variable is the share of cities within a commuting zone that adopted a source-of-income non-discrimination law as outlined in Section 2.4.4.1. Controls are consistent with those in Table 2.3 and Table 2.4. Fixed effects are included and described in the table. Industry FEs are based on a two-digit SIC. Standard errors are all clustered at the commuting zone level and are reported in parentheses. Instrument strength (weak instrument test) is assessed using the Kleibergen–Paap Wald test (Kleibergen and Paap, 2006) for the first-stage estimation and endogeneity with the Durbin–Wu–Hausman test (Durbin, 1954; Wu, 1973; Hausman, 1978) showing whether the ordinary least squares (OLS) estimation is better than IV estimation (null hypothesis). Test statistics are provided, and significance levels are indicated by asterisks.  $*p < .1$ ;  $**p < .05$ ;  $***p < .01$ .

	Dependent Variables		
	Price Informativeness:	Information Asymmetry:	
	Future Cash Flow (1)	Adverse Selection (%) (2)	VCV (3)
$\log(M/A) \times RRI\ NegTheil$ (IV estimated)	0.180** (0.079)		
<i>RRI NegTheil</i> (IV estimated)	−0.987 (0.858)	−16.870*** (5.660)	−3.760* (0.968)
Controls	Yes	Yes	Yes
FEs	Year×State and Firm	Year×Quarter×State, Industry, and CZ	
Weak IV Test Results:			
<i>RRI NegTheil</i>	4.61***	8.16***	9.87***
$\log(M/A) \times RRI\ NegTheil$	6.33***		
Endogeneity Test	3.85**	10.88***	3.06*
Observations	15,171	17,593	91,052
$R^2$	0.830	0.502	0.394

**Table 2.6: Difference-In-Differences Regression**

This table presents the results of difference in difference (DID) regressions examining the effect of changes in racial residential integration (RRI) on stock price informativeness and firm-level information asymmetry following firm relocations. The dependent variable in columns (1) and (2) is future cash flow, proxied by EBFF in year  $t+1$  scaled by total assets in year  $t$ . Columns (3) and (4) examine the volume coefficient of variation (VCV), defined as the standard deviation of trading turnover divided by its quarterly average. Market valuation,  $\log(M/A)$ , is calculated as the natural logarithm of the firm's market capitalization divided by total assets in year  $t$ . *RRI NegTheil* captures the evenness dimension of racial integration (Section 2.2.1). The interaction term  $Post \times \log(M/A)$  captures the change in the association between market valuation and future cash flow due to relocation. Columns (1) and (3) report estimates for firms that moved from a high-RRI to a low-RRI commuting zone, while columns (2) and (4) report estimates for low-to-high relocations. All specifications include firm and year fixed effects (columns (1)–(2)), or firm and year-quarter fixed effects (columns (3)–(4)). Control variables follow those used in Table 2.3. Standard errors are clustered at the firm level and reported in parentheses.  $*p < .1$ ;  $**p < .05$ ;  $***p < .01$ .

	Dependent Variables			
	Price Informativeness:		Information Asymmetry:	
	Future Cash Flow		VCV	
	High-to-Low	Low-to-High	High-to-Low	Low-to-High
	(1)	(2)	(3)	(4)
$Post \times \log(M/A)$	-0.083*** (0.026)	0.024* (0.012)		
Post	-0.048** (0.022)	0.018* (0.011)	0.119* (0.063)	-0.116** (0.055)
Controls	Yes	Yes	Yes	Yes
FEs	Firm and Year		Firm and Year×Quarter	
Firm Compositions:				
Number of Treated Firms	11	13	73	71
Number of Control Firms	18	23	120	126
Observations	399	510	6,096	6,013
$R^2$	0.668	0.739	0.458	0.437

**Table 2.7: RRI, Local Bias, and Local Stock-Picking Ability**

This table presents findings on the relationship between RRI and mutual funds' local bias and local stock-picking performance. Local bias is the difference between the portfolio weight of local stocks and their market portfolio share (Wei and Zhang, 2020). Local Stock-picking ability, a quarterly portfolio-level measure, assesses funds' predictive capacity for local stocks' alphas, computed using Equation 2.3 in Section 2.2.4. Local stocks are defined as those headquartered in the same commuting zone as the fund. The dependent variables are local bias for columns (1)–(2) and local stock-picking ability for columns (3)–(4). *RRI NegTheil* captures the evenness dimension of racial integration. Controls are included in some specifications, with coefficients reported in the Internet Appendix. These controls consist of fund-level controls (i.e., trading style, net flow, turnover ratio, fund load, fund age, expense ratio, total net assets, an indicator for funds managed by one manager, an indicator for funds with more than four managers, the total number of stock holdings, the number of share classes in the fund family, and the total net assets of the fund family) and the same commuting zone controls as in Table 2.3. Standard errors are clustered at the commuting zone level and reported in parentheses. \* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ .

	Local Bias		Local Stock-picking Ability	
	(1)	(2)	(3)	(4)
RRI NegTheil	0.034 (0.035)	0.074 (0.133)	-0.017*** (0.003)	-0.045*** (0.010)
Controls		Yes		Yes
Year $\times$ Quarter $\times$ Style FE		Yes		Yes
Commuting Zone FE	Yes	Yes	Yes	Yes
Observations	49,350	49,350	49,350	49,350
$R^2$	0.094	0.175	0.052	0.070

**Table 2.8: RRI and Local Stock-Picking: Economic Conditions and Manager Race**

This table reports how the relationship between racial residential integration (RRI) and local stock-picking performance varies across economic cycles (panel A) and manager race (panel B). The dependent variable in all columns is local stock-picking ability, defined consistently with Table 2.7. In panel A, column (1) reports estimates for economic expansion periods, and column (2) for recession periods, based on the NBER Business Cycle Dating Committee's classification. Expansions are defined as periods from a trough to the next peak; recessions span from a peak to the next trough. Panel B presents results by manager race. Column (1) includes only funds managed by a single White manager; column (2) includes funds with a single minority manager. Racial identification is based on the NamePrism algorithm, as described in Section 2.5.2.2. RRI is measured by the *RRI NegTheil*, which captures the evenness dimension of racial integration. All regressions control for the variables listed in Table 2.7 (coefficients shown in the Internet Appendix). Commuting zone and year  $\times$  quarter  $\times$  style fixed effects are controlled, and standard errors are clustered at the commuting zone level, reported in parentheses. \* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ .

<b>Dependent Variable: Local Stock-Picking Ability</b>		
<i>Panel A: Economic Conditions</i>		
	Expansion Periods	Recession Periods
	(1)	(2)
RRI NegTheil	-0.039*** (0.010)	-0.119* (0.069)
Controls	Yes	Yes
Year $\times$ Quarter $\times$ Style FE	Yes	Yes
Local Region FE	Yes	Yes
Observations	42,491	6,859
$R^2$	0.060	0.136
<i>Panel B: Manager Race</i>		
	White Managers	Minority Managers
	(1)	(2)
RRI NegTheil	-0.041** (0.019)	-0.318 (0.213)
Controls	Yes	Yes
Year $\times$ Quarter $\times$ Style FE	Yes	Yes
Local Region FE	Yes	Yes
Observations	9,715	777
$R^2$	0.108	0.410

## Appendix to Chapter 2

This appendix is organized into four sections: Section 2A provides concise definitions of all control variables used in the study, complementing the main text. Section B includes supplementary materials, such as graphs, additional tests, and detailed explanations of variables, which reinforce the arguments presented in the main text (2B.1-2B.7). Section C contains the complete or extended versions of tables from the main text, including detailed estimates for all control variables and extra tests mentioned in the main body (2C.1-2C.6). Finally, Section D presents the results of robustness tests (Tables 2D.1-2D.3).

### Appendix 2A.1: Variable definitions

#### Part I: Regional Socioeconomic and Demographic Variables

1. **White population percentage** (*White (%)*). The percentage of white individuals in a specified region (county, commuting zone, local area defined in our study, or state).
2. **African American population percentage** (*African American (%)*). The percentage of African American individuals in a specified region (county, commuting zone, local area defined in our study, or state).
3. **Asian population percentage** (*Asian (%)*). The percentage of Asian individuals in a specified region (county, commuting zone, local area defined in our study, or state).
4. **Hispanic population percentage** (*Hispanic (%)*). The percentage of Hispanic individuals in a specified region (county, commuting zone, local area defined in our study, or state).
5. **Senior citizen ratio** (*Senior Ratio*). The ratio of senior citizens (age  $\geq 65$ ) in the population. Source: U.S. Census Bureau.
6. **Female population ratio** (*Female Ratio*). The ratio of the female (all ages) population. Source: U.S. Census Bureau.
7. **Unemployment ratio** (*Unemployment*). The proportion of individuals who are unemployed. Source: U.S. Bureau of Labor Statistics (county level, adjusted for other levels by population weight).
8. **Educational attainment level** (*Education*). The percentage of adults aged 25 and older with a bachelor's degree or higher. The data for inter-wave years are obtained by interpolating the two nearest available data points. Source: U.S. Census Bureau, SAIPE dataset (county level, adjusted for other levels by population weight).
9. **Total population** (*Total Population*). The natural logarithm of the regional population (county, commuting zone, local area defined in our study, or state). Source: U.S. Census Bureau.
10. **Population density** (*Population Density*). The population divided by the area of a given region (county, commuting zone, local area defined in our study, or state). Source: U.S. Census Bureau.
11. **Median housing value** (*Median Housing Value*). The natural logarithm of median housing value from a compounded value using data from the 2000 U.S. decennial survey

and the Federal Housing Finance Agency. Source: U.S. Census Bureau and Federal Housing Finance Agency (county level, aggregated to other levels).

12. **Median income** (*Median Income*). The natural logarithm of the population-weighted county-level median income for a given area. Source: U.S. Census Bureau (county level, adjusted for other levels by population weight).
13. **Political fragmentation index** (*Political Fragmentation*). This variable is used to control for zoning behavior, highlighting the redundancy of sub-areas and proxying for institutional discrimination as practiced by Lichter, Parisi, and Taquino (2015). In our context, it equals the ratio of the number of census tracts times 5,000 (the average population of a census tract) to the total population in a given area, capturing the extent to which extra census tracts are set up relative to the total population. We use the natural logarithm of this ratio as our political fragmentation index.
14. **Social trust index** (*Trust Index*). A ratio derived from responses to the question 'Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?' The data for inter-wave years are obtained by interpolating the two nearest available data points. More details can be found in Wei and Zhang (2020). Source: World Values Survey (WVS).
15. **Racial diversity index** (*Racial Diversity*). Regional Herfindahl index considering five racial and ethnic shares: White, Asian, Hispanic, African American, and others for corresponding geographic scopes.

## Part II: Regional Capital Market Variables

1. **Regional stock return** (*Regional Ret*). Market value-weighted average return of stocks located in an area (county, commuting zone, local area defined in our study, or state).
2. **Regional market capitalization** (*Regional Mktcap*). Summation of the market capitalization of firms in an area (county, commuting zone, local area defined in our study, or state), taking the form of a natural logarithm in regressions.
3. **Number of regional stocks** (*Regional Nstocks*). The number of publicly listed stocks within an area (county, commuting zone, local area defined in our study, or state). Source: CRSP monthly file.
4. **Regional dividend-paying firm ratio** (*Regional Fpd*). The ratio of firms paying dividends in a given area (county, commuting zone, local area defined in our study, or state).
5. **Regional illiquidity** (*Regional Illiq*). We first calculate the Amihud illiquidity at the firm level, then aggregate it to the regional level (county, commuting zone, local area defined in our study, or state) based on the market capitalization of each firm.
6. **Regional institutional ownership** (*Regional IO*). Fraction of institutional ownership, calculated from Thomson CDA/Spectrum institutional ownership database (13F), aggregating to the regional level (county, commuting zone, local area defined in our study, or state) based on the market fractions of firms. Source: Thomson CDA/Spectrum (13F).

## Part III: Firm-Level Variables (Quarterly)

1. **Institutional ownership ratio** (*IOR*). The fraction of ownership held by institutional investors who hold 5% or more of shares outstanding. Source: Thomson CDA/Spectrum (13F).
2. **Lagged market capitalization** (*Market Cap (t-1)*). The natural logarithm of market capitalization ( $/\$10^6$ ) in the previous quarter.

3. **Inverse trading volume** (*Inverse Volume*). The inverse of total quarterly trading volume (in millions). Source: CRSP stock files.
4. **Lagged stock price** (*Prc (t-1)*). The value is in the form of a natural logarithm. Source: CRSP stock files.
5. **Lagged stock return** (*Ret (t-1)*). The value is in the form of a natural logarithm. Source: CRSP stock files.
6. **Lagged price volatility** (*Volp (t-1)*). The standard deviation of the stock price. Source: CRSP stock files.
7. **Amihud illiquidity measure** (*Illiquidity*). This variable is based on the Amihud (2002) illiquidity measure, which is the absolute return over the dollar stock volume using a daily frequency and then averaged within quarter  $t$ . In our context, we use the  $\log(\text{Amihud illiquidity} \times 1012 + 1)$ .
8. **Analyst coverage** (*Analyst Cover*). Natural logarithm of one plus the number of analysts following the firm in this quarter. Source: I/B/E/S.

#### Part IV: Firm-Level Variables (Annually)

1. **Firm age** (*Age*). The number of years since the firm first appeared in the Compustat database, taking the form of a natural logarithm in regressions. Source: Compustat.
2. **Investment ratio** (*Invest*). The summation of capital expenditures and R&D costs, divided by total assets. Source: Compustat.
3. **Tangibility ratio** (*Tangibility*). The ratio of tangible assets to total assets. Source: Compustat.
4. **Cash holdings ratio** (*Cash Holdings*). The ratio of cash holdings over total assets. Source: Compustat.
5. **Cash flow ratio** (*Cash Flow*). The ratio of Earnings Before Interest and Taxes (EBIT) over total assets.
6. **Leverage ratio** (*Leverage*). The ratio of book debt to total assets. Source: Compustat.
7. **Total assets** (*Total assets*). The natural logarithm of total assets measured in millions of dollars. Source: Compustat.
8. **Sales ratio** (*Sales*). The ratio of sales over total assets. Source: Compustat.
9. **Amihud illiquidity measure** (*Illiquidity*). This variable is based on the Amihud (2002) illiquidity measure, which is the absolute return over the dollar stock volume using a daily frequency and then averaged within year  $t$ . In our context, we use the  $\log(\text{Amihud\_illiquidity} \times 10^{12} + 1)$ .
10. **Book-to-market ratio** (*B/M*). The ratio of book value over market capitalization.
11. **Payout ratio** (*Payout*). The value of total payout over total assets in the previous year. Source: Compustat.
12. **R&D expense disclosure indicator** (*R&D (0/1)*). Equals 1 if the firm reported its R&D cost in the previous year. Source: Compustat.
13. **Advertising expense disclosure indicator** (*Advertising (0/1)*). Equals 1 if the firm reported its advertising expenditure in the previous year. Source: Compustat.

#### Part V: Fund-Level Variables

1. **Number of portfolio stocks** (*Fund Nstocks*). The number of stocks in portfolios is taking the form of a natural logarithm in regressions. Source: CRSP mutual fund data.
2. **Net fund flow ratio** (*Net Flow*). The quarterly flow ratio of mutual funds from the CRSP mutual fund data. The calculation follows Friesen and Sapp (2007).

3. **Fund age** (*Fund Age*). The time length between the initial year of the mutual fund and the current year of the current record is taken in the form of a natural logarithm in regressions. The calculation is based on the CRSP mutual fund data.
4. **Expense ratio** (*Expense Ratio*). The annual expense ratio of mutual funds is from the CRSP mutual fund data.
5. **Total net assets** (*TNA*). The total net assets of mutual funds (in millions) taking the form of a natural logarithm in regressions. The calculation is based on the CRSP mutual fund data and the Thomson Reuters database.
6. **Turnover ratio** (*Turnover*). The annual turnover ratio of the mutual fund.
7. **Fund family total assets** (*Fund Family Assets*). The summation of all total net assets of mutual funds belonging to one asset management firm, taking the form of a natural logarithm in regressions. The calculation is based on the CRSP mutual fund data.
8. **Fund family size** (*Fund Family Size*). The number of different share classes of an asset management firm the mutual fund belongs to, taking the form of the natural logarithm. The calculation is based on the CRSP mutual fund data.
9. **Team-managed fund indicator** (*Team Managed (0/1)*). Equals 1 if the number of managers is  $\geq 4$ , based on the CRSP mutual fund data.
10. **Investment style** (*Style*). Each quarter, we equally divide our sample into four categories: *Growth/value*  $\times$  *Small/Large* based on their portfolio's factor loading. This method comes from Jiang, Zaynutdinova, and Zhang (2021).

## Appendix 2A.2: Supplementary information

### Appendix 2A.2.1: Race and ethnicity classifications

**Table 2A.2.1 Classification for Race and Ethnicity from the US Census Bureau**

Category	Average Count	Percentage
Type 1: Hispanic or Latino (of any race)	53,076,668	17.1%
Type 2: White alone	194,699,100	62.7%
Type 3: African American alone	36,995,090	11.9%
Type 4: Native American and Alaska Native alone	2,200,525	0.7%
Type 5: Asian alone	14,657,800	4.7%
Type 6: Native Hawaiian and Other Pacific Islanders alone	484,662	0.2%
Type 7: Some other races alone	834,954	0.3%
Type 8: Two or more races	7,580,616	2.4%

Note: All statistics come from the 2000-2020 American Decennial Census of Population and Housing

### Appendix 2A.2.2: Sample restrictions for reliable measurement of racial integration

Following established guidance in the literature, we exclude regions where the combined Black, Asian, and Hispanic populations account for less than 5% of total residents at any point in time, as very small minority shares can distort measures of racial residential segregation. This filter enhances the reliability of the RRI measure. Our results are robust to using a higher threshold of 10% or to including all commuting zones without exclusions; these alternative specifications are available upon request. This exclusion affects fewer than 2.5% of observations in both firm-level and fund-level analyses. In the following table, we show how the threshold of our total minority population impacts the record numbers and sample sizes.

Category	No threshold	5% threshold	10% threshold
The number of CZs	706	464	340
Price Informativeness Test (Records)	24,520	23,976	22,866
Price Informativeness Test (CZ Numbers)	197	165	133
Information Asymmetry Test (Records)	150,679	147,191	140,485
Information Asymmetry Test (CZ Numbers)	231	185	144
Mutual Fund Test (Records)	49,437	49,350	48,702
Mutual Fund Test (CZ Numbers)	67	64	59

Note: The record loss from 0 to 5% threshold across all firm-level tests is below 2.3%.

### **Appendix 2A.2.3: Extra information for Section 2.2.3.1 and for Section 2.4.1 - price informativeness**

**(1) Detailed zip code extraction (for Section 2.2.3.1):** For the period 2007–2018, we obtain zip code data from the CRSP/Compustat Merged Database. For 2000–2006, we extracted zip codes from 10-K and 10-Q filings. If a zip code is unavailable, we assign the most recently observed zip code; if that is also missing, we rely on Compustat's latest headquarters listing (Pirinsky and Wang, 2006; Parsons, Sulaeman, and Titman, 2020). We rely exclusively on Compustat headquarters data for only 61 of the 2,479 firms in our final sample (for the information asymmetry test, we rely exclusively on Compustat headquarters data for only 162 of the 5,498 firms in our final sample). These firms, which do not disclose their 10-Q/10-K filings and are not included in the CRSP/Compustat Merged Database, tend to be small and compete for a limited share of overall market capitalization and mutual fund holdings. Additionally, smaller firms are less likely to relocate their headquarters due to the relatively high costs associated with such actions.

**(2) Full control list and variable description (for Section 2.4.1):** These include contemporaneous EBFF-to-asset ratio, leverage, investment, tangible assets, total assets, Cash Holdings, sales-to-asset ratio, stock return volatility, analyst coverage, blockholder ownership, and illiquidity. Appendix A provides detailed definitions of all control variables.

**(3) Board racial diversity:** We also include board racial diversity, which may independently affect market valuation. Following, we measure diversity using director-level race and ethnicity classifications from NamePrism, which assigns directors to one of five categories (White, African American, Asian, Hispanic, or Other) based on name-based probabilities.<sup>36</sup> A director is included in a category if their estimated likelihood of belonging to that group exceeds 50%. Director names are primarily obtained from ISS, with BoardEx used

---

<sup>36</sup> NamePrism has become increasingly popular in finance and economics research (e.g., Pool, Stoffman, and Yonker, 2015; Diamond, McQuade, and Qian, 2019). For a detailed discussion on NamePrism, see Diamond et al. (2019). The application can be found here: <https://www.name-prism.com/>.

as a secondary source. For firm-year observations with incomplete diversity data, we impute the firm's historical average diversity to preserve sample size. Omitting the assignment of historical board diversity for records with insufficient data would result in a loss of approximately 30% of the sample. The imputation approach used is consistent with the inclusion of firm fixed effects.

#### **Appendix 2A.2.4: Extra information for Section 2.2.5 - geographic controls**

**(1) Demographic and socioeconomic characteristics** We include the natural log of population (Christensen and Sarkissian, 2009); median housing value; median household income; population density; education attainment; unemployment Ratio (BLS); the proportion of seniors (age  $\geq 65$ ); and the female share (Shu, Sulaeman, and Yeung, 2012; Wei and Zhang, 2020). Unless directly available at the Commuting zone (CZ) level, county-level series are aggregated to CZs via population-weighted averages each year. These variables capture local economic opportunity, housing affordability, labor-market tightness, and demographic composition, all of which shape information frictions and investor attention.

**(2) Racial composition and diversity.** Following Trounstein (2016), we control for CZ-level racial diversity and the shares of major racial groups (White, African American, Asian, Hispanic), calculated from Census microdata aggregated to CZs. These measures separate the effect of RRI from contemporaneous racial mix and heterogeneity that may independently affect social interaction, political preferences, and capital allocation.

**(3) Social trust:** Following the literature (Wei and Zhang, 2020), we include the social-trust index, derived from the World Values Survey question, "Generally speaking, would you say that most people can be trusted, or that you need to be very careful in dealing with people?" The WVS reports trust for 10 U.S. regions: *New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Rocky Mountain, Northwest, and California*. We assign each county the trust value of its region and aggregate to CZs as the population-weighted average across constituent counties each year. Social trust tends to increase with RRI (Uslaner, 2010; Alesina and Zhuravskaya, 2011; Rothwell, 2012) and is known to affect local information advantages and portfolio choices (Wei and Zhang, 2020); controlling for it helps isolate RRI's effect from broader cooperative norms.

**(4) Institutional discrimination (zoning redundancy).** To capture land-use restrictiveness that may facilitate spatial exclusion independent of social interaction, we follow Lichter et al. (2015) and compute zoning redundancy as:

$$\text{Zoning Redundancy}_{CZ,t} = \frac{\#\{\text{Census Tracts in CZ}\} \times 5,000}{\text{Population}_{CZ,t}}$$

where 5,000 approximates the average tract population. Higher values proxy for finer-grained, potentially exclusionary zoning intensity.

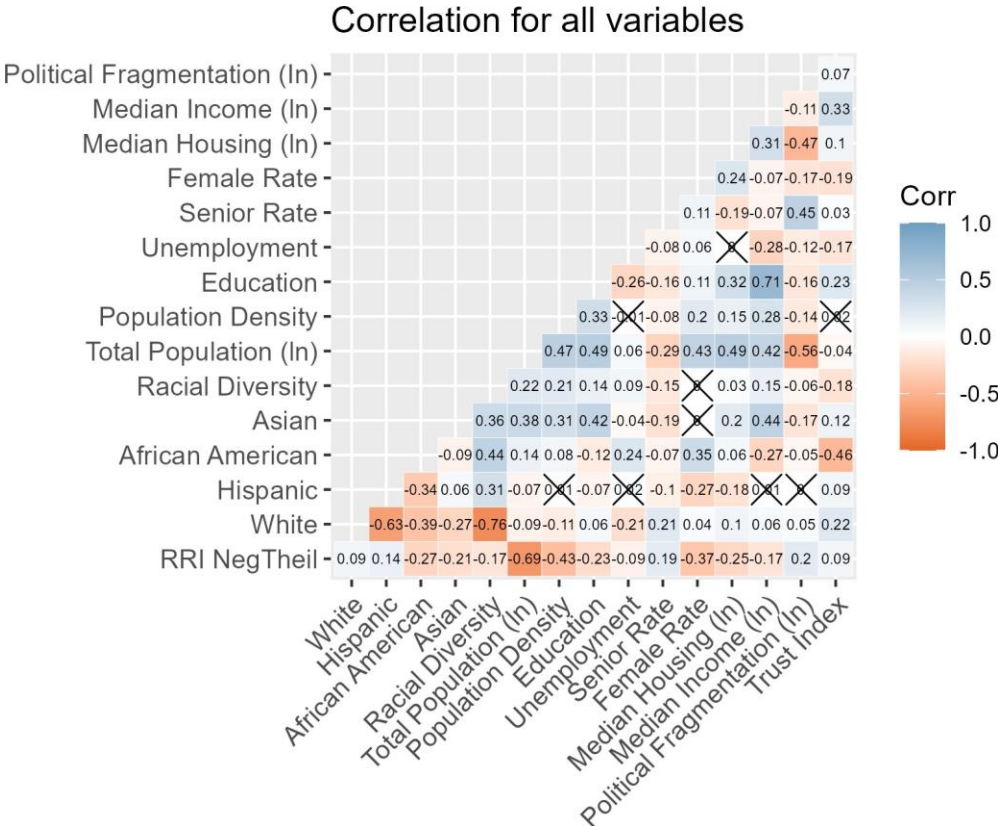
**(5) Local stock-market conditions.** Using CRSP/Compustat, for each CZ-quarter we construct (i) total market capitalization (sum across locally headquartered listed firms), (ii) the number of listed firms, (iii) the fraction of dividend-paying firms, (iv) the value-weighted average return of locally headquartered firms over the horizon matching the dependent variable, and (v) Amihud illiquidity, computed firm-by-firm and aggregated to the CZ by value weights (Bernile et al., 2015). These variables flexibly absorb local market depth, maturity, performance, and trading frictions that could confound the relation between RRI and information outcomes.

**(6) Institutional-investor clustering.** We proxy for institutional clustering with CZ-level institutional ownership, computed as the market-cap-weighted average institutional ownership across locally headquartered firms (from WRDS/13F). Concentrations of institutional capital can enhance price discovery via trading intensity and analyst coverage, potentially improving local information environments (Kacperczyk et al., 2021; Kim et al., 2022).

**(7) Scope and placement.** All geographic controls enter both firm-level and fund-level specifications because they jointly affect firms' information environments and institutions' ability to identify mispricing. Summary statistics are reported in Panel A of Table 2.1.

**Appendix 2A.2.5: Correlation analysis for Section 2.2.5 - geographic controls**

**Figure 2A.1: Correlation Analysis for Demographic Variables**



This picture provides the correlations between different demographic and socioeconomic variables for our sample at the commuting zone level from 2000 to 2020. Definitions of all variables can be found in Appendix A. The 'X' marks an insignificant correlation (insignificant at the 0.05 level).

#### **Appendix 2A.2.6: Extra information for Section 2.4.2 - information asymmetry**

**Full control list and variable descriptions (for Section 2.4.2):** To isolate the effect of RRI, we control for a comprehensive list of firm characteristics known to influence information asymmetry from the latest fiscal year-end, including firm age, leverage, tangible assets, Cash Holdings, book-to-market ratio, sales-to-assets ratio, payout ratio, indicators for R&D and advertising disclosures, institutional block ownership, analyst coverage (Boone and White, 2015; Amiram, Owens, and Rozenbaum, 2016), as well as board racial diversity. For firm-quarter observations with missing board diversity data, we impute values using the historical average diversity of nearby firms within the same commuting zone to minimize data loss. Without assigning the average board diversity from nearby firms, nearly half of the firm-quarter observations would be excluded from both the adverse selection and VCV analyses due to missing data. Imputing board diversity using neighboring firms within the same commuting zone is consistent with the use of commuting zone fixed effects in our empirical specifications, as discussed in Section 2.4.2. We also include market-related variables such as lagged returns, lagged price level, lagged price volatility, and lagged market capitalization (Amiram et al., 2016; Cui et al., 2018). Detailed definitions of these variables are provided in Appendix 2A.1.

#### **Appendix 2A.2.7: Extra information for Section 2.4.2 - information asymmetry**

**Full control list and variable descriptions (for Section 2.5.1):** We control for a range of fund-level characteristics associated with mutual fund performance, including trading style, net flow, turnover ratio, fund load, fund age, expense ratio, total net assets, indicators for single-manager and multi-manager (more than four managers) funds, number of share classes in the fund family, and total net assets of the fund family (Coval and Moskowitz, 2001; Cohen et al., 2008; Kacperczyk et al., 2014; Chuprunin and Sosyura, 2018). Detailed definitions of these variables are provided in Appendix 2A.1.

## Appendix 2A.3: Full tables and extra tests

### Appendix 2A.3.1: RRI and regional social connectedness

To evaluate the relationship between RRI and social networks, we construct the Social Connectedness Index (SCI) following the approach of Bailey et al. (2021). The SCI is derived from county-pair social connectedness data provided by Facebook and is calculated as:

$$SCI_j = \sum_{(c_1, c_2 \in j)} \frac{N_{c_1}}{N_j} \times \frac{N_{c_2}}{N_j} \times SC_{(c_1, c_2)} \times 10^{-12},$$

where  $N_{c_1}$  and  $N_{c_2}$  denote the populations of counties  $c_1$  and  $c_2$  within commuting zone  $j$ , and  $SC_{(c_1, c_2)}$  is the scaled probability (multiplied by  $10^{12}$ ) that two randomly selected individuals from these counties are Facebook friends (Bailey et al., 2018). Thus,  $SCI_j$  reflects the likelihood that two randomly selected individuals within the same commuting zone are socially connected online.

Since Facebook's Social Connectedness Index is based on data collected in October 2021 and does not include earlier periods, we focus on the cross-sectional relationship between racial integration as of 2020 and commuting zone social connectedness in 2021. To improve the reliability of the SCI measure, we exclude commuting zones that both (1) consist of only a single county and (2) are classified as rural or small urban non-metro areas based on the population size of their largest place.<sup>37</sup> These zones tend to be sparsely populated and geographically isolated, offering limited variation in demographic and social conditions. While SCI includes within-county connections, the social network structure in such areas may reflect highly localized clustering rather than meaningful regional connectedness. This filtering approach aligns with that of Bailey et al. (2018), who show that SCI is more robust and interpretable in larger, more interconnected regions. After applying this exclusion, we retain 452 commuting zones for analysis. The SCI ( $\times 10^{12}$ ) has a mean of 9.627 (SD = 14.171); P25

---

<sup>37</sup> Classification follows U.S. Department of Agriculture (USDA) definitions based on the 1990 population of the largest place in a commuting zone: "Small Town/Rural" (<5,000) and "Small Urban Center" (5,000–20,000). See: <https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/documentation/>. This filter removes 11 of the 463 commuting zones with valid RRI values and SCI values.

= 1.535, median = 4.470, and P75 = 11.694.

Building on contact theory, we expect more racially integrated communities to also foster social ties across racial and demographic lines. Using the Social Connectedness Index (SCI) constructed from Facebook data to capture interpersonal networks (Bailey, Cao, Kuchler, Stroebel, and Wong, 2018), we find a positive correlation of 0.43 between RRI in 2020 and SCI in 2021 at the commuting zone level, suggesting that communities with higher RRI also have greater social connectedness.<sup>38</sup> To control for confounding factors, we estimate the following cross-sectional regression:

$$SCI_{j,2021} = \beta_0 + \beta_1 RRI_{j,2020} + \beta X_{j,2020} + \epsilon_{j,2021},$$

where  $SCI_{j,2021}$  denotes social connectedness in commuting zone  $j$ , as defined in Equation (12), and  $RRI_{j,2020}$  is the RRI in commuting zone  $j$  in 2020. The control vector  $\mathbf{X}_{j,2020}$  includes the demographic and socioeconomic variables described in Section 2.2.5. All variables are winsorized at the 1st and 99th percentiles.

As shown in Table 2A.3.1, RRI remains positively and significantly associated with SCI across all model specifications. In column (2), the coefficient of 10.44 implies that a one-standard-deviation increase in RRI (0.10) is associated with a 10.84% increase in SCI, relative to its mean of 9.63. These findings indicate that greater RRI is associated with stronger local social connectedness, supporting the view that racial integration improves economic opportunity by contributing to the breadth and diversity of social networks.

---

<sup>38</sup> To address the concern that the observed correlation between SCI and RRI reflects population size rather than true social connectedness, we orthogonalize both variables with respect to the inverse of the 2020 commuting zone population, as population levels were stable between 2020 and 2021. Empirically,  $SCI$  and  $1/N_{j,2020}$  have a Pearson correlation of 0.95 ( $p < .01$ ), while  $RRI$  and  $1/N_{j,2020}$  show a correlation of 0.42 ( $p < .01$ ). We perform the orthogonalization by estimating:  $SCI$  or  $RRI = \beta_0 + \beta_1(1/N_{j,2020}) + \epsilon_j$ , where  $\epsilon_j$  is the residual component of  $SCI$  or  $RRI$  uncorrelated with population size. The adjusted  $SCI$ – $RRI$  correlation based on the orthogonalized term drops to 0.09 but remains significant at the 5% level.

**Table 2A.3.1: RRI and the Social Connectedness Index**

This table presents the relationship between RRI and the Social Connectedness Index (SCI). The dependent variable in columns (1) and (2) is the level of *SCI*, measured as the probability that two randomly selected individuals within a commuting zone are Facebook friends. Columns (3) and (4) use an orthogonalized version of *SCI*, constructed by regressing *SCI* on the inverse of the 2020 commuting zone population and retaining the residual. All dependent variables are scaled by  $10^{12}$  for readability. The key independent variable, *RRI NegTheil*, captures the evenness dimension of racial integration (Section 2.2.1). Columns (2) and (4) add controls for demographic and socioeconomic characteristics, including racial composition (White, Black, Asian, Hispanic), racial diversity, zoning behavior, median housing value, median income, population size and density, education level, unemployment ratio, share of senior citizens, female population share, and social trust (Section 2.2.5); we also control for the natural log of the number of counties in a commuting zone, *County Number*. Standard errors are in parentheses. \* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ .

	SCI		Orthogonalized SCI	
	(1)	(2)	(3)	(4)
RRI NegTheil	69.10*** (7.074)	10.44** (4.959)		
RRI NegTheil (Orthogonalized)			5.305** (2.672)	5.047** (2.452)
Constant	20.90*** (1.304)	9.683 (35.23)	0.1442 (0.2143)	33.52* (17.83)
Senior Ratio		0.6013 (10.81)		-9.961* (5.426)
Female Ratio		30.83 (24.56)		0.0090 (12.38)
Population Density		0.0303*** (0.0051)		0.0068*** (0.0026)
Total Population (ln)		-6.347*** (0.5213)		-1.210*** (0.2571)
Education		-6.952 (5.697)		-16.82*** (2.858)
Trust Index		1.084 (5.838)		-0.7638 (2.932)
Unemployment		-52.26 (33.28)		-6.434 (16.70)
County Number (ln)		-0.5750 (0.8032)		-0.1152 (0.4032)
Racial Diversity		-3.535 (3.993)		-10.55*** (2.009)
Hispanic (%)		8.570* (4.751)		-5.285** (2.396)
African American (%)		19.88*** (5.379)		15.46*** (2.711)
Asian (%)		77.46*** (19.84)		32.21*** (10.00)
White (%)		3.380 (5.660)		-7.899*** (2.855)
Political Fragmentation		13.08*** (2.763)		0.4755 (1.411)
Median Housing (ln)		-2.820*** (0.1591)		-0.2754*** (0.0821)
Median Income (ln)		8.107*** (2.867)		-0.0556 (1.443)
Observations	452	452	452	452
$R^2$	0.175	0.830	0.009	0.588

**Table 2A.3.2: Complete Version of Table 2.3**

**Panel A: RRI and Price Informativeness**

	Dependent Variable: Future Cash Flow		
	(1)	(2)	(3)
RRI Similarity × log(M/A)	0.043** (0.019)		
RRI NegTheil (White VS Minority) × log(M/A)		0.041** (0.018)	
RRI NegTheil (all) × log(M/A)			0.037* (0.022)
RRI Similarity	-0.118** (0.050)		
RRI NegTheil (White VS Minority)		-0.135*** (0.049)	
RRI NegTheil (all)			-0.150*** (0.046)
log(M/A)	0.061*** (0.01)	0.048*** (0.010)	0.047*** (0.010)
Racial Diversity × log(M/A)	0.009 (0.015)	0.009 (0.015)	0.009 (0.015)
log(M/A) × Board Diversity	0.009** (0.004)	0.009** (0.004)	0.009** (0.004)
Age	0.008** (0.004)	0.008** (0.004)	0.008** (0.004)
Sale	0.022*** (0.004)	0.022*** (0.004)	0.022*** (0.004)
Tangibility	-0.041*** (0.013)	-0.041*** (0.013)	-0.041*** (0.013)
Cash Flow (t)	0.529*** (0.029)	0.528*** (0.029)	0.528*** (0.029)
Invest	-0.108*** (0.022)	-0.107*** (0.022)	-0.107*** (0.022)
Illiquidity	0.0003 (0.0004)	0.0003 (0.0004)	0.0003 (0.0004)
Volp (ln)	-0.004*** (0.0006)	-0.004*** (0.0006)	-0.004*** (0.0006)
Analyst Cover (ln)	-0.003* (0.002)	-0.003* (0.002)	-0.003* (0.002)
Leverage	0.047*** (0.010)	0.047*** (0.010)	0.047*** (0.010)
Cash Holdings	-0.040*** (0.012)	-0.040*** (0.012)	-0.040*** (0.012)
Asset (ln)	-0.012*** (0.003)	-0.012*** (0.003)	-0.012*** (0.003)
IOR	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)
Senior Ratio	-0.009 (0.113)	-0.017 (0.113)	-0.037 (0.113)
Female Ratio	-1.18* (0.518)	1.33** (0.537)	-1.28* (0.521)
Education	0.114* (0.062)	0.131** (0.062)	0.123** (0.058)
Population Density	4.02×10 <sup>-6</sup> (4.09×10 <sup>-6</sup> )	3.83×10 <sup>-6</sup> (4×10 <sup>-6</sup> )	5.13×10 <sup>-6</sup> (4.01×10 <sup>-6</sup> )
Trust Index	0.043** (0.019)	0.044** (0.019)	0.043** (0.019)
Unemployment	0.061 (0.081)	0.062 (0.083)	0.063 (0.085)
Racial Diversity	0.004 (0.058)	-0.003 (0.056)	-0.004 (0.056)
Total Population (ln)	-0.012* (0.007)	-0.013* (0.007)	-0.013* (0.007)
White (%)	-0.029 (0.136)	-0.050 (0.144)	-0.056 (0.145)
Hispanic (%)	0.008	-0.018	-0.026

	(0.122)	(0.127)	(0.128)
African American (%)	0.004	-0.018	-0.031
	(0.121)	(0.126)	(0.126)
Asian (%)	0.085	0.088	0.057
	(0.158)	(0.163)	(0.166)
Political Fragmentation	0.007	0.007	0.007
	(0.021)	(0.021)	(0.021)
Median Housing (ln)	-0.012*	-0.012*	-0.010
	(0.007)	(0.007)	(0.007)
Median Income (ln)	-0.015	-0.020	-0.020
	(0.019)	(0.019)	(0.019)
Regional Illiq	0.007	0.010	-5.35×10 <sup>-5</sup>
	(0.152)	(0.152)	(0.152)
Regional Ret	0.021***	0.021***	0.021***
	(0.007)	(0.007)	(0.007)
Regional IO	0.017**	0.017**	0.016**
	(0.007)	(0.007)	(0.007)
Regional Mtkcap (ln)	-0.006**	-0.006**	-0.006**
	(0.002)	(0.002)	(0.002)
Regional Nstocks (ln)	0.005	0.006	0.006
	(0.004)	(0.004)	(0.004)
Regional Fpd	-0.013*	-0.012*	-0.013*
	(0.007)	(0.007)	(0.007)
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	23,976	23,976	23,976
R <sup>2</sup>	0.82271	0.82278	0.82276

**Panel B: RRI and Information Asymmetry**

Dependent variable:	Adverse Selection (%)					Volume Coefficient of Variation				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
RRI NegTheil	-0.895*** (0.135)	-0.173*** (0.051)	-0.703*** (0.900)	-2.06** (0.647)	-2.05*** (0.647)	-0.11 (0.084)	-0.035 (0.030)	-0.060** (0.027)	-0.800*** (0.278)	-0.781*** (0.275)
RRI NegTheil × IOR					-0.173 (0.231)					-0.413** (0.162)
Age		0.020** (0.008)	-0.069*** (0.009)	0.021*** (0.008)	0.021*** (0.008)		-0.045*** (0.005)	-0.040*** (0.004)	-0.034*** (0.004)	-0.034*** (0.004)
Inversed Volume		0.678*** (0.030)	0.672*** (0.034)	0.677*** (0.030)	0.677*** (0.030)		0.597*** (0.035)	0.591*** (0.033)	0.569*** (0.032)	0.569*** (0.032)
Sale		0.007 (0.008)	0.017 (0.012)	0.006 (0.009)	0.006 (0.009)		-0.014*** (0.004)	-0.026*** (0.004)	-0.022*** (0.004)	-0.022*** (0.004)
Tangibility		0.001 (0.027)	0.118*** (0.044)	-0.021 (0.031)	-0.021 (0.031)		-0.075*** (0.014)	-0.083*** (0.019)	-0.085*** (0.019)	-0.086*** (0.019)
B/M		0.045*** (0.009)	0.082*** (0.014)	0.042*** (0.010)	0.042*** (0.010)		-0.026*** (0.005)	-0.028*** (0.005)	-0.021*** (0.005)	-0.021*** (0.005)
Leverage		0.224*** (0.033)	0.302*** (0.048)	0.211*** (0.039)	0.211*** (0.039)		0.019 (0.021)	0.015 (0.021)	0.005 (0.021)	0.005 (0.021)
Cash Holdings		-0.037 (0.033)	-0.140*** (0.032)	-0.046 (0.031)	-0.046 (0.031)		0.014 (0.023)	-0.016 (0.025)	-0.020 (0.025)	-0.020 (0.025)
Payout		0.003 (0.003)	0.011** (0.005)	0.002 (0.003)	0.002 (0.003)		0.002* (0.001)	0.001 (0.001)	0.001 (0.0010)	0.001 (0.001)
R&D (0/1)		-0.009 (0.012)	-0.006 (0.018)	-0.011 (0.014)	-0.011 (0.014)		0.0002 (0.007)	-0.007 (0.007)	-0.007 (0.006)	-0.007 (0.006)
Advertising		-0.002 (0.008)	-0.086*** (0.014)	0.007 (0.009)	0.008 (0.009)		0.016*** (0.006)	-0.005 (0.007)	2.88 × 10 <sup>-5</sup> (0.007)	7.75 × 10 <sup>-5</sup> (0.007)
Ret (t-1)		-0.059*** (0.015)	-0.090*** (0.017)	-0.062*** (0.015)	-0.062*** (0.015)		-0.033*** (0.004)	-0.014*** (0.004)	-0.032*** (0.004)	-0.032*** (0.004)
Pre (t-1)		0.007 (0.005)	-0.014** (0.006)	0.008* (0.005)	0.008* (0.005)		0.017*** (0.003)	0.017*** (0.002)	0.018*** (0.003)	0.018*** (0.003)
Volp (t-1)		-0.033*** (0.008)	0.036*** (0.008)	-0.030*** (0.008)	-0.030*** (0.008)		-0.018*** (0.005)	-0.013*** (0.005)	-0.016*** (0.005)	-0.016*** (0.005)
Analyst Cover (ln)		-0.018** (0.008)	-0.041*** (0.008)	-0.018** (0.008)	-0.018** (0.008)		-0.053*** (0.004)	-0.055*** (0.003)	-0.052*** (0.003)	-0.052*** (0.003)
IOR		-0.065* (0.039)	-0.409*** (0.043)	-0.087** (0.041)	-0.143 (0.088)		-0.054*** (0.018)	-0.070*** (0.015)	-0.053*** (0.018)	-0.182*** (0.057)
diversity		0.040 (0.029)	0.132*** (0.039)	0.084*** (0.028)	0.084*** (0.028)		0.014 (0.012)	-0.008 (0.011)	0.010 (0.010)	0.010 (0.011)
Market Cap (t-1)		-0.032*** (0.006)	-0.004 (0.006)	-0.033*** (0.006)	-0.033*** (0.006)		-0.099*** (0.003)	-0.101*** (0.003)	-0.101*** (0.003)	-0.101*** (0.003)
Political Fragmentation				-0.056 (0.202)	-0.056 (0.202)				-0.107 (0.068)	-0.107 (0.068)
Racial Diversity				0.679 (0.561)	0.684 (0.561)				0.284 (0.234)	0.286 (0.236)

White (%)				2.71 (2.08)	2.70 (2.08)			0.473 (0.666)	0.463 (0.667)
Hispanic (%)				1.85 (1.69)	1.84 (1.69)			0.157 (0.613)	0.117 (0.613)
African American (%)				1.05 (2.17)	1.05 (2.17)			-0.239 (0.685)	-0.267 (0.682)
Asian (%)				4.20* (2.45)	4.18* (2.46)			0.712 (0.927)	0.673 (0.931)
Total Population (ln)				-0.131 (0.205)	-0.127 (0.205)			-0.117 (0.100)	-0.106 (0.101)
Senior Ratio				0.965 (2.05)	0.967 (2.05)			-0.737 (0.835)	-0.757 (0.833)
Female Ratio				13.1** (5.82)	13.1** (5.82)			5.49** (2.51)	5.38** (2.53)
Education				1.87** (0.777)	1.84** (0.777)			0.717** (0.362)	0.709* (0.360)
Population Density				-0.0006 (0.0004)	-0.0006 (0.0004)			0.0006*** (0.0001)	0.0005*** (0.0001)
Trust Index				0.436** (0.172)	0.438** (0.172)			0.032 (0.071)	0.041 (0.072)
Median Housing (ln)				0.092 (0.064)	0.092 (0.064)			0.025 (0.024)	0.026 (0.024)
Median Income (ln)				-0.289 (0.238)	-0.286 (0.238)			-0.078 (0.088)	-0.077 (0.087)
Unemployment				0.145 (0.798)	0.139 (0.798)			-0.207 (0.244)	-0.205 (0.246)
Regional Illiq				-2.25 (1.37)	-2.26 (1.37)			2.04*** (0.550)	2.04*** (0.551)
Regional Ret				0.006 (0.051)	0.006 (0.051)			-0.012 (0.016)	-0.012 (0.016)
Regional Mktcap (ln)				-0.033 (0.029)	-0.033 (0.029)			0.010 (0.007)	0.010 (0.007)
Regional IO				-0.070 (0.067)	-0.069 (0.067)			-0.003 (0.026)	-0.002 (0.026)
Regional Nstocks (ln)				-0.0009 (0.044)	-0.0008 (0.044)			-0.012 (0.015)	-0.012 (0.015)
Regional Fpd				0.121* (0.062)	0.121* (0.062)			-0.026 (0.023)	-0.026 (0.023)
Constant	0.125*** (0.045)					0.810*** (0.028)			
Year × Quarter FE		Yes		Yes	Yes		Yes	Yes	Yes
Industry FE			Yes	Yes	Yes		Yes	Yes	Yes
Commuting Zone FE									
Observations	29,088	29,088	29,088	29,088	29,088	147,191	147,191	147,191	147,191
R <sup>2</sup>	0.01814	0.44492	0.16759	0.43314	0.43316	0.09040	0.35617	0.35143	0.36708

**Table 2A.3.3: Complete Version of Table 2.4**

**Panel A: Sorted by Size**

Firms are sorted by the median value of the total assets at each time point. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Dependent Variables: Groups:	EBIT(t+1)/AT(t)		Adverse Selection (%)		VCV	
	(1) Smaller	(2) Larger	(3) Smaller	(4) Larger	(5) Smaller	(6) Larger
log(M/B) × RRI NegTheil	0.052* (0.028)	0.007 (0.021)				
RRI NegTheil	-0.175*** (0.058)	-0.088* (0.049)	-2.45** (1.18)	-1.01* (0.62)	-1.52*** (0.579)	-0.397 (0.244)
Age	0.010* (0.006)	0.009** (0.004)	0.028** (0.013)	0.002 (0.009)	-0.057*** (0.008)	-0.054*** (0.004)
Sale	-0.0009 (0.005)	-0.0004 (0.004)	-0.003 (0.012)	0.012 (0.009)	0.013* (0.007)	0.009** (0.004)
Tangibility	-0.032* (0.018)	-0.057*** (0.016)	-0.11** (0.052)	-0.003 (0.027)	-0.095** (0.037)	-0.030** (0.015)
Analyst Cover (ln)	-0.003 (0.003)	-0.004 (0.002)	-0.032*** (0.012)	-0.006 (0.006)	-0.050*** (0.006)	-0.050*** (0.003)
Leverage	0.050*** (0.013)	0.052*** (0.010)	-0.247*** (0.072)	-0.102* (0.058)	-0.002 (0.033)	0.050*** (0.017)
Cash Holdings	-0.044*** (0.011)	-0.010 (0.010)	-0.073** (0.037)	-0.027 (0.025)	-0.033 (0.023)	-0.012 (0.014)
IOR	-0.006 (0.011)	-0.014* (0.008)	-0.105* (0.058)	-0.054 (0.035)	0.093*** (0.027)	0.005 (0.022)
Racial Diversity	0.046 (0.071)	0.017 (0.079)	1.77* (0.990)	-0.329 (0.459)	0.705* (0.370)	-0.098 (0.172)
Board Diversity	-0.012* (0.065)	-0.008 (0.065)	0.060 (0.040)	0.077** (0.032)	0.032 (0.020)	-0.013 (0.012)
Senior Ratio	-0.020 (0.178)	0.107 (0.162)	1.89 (5.12)	1.99 (1.86)	-1.58 (1.29)	0.640 (0.772)
Female Ratio	0.184** (0.805)	-0.400 (0.582)	6.06 (10.4)	-8.60 (6.10)	1.29* (4.39)	0.578 (2.31)
Education	0.050 (0.090)	0.248*** (0.075)	0.710 (1.29)	1.52** (0.784)	1.22** (0.535)	0.157 (0.275)
Population Density	9.81 × 10 <sup>-6</sup> (6.14 × 10 <sup>-6</sup> )	-4.68 × 10 <sup>-6</sup> *** (4.45 × 10 <sup>-6</sup> )	-9.01e-6* (0.0006)	0.0002 (0.0003)	0.0001*** (0.0003)	0.0875 × 10 <sup>-5</sup> (0.0001)
Trust Index	0.022 (0.025)	0.034 (0.030)	0.306 (0.306)	0.224* (0.117)	-0.053 (0.104)	0.120** (0.068)
Unemployment	0.046 (0.113)	0.113 (0.113)	1.06 (1.66)	-0.203 (0.235)	0.293 (0.235)	-0.050 (0.167)
White (%)	0.0008 (0.191)	0.043 (0.173)	8.71** (4.13)	-0.826 (1.83)	0.359 (1.61)	-0.322 (0.649)
Hispanic (%)	-0.009 (0.165)	0.022 (0.155)	6.94* (3.87)	-1.68 (1.40)	1.26 (1.17)	-0.510 (0.549)
African American (%)	-0.006 (0.179)	-0.002 (0.145)	6.10 (3.86)	-1.93 (1.56)	0.149 (1.28)	-0.055 (0.606)
Asian (%)	0.038 (0.228)	0.259 (0.197)	11.7* (4.67)	-9.651 (2.11)	1.43 (1.68)	0.414 (0.804)
Total Population (ln)	0.006 (0.010)	-0.006 (0.008)	0.029 (0.342)	0.135 (0.210)	-0.004 (0.162)	-0.013 (0.074)
Political Fragmentation	0.015 (0.015)	-0.025 (0.016)	0.021 (0.074)	-0.167 (0.054)	-0.050 (0.050)	-0.082 (0.028)
Median Housing (ln)	-0.012 (0.008)	0.0002 (0.009)	0.161 (0.107)	0.056 (0.071)	0.048 (0.040)	0.015 (0.019)
Median Income (ln)	-0.005 (0.028)	-0.066** (0.023)	-0.816** (0.357)	-0.011 (0.230)	0.022 (0.150)	-0.089 (0.079)
Regional Illiq	-0.0007 (0.228)	0.035 (0.147)	-0.365 (1.87)	0.056 (1.46)	-0.021 (0.791)	0.138 (0.598)
Regional IOR	0.014 (0.010)	0.037*** (0.010)	-0.074 (0.074)	0.069** (0.055)	-0.016 (0.027)	-0.018 (0.015)
Regional IO	0.012 (0.012)	-0.009 (0.009)	0.106 (0.043)	0.068 (0.052)	0.039 (0.039)	0.021 (0.021)
Regional Mktcap (ln)	-0.004 (0.004)	-0.006** (0.003)	-0.043 (0.040)	-0.024 (0.028)	-0.007 (0.011)	0.007 (0.007)

Regional Nstocks (ln)	0.010*	-0.001	0.061	0.032	-0.020	-0.005
	(0.007)	(0.004)	(0.036)	(0.029)	(0.020)	(0.010)
Regional Fpd	-0.020**	0.003	0.170	0.086*	0.050	0.027
	(0.012)	(0.002)	(0.103)	(0.054)	(0.012)	(0.034)
log(M/B)	0.056***	0.038***				
	(0.016)	(0.011)				
log(M/B) × Racial Diversity	0.009	0.009				
	(0.020)	(0.017)				
log(M/B) × Board Diversity	0.019**	0.003				
	(0.007)	(0.006)				
cashflow	0.523***	0.471***				
	(0.029)	(0.037)				
Invest	-0.151***	-0.026				
	(0.025)	(0.044)				
Illiquidity	0.0005	0.0004				
	(0.0008)	(0.0005)				
Asset (ln)	-0.011**	-0.017***				
	(0.005)	(0.003)				
Volp (ln)	-0.005**	-0.003***				
	(0.0009)	(0.0007)				
Inverse Volume			0.797***	0.744***	0.318***	1.11***
			(0.034)	(0.056)	(0.032)	(0.037)
B/M			0.046**	0.023*	-0.014	-0.023***
			(0.017)	(0.012)	(0.011)	(0.007)
Payout			0.004	0.0006	0.002	-0.002
			(0.006)	(0.003)	(0.002)	(0.001)
R&D (0/1)			-0.030	-0.005	-0.041***	0.029***
			(0.020)	(0.014)	(0.011)	(0.004)
Advertising (0/1)			0.011	0.008	-0.017	0.016***
			(0.012)	(0.010)	(0.013)	(0.005)
Ret (t-1)			-0.058***	-0.088***	-0.046***	-0.005
			(0.021)	(0.021)	(0.006)	(0.005)
Pre (t-1)			0.010*	0.013	0.0009	0.001
			(0.005)	(0.009)	(0.003)	(0.004)
Volp (t-1)			-0.007**	-0.049**	-0.0004	-0.003
			(0.013)	(0.007)	(0.007)	(0.002)
Market Cap (t-1)			-0.020*	-0.056***	-0.169***	-0.066***
			(0.010)	(0.006)	(0.008)	(0.004)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Quarter FE			Yes	Yes	Yes	Yes
Industry FE			Yes	Yes	Yes	Yes
Commuting Zone FE			Yes	Yes	Yes	Yes
Observations	11,871	11,871	14,523	14,523	73,578	73,576
R <sup>2</sup>	0.85503	0.76336	0.39320	0.54832	0.28425	0.34137

**Panel B: Sorted by Analyst Coverage**

Firms are sorted by the median value of the analyst coverage (number of analysts following the stock) at each time point. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Dependent Variables:	EBIT(t+1)/AT(t)		Adverse Selection (%)		VCV	
	(1)	(2)	(3)	(4)	(5)	(6)
Groups:	Fewer Analyst	More Analyst	Fewer Analyst	More Analyst	Fewer Analyst	More Analyst
log(M/B) × RRI NegTheil	0.041*	0.015				
	(0.023)	(0.023)				
RRI NegTheil	-0.067	-0.135*	-2.62**	-1.77**	-1.11**	-0.429*
	(0.044)	(0.073)	(1.09)	(0.691)	(0.521)	(0.250)
Age	0.012**	0.011*	0.039***	0.012	-0.041***	-0.039***
	(0.006)	(0.006)	(0.015)	(0.009)	(0.008)	(0.003)
Sale	0.017***	0.025***	-0.004	0.024***	-0.035***	-0.011***
	(0.005)	(0.006)	(0.012)	(0.009)	(0.007)	(0.004)
Tangibility	-0.017	-0.085***	-0.020	0.025	-0.138***	-0.052***
	(0.013)	(0.020)	(0.062)	(0.034)	(0.028)	(0.015)
Analyst Cover (ln)	-0.003	-0.001	-0.045***	0.021*	-0.117***	0.005
	(0.003)	(0.005)	(0.013)	(0.011)	(0.007)	(0.007)
Leverage	0.054***	0.059***	0.224***	0.171***	-0.0003	0.014

	(0.013)	(0.015)	(0.072)	(0.030)	(0.037)	(0.014)
Cash Holdings	-0.022	-0.035**	-0.060	-0.012	-0.082**	0.021*
	(0.015)	(0.016)	(0.052)	(0.029)	(0.040)	(0.012)
IOR	-0.008	-0.009	-0.131**	-0.016	0.073**	-0.046***
	(0.010)	(0.010)	(0.066)	(0.034)	(0.037)	(0.015)
Racial Diversity	0.040	0.051	1.32	0.756	0.773*	-0.007
	(0.068)	(0.077)	(0.999)	(0.604)	(0.451)	(0.205)
Board Diversity	-0.0003	-0.006	0.040	0.086***	0.011	-0.012
	(0.004)	(0.005)	(0.063)	(0.028)	(0.025)	(0.012)
Senior Ratio	0.028	0.104	-1.87	3.21	-1.18	-0.196
	(0.144)	(0.252)	(3.08)	(2.05)	(1.43)	(0.811)
Female Ratio	-1.11	-1.04	13.6	5.37	11.5**	-0.376
	(0.713)	(0.798)	(8.91)	(7.94)	(4.71)	(2.89)
Education	0.058	0.312***	1.06	1.34	1.06*	-0.105
	(0.079)	(0.078)	(1.37)	(0.846)	(0.572)	(0.232)
Population Density	$1.34 \times 10^{-5**}$	$4.66 \times 10^{-6}$	-0.002***	0.0001	0.0009***	$-1.82 \times 10^{-6}$
	( $5.65 \times 10^{-6}$ )	( $5.31 \times 10^{-6}$ )	(0.0006)	(0.0003)	(0.0002)	(0.0001)
Trust Index	0.011	0.056*	0.285	0.309*	0.039	0.048
	(0.021)	(0.029)	(0.446)	(0.166)	(0.119)	(0.076)
Unemployment	0.073	-0.010	0.516	0.347	0.351	-0.445
	(0.109)	(0.114)	(1.38)	(0.894)	(0.409)	(0.296)
White (%)	0.103	-0.041	8.29**	1.36	1.81	-0.447
	(0.188)	(0.303)	(3.63)	(2.18)	(1.27)	(0.697)
Hispanic (%)	0.082	-0.030	5.80**	0.503	1.03	-0.815
	(0.171)	(0.275)	(2.86)	(2.06)	(1.13)	(0.725)
African American (%)	0.117	-0.142	5.93	-1.26	0.112	-0.342
	(0.179)	(0.263)	(3.75)	(2.13)	(1.12)	(0.794)
Asian (%)	0.090	0.208	11.2*	2.68	2.09	-0.025
	(0.235)	(0.320)	(4.44)	(2.48)	(1.76)	(0.831)
Total Population (ln)	-0.019**	-0.003	0.230	-0.140	-0.119	-0.091
	(0.007)	(0.011)	(0.333)	(0.227)	(0.177)	(0.076)
Political Fragmentation	-0.015	-0.003	0.070	-0.019	-0.025	-0.072
	(0.027)	(0.036)	(0.322)	(0.170)	(0.136)	(0.061)
Median Housing (ln)	-0.014	-0.001	0.019	-0.009	0.047	0.038**
	(0.008)	(0.010)	(0.101)	(0.070)	(0.045)	(0.021)
Median Income (ln)	-0.000	-0.056**	-0.275	-0.008	-0.047	-0.104
	(0.024)	(0.026)	(0.377)	(0.222)	(0.154)	(0.071)
Regional Illiq	0.026	-0.138	-3.679	-0.813	-1.154	-0.071
	(0.226)	(0.379)	(1.89)	(1.87)	(0.818)	(0.624)
Regional Ret	0.014	0.037***	-0.079	0.087*	1.528	0.016
	(0.008)	(0.013)	(0.100)	(0.071)	(0.028)	(0.024)
Regional IO	0.015*	0.007	-0.002	-0.128*	0.029	-0.036*
	(0.009)	(0.004)	(0.101)	(0.073)	(0.042)	(0.020)
Regional Mktcap (ln)	-0.004	-0.009**	-0.018	-0.057*	0.008	0.005
	(0.002)	(0.003)	(0.035)	(0.031)	(0.012)	(0.006)
Regional Nstocks (ln)	0.008*	-0.004	0.015	0.008	-0.025	-0.007
	(0.005)	(0.007)	(0.069)	(0.041)	(0.023)	(0.013)
Regional Fpd	-0.013	-0.016	0.215*	0.070	-0.020	-0.034*
	(0.011)	(0.011)	(0.117)	(0.066)	(0.038)	(0.018)
log(M/B) × Racial Diversity	0.012	0.006				
	(0.018)	(0.024)				
log(M/B) × Board Diversity	0.010	0.010				
	(0.008)	(0.007)				
log(M/B)	0.043***	0.047***				
	(0.012)	(0.014)				
cashflow	0.538***	0.495***				
	(0.036)	(0.045)				
Invest	-0.124***	-0.060*				
	(0.033)	(0.035)				
Illiquidity	0.0005	-0.0009				
	(0.0005)	(0.0010)				
Asset (ln)	-0.014***	-0.015***				
	(0.004)	(0.003)				
Volp (ln)	-0.004***	-0.005***				
	(0.0010)	(0.0008)				
Volp (t-1)			-0.014	-0.064***	-0.011	-0.003
			(0.016)	(0.006)	(0.007)	(0.003)

Inverse Volume			0.732***	0.980***	0.431***	1.10***
			(0.039)	(0.145)	(0.039)	(0.095)
B/M			0.021	0.076***	-0.029***	-0.040***
			(0.018)	(0.013)	(0.008)	(0.005)
Payout			-0.009	0.003	0.003	-0.002**
			(0.006)	(0.004)	(0.002)	(0.0010)
R&D (0/1)			-0.022	-0.028**	-0.027***	0.005
			(0.022)	(0.013)	(0.009)	(0.006)
Advertising (0/1)			0.005	0.009	-0.015	0.012***
			(0.017)	(0.009)	(0.013)	(0.004)
Ret (t-1)			-0.103***	-0.057***	-0.040***	-0.016***
			(0.033)	(0.019)	(0.008)	(0.005)
Pre (t-1)			0.005	0.005	0.011***	0.015***
			(0.008)	(0.005)	(0.003)	(0.003)
Market Cap (t-1)			-0.039***	-0.029***	-0.146***	-0.084***
			(0.010)	(0.007)	(0.004)	(0.003)
Year FE	Yes	Yes				
Firm FE	Yes	Yes				
Year × Quarter FE			Yes	Yes	Yes	Yes
Industry FE			Yes	Yes	Yes	Yes
Commuting Zone FE			Yes	Yes	Yes	Yes
Observations	11,257	11,356	9,740	12,342	67,105	68,317
R <sup>2</sup>	0.86406	0.81285	0.40924	0.53005	0.31076	0.27311

### Panel C: Sorted by Age

Firms are sorted by the median value of the age of the firm. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Dependent Variables:	EBIT(t+1)/AT(t)		Adverse Selection (%)		VCV	
	(1)	(2)	(3)	(4)	(5)	(6)
Groups:	Younger	Older	Younger	Older	Younger	Older
log(M/B) × RRI NegTheil	0.031 (0.029)	0.016 (0.018)				
RRI NegTheil	-0.315*** (0.100)	-0.026 (0.030)	-3.38*** (1.03)	-0.752 (0.811)	-1.30*** (0.431)	-0.378 (0.342)
Age	0.012 (0.008)	0.038** (0.016)	-0.035** (0.014)	0.073*** (0.023)	-0.080*** (0.008)	-0.016* (0.010)
Sale	0.029*** (0.005)	0.021*** (0.005)	0.010 (0.011)	0.006 (0.012)	-0.035*** (0.006)	-0.004 (0.005)
Tangibility	-0.034 (0.021)	-0.060*** (0.015)	-0.004 (0.042)	-0.073 (0.048)	-0.120*** (0.031)	-0.035* (0.021)
Analyst Cover (ln)	-0.002 (0.003)	-0.006*** (0.002)	-0.022*** (0.007)	-0.011 (0.012)	-0.063*** (0.004)	-0.041*** (0.005)
Leverage	0.053*** (0.016)	0.048*** (0.009)	0.151*** (0.044)	0.286*** (0.078)	-0.022 (0.023)	0.030 (0.026)
Cash Holdings	-0.036** (0.017)	-0.034*** (0.010)	-0.068** (0.034)	-0.008 (0.061)	-0.039 (0.026)	-0.015 (0.026)
IOR	-0.009 (0.011)	-0.011 (0.011)	-0.035 (0.040)	-0.108* (0.059)	0.003 (0.027)	-0.038*** (0.026)
Racial Diversity	0.003 (0.077)	0.088 (0.077)	1.30 (0.951)	-0.257 (0.935)	0.043 (0.299)	0.394 (0.249)
Board Diversity	-0.016** (0.007)	0.002 (0.004)	0.090*** (0.041)	0.083* (0.042)	0.008 (0.017)	-0.001 (0.014)
Senior Ratio	-0.094 (0.173)	-0.152 (0.146)	-0.959 (2.97)	0.981 (3.23)	-0.960 (1.32)	-1.33 (1.10)
Female Ratio	-0.857 (0.857)	-0.334 (0.698)	16.1 (9.93)	6.31 (9.06)	7.37 (3.78)	6.67** (2.90)
Education	0.180* (0.109)	0.090 (0.056)	3.33*** (1.03)	0.886 (1.06)	1.30** (0.503)	0.052 (0.367)
Population Density	-4.04 × 10 <sup>-6**</sup> (7.38 × 10 <sup>-6</sup> )	-0.5 × 10 <sup>-5***</sup> (3.93 × 10 <sup>-6</sup> )	0.0003 (0.0005)	-0.001* (0.0006)	0.0004***-6.04 × 10 <sup>-5</sup> (0.0002)	(0.0002)
Trust Index	0.042* (0.025)	0.025 (0.020)	0.431** (0.167)	0.466* (0.273)	-0.008 (0.104)	0.014 (0.078)
Unemployment	0.015 (0.150)	0.072 (0.070)	-0.176 (0.932)	-0.753 (0.997)	-0.574* (0.318)	-0.100 (0.286)
White (%)	-0.084	0.277	2.36	2.23	-0.058	0.440

	(0.316)	(0.213)	(3.83)	(2.80)	(1.11)	(0.842)
Hispanic (%)	-0.044	0.213	0.882	1.36	0.577	-0.649
	(0.277)	(0.199)	(3.14)	(2.31)	(0.983)	(0.722)
African American (%)	-0.101	0.135	-0.738	1.89	-0.954	-0.568
	(0.292)	(0.183)	(3.89)	(2.71)	(1.19)	(0.865)
Asian (%)	0.177	0.200	3.21	5.09	-0.542	1.46
	(0.366)	(0.233)	(4.10)	(3.42)	(1.39)	(1.12)
Total Population (ln)	-0.033**	0.010	-0.582*	0.421	-0.244*	0.011
	(0.013)	(0.008)	(0.326)	(0.301)	(0.137)	(0.109)
Political Fragmentation	-0.006	-0.004	-0.049	-0.066	-0.086	-0.154*
	(0.036)	(0.024)	(0.264)	(0.308)	(0.114)	(0.084)
Median Housing (ln)	-0.006	-0.017**	0.220**	-0.011	0.015	0.028
	(0.010)	(0.008)	(0.086)	(0.091)	(0.038)	(0.024)
Median Income (ln)	-0.039	-0.011	-0.641**	0.002	-0.066	-0.026
	(0.037)	(0.022)	(0.367)	(0.245)	(0.125)	(0.092)
Regional Illiq	-0.551**	0.233	-0.891	-2.57	2.89***	1.34**
	(0.272)	(0.149)	(1.97)	(1.57)	(0.882)	(0.663)
Regional Ret	0.033***	0.019**	-0.112*	0.089	-0.029	0.005
	(0.009)	(0.009)	(0.059)	(0.079)	(0.026)	(0.019)
Regional IO	0.019	0.007	-0.009	-0.157**	0.059	-0.047
	(0.015)	(0.009)	(0.105)	(0.076)	(0.038)	(0.033)
Regional Mktcap (ln)	-0.007*	-0.004**	-0.047	0.010	-0.005	0.016
	(0.005)	(0.002)	(0.034)	(0.033)	(0.011)	(0.011)
Regional Nstocks (ln)	0.011	-0.004	0.009	-0.030	-0.018	-0.009
	(0.007)	(0.005)	(0.056)	(0.050)	(0.021)	(0.019)
Regional Fpd	-0.016	-0.011	0.232**	0.042	-0.026	-0.018
	(0.013)	(0.008)	(0.099)	(0.088)	(0.041)	(0.029)
log(M/B) × Racial Diversity	0.006	0.013				
	(0.021)	(0.017)				
log(M/B) × Board Diversity	0.015*	0.006				
	(0.008)	(0.008)				
log(M/B)	0.051***	0.035***				
	(0.015)	(0.011)				
Invest	-0.127***	-0.119**				
	(0.028)	(0.055)				
Illiquidity	0.004***	-0.0005				
	(0.001)	(0.0005)				
Asset (ln)	-0.012***	-0.014***				
	(0.004)	(0.004)				
Volp (ln)	-0.006***	-0.002***				
	(0.0010)	(0.0006)				
Volp (t-1)			-0.058***	-0.003	-0.023***	-0.005
			(0.008)	(0.011)	(0.008)	(0.004)
Inverse Volume			0.716***	0.639***	0.371***	0.670***
			(0.054)	(0.035)	(0.072)	(0.032)
B/M			0.044***	0.031**	-0.031***	-0.018**
			(0.016)	(0.015)	(0.006)	(0.009)
Payout			0.004	0.002	0.002	-0.0003
			(0.004)	(0.005)	(0.001)	(0.001)
R&D (0/1)			-0.017	-0.003	-0.035***	0.012
			(0.014)	(0.024)	(0.011)	(0.008)
Advertising (0/1)			0.009	0.011	-0.007	0.003
			(0.012)	(0.016)	(0.009)	(0.007)
Ret (t-1)			-0.051**	-0.100***	-0.035***	-0.025***
			(0.020)	(0.025)	(0.006)	(0.009)
Pre (t-1)			0.005	0.002	0.019***	0.012***
			(0.005)	(0.010)	(0.003)	(0.004)
Market Cap (t-1)			-0.009	-0.056***	-0.111***	-0.090***
			(0.008)	(0.008)	(0.005)	(0.003)
Year FE	Yes	Yes				
Firm FE	Yes	Yes				
Year × Quarter FE			Yes	Yes	Yes	Yes
Industry FE			Yes	Yes	Yes	Yes
Commuting Zone FE			Yes	Yes	Yes	Yes
Observations	11,551	11,376	13,988	14,113	70,909	71,179
R <sup>2</sup>	0.84907	0.76413	0.44404	0.45452	0.29711	0.44802

**Table 2A.3.4: Extended Version of Table 2.5 and Extra Tests**

**Panel A: Extended Table 2.5 Instrumental Variable Regression**

IV Stages: Dependent Variables:	Stock Price Informativeness Test			Information Asymmetry Test			
	First RRI (1)	First RRI × log(M/A) (2)	Second Future Cash Flow (3)	First RRI (4)	Second Adverse Selection (%) (5)	First RRI (6)	Second VCV (7)
RRI NegTheil × log(M/A)			0.180** (0.079)				
RRI NegTheil			-0.987 (0.858)		-16.9*** (5.66)		-3.76* (2.12)
Share of SOP elites × log(M/A)	-0.005 (0.025)	4.04*** (1.34)					
Share of SOP elites	1.05*** (0.347)	-2.73*** (0.817)		0.657*** (0.230)		0.565*** (0.180)	
Number of Vouchers (ln)	0.010** (0.004)	0.0005 (0.009)	0.007 (0.010)	0.011** (0.005)	0.220* (0.116)	0.008** (0.003)	0.061** (0.029)
log(M/A) × Number of Vouchers (ln)	4.62 × 10 <sup>-5</sup> (0.0003)	-0.051** (0.021)	0.005 (0.005)				
State × Year FE	Yes	Yes	Yes				
Firm FE	Yes	Yes	Yes				
State × Year × Quarter FE				Yes	Yes	Yes	Yes
Industry FE				Yes	Yes	Yes	Yes
Commuting Zone FE				Yes	Yes	Yes	Yes
Observations	15,171	15,171	15,171	17,503	17,503	91,052	91,052
R <sup>2</sup>	0.99854	0.98680	0.83043	0.99866	0.50245	0.99819	0.39473
F-test (1st stage), stat.	266.42	338.60		799.67		3,259.5	
F-test (1st stage), stat., RRI NegTheil			266.42		799.67		3,259.5
F-test (1st stage), stat., RRI NegTheil × log(M/A)			338.60				
Wald (1st stage), stat., RRI NegTheil			4.6194		8.1642		9.8732
Wald (1st stage), stat., RRI NegTheil × log(M/A)			6.3346				
Wald (1st stage), p-value, RRI NegTheil			0.00988		0.00428		0.00168
Wald (1st stage), p-value, RRI NegTheil × log(M/A)			0.00178				
Wu-Hausman, stat.			3.8599		10.876		3.0629
Wu-Hausman, p-value			0.02109		0.00098		0.08010

**Panel B: Instrumental Variable Regression with 2-year Lag**

IV Stages: Dependent Variables:	Stock Price Informativeness Test			Information Asymmetry Test			
	First RRI	First RRI × log(M/A)	Second Future Cash Flow	First RRI	Second Adverse Selection (%)	First RRI	Second VCV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
RRI NegTheil × log(M/A)			0.214* (0.112)				
RRI NegTheil			-1.00 (0.701)		-11.4*** (4.32)		-3.93* (2.07)
Share of SOP elites × log(M/A)	-0.011 (0.026)	4.17*** (1.37)					
Share of SOP elites	1.02*** (0.340)	-2.50*** (0.685)		0.962*** (0.288)		0.767*** (0.249)	
Number of Vouchers (ln)	0.010** (0.004)	-0.001 (0.009)	0.016* (0.009)	0.011** (0.005)	0.175** (0.083)	0.009** (0.003)	0.076** (0.034)
log(M/A) × Number of Vouchers (ln)	3.04 × 10 <sup>-5</sup> (0.0003)	-0.050** (0.021)	0.007 (0.006)				
State × Year FE	Yes	Yes	Yes				
Firm FE	Yes	Yes	Yes				
State × Year × Quarter FE				Yes	Yes	Yes	Yes
Industry FE				Yes	Yes	Yes	Yes
Commuting Zone FE				Yes	Yes	Yes	Yes
Observations	15,171	15,171	15,171	17,503	17,503	91,052	91,052
R <sup>2</sup>	0.99852	0.98674	0.82985	0.99869	0.50654	0.99820	0.39470
F-test (1st stage), stat.	316.22	329.00		1,260.5		4,148.7	
F-test (1st stage), stat., RRI NegTheil			316.22		1,260.5		4,148.7
F-test (1st stage), stat., RRI NegTheil × log(M/A)			329.00				
Wald (1st stage), stat.	4.4736	7.4294		11.165		9.5254	
Wald (1st stage), stat., RRI NegTheil			4.4736		11.165		9.5254
Wald (1st stage), stat., RRI NegTheil × log(M/A)			7.4294				
Wald (1st stage), p-value	0.01142	0.00060		0.00084		0.00203	
Wald (1st stage), p-value, RRI NegTheil			0.01142		0.00084		0.00203
Wald (1st stage), p-value, RRI NegTheil × log(M/A)			0.00060				
Wu-Hausman, stat.			5.3674		6.7913		4.2486
Wu-Hausman, p-value			0.00468		0.00917		0.03929

**Panel C: Instrumental Variable Regression Removing Non-relocation Constraint**

IV Stages: Dependent Variables:	Stock Price Informativeness Test			Information Asymmetry Test			
	First RRI (1)	First RRI × log(M/A) (2)	Second Future Cash Flow (3)	First RRI (4)	Second Adverse Selection (%) (5)	First RRI (6)	Second VCV (7)
RRI NegTheil × log(M/A)			0.118** (0.058)				
RRI NegTheil			-1.23*** (0.286)		-15.0*** (5.61)		-4.00* (2.18)
Share of SOP elites × log(M/A)	0.047 (0.046)	4.20*** (1.19)					
Share of SOP elites	1.72 (0.239)	-1.37* (0.698)		0.667*** (0.232)		0.552*** (0.171)	
log(M/A) × Number of Vouchers (ln)	0.0006 (0.0006)	-0.054*** (0.019)	0.003 (0.004)				
Number of Vouchers (ln)	0.002 (0.003)	0.001 (0.008)	0.007 (0.008)	0.014*** (0.005)	0.198* (0.114)	0.009*** (0.003)	0.058* (0.031)
State × Year FE	Yes	Yes	Yes				
Firm FE	Yes	Yes	Yes				
State × Year × Quarter FE				Yes	Yes	Yes	Yes
Industry FE				Yes	Yes	Yes	Yes
Commuting Zone FE				Yes	Yes	Yes	Yes
Observations	16,926	16,926	16,926	20,199	20,199	104,428	104,428
R <sup>2</sup>	0.99644	0.98620	0.82914	0.99854	0.48586	0.99817	0.38643
F-test (1st stage), stat.	539.54	397.58		895.13		3,497.3	
F-test (1st stage), stat., RRI NegTheil			539.54		895.13		3,497.3
F-test (1st stage), stat., RRI NegTheil × log(M/A)			397.58				
Wald (1st stage), stat.	28.904	6.1932		8.2318		10.360	
Wald (1st stage), stat., RRI NegTheil			28.904		8.2318		10.360
Wald (1st stage), stat., RRI NegTheil × log(M/A)			6.1932				
Wald (1st stage), p-value	2.97 × 10 <sup>-13</sup>	0.00205		0.00412		0.00129	
Wald (1st stage), p-value, RRI NegTheil			2.97 × 10 <sup>-13</sup>		0.00412		0.00129
Wald (1st stage), p-value, RRI NegTheil × log(M/A)			0.00205				
Wu-Hausman, stat.			6.5502		8.6394		3.5803
Wu-Hausman, p-value			0.00143		0.00329		0.05847

**Panel D: Instrumental Variable Regression without Controlling for Vouchers**

IV Stages: Dependent Variables:	Stock Price Informativeness Test			Information Asymmetry Test			
	First RRI	First RRI × log(M/A)	Second Future Cash Flow	First RRI	Second Adverse Selection (%)	First RRI	Second VCV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
RRI NegTheil x log(M/A)			0.234* (0.125)				
RRI NegTheil			-1.09 (0.915)		-16.8*** (5.66)		-3.67* (2.12)
Share of SOP elites × log(M/A)	-0.014 (0.026)	2.59** (1.21)					
Share of SOP elites	1.01*** (0.365)	-1.33 (0.804)		0.660*** (0.235)		0.571*** (0.185)	
State × Year FE	Yes	Yes	Yes				
Firm FE	Yes	Yes	Yes				
State × Year × Quarter FE				Yes	Yes	Yes	Yes
Industry FE				Yes	Yes	Yes	Yes
Commuting Zone FE				Yes	Yes	Yes	Yes
Observations	15,171	15,171	15,171	17,503	17,503	91,052	91,052
R <sup>2</sup>	0.99845	0.98294	0.82836	0.99863	0.50230	0.99815	0.39471
F-test (1st stage), stat.	238.94	108.34		790.27		3,270.4	
F-test (1st stage), stat., RRI NegTheil			238.94		790.27		3,270.4
F-test (1st stage), stat., RRI NegTheil x log(M/A)			108.34				
Wald (1st stage), stat.	3.8399	4.0740		7.8959		9.5071	
Wald (1st stage), stat., RRI NegTheil			3.8399		7.8959		9.5071
Wald (1st stage), stat., RRI NegTheil x log(M/A)			4.0740				
Wald (1st stage), p-value	0.02152	0.01703		0.00496		0.00205	
Wald (1st stage), p-value, RRI NegTheil			0.02152		0.00496		0.00205
Wald (1st stage), p-value, RRI NegTheil x log(M/A)			0.01703				
Wu-Hausman, stat.						3.1235	
Wu-Hausman, p-value						0.07717	

**Table 2A.3.5: Extended Version of Table 2.6**

This table presents the results of DiD regressions examining the effect of changes in racial residential integration (RRI) on stock price informativeness and firm-level information asymmetry following firm relocation. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Dependent variable:	Stock Price Informativeness		Information Asymmetry	
	High-to-Low	Low-to-High	High-to-Low	Low-to-High
	(1) Future Cash Flow	(2) Future Cash Flow	(3) VCV	(4) VCV
Post × log(M/A)	-0.083*** (0.026)	0.024* (0.012)		
Post	-0.043** (0.022)	0.018* (0.011)	0.119* (0.063)	-0.116** (0.055)
log(M/A) × Treated	0.088*** (0.029)	0.034** (0.017)		
log(M/A)	-0.072 (0.043)	0.022 (0.035)		
log(M/A) × Racial Diversity	0.186*** (0.062)	0.042 (0.051)		
log(M/A) × Board Diversity	0.005 (0.035)	-0.046 (0.031)		
Age	0.031 (0.035)	0.007 (0.021)	-0.223*** (0.063)	-0.179*** (0.051)
RRI NegTheil	0.344* (0.196)	-0.011 (0.095)	0.254 (0.266)	0.176 (0.263)
Analyst Cover (ln)	-0.106 (0.012)	-0.017* (0.009)	-0.050*** (0.012)	-0.022* (0.012)
Sale	0.103*** (0.035)	0.004 (0.022)	0.032 (0.035)	-0.027 (0.025)
Tangibility	-0.066 (0.312)	-0.073 (0.072)	0.004 (0.115)	0.092 (0.140)
cashflow	0.280*** (0.090)	0.487*** (0.072)		
Invest	-0.024 (0.181)	-0.021 (0.130)		
Illiquidity	0.003 (0.003)	0.007 (0.018)		
Volp (ln)	0.006* (0.004)	-0.005** (0.002)		
Board Diversity	-0.007 (0.028)	-0.044** (0.014)	0.045 (0.054)	0.117** (0.047)
Leverage	0.102 (0.066)	-0.071 (0.055)	0.103 (0.075)	-0.033 (0.067)
Cash Holdings	0.004 (0.057)	0.001 (0.028)	-0.023 (0.073)	-0.034 (0.070)
Asset (ln)	-0.029 (0.026)	-0.006 (0.012)		
IOR	-0.048 (0.050)	0.028 (0.027)	-0.044 (0.079)	0.015 (0.084)
Senior Ratio	-0.357 (0.752)	0.398 (0.251)	-0.501 (0.640)	-0.443 (0.908)
Female Ratio	-2.88* (1.44)	-2.96** (1.45)	2.37 (4.47)	-6.12 (5.15)
Education	-0.474* (0.259)	0.266* (0.144)	0.327 (0.562)	1.69*** (0.498)
Population Density	$8.79 \times 10^{-5}$ ** ( $3.78 \times 10^{-5}$ )	$2.05 \times 10^{-5}$ * ( $1.12 \times 10^{-5}$ )	$5.23 \times 10^{-5}$ ( $3.66 \times 10^{-5}$ )	$5.88 \times 10^{-5}$ ( $3.68 \times 10^{-5}$ )
Trust Index	-0.296*** (0.094)	0.131** (0.053)	0.511 (0.333)	-0.170 (0.243)
Unemployment	0.315 (0.443)	0.141 (0.316)	-0.381 (0.920)	0.165 (0.962)
Racial Diversity	-0.173 (0.162)	0.035 (0.117)	0.310 (0.422)	-0.478 (0.489)
White (%)	-0.660 (0.401)	-0.465 (0.473)	-0.970 (0.926)	-0.067 (2.39)

Hispanic (%)	-0.768 (0.957)	-0.494 (0.505)	-0.178 (0.750)	0.323 (2.12)
African American (%)	-0.148 (0.942)	-0.235 (0.529)	-0.672 (0.725)	0.293 (2.08)
Asian (%)	-0.933 (1.18)	-0.791 (0.584)	-0.921 (1.09)	-0.133 (2.33)
Total Population (ln)	-0.034 (0.020)	0.002 (0.013)	-0.031 (0.033)	0.039 (0.044)
Median Housing (ln)	0.021 (0.035)	0.016 (0.024)	-0.038 (0.059)	0.058 (0.072)
Median Income (ln)	0.001 (0.092)	-0.053 (0.059)	0.185 (0.207)	-0.354* (0.191)
Regional Illiq	0.069 (0.470)	0.044 (0.470)	-0.065 (1.41)	-1.62 (1.75)
Regional Ret	0.069*** (0.023)	0.034 (0.025)	0.099 (0.062)	0.122* (0.068)
Regional IO	0.025 (0.045)	0.002 (0.027)	0.053 (0.087)	-0.040 (0.074)
Political Fragmentation	-0.122 (0.099)	-0.025 (0.059)	0.110 (0.162)	-0.103 (0.147)
Regional Mktcap (ln)	-0.008 (0.008)	-0.0007 (0.005)	-0.002 (0.013)	-0.035* (0.019)
Regional Nstocks (ln)	0.047** (0.023)	-0.008 (0.010)	0.008 (0.032)	-0.014 (0.031)
Regional Fpd	-0.021 (0.038)	-0.008 (0.022)	-0.041 (0.057)	0.036 (0.054)
Inversed Volume			-0.105 (0.087)	-0.056 (0.074)
B/M			-0.046** (0.022)	-0.034 (0.023)
Payout			-0.002 (0.004)	-0.001 (0.005)
R&D (0/1)			-0.005 (0.043)	-0.079*** (0.029)
Advertising			-0.028 (0.038)	0.010 (0.035)
Ret (t-1)			-0.035 (0.022)	-0.002 (0.022)
Pre (t-1)			-0.022 (0.022)	-0.020 (0.023)
Volp (t-1) (ln)			-0.026** (0.013)	-0.025** (0.011)
Market Cap (t-1)			-0.089*** (0.025)	-0.083*** (0.025)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes		
Year × Quarter FE			Yes	Yes
Observations	399	510	6,013	6,096
R <sup>2</sup>	0.66840	0.73933	0.43665	0.45754

**Table 2A.3.6: Complete Versions of Table 2.7 and Table 2.8**

	Local Bias		LSP (%)		Different Economic Conditions LSP (%)		Different Manager Race LSP (%)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Expansions	Recessions	White Managers	Minority Managers
RRI NegTheil	0.034 (0.035)	0.074 (0.183)	-1.74*** (0.290)	-4.51*** (1.01)	-3.90*** (0.999)	-11.9* (6.87)	-4.05** (1.90)	-31.8 (21.3)
Female Ratio		-0.123 (1.46)		-13.1 (13.1)	-3.56 (11.2)	-41.1 (78.1)	-5.32 (17.3)	351.7** (162.4)
Female (0/1)							-0.011 (0.044)	0.068 (0.049)
Work Exp (ln)							0.038 (0.027)	0.094 (0.062)
Team Managed (0/1)		-0.0006 (0.002)		-0.006 (0.004)	0.004 (0.004)	-0.066*** (0.018)		
Single Manager		0.002 (0.002)		-0.002 (0.005)	0.004 (0.005)	-0.038* (0.022)		
Expense Ratio		-0.403 (0.566)		0.045 (1.25)	-2.57** (1.13)	16.0*** (3.78)	0.472 (2.58)	11.7** (4.44)
Net Flow		0.003 (0.002)		-0.020 (0.018)	-0.017 (0.016)	0.002 (0.085)	-0.081* (0.045)	-0.170 (0.209)
Total Load		0.091 (0.118)		0.175 (0.180)	0.336* (0.198)	-0.942** (0.469)	0.314 (0.337)	-1.80*** (0.502)
Turnover		0.0004 (0.002)		0.001 (0.006)	-0.002 (0.006)	0.016 (0.014)	0.004 (0.010)	-0.027 (0.044)
Fund Age (ln)		0.001 (0.002)		0.0001 (0.004)	0.0008 (0.005)	-0.006 (0.008)	0.010 (0.010)	0.023 (0.044)
TNA (ln)		-0.0001 (0.0006)		0.002* (0.001)	0.0004 (0.001)	0.009 (0.008)	-0.004 (0.003)	0.016 (0.021)
Nstocks (ln)		-0.011*** (0.001)		0.005 (0.009)	0.004 (0.007)	0.013 (0.028)	0.004 (0.019)	0.006 (0.030)
Fund Family Asset (ln)		0.0004 (0.0008)		0.007** (0.003)	0.006* (0.003)	0.015* (0.008)	0.015* (0.008)	0.003 (0.025)
Fund Family Size (ln)		-0.003 (0.002)		-0.013** (0.006)	-0.010* (0.006)	-0.028* (0.017)	-0.022 (0.015)	0.002 (0.063)
Senior Ratio		0.448 (0.320)		-1.24 (2.75)	-2.37 (2.31)	0.408 (13.1)	1.05 (5.58)	-12.1 (73.1)
Education		-0.112 (0.105)		3.17*** (1.07)	1.98* (1.06)	11.3** (4.66)	0.208 (1.63)	58.8** (24.9)
Population Density		-1.46 × 10 <sup>-5</sup> (3.46 × 10 <sup>-5</sup> )		-0.0009** (0.0004)	-0.0006* (0.0003)	-0.003 (0.002)	-0.002** (0.0007)	0.006 (0.007)
Trust Index		0.017 (0.031)		0.105 (0.163)	0.007 (0.244)	-0.241 (2.21)	0.268 (0.331)	1.94* (0.996)
Unemployment		-0.030 (0.082)		0.62* (0.940)	1.21 (0.752)	6.06 (4.11)	1.68 (1.28)	-10.2 (7.04)
Racial Diversity Total		0.105 (0.203)		-1.79 (1.19)	-0.838 (1.41)	-7.02 (4.59)	-0.618 (1.77)	-18.7 (17.2)
Population (ln)		0.008 (0.033)		-0.429 (0.357)	-0.411 (0.301)	-3.28* (1.76)	-0.629 (0.567)	-3.00 (12.3)
White (%)		0.569 (0.463)		-9.97*** (3.15)	-7.75** (3.55)	-30.3*** (8.89)	-14.2*** (4.69)	-59.3 (88.6)
Hispanic (%)		0.447 (0.407)		-12.6*** (2.85)	-11.3*** (3.25)	-29.1*** (8.51)	-16.7*** (4.30)	-69.1 (101.0)
African American (%)		0.465 (0.338)		-9.93*** (3.30)	-8.42*** (3.11)	-26.9** (11.8)	-19.6*** (5.41)	-56.1 (113.0)
Asian (%)		0.982* (0.532)		-6.79** (3.27)	-5.05 (4.43)	-18.7 (16.1)	-7.38 (5.20)	-120.7 (115.8)
Political Fragmentation		0.041 (0.068)		-0.893*** (0.287)	-0.790*** (0.238)	-2.27* (1.18)	-1.09* (0.631)	-2.23 (9.13)
Median Housing (ln)		-0.003 (0.005)		-0.203*** (0.045)	-0.220*** (0.047)	0.774*** (0.243)	-0.288*** (0.063)	0.273 (0.256)

Median		-0.042*		-0.266	0.029	-4.24***	0.228	-0.960
Income (ln)		(0.024)		(0.275)	(0.312)	(1.33)	(0.358)	(2.15)
Regional Ret		0.007		-0.172*	-0.097	-0.396**	-0.086	-1.48***
		(0.004)		(0.096)	(0.093)	(0.173)	(0.136)	(0.441)
Regional Illiq		-0.115		-8.24	-11.6*	-21.0	4.79	-60.4**
		(0.365)		(7.92)	(5.95)	(22.6)	(10.3)	(23.5)
Regional Fpd		0.001		0.252***	0.204***	0.217	-0.153	-0.637
		(0.009)		(0.088)	(0.075)	(0.520)	(0.168)	(0.882)
Regional		0.002		0.047	0.040	-0.123	-0.0004	-1.33
Nstocks (ln)		(0.007)		(0.050)	(0.054)	(0.204)	(0.088)	(0.827)
Regional		-0.095***		-0.147***	-0.142***	-0.262	-0.133*	-0.655
Mktcap (ln)		(0.003)		(0.048)	(0.049)	(0.164)	(0.070)	(0.313)
Regional IO		0.019		0.161	0.083	0.905	0.189	1.73*
		(0.011)		(0.135)	(0.097)	(0.741)	(0.180)	(0.951)
Commuting	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zone FE								
Year × Quarter		Yes		Yes	Yes	Yes	Yes	Yes
× Style FE								
Observations	49,350	49,350	49,350	49,350	42,491	6,859	9,715	777
R <sup>2</sup>	0.00369	0.17496	0.01153	0.07153	0.06046	0.13563	0.10843	0.41067

## Appendix 2A.4: Extra robustness tests

### Appendix 2A.4.1 Different local community definitions and the ACS annual survey

**(1) Alternative definitions of local communities:** While commuting zones (CZs) capture travel patterns effectively, they may overlook two important aspects. First, CZs often span multiple counties, which can differ in politically relevant characteristics such as policy preferences and governance. Counties may capture localized variation in regional variables that CZs miss. Second, commuting patterns have evolved over recent decades, with more cross-zone commuting potentially blurring CZ boundaries.

To assess the sensitivity of our results to the choice of geographic unit, we begin with a county-level analysis to test whether a narrower definition of locality yields stronger or weaker associations. We then expand the analysis by constructing alternative firm-level and fund-level RRI measures at both the local-region and state levels to capture broader local environment and cross-zone commuting. The local region is defined as all counties within a 40-mile radius of a firm's/fund's headquarters and has more counties than CZs on average.<sup>39</sup> This definition typically includes more counties than a CZ and is intended to capture a wider swath of the area surrounding a firm. To construct a local region, we first identify all zip codes located within a 40-mile radius of a firm's/fund's headquarters and map zip codes to counties based on U.S. Department of Housing and Urban Development data to obtain the included counties. We then compute RRI for a local region using census tract data from all its counties. We refer to this as a firm-level local-region RRI, as it reflects the racial integration characteristics of the broader area surrounding a firm's/a fund's headquarters rather than just its home CZ. We use the same set of counties to construct all other local region-level variables by aggregating county-level characteristics.<sup>40</sup> Similarly, we construct state-level RRI measures and corresponding controls

---

<sup>39</sup> To define the residential region of a firm's employees, we use a conservative radius of 40 miles. According to the Transportation Statistics Annual Report (2023), the average daily distance traveled by Americans is 37 miles, a figure that has steadily increased over the past two decades, as reported by the Bureau of Transportation Statistics. Given that commuting routes are rarely linear, a 40-mile radius is a conservative estimate, suggesting that actual commuting distances may exceed this threshold.

<sup>40</sup> We use the zipcodeR function in R to implement this radius-based approach. This function is developed based

at the state level. Fixed effects and standard errors are adjusted accordingly.

The first three columns from panel A of Table 2A.4.1 show that the stock price informativeness results remain robust when using local region and state-level RRI measures, but are not robust when using county-level RRI. The first nine columns from panel B of Table 2A.4.1 show that the information asymmetry results are also mixed: the adverse selection results remain robust across all geographic definitions, while the VCV results are no longer statistically significant at the county or state level, and the local stock picking ability results are no longer statistically significant at the county level. The lack of robustness at the county level likely reflects its limited spatial coverage, especially since commuting zones can span multiple counties, which may not capture the broader social and informational networks relevant to firms. At the state level, the muted effects of RRI on VCV may reflect offsetting dynamics across investor types. On one hand, greater racial integration could reduce the advantage of institutional investors who rely on local information advantages, weakening their participation in informed trading, which is also supported by the significant state level mutual fund results. On the other hand, integration at the state level encompasses a larger population of local retail investors, who may become marginally more informed through enhanced access to public information and social diffusion. This increase in local retail informed trading can offset the decline in institutional activity. As a result, the relationship between RRI and trading volume-based measures of information asymmetry, such as VCV, becomes less precise and harder to detect.

**(2) RRI measure using annual data:** Our baseline RRI measure relies on decennial U.S. Census data, requiring interpolation to generate annual measures. This interpolation may introduce forward-looking bias if future values affect current estimates. While this approach is

---

on the uszipcode documentation, which provides further details at <https://uszipcode.readthedocs.io/index.html>. The documentation uses data from the U.S. Census Bureau website ([data.census.gov](https://data.census.gov)), incorporating static statistics from the 2010 Census and updated demographic information from 2020. To map zip codes to counties, we use the 2020 Q1 version of the zip code county crosswalk, available at: [https://www.huduser.gov/portal/datasets/usps\\_crosswalk.html](https://www.huduser.gov/portal/datasets/usps_crosswalk.html)

consistent with prior studies using cultural variables or analyzing residential patterns (e.g., Trounstine, 2016), it raises concerns about measurement precision. To address this, we conduct a robustness test using annual data from the ACS, which contains data starting in 2005. The ACS reports racial composition by Public Use Microdata Areas (PUMAs), each containing at least 100,000 people.<sup>41</sup> For context, in 2020, there were 74,001 census tracts and 2,378 PUMAs, implying that each PUMA covers approximately 31 census tracts on average. Since our commuting zone RRI is constructed using census tract data, the switch to PUMA-level data for the ACS reduces geographic granularity. Because RRI captures variation in racial distribution across subunits, the larger size of PUMAs relative to census tracts may obscure local variation and limit measurement precision.

Using ACS data, we compute annual state-level RRI, since PUMAs often do not align well with CZ boundaries, and assign values to firms based on their headquarters locations.<sup>42</sup> We then replicate our tests of stock price informativeness, information asymmetry (focusing on the adverse selection component of the bid-ask spread), and local stock picking ability. The results remain consistent with our baseline findings and are reported in Table 2A.4.1.

---

<sup>41</sup> While ACS data on race are also available at the county level, which is a smaller geographic unit than a PUMA, such data are reported only for counties with populations of at least 65,000. As a result, a substantial portion of our sample lacks county-level data, leading to significant missing values.

<sup>42</sup> Many PUMAs span multiple commuting zones, making it difficult to determine which commuting zone a given PUMA belongs to (Strahof, 2023). While a probability-based method could be used to estimate aggregated variables, like total population, by assigning PUMAs to commuting zones based on the proportion of their geographic overlap, this approach is not appropriate for our RRI measure. Since our RRI captures how the racial composition of smaller predefined regions deviates from the overall composition of the larger area, it requires that each smaller region be fully contained within the commuting zone to ensure accuracy.

**Table 2A.4.1: Alternative Community Definitions and Annual ACS Results**

In the table presented, we replicate baseline tests in the main body using three definitions of local regions—county, commuting zone, and state—as well as ACS one-year survey data, which has been available annually since 2005. Panel A reports results for the stock price informativeness test. The dependent variable in all columns is future cash flow, proxied by EBIT in year  $t+1$ , scaled by total assets in year  $t$ . The firm's market valuation, defined as  $\log(M/A)$ , is the natural log of market capitalization divided by total assets in year  $t$ . Panel B examines information asymmetry and local stock-picking ability. Columns (1), (4), (7), and (10) test adverse selection (*AdSel*), defined as the portion of the bid-ask spread that compensates market makers for risks associated with informed trading. Columns (2), (5), (8), and (11) report results for the volume coefficient of variation (*VCV*), calculated as the standard deviation of trading turnover divided by its average over a given quarter. Columns (3), (6), (9), and (12) assess local stock-picking ability (*LSP*). Across both panels, the key independent variable is RRI NegTheil, a measure of the evenness dimension of racial residential integration. Firm-level regressions include controls consistent with baseline models, adjusted to match the corresponding local definitions. Fund-level regressions follow the baseline model, with all control variables applied at the relevant geographic level. Standard errors are clustered at the county, local region, and state levels, depending on the test. Note: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\*  $p < 0.01$ .

**Panel A: Stock Informativeness Results**

	Dependent Variable: Future Cash Flow			
	County (1)	Local Region (2)	State (Decennial) (3)	State (ACS, Annual) (4)
$\log(M/A) \times \text{RRI NegTheil}$	0.001 (0.013)	0.031** (0.014)	0.042*** (0.017)	0.050** (0.021)
$\log(M/A) \times \text{Racial Diversity}$	0.025* (0.014)	0.013 (0.011)	-0.012 (0.016)	-0.004 (0.022)
$\log(M/A)$	0.028*** (0.008)	0.045*** (0.008)	0.062*** (0.011)	0.060*** (0.013)
$\log(M/A) \times \text{Board Racial Diversity}$	0.010** (0.004)	0.007 (0.005)	0.010* (0.006)	0.010 (0.007)
RRI NegTheil	-0.053 (0.032)	-0.119*** (0.032)	-0.061 (0.041)	-0.082* (0.047)
Racial Diversity	0.013 (0.030)	0.046 (0.029)	0.098* (0.052)	-0.113 (0.046)
Age	0.007* (0.004)	0.009*** (0.003)	0.010** (0.004)	0.010 (0.008)
Board Diversity	-0.007** (0.003)	-0.005 (0.003)	-0.006** (0.003)	-0.007 (0.004)
Sale	0.021*** (0.004)	0.018*** (0.003)	0.019*** (0.004)	0.022*** (0.005)
Tangibility	-0.041*** (0.013)	-0.039*** (0.011)	-0.049*** (0.010)	-0.063*** (0.017)
cashflow	0.510*** (0.028)	0.501*** (0.023)	0.507*** (0.030)	0.481*** (0.040)
Invest	-0.113*** (0.025)	-0.112*** (0.022)	-0.121*** (0.018)	-0.111*** (0.034)
Illiquidity	0.0002 (0.0006)	0.0006 (0.0007)	-0.0001 (0.0009)	-0.0001 (0.0010)
Volp (ln)	-0.004*** (0.0006)	-0.004*** (0.0005)	-0.004*** (0.0007)	-0.003*** (0.0009)
Analyst Cover (ln)	-0.002 (0.002)	-0.005*** (0.002)	-0.004** (0.002)	-0.005** (0.002)
Leverage	0.044*** (0.007)	0.044*** (0.007)	0.044*** (0.0006)	0.044*** (0.008)
Cash Holdings	-0.010 (0.013)	-0.008 (0.009)	-0.0002 (0.010)	-0.001 (0.014)
Asset (ln)	-0.010*** (0.003)	-0.014*** (0.002)	-0.009*** (0.003)	-0.009*** (0.003)
IOR	-0.006 (0.006)	-0.007 (0.005)	-0.008 (0.008)	-0.008 (0.008)

Senior Ratio	0.140 (0.098)	-0.033 (0.123)	-0.074 (0.192)	0.408 (0.246)
Female Ratio	-1.19*** (0.364)	-0.561 (0.403)	0.435 (0.485)	-1.81*** (0.637)
Education	0.038 (0.031)	0.093** (0.042)	-0.053 (0.071)	-0.027 (0.063)
Population Density	$6.56 \times 10^{-7}$ ( $5.02 \times 10^{-7}$ )	$4.41 \times 10^{-6}$ ( $8.44 \times 10^{-6}$ )	$-3.96 \times 10^{-5}$ ( $3.2 \times 10^{-5}$ )	$2.98 \times 10^{-5}$ ( $3.21 \times 10^{-5}$ )
Trust Index	0.039** (0.020)	0.042** (0.017)	0.020 (0.023)	0.057** (0.023)
Unemployment	0.130 (0.093)	0.051 (0.068)	-0.098 (0.092)	-0.120 (0.096)
Total Population (ln)	-0.0009 (0.005)	-0.012* (0.006)	-0.017 (0.012)	-0.016 (0.013)
White (%)	0.047 (0.127)	0.135 (0.131)	0.194 (0.204)	0.002 (0.148)
Hispanic (%)	0.035 (0.130)	0.102 (0.123)	0.456** (0.200)	0.086 (0.084)
African American (%)	0.076 (0.145)	0.088 (0.119)	0.371** (0.179)	0.158** (0.067)
Asian (%)	0.096 (0.157)	0.166 (0.141)	0.664*** (0.246)	0.043 (0.135)
Regional Ret	0.024*** (0.006)	0.019*** (0.005)	0.022* (0.012)	-0.003 (0.010)
Regional Illiq	0.058 (0.066)	-0.027 (0.105)	-0.143 (0.304)	-0.365 (0.689)
Regional gpd	-0.003 (0.006)	-0.012 (0.007)	-0.014 (0.012)	-0.023 (0.015)
Regional Nstocks (ln)	0.005 (0.004)	0.001 (0.004)	-0.0004 (0.007)	-0.005 (0.006)
Regional Mktcap (ln)	-0.006** (0.002)	-0.005** (0.002)	-0.005 (0.003)	0.003 (0.006)
Regional IO	0.022*** (0.008)	0.017** (0.009)	0.024 (0.016)	0.014 (0.013)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	32,625	32,625	32,628	9,014
R <sup>2</sup>	0.82369	0.79437	0.81714	0.83250

**Panel B: Stock Information Asymmetry and Fund Local Investing Results**

	County			Local Region			State (Decennial)			State (ACS, Annual)		
	(1) AdvSel (%)	(2) VCV	(3) LSP (%)	(4) AdvSel (%)	(5) VCV	(6) LSP (%)	(7) AdvSel (%)	(8) VCV	(9) LSP (%)	(10) AdvSel (%)	(11) VCV	(12) LSP (%)
RRI NegTheil	-1.39*** (0.420)	-0.115 (0.196)	-0.667 (0.793)	-2.43*** (0.810)	-0.599** (0.290)	-3.83** (1.70)	-1.57** (0.680)	-0.195 (0.412)	-11.5*** (3.05)	-0.749** (0.355)	0.021 (0.183)	-10.8*** (2.88)
Age	0.018** (0.009)	-0.033*** (0.004)		0.024** (0.010)	-0.034*** (0.004)		0.020** (0.008)	-0.036*** (0.004)		0.012** (0.005)	-0.038*** (0.005)	
Inversed Volume	0.702*** (0.034)	0.570*** (0.035)		0.748*** (0.035)	0.579*** (0.038)		0.696*** (0.024)	0.563*** (0.028)		0.789*** (0.023)	0.683*** (0.047)	
Board Diversity	0.078** (0.030)	0.008 (0.013)		0.075** (0.035)	0.004 (0.012)		0.086*** (0.024)	-0.006 (0.009)		0.022 (0.027)	-0.001 (0.013)	
Sale	0.004 (0.009)	-0.022*** (0.005)		0.009 (0.010)	-0.028*** (0.005)		0.003 (0.010)	-0.025*** (0.004)		-0.003 (0.007)	-0.031*** (0.004)	
Tangibility	0.007 (0.032)	-0.086*** (0.023)		0.007 (0.039)	-0.073*** (0.022)		-0.039 (0.035)	-0.093*** (0.022)		-0.035 (0.022)	-0.116*** (0.029)	
B/M	0.037*** (0.011)	-0.017*** (0.005)		0.055*** (0.014)	-0.024*** (0.005)		0.038*** (0.010)	-0.016*** (0.004)		-0.034*** (0.008)	-0.023*** (0.006)	
Leverage	0.218*** (0.045)	0.004 (0.020)		0.208*** (0.044)	2.22 × 10 <sup>-5</sup> (0.018)		0.231*** (0.046)	0.021 (0.013)		0.028 (0.028)	0.008 (0.015)	
Cash Holdings	-0.032 (0.032)	-0.024 (0.023)		-0.035 (0.037)	-0.028 (0.021)		-0.031 (0.032)	-0.011 (0.027)		-0.017 (0.032)	-0.0006 (0.034)	
Payout	0.002 (0.003)	0.001 (0.001)		0.001 (0.003)	0.0005 (0.001)		0.003 (0.003)	0.0008 (0.001)		0.001 (0.001)	-0.0008 (0.001)	
R&D (0/1)	-0.012 (0.014)	-0.011 (0.007)		-0.013 (0.015)	-0.009 (0.007)		-0.013 (0.014)	-0.004 (0.008)		0.0007 (0.014)	-0.003 (0.008)	
Advertising	0.003 (0.009)	0.0007 (0.006)		0.015 (0.011)	0.0008 (0.006)		-0.002 (0.010)	0.002 (0.006)		-0.012 (0.008)	0.006 (0.007)	
IOR	-0.080** (0.037)	-0.061*** (0.015)		-0.081** (0.036)	-0.064*** (0.016)		-0.087** (0.043)	-0.107*** (0.014)		-0.068** (0.025)	-0.163*** (0.012)	
Ret (t-1)	-0.068*** (0.015)	-0.042*** (0.005)		-0.070*** (0.015)	-0.038*** (0.005)		-0.072*** (0.016)	-0.030*** (0.005)		0.002 (0.009)	-0.040*** (0.007)	
Pre (t-1) (ln)	0.006 (0.005)	0.016*** (0.002)		0.010* (0.005)	0.018*** (0.003)		0.004 (0.004)	0.018*** (0.002)		0.010** (0.004)	0.020*** (0.002)	
Volp (t-1)	-0.059** (0.029)	0.012 (0.019)		-0.058** (0.029)	0.008 (0.017)		-0.193** (0.092)	-0.349*** (0.058)		-0.111* (0.060)	-0.165** (0.065)	
Market Cap (t-1)	-0.043*** (0.006)	-0.106*** (0.003)		-0.042*** (0.005)	-0.108*** (0.003)		-0.043*** (0.005)	-0.111*** (0.004)		-0.035*** (0.004)	-0.102*** (0.004)	
log(analysts+1)	-0.014* (0.008)	-0.054*** (0.003)		-0.011 (0.007)	-0.052*** (0.003)		-0.017*** (0.006)	-0.043*** (0.004)		-0.026*** (0.006)	-0.046*** (0.004)	
Senior Ratio	0.860	-0.372	2.03	0.078	-1.62*	-2.75	-1.97	-1.13	9.72	-0.059	-1.41	10.4

	(1.22)	(0.536)	(1.82)	(2.58)	(0.982)	(2.73)	(2.30)	(0.817)	(9.43)	(1.83)	(1.04)	(9.62)
Female Ratio	5.94	3.83**	-4.09	8.63	4.68	1.67	16.9	6.82*	-20.6	10.3	2.26	-15.2
	(3.68)	(1.49)	(7.24)	(7.79)	(3.22)	(15.1)	(10.4)	(4.02)	(34.2)	(7.66)	(3.71)	(33.3)
Education	1.38**	0.028	1.97**	1.26	0.263	1.00	1.72	0.493	-0.739	1.40**	0.507	-1.76
	(0.683)	(0.264)	(0.850)	(1.11)	(0.625)	(1.60)	(1.30)	(0.551)	(2.79)	(0.664)	(0.352)	(3.10)
Population Density	0.584	0.781***	-0.677	-15.9*	5.63	-12.7	-111.8***	-11.7	65.4	-8.87	-8.09	88.4*
	(0.506)	(0.211)	(0.623)	(8.54)	(4.06)	(13.9)	(35.2)	(10.6)	(62.5)	(9.45)	(5.91)	(51.6)
Trust Index	0.488***	0.013	0.083	0.207	0.009	0.358	0.479***	0.011	0.829	0.185	-0.004	0.820
	(0.185)	(0.070)	(0.175)	(0.167)	(0.071)	(0.249)	(0.156)	(0.061)	(0.549)	(0.134)	(0.088)	(0.547)
Unemployment	-0.307	-0.459**	1.09**	-0.071	-0.374	2.14***	-0.919	-0.241	4.69***	-0.369	-0.324	5.03***
	(0.986)	(0.210)	(0.523)	(0.753)	(0.253)	(0.817)	(0.696)	(0.234)	(1.05)	(0.455)	(0.312)	(0.998)
Total Population (ln)	-0.070	-0.085	-0.361	-0.202	-0.070	-0.345	-0.077	-0.114	-2.32**	0.005	-0.008	-2.22**
	(0.123)	(0.068)	(0.310)	(0.273)	(0.101)	(0.561)	(0.299)	(0.116)	(1.03)	(0.154)	(0.119)	(0.992)
White (%)	0.076	-0.698	-4.04	-0.958	0.102	-8.10	-1.97	1.33	-5.96	0.698	0.032	-4.91
	(1.43)	(0.546)	(3.88)	(2.18)	(0.885)	(5.65)	(2.76)	(0.958)	(6.83)	(0.526)	(0.319)	(6.58)
Hispanic (%)	-0.103	-0.302	-5.88	0.061	-0.508	-11.1**	-0.946	1.73*	-14.4**	-0.033	0.199	-15.1**
	(1.36)	(0.505)	(3.54)	(2.19)	(0.866)	(4.51)	(2.89)	(1.01)	(6.50)	(0.557)	(0.274)	(6.50)
African American (%)	-0.600	-0.915	-4.55	0.736	-1.40	-12.1***	-2.26	0.890	-17.0*	-0.225	-0.012	-16.0
	(1.57)	(0.597)	(3.54)	(2.50)	(0.923)	(4.66)	(2.45)	(0.909)	(9.65)	(0.470)	(0.215)	(9.62)
Asian (%)	0.623	0.250	-2.28	3.18	0.830	-4.45	4.63	2.53**	-13.6	1.04	1.07	-13.8*
	(1.41)	(0.568)	(2.57)	(2.50)	(1.05)	(7.54)	(3.41)	(1.23)	(8.11)	(1.13)	(1.02)	(8.09)
Racial Diversity	-0.317	0.247*	0.767	-0.790	0.418*	-0.204	0.482	0.440	-2.26	0.524	0.156	-2.10
	(0.361)	(0.130)	(1.18)	(0.590)	(0.244)	(1.73)	(0.794)	(0.309)	(1.78)	(0.367)	(0.235)	(1.81)
Political Fragmentation	-0.160*	-0.042	-0.045	-0.223	-0.040	-0.426	0.059	-0.058	-0.213	0.026	-0.064*	-0.191
	(0.085)	(0.041)	(0.173)	(0.216)	(0.100)	(0.366)	(0.281)	(0.108)	(0.775)	(0.074)	(0.038)	(0.817)
Median Housing (ln)	0.030	-0.006	-0.028***	0.007	0.024	-0.116	0.112	0.030	-0.268***	-0.007	-0.017	-0.275***
	(0.019)	(0.007)	(0.010)	(0.082)	(0.026)	(0.083)	(0.075)	(0.022)	(0.082)	(0.030)	(0.023)	(0.082)
Median Income (ln)	-0.082	0.038	0.025	-0.302	-0.094	0.506	-0.658**	-0.066	0.978	-0.102	0.001	1.19
	(0.134)	(0.053)	(0.129)	(0.251)	(0.084)	(0.323)	(0.290)	(0.093)	(0.869)	(0.173)	(0.132)	(0.835)
Regional Ret	-0.005	0.0006	-0.326***	0.056	-0.008	-0.145	0.086	-0.001	-0.210	0.087	0.075**	-0.225*
	(0.034)	(0.010)	(0.087)	(0.061)	(0.018)	(0.097)	(0.060)	(0.018)	(0.138)	(0.061)	(0.029)	(0.132)
Regional Illiq	-0.199*	0.108	-7.56*	-1.53	1.58***	-3.56	-0.923	2.80**	-38.0	1.58	1.98	-38.1
	(0.115)	(0.067)	(4.21)	(1.09)	(0.502)	(4.35)	(4.03)	(1.22)	(24.8)	(4.49)	(1.85)	(23.7)
Regional Fpd	0.005	-0.037**	0.085	0.095	-0.056*	0.254***	-0.056	-0.074*	-0.167	0.025	-0.038	-0.170
	(0.045)	(0.016)	(0.062)	(0.083)	(0.029)	(0.097)	(0.127)	(0.040)	(0.208)	(0.075)	(0.045)	(0.210)
Regional Nstocks (ln)	-0.032	-0.007	0.024	0.024	-0.007	0.036	-0.030	0.032*	0.136	0.034	0.059***	0.135
	(0.028)	(0.010)	(0.054)	(0.050)	(0.018)	(0.052)	(0.069)	(0.017)	(0.223)	(0.035)	(0.017)	(0.201)
Regional Mktcap (ln)	-0.008	0.002	-0.069***	-0.062**	-0.002	-0.064*	-0.024	-0.001	-0.058	-0.002	-0.015	-0.038
	(0.017)	(0.006)	(0.023)	(0.027)	(0.009)	(0.034)	(0.060)	(0.011)	(0.068)	(0.023)	(0.013)	(0.063)
Regional IO	-0.042	-0.019	0.125*	-0.012	-0.029	0.034	0.232***	0.014	0.302	0.022	-0.003	0.324*
	(0.047)	(0.019)	(0.073)	(0.092)	(0.032)	(0.110)	(0.080)	(0.032)	(0.189)	(0.065)	(0.044)	(0.191)
Expense Ratio			0.722			-1.23			-3.54			-3.47
			(1.41)			(1.63)			(2.48)			(2.47)

Net Flow			0.004 (0.016)			-0.002 (0.029)			0.0009 (0.022)			0.0008 (0.022)
Total Load			-0.111 (0.173)			-0.040 (0.241)			0.380 (0.393)			0.381 (0.392)
Turnover			0.003 (0.008)			-0.037*** (0.011)			-0.016** (0.007)			-0.016** (0.007)
Fund Age (ln)			0.002 (0.004)			-0.005 (0.005)			0.0008 (0.005)			0.0008 (0.005)
TNA (ln)			0.003*** (0.0010)			-0.0009 (0.002)			-0.002 (0.002)			-0.002 (0.002)
Nstocks (ln)			0.012 (0.008)			0.007 (0.007)			0.0003 (0.006)			0.0002 (0.006)
Fund Family Asset (ln)			0.001 (0.004)			0.008 (0.006)			0.014* (0.007)			0.014* (0.007)
Fund Family Size (ln)			-0.004 (0.005)			-0.018* (0.010)			-0.026** (0.011)			-0.026** (0.011)
Team Managed (0/1)			-0.007 (0.004)			-0.003 (0.008)			-0.002 (0.005)			-0.002 (0.005)
Single Manager			-0.003 (0.005)			-0.004 (0.010)			-0.007 (0.012)			-0.007 (0.012)
Year × Quarter FE	Yes	Yes		Yes	Yes		Yes	Yes		Yes	Yes	
Industry FE	Yes	Yes		Yes	Yes		Yes	Yes		Yes	Yes	
Year × Quarter × Style FE			Yes			Yes			Yes			Yes
Local Community FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,800	141,887	37,380	27,342	141,888	64,607	29,821	150,300	56,743	16,221	108,875	56,743
R <sup>2</sup>	0.44635	0.37049	0.08414	0.47812	0.38143	0.13827	0.42690	0.36740	0.08421	0.46920	0.35396	0.08436

## Appendix 2A.4.2 Alternative RRI measures

In our baseline analysis, we construct a commuting zone-level RRI measure using a group-weighted negative Theil index. At the CZ-level, we compute each tract-level entropy-based diversity between White residents and each of the three largest minority groups—Black, Hispanic, and Asian—and compare it to the corresponding CZ-level entropy-based diversity of its. The population-weighted pair-wise difference would be the group-specific Theil index. We then aggregate these three indices into a single RRI measure by weighting each one by its population share within the combined minority population of the commuting zone. This approach adjusts for regional differences in minority composition—for example, the prominence of Hispanic populations in California—and allows for consistent comparisons across areas with different dominant minority groups. However, it excludes smaller minority groups and limits tract-level diversity to pairwise comparisons with White residents, even though entropy can capture the full distribution of all racial groups simultaneously. To ensure that our findings are not sensitive to this specific construction, we conduct robustness checks using three alternative measures that differ from the baseline in how they summarize and compare racial composition.

First, we compute a negative Theil index comparing White residents to the combined non-White population. This simplifies the baseline construction by eliminating group-specific weighting and treating all minority residents as a single group. It retains the entropy-based structure but reduces dimensionality. We name it as *RRI NegTheil(White/Non-white)*.

Second, we adopt a multi-group entropy approach following Trounstein (2016), which uses a single Theil index to capture the full distribution across all major racial categories simultaneously.<sup>43</sup> Unlike the baseline, this measure does not rely on pairwise comparisons or

---

<sup>43</sup> Specifically, we calculate tract-level entropy as:  $E_i = \sum_r P_i^r \times \log(1/P_i^r)$ , where  $P_i^r$  represent the proportion of racial group  $r$  in census tract  $i$ , and  $i$  includes five categories: White, Black, Hispanic, Asian, and Other (Trounstein, 2016) We apply the same formula at the commuting zone level to compute overall racial diversity for each zone.

separate weights, which allows a more integrated view of diversity but may introduce sensitivity to tracts with extreme racial compositions. We name it as *RRI NegTheil(All)*.

Third, we use a similarity index, defined as the negative of the dissimilarity index, which is one of the earliest and most widely used measures of residential evenness (Jahn, Schmit, and Schrag, 1947). While conceptually aligned with the goal of capturing residential evenness, it differs mathematically from the Theil-based measures. Specifically, it relies on the absolute value of differences between the racial composition of tracts and their broader reference area, rather than entropy-based calculations. The similarity index is defined as:

$$\text{RRI Similarity}_{j,k} = 1 - \frac{1}{2} \sum_i \left| \frac{m_{i,k}}{M_{j,k}} - \frac{w_i}{W_j} \right|,$$

where  $m_{i,k}$  and  $w_i$  denote the number of racial minority  $k$  and white people, respectively, in census tract  $i$  within commuting zone  $j$ .  $M_{j,k}$  and  $W_j$  represent the total number of racial minority  $k$  and white people, respectively, in commuting zone  $j$ . To illustrate the intuition behind the similarity index, consider a local area  $j$  with two census tracts ( $i = 1, 2$ ) and two racial groups: the majority  $k$  and white people, respectively. If commuting zone  $j$  is perfectly segregated, all minority members reside in the first tract ( $\frac{m_{1,k}}{M_{j,k}} = 1$  and  $\frac{w_1}{W_j} = 0$ ), while all majority members are in the second ( $\frac{m_{2,k}}{M_{j,k}} = 0$  and  $\frac{w_2}{W_j} = 1$ ), yielding a similarity index of  $-1$  (RRI Similarity =  $-1$ ). Conversely, in a perfectly integrated commuting zone with equal distribution of both groups across tracts ( $\frac{m_{i,k}}{M_{j,k}} = \frac{w_i}{W_j} = 0.5, i = 1, 2$ ), the similarity index takes on a value of  $0$  (RRI Similarity =  $0$ ). This example represents the two extreme cases. In reality, no census tract is perfectly integrated or segregated; therefore,  $\text{RRI Similarity} \in (-1, 0)$ .

As in constructing our negative Theil index, we calculate three  $\text{RRI Similarity}_j^{(k)}$  using the above equation, considering each of the three minority racial groups (African American, Asian, and Hispanic) in conjunction with the majority group (white). Subsequently, we aggregate these race-specific similarity indices for state  $j$  based on the relative sizes of the three

racial minority groups, employing the formula:

$$\text{RRI Similarity}_j = \sum_k P_k \times \text{RRI Similarity}_j^{(k)},$$

where  $k$  denotes minority group  $k$ , including African American, Asian, and Hispanic populations, and  $P_k$  is the proportion of minority  $k$  relative to the combined African American, Asian, and Hispanic populations. All results for firm-level tests and fund-level tests are reported in Table 2A.4.2. Across all three measures, our main results on stock price informativeness and information asymmetry remain consistent. Full results are reported in Appendix 2A.4.

**Table 2A.4.2: Results of Alternative Evenness Measure**

**Panel A: Stock Informativeness Results**

	Dependent Variable: Future Cash Flow		
	(1)	(2)	(3)
RRI Similarity × log(M/A)	0.043** (0.019)		
RRI NegTheil (White VS Minority) × log(M/A)		0.041** (0.018)	
RRI NegTheil (all) × log(M/A)			0.037* (0.022)
RRI Similarity	-0.118** (0.050)		
RRI NegTheil (White VS Minority)		-0.135*** (0.049)	
RRI NegTheil (all)			-0.150*** (0.046)
log(M/A)	0.061*** (0.013)	0.048*** (0.010)	0.047*** (0.010)
Racial Diversity × log(M/A)	0.009 (0.015)	0.009 (0.015)	0.009 (0.015)
log(M/A) × Board Diversity	0.009** (0.004)	0.009** (0.004)	0.009** (0.004)
Age	0.008** (0.004)	0.008** (0.004)	0.008** (0.004)
Sale	0.022*** (0.004)	0.022*** (0.004)	0.022*** (0.004)
Tangibility	-0.041*** (0.013)	-0.041*** (0.013)	-0.041*** (0.013)
Cash Flow (t)	0.529*** (0.029)	0.528*** (0.029)	0.528*** (0.029)
Invest	-0.108*** (0.022)	-0.107*** (0.022)	-0.107*** (0.022)
Illiquidity	0.0003 (0.0004)	0.0003 (0.0004)	0.0003 (0.0004)
Volp (ln)	-0.004*** (0.0006)	-0.004*** (0.0006)	-0.004*** (0.0006)
Analyst Cover (ln)	-0.003* (0.002)	-0.003* (0.002)	-0.003* (0.002)
Leverage	0.047*** (0.010)	0.047*** (0.010)	0.047*** (0.010)
Cash Holdings	-0.041*** (0.012)	-0.040*** (0.012)	-0.040*** (0.012)
Asset (ln)	-0.012*** (0.003)	-0.012*** (0.003)	-0.012*** (0.003)
IOR	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)
Senior Ratio	-0.009 (0.113)	-0.017 (0.113)	-0.037 (0.113)
Female Ratio	-1.18** (0.518)	-1.33** (0.537)	-1.28** (0.521)
Education	0.114* (0.062)	0.131** (0.062)	0.123** (0.058)
Population Density	$4.02 \times 10^{-6}$ ( $4.09 \times 10^{-6}$ )	$3.83 \times 10^{-6}$ ( $4 \times 10^{-6}$ )	$5.13 \times 10^{-6}$ ( $4.01 \times 10^{-6}$ )
Trust Index	0.043** (0.019)	0.044** (0.019)	0.043** (0.019)
Unemployment	0.061 (0.081)	0.062 (0.083)	0.063 (0.085)
Racial Diversity	0.004 (0.058)	-0.003 (0.056)	-0.004 (0.056)
Total Population (ln)	-0.012* (0.007)	-0.013* (0.007)	-0.013* (0.007)
White (%)	-0.029 (0.136)	-0.050 (0.144)	-0.056 (0.145)

Hispanic (%)	0.008 (0.122)	-0.018 (0.127)	-0.026 (0.128)
African American (%)	0.004 (0.121)	-0.018 (0.126)	-0.031 (0.126)
Asian (%)	0.085 (0.158)	0.088 (0.163)	0.057 (0.166)
Political Fragmentation	0.007 (0.021)	0.007 (0.021)	0.007 (0.021)
Median Housing (ln)	-0.012* (0.007)	-0.012* (0.007)	-0.010 (0.007)
Median Income (ln)	-0.015 (0.020)	-0.020 (0.019)	-0.020 (0.019)
Regional Illiq	-0.007 (0.152)	0.010 (0.152)	$-5.35 \times 10^{-5}$ (0.152)
Regional Ret	0.021*** (0.007)	0.021*** (0.007)	0.021*** (0.007)
Regional IO	0.017** (0.007)	0.017** (0.007)	0.016** (0.007)
Regional Mktcap (ln)	-0.006** (0.002)	-0.006** (0.002)	-0.006** (0.002)
Regional Nstocks (ln)	0.005 (0.004)	0.006 (0.004)	0.006 (0.004)
Regional Fpd	-0.013* (0.007)	-0.012* (0.007)	-0.013* (0.007)
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	23,976	23,976	23,976
R <sup>2</sup>	0.82271	0.82278	0.82276

**Panel B: Stock Information Asymmetry and Fund Local Investing Results**

	Dependent Variables:								
	Adverse Selection (%)			VCV			Local Stock-picking Ability (%)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RRI Similarity	-1.94** (0.749)			-0.792*** (0.275)			-4.83*** (1.33)		
RRI Similarity × IOR	-0.363 (0.274)			-0.494*** (0.188)					
RRI NegTheil (White VS Minority)		-2.18*** (0.510)			-0.363 (0.257)			-4.33*** (0.793)	
RRI NegTheil (White VS Minority) × IOR		-0.394 (0.285)			-0.630*** (0.168)				
RRI NegTheil (All)			-2.83*** (0.492)			-0.487* (0.276)			-5.05*** (0.710)
RRI NegTheil (All) × IOR			-0.357 (0.288)			-0.645*** (0.180)			
IOR	-0.304* (0.158)	-0.205** (0.084)	-0.192** (0.083)	-0.334*** (0.109)	-0.221*** (0.049)	-0.218*** (0.050)			
Age	0.020*** (0.008)	0.020** (0.008)	0.019** (0.008)	-0.033*** (0.004)	-0.033*** (0.004)	-0.033*** (0.004)			
Inversed Volume	0.691*** (0.030)	0.691*** (0.031)	0.691*** (0.030)	0.573*** (0.031)	0.574*** (0.031)	0.574*** (0.031)			
Sale	0.004 (0.009)	0.004 (0.009)	0.005 (0.009)	-0.024*** (0.004)	-0.024*** (0.004)	-0.024*** (0.004)			
Tangibility	-0.025 (0.031)	-0.023 (0.031)	-0.023 (0.031)	-0.088*** (0.020)	-0.088*** (0.020)	-0.088*** (0.020)			
B/M	0.046*** (0.010)	0.046*** (0.011)	0.046*** (0.010)	-0.018*** (0.005)	-0.018*** (0.005)	-0.018*** (0.005)			
Leverage	0.216*** (0.038)	0.216*** (0.038)	0.215*** (0.038)	0.009 (0.021)	0.009 (0.021)	0.009 (0.021)			
Cash Holdings	-0.045 (0.031)	-0.045 (0.031)	-0.046 (0.031)	-0.021 (0.026)	-0.021 (0.026)	-0.021 (0.026)			
Payout	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)	0.001 (0.0010)	0.001 (0.0010)	0.001 (0.0010)			
R&D (0/1)	-0.011 (0.014)	-0.011 (0.014)	-0.012 (0.014)	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.006)			
Advertising	0.009 (0.009)	0.009 (0.009)	0.008 (0.009)	0.0003 (0.007)	6.78 × 10 <sup>-5</sup> (0.007)	1.84 × 10 <sup>-5</sup> (0.007)			
Ret (t-1)	-0.073*** (0.015)	-0.074*** (0.015)	-0.074*** (0.015)	-0.039*** (0.004)	-0.039*** (0.004)	-0.039*** (0.004)			
Pre (t-1) (ln)	0.006 (0.004)	0.006 (0.004)	0.006 (0.004)	0.016*** (0.002)	0.016*** (0.002)	0.016*** (0.002)			
Volp (t-1) (ln)	-0.059**	-0.058**	-0.057**	0.006	0.005	0.005			

	(0.027)	(0.027)	(0.027)	(0.017)	(0.017)	(0.017)			
log(analysts+1)	-0.016**	-0.016**	-0.016**	-0.052***	-0.052***	-0.052***			
	(0.008)	(0.008)	(0.008)	(0.003)	(0.003)	(0.003)			
Market Cap (t-1)	-0.041***	-0.041***	-0.041***	-0.105***	-0.105***	-0.105***			
	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)			
Senior Ratio	1.32	0.903	0.521	-0.722	-0.588	-0.639	0.269	-2.58	-2.27
	(2.14)	(2.01)	(1.93)	(0.824)	(0.834)	(0.844)	(2.94)	(2.66)	(2.57)
Female Ratio	12.9**	12.6**	13.1**	5.78**	5.68**	5.82**	-17.9	-11.8	-9.97
	(5.82)	(5.77)	(5.64)	(2.62)	(2.56)	(2.59)	(13.8)	(13.0)	(12.2)
Education	1.53**	1.88**	1.88**	0.673*	0.594*	0.582*	3.31***	2.97***	2.91***
	(0.762)	(0.774)	(0.727)	(0.358)	(0.349)	(0.344)	(1.09)	(1.11)	(1.05)
Population Density (10k/km <sup>2</sup> )	-8.82***	-6.16*	-5.65*	4.21***	3.92***	4.18***	-10.9***	-10.1**	-11.5***
	(3.38)	(3.40)	(3.13)	(1.43)	(1.41)	(1.30)	(3.80)	(4.00)	(3.80)
Trust Index	0.438**	0.481***	0.436**	0.043	0.063	0.055	0.128	0.130	0.076
	(0.173)	(0.175)	(0.172)	(0.071)	(0.074)	(0.073)	(0.168)	(0.152)	(0.150)
Unemployment	0.357	0.426	0.339	-0.136	-0.132	-0.142	1.52	1.84**	1.43
	(0.797)	(0.805)	(0.808)	(0.247)	(0.251)	(0.253)	(0.967)	(0.886)	(0.931)
White (%)	3.67*	3.13	2.89	0.685	0.580	0.575	-9.71***	-10.3***	-9.84***
	(2.04)	(2.11)	(2.07)	(0.642)	(0.671)	(0.666)	(3.29)	(3.26)	(2.93)
Hispanic (%)	2.55	1.50	1.55	0.330	0.160	0.202	-12.0***	-13.9***	-13.0***
	(1.66)	(1.76)	(1.71)	(0.581)	(0.640)	(0.617)	(2.87)	(3.08)	(2.62)
African American (%)	2.10	1.54	1.08	-0.0005	-0.154	-0.191	-8.30**	-9.58***	-10.4***
	(2.16)	(2.08)	(2.06)	(0.652)	(0.691)	(0.692)	(3.31)	(3.34)	(2.99)
Asian (%)	5.63**	4.87**	3.68	1.08	1.19	0.959	-6.17*	-6.76**	-7.42**
	(2.39)	(2.43)	(2.41)	(0.884)	(0.918)	(0.926)	(3.46)	(3.24)	(2.83)
Racial Diversity	0.950*	0.806	0.748	0.375	0.372	0.372	-1.65	-1.43	-1.34
	(0.559)	(0.558)	(0.525)	(0.232)	(0.232)	(0.234)	(1.27)	(1.23)	(1.15)
Total Population (ln)	-0.004	-0.112	-0.133	-0.080	-0.067	-0.073	-0.211	-0.602*	-0.477
	(0.206)	(0.194)	(0.182)	(0.100)	(0.102)	(0.099)	(0.350)	(0.332)	(0.312)
Political Fragmentation	-0.079	-0.051	-0.045	-0.120*	-0.091	-0.091	-0.928***	-0.932***	-0.907***
	(0.205)	(0.192)	(0.187)	(0.068)	(0.069)	(0.070)	(0.296)	(0.278)	(0.268)
Median Housing (ln)	0.086	0.089	0.083	0.025	0.031	0.031	-0.223***	-0.199***	-0.212***
	(0.064)	(0.061)	(0.061)	(0.024)	(0.025)	(0.025)	(0.046)	(0.047)	(0.046)
Median Income (ln)	-0.242	-0.357	-0.408*	-0.066	-0.066	-0.072	-0.316	-0.287	-0.353
	(0.237)	(0.242)	(0.239)	(0.088)	(0.088)	(0.089)	(0.280)	(0.270)	(0.263)
Regional Illiq	-2.41*	-2.04	-2.00	1.96***	2.02***	2.04***	-8.16	-7.47	-7.72
	(1.38)	(1.37)	(1.36)	(0.561)	(0.557)	(0.557)	(7.84)	(8.03)	(7.98)
Regional Ret	0.011	0.011	0.011	-0.011	-0.011	-0.011	-0.178*	-0.170*	-0.174*
	(0.051)	(0.051)	(0.052)	(0.016)	(0.016)	(0.016)	(0.096)	(0.096)	(0.096)
Regional Mktcap (ln)	-0.040	-0.037	-0.036	0.008	0.008	0.009	-0.130***	-0.148***	-0.147***
	(0.030)	(0.030)	(0.030)	(0.008)	(0.007)	(0.007)	(0.047)	(0.047)	(0.044)
Regional IO	-0.049	-0.052	-0.067	0.006	0.003	-0.0009	0.235*	0.151	0.120
	(0.064)	(0.065)	(0.065)	(0.026)	(0.025)	(0.025)	(0.133)	(0.141)	(0.135)

Regional Nstocks (ln)	0.004 (0.044)	0.006 (0.044)	0.011 (0.044)	-0.011 (0.015)	-0.013 (0.015)	-0.012 (0.015)	0.075 (0.052)	0.039 (0.050)	0.044 (0.051)
Regional Fgpd	0.121* (0.062)	0.126** (0.063)	0.120* (0.063)	-0.024 (0.023)	-0.025 (0.023)	-0.026 (0.023)	0.250*** (0.084)	0.236*** (0.083)	0.203** (0.080)
Expense Ratio							0.004 (1.26)	-0.038 (1.27)	0.019 (1.26)
Net Flow							-0.021 (0.018)	-0.020 (0.018)	-0.021 (0.018)
Total Load							0.170 (0.180)	0.188 (0.180)	0.193 (0.180)
Turnover							0.0009 (0.006)	0.001 (0.006)	0.001 (0.006)
Fund Age (ln)							0.0005 (0.004)	0.0001 (0.004)	6.73 × 10 <sup>-5</sup> (0.004)
TNA (ln)							0.002* (0.001)	0.002* (0.001)	0.002* (0.001)
Nstocks (ln)							0.004 (0.009)	0.005 (0.009)	0.005 (0.009)
Fund Family Asset (ln)							0.007** (0.003)	0.007** (0.003)	0.007** (0.003)
Fund Family Size (ln)							-0.013** (0.006)	-0.014** (0.006)	-0.013** (0.006)
Team Managed (0/1)							-0.006 (0.004)	-0.006 (0.004)	-0.006 (0.004)
Single Manager							-0.003 (0.005)	-0.003 (0.005)	-0.002 (0.005)
Year × Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes			
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes			
local region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Quarter × Style FE							Yes	Yes	Yes
Observations	29,088	29,088	29,088	147,191	147,191	147,191	49,350	49,350	49,350
R <sup>2</sup>	0.43172	0.43208	0.43239	0.36667	0.36670	0.36669	0.07137	0.07163	0.07185

## Appendix 2A.4.3 Placebo test and other robustness checks for funds

**Table 2A.4.3: Placebo Test and Other Robustness Checks at the Fund Level**

This table presents fund-level results for index funds (columns (1)–(2)) as a placebo test, and alternative model specifications (columns (3)–(5)). Column (3) excludes funds headquartered in Massachusetts, New York, and California. Column 4 includes fund fixed effects. Column (5) adds controls for past local and nonlocal stock-picking performance. The dependent variable in column (1) is local bias, defined as the difference between a fund’s portfolio weight on local stocks and the local stocks’ weight in the market portfolio (Wei and Zhang, 2020). Columns (2)–(5) use local stock-picking ability (LSP) as the dependent variable. *LSP* is a quarterly, portfolio-level measure that captures a fund’s ability to predict local stock alphas. Local stocks are firms headquartered in the same commuting zone as the fund. RRI NegTheil measures the evenness aspect of racial residential integration. Standard errors are clustered at the commuting-zone level and reported in parentheses. Note: Standard errors are shown in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

	Index Funds		Removing MA, NY, CA	Fund FE	Extra Controls
	(1) Local Bias	(2) LSP(%)	(3) LSP(%)	(4) LSP(%)	(5) LSP(%)
RRI NegTheil	−0.166 (0.296)	−0.362 (1.28)	−2.73** (1.28)	−4.23*** (1.26)	−4.79*** (1.03)
Lagged Stock-picking Ability (%)					0.009 (0.016)
Non Local Stock Picking (%)					0.022*** (0.003)
Expense Ratio	−0.300 (0.724)	−1.63 (2.53)	0.143 (1.70)	−0.359 (1.90)	0.137 (1.20)
Net Flow	0.004 (0.003)	0.003 (0.012)	−0.013 (0.028)	−0.044* (0.025)	−0.022 (0.016)
Total Load	−0.256 (0.196)	0.556 (0.748)	0.118 (0.382)	−0.007 (0.316)	0.213 (0.183)
Turnover	0.0008 (0.0007)	0.0009 (0.003)	−0.003 (0.007)	−0.002 (0.004)	0.003 (0.007)
Fund Age (ln)	0.003 (0.003)	0.009 (0.009)	0.004 (0.006)	0.013 (0.016)	0.0006 (0.004)
TNA (ln)	−0.001 (0.002)	−0.005 (0.004)	0.0004 (0.002)	−0.013*** (0.003)	0.002** (0.001)
Nstocks (ln)	−0.001 (0.007)	−0.016 (0.017)	0.004 (0.007)	−0.006 (0.013)	0.005 (0.010)
Fund Family Asset (ln)	−0.004* (0.002)	−0.013 (0.008)	0.006 (0.006)	−0.0002 (0.005)	0.007** (0.003)
Fund Family Size (ln)	0.005 (0.003)	0.013 (0.011)	−0.011 (0.011)	−0.004 (0.006)	−0.011* (0.006)
Team Managed (0/1)	−0.006* (0.003)	−0.019 (0.011)	−0.007 (0.007)	−0.007 (0.005)	−0.005 (0.004)
Single Manager	0.001 (0.003)	−0.0004 (0.011)	−0.003 (0.009)	−0.002 (0.006)	−0.001 (0.005)
Senior Ratio	−0.993 (0.968)	−9.50 (8.40)	−0.423 (2.30)	−2.72 (2.17)	−1.58 (2.72)

Female Ratio	3.41 (4.62)	54.2*** (17.9)	-15.6 (12.2)	-15.4 (13.1)	-14.5 (13.4)
Education	-0.736 (0.443)	3.98** (1.63)	1.48 (1.04)	3.40*** (1.22)	3.31*** (1.08)
Population Density (10k/km <sup>2</sup> )	7.13*** (2.22)	-17.3 (13.3)	18.1 (16.4)	-10.6** (4.33)	-8.63** (3.60)
Trust Index	0.064 (0.051)	0.450 (0.365)	-0.070 (0.180)	0.266* (0.139)	0.084 (0.162)
Unemployment	0.141 (0.167)	1.13 (1.79)	1.30* (0.735)	1.55 (0.969)	1.54 (1.01)
White (%)	0.677 (1.43)	-10.7* (5.85)	-1.27 (3.10)	-11.8*** (3.00)	-10.3*** (2.92)
Hispanic (%)	0.655 (1.68)	-20.1*** (7.29)	-5.94** (2.71)	-13.9*** (3.06)	-12.5*** (2.65)
African American (%)	0.664 (2.40)	-27.3** (11.0)	-4.67 (2.94)	-12.9*** (3.68)	-9.48*** (3.32)
Asian (%)	0.336 (1.68)	-13.5** (6.43)	-2.06 (4.00)	-9.54*** (3.13)	-7.28** (3.15)
Racial Diversity	0.199 (0.531)	2.56 (2.84)	2.87** (1.32)	-2.27* (1.15)	-2.10* (1.13)
Total Population (ln)	-0.258* (0.152)	0.653* (0.356)	0.011 (0.339)	-0.283 (0.394)	-0.502 (0.317)
Political Fragmentation	0.005 (0.159)	-0.981 (0.609)	-0.437 (0.353)	-0.890** (0.357)	-1.09*** (0.298)
Median Housing (ln)	0.017 (0.015)	-0.042 (0.128)	-0.198* (0.106)	-0.175*** (0.054)	-0.212*** (0.045)
Median Income (ln)	-0.111** (0.050)	0.601 (0.655)	-0.440 (0.279)	-0.168 (0.302)	-0.267 (0.267)
Regional Ret	0.016** (0.007)	0.061 (0.088)	-0.157 (0.107)	-0.167* (0.089)	-0.181* (0.097)
Regional Illiq	-0.266 (0.166)	-6.94 (5.24)	-12.6 (8.85)	-8.68 (8.51)	-9.98 (8.05)
Regional Fpd	-1.51×10 <sup>-5</sup> (0.016)	0.034 (0.186)	0.178** (0.078)	0.223** (0.092)	0.266*** (0.091)
Regional Nstocks (ln)	-0.007 (0.009)	0.026 (0.108)	0.050 (0.056)	0.032 (0.055)	0.039 (0.050)
Regional Mktcap (ln)	-0.038*** (0.011)	-0.016 (0.062)	-0.127*** (0.045)	-0.161*** (0.058)	-0.149*** (0.050)
Regional IO	0.036 (0.026)	0.003 (0.295)	-0.088 (0.109)	0.211 (0.145)	0.160 (0.144)
Year × Quarter × Style FE	Yes	Yes	Yes		Yes
Commuting Zone FE	Yes	Yes	Yes		Yes
Year × Quarter FE				Yes	
Fund FE				Yes	
Commuting Zone FE	Yes	Yes	Yes	Yes	Yes
Observations	5,437	5,437	22,745	49,350	47,650
R <sup>2</sup>	0.53603	0.18400	0.06938	0.09508	0.08245

## Chapter 3: Racial Integration in Social Networks and Asset Pricing

### Efficiency

#### 3.1 Introduction

Financial markets are fundamentally social ecosystems. A vast and growing literature in social finance establishes that investment decisions are not made in isolation but are profoundly shaped by peer interactions, word-of-mouth communication, and social networks (Antweiler and Frank, 2004; Ivković and Weisbenner, 2007; Brown, Ivković, Smith, and Weisbenner, 2008; Kaustia and Knüpfer, 2012; Pool et al., 2015; Bailey et al., 2018; Hirshleifer, 2020; Kuchler and Stroebl, 2021). From the diffusion of stock market participation (Hong, Kubik, and Stein, 2004) to the transmission of sentiment (Shiller, 2017), the structure of social ties governs the flow of economic information.<sup>44</sup> In this paradigm, the network functions as the market's nervous system, determining how quickly and accurately private signals are aggregated into prices (Duffie and Manso, 2007; Han and Yang, 2013; Ozsoylev and Walden, 2011; Hirshleifer et al., 2025).

However, existing theoretical models of social learning typically rely on a simplifying assumption: communication is homogeneous. Standard frameworks, such as Han and Yang (2013) or Ozsoylev and Walden (2011), posit a world in which information flows with equal friction, or none at all, between any two connected agents. This colorblind network topology ignores a pervasive feature of human social structure: stratification driven by homophily. In reality, social networks are deeply segmented by race, ethnicity, and geography (McPherson et al., 2001). Information often circulates rapidly within demographic clusters but encounters significant friction when crossing group boundaries. By neglecting this heterogeneity, standard models may overstate

---

<sup>44</sup> Hong et al. (2004) demonstrate a social *multiplier* in stock market participation, where the density of local social ties, promotes peer-to-peer information sharing and lowers the fixed costs of market entry. Complementing this, Shiller (2017) identifies social networks as the primary conduit for narrative contagion, wherein the topology of social connections governs the epidemic-like spread of economic sentiment and the subsequent formation of asset bubbles.

market efficiency and fail to capture the distributional consequences of social segregation. The empirical findings from Chapter 2 also support this concern.

This chapter bridges this gap by asking a fundamental question: *How does racial integration affect asset pricing efficiency and investor welfare?* Conceptually, the framework extends naturally beyond race: any demographic attribute that generates homophily-driven stratification and heterogeneous communication frictions across groups can be incorporated into the model. We focus specifically on racial integration rather than other forms of social distance for two primary reasons: significance and measurability. First, race remains one of the most salient and persistent cleavages in U.S. social and economic life (Acharya, Blackwell, and Sen, 2016; Derenoncourt, 2022; Abascal, Armenta, Halm, and Hopkins, 2024). As Alesina and La Ferrara (2002) document, racial heterogeneity is a primary driver of reduced social trust, effectively narrowing the radius of trust within demographic groups (Fukuyama, 1995). This fragmentation has profound economic consequences. By linking these insights, we posit that racial segregation fragments social networks in ways that impede the free flow of financial information.

Second, unlike more abstract notions of cultural distance, racial integration can be measured using racial residential integration (RRI), which provides plausibly exogenous variation in social structure and has been shown to increase cross-group interaction (e.g., Chetty, Hendren, Kline, and Saez, 2014). By modeling this specific friction, we provide a tractable framework for analyzing how the geometry of social diversity shapes the “physics” of price formation.

To analyze the equilibrium relationship between racial integration and market efficiency, we develop a noisy rational expectations equilibrium (REE) model embedded in a racially stratified islands network.<sup>45</sup> The economy consists of a racial majority and a racial minority, whose social

---

<sup>45</sup> Isolated island models are a standard framework for analyzing the effects of localized social interactions in large-scale networks. For foundational treatments, see Jackson (2008) and Wu (2025).

interactions are constrained by the degree of social integration (e.g., residential and professional integration). We introduce a key structural parameter,  $\rho$  (racial integration), which endogenizes the impact of social structure on investment decisions along two dimensions.

First, signal transmission:  $\rho$  governs the fidelity of cross-group communication. While within-group signals are transmitted with partial precision, signals that cross racial boundaries are subject to an additional “trust tax” or transmission bias that further degrades their quality. This mechanism captures the sociological reality that intergroup communication is often noisier or costlier due to implicit bias or the absence of shared context (Lazear, 1999; Dovidio, Gaertner, Kawakami, and Hodson, 2002).

Second, information acquisition costs:  $\rho$  determines the marginal cost of acquiring private signals. We assume that the cost of sourcing and verifying private information varies across racial groups and decreases with social integration. This specification reflects the role of social networks as informational infrastructures, where social proximity substantially reduces search and verification costs (Hong et al., 2005; Pool et al., 2015). Moreover, the racial disparity in acquisition costs captures the idea that, in segregated environments, exclusion from central social nodes acts as a structural barrier that raises the shadow price of information for minority investors (Loury, 1997; DiMaggio and Garip, 2012). Together, these features allow the model to characterize how network structure and demographic segmentation jointly shape the supply of information and the informativeness of asset prices.

Our analysis yields three main results. First, in a baseline setting with exogenous information, racial integration acts as a positive shock to price informativeness. By lowering transmission frictions between the larger majority cluster and the smaller minority cluster, integration unlocks a latent pool of diverse signals that were previously sequestered. This enhanced aggregation reduces aggregate uncertainty, leading to a strictly lower cost of equity and higher trading volume. This

result reinforces the classic insight that prices are only as informative as the market's ability to aggregate dispersed signals under strategic trading. By easing cross-cluster communication frictions, integration strengthens information aggregation (Grossman and Stiglitz, 1980; Kyle, 1985), reduces the required return for bearing information risk (Easley and O'Hara, 2004; Lambert, Leuz, and Verrecchia, 2012), and increases trading activity through greater informed participation and lower informational frictions (Glosten and Milgrom, 1985; Hasbrouck, 1991).

Second, when we endogenize information acquisition, we uncover a novel cost-reduction channel. While standard theory predicts that more informative prices crowd out private research (the Grossman–Stiglitz effect), we show that integration lowers the marginal cost of acquiring soft information, particularly for minority investors who previously faced high search frictions. This cost reduction dominates free-riding incentives, resulting in a net increase in information production, market liquidity, and price informativeness.

Third, we document a powerful redistributive effect. Segregated markets exhibit a structural welfare gap: minority investors face information poverty relative to the majority, trade on noisier beliefs, and capture fewer informational rents. Integration substantially narrows this gap. By reducing cross-cluster communication frictions and broadening access to the aggregate signal set, integration allows minority investors to compete on a more level informational footing, shifting informational rents away from the dominant group toward previously marginalized participants (Verrecchia, 1982; Kyle, 1985). This redistribution is even more pronounced when information acquisition is endogenous, as lower communication frictions further reduce the effective cost of becoming informed for minority traders.

This chapter contributes to several strands of the literature. First, we contribute to research on social finance and information diffusion. While prior studies show that social networks facilitate information transfer (e.g., Cohen et al., 2008; Kuchler and Stroebel, 2021) and that network

topology affects price efficiency (Colla and Mele, 2010; Ozsoylev and Walden, 2011), these models typically assume demographically homogeneous nodes. We advance this literature by explicitly modeling racial homophily, the empirical tendency for social ties to form within demographic groups, as a fundamental network friction (McPherson et al., 2001). By formalizing how in-group social capital and intergroup communication barriers shape the cost and precision of information, we show that racial integration is a first-order determinant of market quality. This bridges the gap between abstract network models and the empirical reality of racially segmented financial participation and information access (e.g., Loury, 1977; Hong et al., 2004).

Second, our model offers new insights into policy debates on diversity and market efficiency. While much of the finance literature frames diversity through governance, representation, and firm outcomes (Adams and Ferreira, 2009; Ahern and Dittmar, 2012; Gompers and Wang, 2017; Bernile et al., 2018), our mechanism highlights the importance of network inclusion: whether minority investors are connected to the dominant information cluster such that their signals and inferences can be pooled and transmitted. In our model, integration does not merely reallocate surplus; it expands the effective aggregate information set available to the market, generating more informative prices and a lower cost of capital for all firms. This result aligns with the diversity dividend hypothesis while extending it to the mechanics of price discovery in secondary markets.

The remainder of the chapter is organized as follows. Section 3.2 presents the model economy and information structure. Section 3.3 analyzes equilibrium outcomes under exogenous information. Section 3.4 extends the model to endogenous information acquisition. Section 3.5 concludes.

## **3.2 Model**

### 3.2.1 Economy and agents

Following Han and Yang (2013), we consider a static, one-period economy populated by a finite number of traders who exchange two financial assets. The first asset is risk-free, serves as the numeraire, and offers a gross return normalized to one. The second asset is risky, interpretable as a stock or market index, and delivers a random terminal payoff  $\tilde{v}$ , given by:

$$\tilde{v} \sim \mathcal{N}(\bar{v}, \tau_v^{-1}), \quad (12)$$

where  $\tau_v$  is the inverse of the variance (precision) of the payoff.

The social structure of the economy is modeled as a collection of segmented markets, following the island topology described by Jackson (2008) and Wu (2025). The economy is divided into  $G$  distinct communities (or “islands”), which can be interpreted as segmented markets such as cities or commuting zones. Each community contains  $N$  traders. Within each community, traders are stratified into two demographic groups: a racial majority group ( $W$ ) and a racial minority group ( $M$ ). Let  $\mu \in (0.5, 1)$  denote the fraction of traders belonging to the majority group. Consequently, each island contains  $\mu N$  majority traders and  $(1 - \mu)N$  minority traders.

This demographic structure is central to modeling frictions in information diffusion arising from homophily, the tendency for individuals to associate disproportionately with similar others (McPherson et al., 2001). Traders act competitively and maximize expected utility over terminal wealth. We assume all traders exhibit Constant Absolute Risk Aversion (CARA), with preferences given by:

$$U_i = -e^{-\gamma W_i}, \quad (13)$$

where  $\gamma$  is the coefficient of risk aversion and  $W_i$  denotes the terminal wealth of trader  $i$ . This preference specification is standard in the noisy rational expectations literature (e.g., Hellwig, 1980), as it ensures that equilibrium demand functions are linear in prices and information signals.

Prior to trading, agents differ in their informational status. A trader may choose to remain uninformed, relying solely on prices and socially transmitted signals, or become informed by

additionally observing a private signal. Let  $\phi_W$  denote the fraction of informed traders in the racial majority group and  $\phi_M$  the fraction of informed traders in the racial minority group. These fractions,  $(\phi_W, \phi_M)$ , are key state variables that determine the aggregate quantity and composition of private information injected into each local market.

### 3.2.2 Information structure and heterogeneous diffusion

The core feature of our model is heterogeneity in how information diffuses across social boundaries. Information enters the economy through private acquisition and subsequently propagates through the social network via communication channels whose quality depends on the racial composition of the interacting pair.

#### 3.2.2.1 Private information acquisition

If a trader  $i$  chooses to become informed, they incur a cost  $c_r$  (which depends on their group  $r \in \{W, M\}$ ) to observe a private signal  $\tilde{s}_i$  regarding the asset's liquidating value:

$$\tilde{s}_i = \tilde{v} + \tilde{\varepsilon}_i, \quad \tilde{\varepsilon}_i \sim \mathcal{N}(0, \tau_s^{-1}), \quad (14)$$

where  $\tilde{\varepsilon}_i$  is the random noise embedded in the signal, which is normally distributed with a mean of zero and a variance of  $\tau_s^{-1}$ . The parameter  $\tau_s$  measures the precision of the signal or the quality of original research or "hard" information. The decision of whether to acquire a private information signal is made before an investor engages in trading the risky asset.

#### 3.2.2.2 Social communication and diffusion noise

Following Han and Yang (2013), traders update their beliefs through social communication by exchanging information with other investors within the same community (island). We depart from the standard assumption of homogeneous communication by allowing the precision of social

signals to depend on the group identities of the sender and receiver.

For communication between traders belonging to the same racial group, the social signal received by trader  $i$  from trader  $j$  is given by:

$$\tilde{y}_{i,j} = \tilde{s}_j + \tilde{\eta}_{i,j} = \tilde{v} + \tilde{\varepsilon}_j + \tilde{\eta}_{i,j}, \quad \tilde{\eta}_{i,j} \sim \mathcal{N}(0, \tau_\eta^{-1}), \quad (15)$$

where  $\tilde{\eta}_{i,j}$  represents normally distributed transmission noise, which is normally distributed with the mean of zero and the variance  $\tau_\eta^{-1}$ .  $\tau_\eta$  measures the precision of the social transmission of the signal  $\tilde{s}_j$ . Given independent noise in the original signal  $\tilde{s}_j$  based on (14), the conditional variance of  $\tilde{y}_{i,j}$  based on (15) (conditional on the true pay-off) is  $\tau_s^{-1} + \tau_\eta^{-1}$ , implying a combined precision of  $\tau_y = \frac{\tau_s \times \tau_\eta}{\tau_\eta + \tau_s}$ . This reflects how social signals degrade as they diffuse, even within homogenous groups.

Communication across racial groups is subject to additional frictions, which we model as a degradation in signal fidelity. This assumption is supported by a substantial interdisciplinary literature. Sociological research documents that individuals disproportionately associate with others who share their identity, a phenomenon known as homophily, which leads to racially homogeneous networks that restrict the frequency and reach of intergroup information exchange (McPherson et al., 2001). Even when cross-racial communication occurs, the quality of information is often diminished by transmission bias and trust deficits. Social identity theory suggests that in-group favoritism causes agents to perceive out-group signals as less reliable (Tajfel, Turner, Austin, and Worchel, 2001).

Structural differences in group identity therefore create communication costs that reduce the value of diverse information sets, as formalized by Lazear (1999). Moreover, subtle transmission biases further interfere with communication: Dovidio et al. (2002) show that in interracial interactions, conflicts between verbal cues and nonverbal signals, arising from implicit bias, create misperceptions and distrust that impede information flow.

To quantify these frictions, we introduce a racial integration parameter  $\rho \in (0,1)$ , which governs the effectiveness of cross-group information transmission. A value of  $\rho=1$  represents full integration, where signals across groups are as precise as within-group communication. A value of  $\rho=0$  corresponds to complete segregation, where no usable cross-group signals are transmitted. Under this framework, we explicitly model the total combined noise for a cross-group signal  $\tilde{y}'_{i,j}$  such that its effective precision is linearly scaled by  $\rho$  with  $\text{Precision}(y'_{ij}|\tilde{v}) = \rho\tau_y = \rho \frac{\tau_s \times \tau_\eta}{\tau_\eta + \tau_s}$ .

Formally, we model  $\tilde{y}'_{i,j}$  as:

$$\tilde{y}'_{i,j} = \tilde{s}_j + \tilde{\eta}'_{i,j} = \tilde{v} + \tilde{\varepsilon}_j + \tilde{\eta}'_{i,j}, \quad \tilde{\varepsilon}_j + \tilde{\eta}'_{i,j} \sim \mathcal{N}(0, (\rho\tau_y)^{-1}). \quad (16)$$

Modelling the total combined precision of the cross-group signal ( $\tilde{\varepsilon}_j + \tilde{\eta}'_{i,j}$ ) directly as  $\rho\tau_y$  is essential for maintaining the tractability and readability of the equilibrium results. (16) ensures the  $\rho$  acts as a linear scaling factor for information quality, which significantly simplifies the aggregate logic in the late stage.

Finally, the market is completed by the presence of noise traders who are purely liquidity providers and do not trade based on their information. Therefore, we assume their racial/ethnic identity is irrelevant to their trading behavior. The per capita stock supply from these noise traders,  $\tilde{x}$ , introduce the uncertainty in the market-clearing condition, preventing prices from fully revealing all private information (Grossman and Stiglitz, 1980). Noise trader's supply is normally distributed:

$$\tilde{x} \sim \mathcal{N}(\bar{x}, \tau_x^{-1}). \quad (17)$$

Here,  $\bar{x}$  is the mean of the supply and  $\tau_x^{-1}$  is the variance of the supply.

### 3.2.3 Price-implied signal extraction

In this economy, traders rationally anticipate that the equilibrium price  $\tilde{p}$  aggregates dispersed private and social information and therefore serves as a public signal. In other words, agents can

use the observed price as a sufficient statistic to infer information about the asset's future payoff. Given the assumptions of normally distributed payoffs and Constant Absolute Risk Aversion (CARA) preferences, we conjecture a linear rational expectations equilibrium (REE) price function that is affine in the fundamental asset value  $\tilde{v}$  and the net supply shock from noise traders  $\tilde{x}$ :

$$\tilde{p} = \alpha_0 + \alpha_v \tilde{v} - \alpha_x \tilde{x}. \quad (18)$$

Here,  $\alpha_0$  is the price intercept,  $\alpha_v$  measures the sensitivity of the price to the fundamental payoff, and  $\alpha_x$  captures the price impact of noise trader supply (Hellwig, 1980).

To extract information about future payoffs from the observed price, we rearrange equation (18) to express the fundamental value as a function of price and noise demand:

$$\tilde{v} = \frac{\tilde{p} - \alpha_0 + \alpha_x \tilde{x}}{\alpha_v}. \quad (19)$$

Because the noise demand  $\tilde{x}$  is not observable to traders, their best estimate of it is the average level  $\bar{x}$ . Substituting  $\bar{x}$  from  $\tilde{x}$ 's distribution (17) into equation (19), we can get the price-implied signal ( $\tilde{\xi}$ ):

$$\tilde{\xi} = \frac{\tilde{p} - \alpha_0 + \alpha_x \bar{x}}{\alpha_v} = \tilde{v} - \frac{\alpha_x}{\alpha_v} (\tilde{x} - \bar{x}). \quad (20)$$

Price informativeness is defined as the precision of the price-implied signal,  $\tau_\pi$ , which can be written as based on equation (20):

$$\tau_\pi = \left( \frac{\alpha_v}{\alpha_x} \right)^2 \tau_x. \quad (21)$$

The price informativeness  $\tau_\pi$  quantifies the quality of public signal embedded in the price and serves as a central metric for market efficiency in our model.

### 3.3 Financial Market Equilibrium under Exogenous Private Signal Acquiring Design

In this section, we analyze the rational expectations equilibrium while holding the distribution of private information fixed. Specifically, we assume that the fractions of informed traders in the racial majority and racial minority groups are exogenously given. That is,  $\phi_W$  and  $\phi_M$  are treated as fixed parameters. Our objective is to derive closed-form expressions for the equilibrium price coefficients and to examine how frictions in cross-group communication, governed by the racial integration parameter  $\rho$ , affect aggregate market outcomes, including price informativeness, the cost of equity, liquidity, welfare, and trading volume.

Traders maximize expected utility, as defined by (13), over terminal wealth. Under the assumption of normally distributed payoffs, the optimal demand for the risky asset is linear in the expected excess return and inversely proportional to the conditional variance of the asset payoff. Formally, the optimal demand  $q_i$  for trader  $i$ , with information set  $\mathcal{F}_{i,r}$  (where  $r \in \{W, M\}$  denotes the trader's racial group), is given by:

$$q_i = \frac{E_i[\tilde{v} | \mathcal{F}_{i,r}] - \tilde{p}}{\gamma \cdot \text{Var}_i(\tilde{v} | \mathcal{F}_{i,r})}, \quad (22)$$

The information set  $\mathcal{F}_{i,r}$  includes the trader's private signal (if informed), the set of social signals received from other traders within the same island, and the public signal inferred from the equilibrium price.

Traders update beliefs using Bayesian inference. Because all signals are normally distributed and conditionally independent given  $\tilde{v}$ , posterior precision equals the sum of the precisions of all observed signals, and the posterior mean is a precision-weighted average of these signals (DeGroot, 1970; Hellwig, 1980).<sup>46</sup> As a result, we can have the optimal demand for majority and minority traders by substituting signals depicted by (12) and (15) – (17) into equation (22):

---

<sup>46</sup> Formally, for Bayesian updating with all normally distributed signals, the posterior expectation and variance can be expressed as posterior mean =  $\frac{\tau_{\text{prior}}(\text{Prior mean}) + \sum \tau_{\text{signal } i} \text{signal } i}{\tau_{\text{prior}} + \sum \tau_{\text{signal } i}}$  and posterior variance =  $\frac{1}{\tau_{\text{prior}} + \sum \tau_{\text{signal } i}}$ .

$$\begin{aligned}
q_{W,i} &= d_i(\tilde{p}; \mathcal{F}_{i,W}) = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \tau_\pi \tilde{\xi} + \tau_y \sum_{j=1}^{\phi_W \mu N} \tilde{y}_j + \rho \tau_y \sum_{j=1}^{\phi_M (1-\mu) N} \tilde{y}_j' + \mathbf{1}_{\{i=IN\}} (\tau_s \tilde{s}_i - \tau_y \tilde{y}_j) \right] \\
&\quad - \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \phi_W \mu N \tau_y + \rho \phi_M (1-\mu) N \tau_y + \mathbf{1}_{\{i=IN\}} (\tau_s - \tau_y)], \\
q_{M,i} &= d_i(\tilde{p}; \mathcal{F}_{i,M}) = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \tau_\pi \tilde{\xi} + \rho \tau_y \sum_{j=1}^{\phi_W \mu N} \tilde{y}_j' + \tau_y \sum_{j=1}^{\phi_M (1-\mu) N} \tilde{y}_j + \mathbf{1}_{\{i=IN\}} (\tau_s \tilde{s}_i - \tau_y \tilde{y}_j) \right] \\
&\quad - \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \rho \phi_W \mu N \tau_y + \phi_M (1-\mu) N \tau_y + \mathbf{1}_{\{i=IN\}} (\tau_s - \tau_y)]. \tag{23}
\end{aligned}$$

Here,  $\mathbf{1}_{\{i=IN\}}$  is equal to 1 if trader  $i$  is informed and equal to 0 otherwise.

The equilibrium price  $\tilde{p}$ , is determined by the market-clearing condition, which equates the aggregate demand from rational traders to the random net supply generated by noise traders. Building on Schneider (2009), we adopt the large economy assumption ( $G \rightarrow \infty$ ) in formulating the market-clearing condition. This asymptotic assumption, standard in the literature, ensures that idiosyncratic noise terms associated with individual signals and local transmission errors wash out in the aggregate. The market-clearing condition is expressed as:

$$\lim_{G \rightarrow \infty} \frac{1}{G} \sum_{g=1}^G \left[ \frac{1}{N} \sum_{i \in W} d_i(\tilde{p}; \mathcal{F}_{i,W}) + \frac{1}{N} \sum_{i \in M} d_i(\tilde{p}; \mathcal{F}_{i,M}) \right] = \tilde{x}. \tag{24}$$

By substituting the optimal demands (23) into equation (24) and matching coefficients with the conjectured price function (18), we can solve for the equilibrium price coefficients  $\alpha_0$ ,  $\alpha_v$ , and  $\alpha_x$ .

The next proposition presents this result formally.

**Proposition 1:** *There exists a unique linear rational expectations equilibrium in the financial market, with price function:*

$$\tilde{p} = \alpha_0 + \alpha_v \tilde{v} - \alpha_x \tilde{x},$$

where

$$\alpha_0 = \frac{\tau_v \bar{v} + (\alpha_v/\alpha_x)\tau_x \bar{x}}{\tau_v + \tau_\pi + \tau_{signal}},$$

$$\alpha_v = \frac{\tau_\pi + \tau_{signal}}{\tau_v + \tau_\pi + \tau_{signal}},$$

$$\alpha_x = \frac{(\alpha_v/\alpha_x)\tau_x + \gamma}{\tau_v + \tau_\pi + \tau_{signal}},$$

where

$$\tau_{signal} = \tau_y[\rho N(\mu - \mu^2)(\phi_{MI} + \phi_{MA}) + (1 - \mu)(N - N\mu - 1)\phi_{MI} + \mu(N\mu - 1)\phi_{MA}] + \tau_s[\mu\phi_{MA} + (1 - \mu)\phi_{MI}],$$

$$\frac{\alpha_v}{\alpha_x} = \frac{\tau_{signal}}{\gamma} \text{ and } \tau_\pi = \tau_x \frac{\tau_{signal}^2}{\gamma^2}$$

### 3.3.1 Information structure and heterogeneous diffusion

Market efficiency is commonly measured by price informativeness, which captures how accurately the current market price reflects an asset's future liquidating value. As shown in equation (20), price informativeness  $\tau_\pi$  is determined by the squared signal-to-noise ratio  $(\alpha_v/\alpha_x)^2$ . In equilibrium, this ratio is monotonically increasing in the effective aggregate signal precision, denoted by  $\tau_{signal}$ . This term represents the total quality of information aggregated by the price system from all private and social channels in the economy. Rearranging the expression for  $\tau_{signal}$  from Proposition 1 yields the following decomposition:

$$\begin{aligned} \tau_{signal} = & \underbrace{\tau_y \rho N(\mu - \mu^2)(\phi_W + \phi_M)}_{(1) \text{ Cross-group social signals}} + \underbrace{\tau_y N \mu^2 \phi_W}_{(2) \text{ Majority in-group}} + \underbrace{\tau_y N(1 - \mu)^2 \phi_M}_{(3) \text{ Minority in-group}} \\ & + \underbrace{(\tau_s - \tau_y)(\mu\phi_W + (1 - \mu)\phi_M)}_{(4) \text{ Direct private signals}}. \end{aligned} \quad (25)$$

Equation (25) decomposes the economy's total information endowment into four distinct channels. The first term, cross-group social signals, is central to our analysis. It captures the informational value of signals transmitted between the racial majority and minority investors. The magnitude of this component depends on the mass of cross-group links, proportional to  $\mu(1 - \mu)$ , and is linearly

scaled by the racial integration parameter  $\rho$ , which governs the efficiency of intergroup diffusion.

The second and third terms represent in-group social signals circulating within racially homogeneous clusters of majority and minority investors, respectively. These components operate at the baseline social transmission precision  $\tau_y$  and scale with the square of group size ( $\mu^2$  and  $(1-\mu)^2$ ), reflecting the combinatorial structure of within-group network interactions. The fourth term captures the residual precision of direct private signals. Because social transmission introduces noise ( $\tau_y < \tau_s$ ), the price aggregates the undiluted portion of private information,  $\tau_s - \tau_y$ , held by informed traders prior to diffusion through the network.

The central theoretical prediction of our model concerns the role of racial integration. Inspection of equation (25) reveals that the integration parameter  $\rho$  enters exclusively through the cross-group term. Differentiating  $\tau_{\text{signal}}$  with respect to  $\rho$  yields:

$$\frac{\partial \tau_{\text{signal}}}{\partial \rho} = \tau_y N(\mu - \mu^2)(\phi_W + \phi_M) > 0. \quad (26)$$

Because  $\mu \in (0.5, 1)$ , the term  $\mu - \mu^2$  is strictly positive. Given the non-negativity of the remaining parameters within equation (26), an increase in racial integration unambiguously raises price informativeness. Intuitively, integration improves market efficiency by amplifying the influence of informed signals across racial groups and increasing each group's effective access to information, holding the total number of signals fixed.

The economic intuition underlying this result aligns with the mechanics of social transmission bias described by Hirshleifer (2020). In a segregated economy (low  $\rho$ ), valid signals acquired by minority investors are discounted or lost when transmitted to majority investors, and vice versa, due to communication frictions and trust deficits, as documented by Guiso, Sapienza, and Zingales (2008). This process effectively removes a portion of the economy's information endowment from the aggregate price. As integration increases ( $\rho \rightarrow 1$ ), these "translation costs" fall, allowing the market to aggregate information from diverse social clusters more effectively, a mechanism

formalized by Lazear (1999).

Finally, the marginal benefit of racial integration is proportional to  $N(\mu - \mu^2)$ . This implies that the gains from integration are maximized when the population is most diverse ( $\mu \approx 0.5$ ) and when the social network is dense (large  $N$ ). Consequently, racial integration is particularly valuable for price discovery in modern, diverse economies characterized by high connectivity.

### 3.3.2 Cost of equity and racial integration

The cost of equity reflects the market's assessment of risk and is defined here as the expected excess return required by traders to hold the risky asset. In a noisy rational expectations framework, this risk premium is determined not only by the volatility of fundamental cash flows but also by the residual uncertainty that remains after all available information has been incorporated into prices.

Formally, the ex-ante cost of equity follows from the market-clearing condition and can be written as:

$$E(\tilde{v} - \tilde{p}) = \frac{\gamma \bar{x}}{\tau_v + \tau_\pi + \tau_{\text{signal}}}, \quad (27)$$

Here,  $\tau_{\text{signal}}$  is defined in equation (25), and price informativeness satisfies  $\tau_\pi = \tau_x \tau_{\text{signal}}^2 / \gamma^2$ , which is increasing in  $\tau_{\text{signal}}$ . The denominator,  $\tau_v + \tau_\pi + \tau_{\text{signal}}$ , represents the posterior precision of the asset payoff. It captures the combined informativeness of prior beliefs ( $\tau_v$ ), the price-implied signal ( $\tau_\pi$ ), and aggregated private and social information ( $\tau_{\text{signal}}$ ). For given risk aversion  $\gamma$  and expected noise trader supply  $\bar{x}$ , this posterior precision governs the equilibrium compensation required for bearing risk.

Because the cost of equity is decreasing in total posterior precision, the relationship between racial integration and the cost of equity depends on how the integration parameter  $\rho$  affects the overall information content available to the market. Differentiating the expected excess return (27)

with respect to  $\rho$  yields:

$$\frac{\partial E[\tilde{v} - P]}{\partial \rho} = -\frac{\gamma \bar{x}}{(\tau_{\text{total}})^2} \left( \frac{\partial \tau_{\text{total}}}{\partial \rho} \right) = -\frac{\gamma \bar{x}}{(\tau_{\text{total}})^2} \left( \frac{2\tau_{\text{signal}}}{\gamma^2} + 1 \right) N(\mu - \mu^2)(\phi_W + \phi_M) < 0.$$

Here, all terms except for the negative one on the right-hand side are strictly positive by construction, implying that higher integration reduces the cost of equity.

### 3.3.3 Market liquidity and racial integration

We define market liquidity as the market's ability to absorb aggregate supply shocks with minimal price impact. Following Kyle (1985) and Vives (2008), we measure liquidity by the inverse of the price impact coefficient,  $1/\alpha_x$ . This method captures market depth: a higher value of  $1/\alpha_x$  implies that equilibrium prices are less sensitive to random noise trader demand  $\tilde{x}$ , allowing large trades to be executed with smaller price concessions.

Following Han and Yang (2013) and using the equilibrium conditions in Proposition 1, liquidity can be decomposed into two fundamental components:

$$\frac{1}{\alpha_x} = \frac{\alpha_v}{\alpha_x} + \frac{\tau_v}{(\alpha_v/\alpha_x)\tau_x + \gamma}. \quad (28)$$

The first component  $(\alpha_v/\alpha_x)$  reflects the signal-to-noise ratio of the price and captures the effectiveness of combined signals. As the market becomes more informative with a higher  $\alpha_v/\alpha_x$ , rational traders face lower posterior uncertainty. This gives them the confidence to take larger, more aggressive positions against deviations in the asset price. This aggressive speculation effectively absorbs the noise trader demand, thereby deepening the market and increasing liquidity.

The second term of equation (28) captures the liquidity reduction caused by adverse selection. In a Rational Expectations Equilibrium (REE), uninformed or less-informed traders are wary of trading against those with superior information. As the speculative intensity  $(\alpha_v/\alpha_x)$  rises, the price becomes more efficient, but the relative risk of adverse selection for those relying on priors or

noisier signals can shift. This reflects the standard microstructure intuition that higher information asymmetry (or more aggressive informed trading) can crowd out the liquidity provision from risk-bearing motives. Mathematically, with more informed trading ( $\alpha_v/\alpha_x$  is higher), market makers may reduce their trading intensity, displayed as a lower  $\frac{\tau_v}{(\alpha_v/\alpha_x)\tau_x + \gamma}$ .<sup>47</sup>

The net effect of racial integration ( $\rho$ ) on liquidity depends on the trade-off between these two forces. Integration unambiguously improves the speculative component by raising total information precision ( $\tau_{signal}$ ). However, this increase in speculative aggressiveness simultaneously exerts downward pressure on the adverse selection component. Differentiating equation (28) reveals that the positive speculative effect dominates the negative adverse selection effect if and only if the market is sufficiently efficient. In other words,  $\frac{\partial(1/\alpha_x)}{\partial(\alpha_v/\alpha_x)} > 0$  can be achieved if and only if:

$$\frac{\alpha_v}{\alpha_x} > \sqrt{\frac{\tau_v}{\tau_x}} - \frac{\gamma}{\tau_x}. \quad (29)$$

The intuition here is that only when the price informativeness is high enough ( $> \sqrt{\frac{\tau_v}{\tau_x}} - \frac{\gamma}{\tau_x}$ ), liquidity will increase with higher price informativeness ( $\frac{\partial(1/\alpha_x)}{\partial(\alpha_v/\alpha_x)} > 0$ ). To obtain the relationship

between racial integration and liquidity, we further substitute the results from Proposition 1 into equation (18). We have the claim that  $\frac{\partial(1/\alpha_x)}{\partial\rho} = \frac{\partial(1/\alpha_x)}{\partial(\alpha_v/\alpha_x)} \tau_y N(\mu - \mu^2)(\phi_W + \phi_M) > 0$ , if and only if  $\frac{\alpha_v}{\alpha_x} = \frac{\tau_y[\rho N(\mu - \mu^2)(\phi_{MI} + \phi_{MA}) + (1 - \mu)(N - N\mu - 1)\phi_{MI} + \mu(N\mu - 1)\phi_{MA}] + \tau_s[\mu\phi_{MA} + (1 - \mu)\phi_{MI}]}{\gamma} > \sqrt{\frac{\tau_v}{\tau_x}} - \frac{\gamma}{\tau_x}$ . In other words,

racial integration  $\rho$  enhances liquidity when the price is already informative enough to mitigate the worst effects of adverse selection. Thus, in most realistic settings, especially in developed and

---

<sup>47</sup> This component also increase with  $\tau_v$ , which means when the prior uncertainty is lower ( $\tau_v$  is high) the market maker will be more willing to provide liquidity as the potential risk because now both inventory and trading lose against informed traders reduce.

relatively efficient capital markets, racial integration tends to improve market liquidity.

### 3.3.4 Trading volume, welfare, and racial integration

#### 3.3.4.1 Analytical framework

Having established the effects of racial integration on price efficiency and liquidity, we now examine its implications for trading activity and investor welfare. Together, these metrics provide a comprehensive perspective on how social structure influences both market functioning, as reflected in trading volume, and the distribution of gains among market participants, as captured by welfare.

Trading volume is defined as half the sum of the absolute turnover of all market participants. In our large economy limit, expected trading volume  $V$  is given by the aggregate absolute demand of rational traders plus the absolute demand of noise traders:

$$V \equiv q(\tilde{v}, \tilde{x}) = \frac{1}{2} \left( \lim_{G \rightarrow \infty} \frac{1}{G} \sum_{g=1}^G \left[ \frac{1}{N} \sum_{i=1}^N |d_i(\tilde{p}; \mathcal{F}_{i,r})| \right] + |\tilde{x}| \right). \quad (30)$$

While the closed-form expression for  $V$  is analytically complex and challenging to trace due to the folding of normal distributions, the mechanism driving volume is theoretically clear. Trading volume in REE models is driven by two factors: the magnitude of noise trading shocks and the aggressiveness with which rational traders respond to price displacements.

We measure investor welfare using the ex-ante certainty equivalent (CE) of terminal wealth. For a trader  $I$  with CARA preferences, the CE is determined by the reduction in posterior uncertainty:

$$\begin{aligned}
& \text{Minority Group} \left\{ \begin{aligned} CE_{UN,M} &= \frac{1}{\gamma} \log \left( \sqrt{\frac{Var(\tilde{v} - \tilde{p})}{Var(\tilde{v} - \tilde{p} | \mathcal{F}_{UN,M})}} \right) + \frac{E[\tilde{v} - \tilde{p}]^2}{2\gamma Var(\tilde{v} - \tilde{p})} \\ CE_{IN,M} &= \frac{1}{\gamma} \log \left( \sqrt{\frac{Var(\tilde{v} - \tilde{p})}{Var(\tilde{v} - \tilde{p} | \mathcal{F}_{IN,M})}} \right) + \frac{E[\tilde{v} - \tilde{p}]^2}{2\gamma Var(\tilde{v} - \tilde{p})} \end{aligned} \right. , \\
& \text{Majority Group} \left\{ \begin{aligned} CE_{UN,W} &= \frac{1}{\gamma} \log \left( \sqrt{\frac{Var(\tilde{v} - \tilde{p})}{Var(\tilde{v} - \tilde{p} | \mathcal{F}_{UN,W})}} \right) + \frac{E[\tilde{v} - \tilde{p}]^2}{2\gamma Var(\tilde{v} - \tilde{p})} \\ CE_{IN,W} &= \frac{1}{\gamma} \log \left( \sqrt{\frac{Var(\tilde{v} - \tilde{p})}{Var(\tilde{v} - \tilde{p} | \mathcal{F}_{IN,W})}} \right) + \frac{E[\tilde{v} - \tilde{p}]^2}{2\gamma Var(\tilde{v} - \tilde{p})} \end{aligned} \right. . \quad (31)
\end{aligned}$$

Here  $Var(\tilde{v} - \tilde{p})$  and  $E[\tilde{v} - \tilde{p}]$  are the unconditional prior variance and mean without learning any signals, so the improvement in the welfare (certainty equivalents) comes from the reduction in the posterior variance (the increase in the precision of perceived risk). As price is an observable constant for all traders, the key term,  $Var(\tilde{v} - \tilde{p} | \mathcal{F}_{i,r})$  equals  $Var(\tilde{v} | \mathcal{F}_{IN,W})$ . For the welfare of noise traders, similar to Han and Yang (2013), we follow the literature (e.g., Chowdhry and Nanda 1991; Leland 1992; Easley, O'Hara, and Yang 2016) to use the expected revenue since they seek to meet their liquidity demand at the least opportunity cost:

$$CE_{Nosie} = E[(\tilde{p} - \tilde{v})\tilde{x}] = -E[(\tilde{v} - \tilde{p})\tilde{x}] - \alpha_x Var(\tilde{x}). \quad (32)$$

Here, the welfare of noise traders will increase with the liquidity and decrease with the cost of capital.

The key metric of interest is not only the aggregate welfare, but also the welfare gap ( $DCE_{IN} = CE_{IN,W} - CE_{IN,M}$ ). In a segregated economy, minority traders face a structural information disadvantage due to their smaller group size and limited access to the majority's signal pool. A comparative static of this welfare gap regarding racial integration can be analyzed by substituting corresponding components from equation system (31) into  $CE_{IN,W} - CE_{IN,M}$ , we have:

$$DCE_{IN} = \frac{1}{2\gamma} \log \left( \frac{\tau_v + \tau_\pi + \phi_W \mu N \tau_y + \rho \phi_M (1 - \mu) N \tau_y + \tau_s - \tau_y}{\tau_v + \tau_\pi + \rho \phi_W \mu N \tau_y + \phi_M (1 - \mu) N \tau_y + \tau_s - \tau_y} \right). \quad (33)$$

Under the assumption that  $\phi_W = \phi_M$  as exogenously given, the denominator of expression (33) is always larger than the numerator because  $\phi_W \mu N \tau_y > \phi_M (1 - \mu) N \tau_y$  and  $\rho \in (0,1)$ . And this difference is going to shrink as  $\rho$  approaches 1 and reaches the minimum level of zero if we ideally set  $\rho$  to be one. The result shows that racial integration can shrink the welfare disparities across different racial groups. Similarly, this pattern also applies to the unformed traders.

#### 3.3.4.2 Numerical calibration and results

Unlike the closed-form expressions available for price efficiency, the equilibrium expressions for trading volume and ex ante welfare involve additional analytical complexity. Trading volume depends on a folded normal distribution, while welfare expressions feature logarithmic ratios of posterior variances. To characterize how racial integration affects trading volume and investor welfare, we therefore rely on a numerical calibration using standard market parameters. Importantly, this calibration is not intended to match observed market microstructure. Rather, its purpose is to isolate comparative statics, specifically, how changes in racial integration affect equilibrium outcomes.

To perform the numerical analysis, we adopt a one-year time horizon and select parameter values consistent with the established asset pricing literature. We normalize the unconditional expected asset payoff to 1. Following Han and Yang (2013), we specify the ex-ante payoff precision as  $\tau_v = 25$ , corresponding to an annual return volatility of approximately 20%. The coefficient of absolute risk aversion is set to  $\gamma = 2$ . For the supply side, we normalize the per capita aggregate supply to unity ( $\bar{x} = 1$ ) and set the precision of noise trader demand to  $\tau_x = 10$ . This parameterization implies an annual liquidity supply volatility of roughly 30% of the total outstanding shares. Regarding the information structure, we adhere to the calibration of Gennotte

and Leland (1990) by fixing the rational trader's signal-to-noise ratio at 0.2, which yields a private signal precision of  $\tau_s = 5$ . To capture the noise introduced by social transmission, we assume that communicating a signal through the network reduces its precision by 50% under conditions of perfect integration. This implies a social signal precision of  $\tau_y = 2.5$  (derived from a transmission noise precision of  $\tau_\eta = 5$ ). In the exogenous information setting analyzed in this section, we hold the fraction of informed traders constant across groups at  $\phi_W = \phi_M = 0.73$  following Han and Yang (2013).

Finally, we incorporate demographic constraints to reflect the U.S. context. We set the size of each community network to  $N = 20$  traders and fix the population share of the majority group at  $\mu = 0.63$ , approximating the average proportion of the White population in the United States between 2000 and 2020 as reported by the U.S. Census Bureau. This calibration ensures that our analysis of racial integration operates within a realistically stratified social topology. We present the equilibrium outcomes in Figure 1, which plots key market variables against the racial integration parameter  $\rho$  on the horizontal axis. Panels (a1)–(a4) correspond to the exogenous information environment and isolate the structural effects of integration when information acquisition is held fixed.

[ Insert Figure 3.1 Here]

Panel (a1) shows that the expected trading volume defined by equation (30) is monotonically increasing in racial integration. The underlying mechanism is a reduction in posterior uncertainty: higher integration increases effective aggregate signal precision, leading rational investors to perceive lower conditional risk. This reduction in risk lowers the compensation required for holding inventory, encouraging traders to respond more aggressively to signal realizations for any given price deviation. As a result, the market clears with higher aggregate turnover.

Panel (a2) illustrates that the welfare of noise traders, defined by (32), improves (i.e., becomes less negative) as integration increases. Noise traders, who trade for liquidity reasons independent

of fundamentals, effectively pay a cost equal to their expected losses to informed traders. As racial integration improves price informativeness, it reduces the cost of equity, as established in Section 3.3.2. A lower cost of equity implies a smaller risk premium is required to induce rational agents to absorb liquidity demand. Consequently, integration reduces the “cost of liquidity” faced by uninformed participants.

Panel (a3) shows that the aggregate welfare of rational (uninformed) traders declines with greater integration. This outcome reflects the zero-sum nature of trading gains in a fixed-information environment. Because rational traders earn excess returns by providing liquidity to noise traders, improvements in price informativeness and the associated reduction in risk compress the equilibrium risk premium and thus reduce their expected gains.

Finally, panel (a4) decomposes welfare effects by racial group. Although welfare declines for both Majority and Minority uninformed traders due to the general compression of risk premia, the magnitude of the decline differs across groups. Minority traders, who face a structural information disadvantage in the segregated equilibrium ( $\rho \approx 0$ ), experience a relative improvement as integration increases. The convergence of the two welfare paths in panel (a4) confirms that racial integration narrows the welfare gap. By reducing frictions in cross-group communication, integration diminishes the informational advantage of the majority and yields a more equitable distribution of the remaining market surplus.

### **3.4 Financial Market Equilibrium under Endogenous Information Acquisition**

In the previous section, we analyzed market outcomes while holding the distribution of information fixed. We now relax this assumption to examine the equilibrium production of information. In this more general framework, agents make an ex-ante decision to become informed by weighing the marginal benefit of acquiring a private signal against its cost.

This endogenous setting introduces a critical feedback loop. While social integration improves price efficiency, as shown in Section 3.3, it also alters incentives for acquiring costly private information. By enabling agents to free-ride on social signals and informative prices, greater integration may crowd out private information production. At the same time, integration can lower the cost of acquiring information, particularly for disadvantaged groups. The net effect on market efficiency, therefore, depends on the interaction of these opposing forces.

### 3.4.1 Asymmetric information acquisition cost and acquisition decision

Prior to the trading round, each trader  $i$  belonging to racial group  $r \in \{W, M\}$  decides whether to incur a cost  $c_r$  to observe a private signal  $s_i$ . This decision is made by maximizing ex ante expected utility. The net benefit of becoming informed, denoted by  $B_r(\phi_W, \phi_M; \rho) \equiv CE_{IN,r} - CE_{UN,r}$ , is defined as the difference between the certainty equivalents associated with the informed and uninformed states.

A rational trader acquires information if and only if the expected benefit exceeds the acquisition cost, i.e.,  $B_r \geq c_r$ . In an interior equilibrium, where a fraction  $0 < \phi_r^* < 1$  of traders in group  $r$  becomes informed, the marginal benefit must equal the marginal cost:<sup>48</sup>

$$\mathcal{B}_r(\phi_W^*, \phi_M^*; \rho) = c_r(\rho) \quad \text{for } r \in \{W, M\}. \quad (34)$$

Here a distinct feature of our model is that the cost of information acquisition  $c_r(\rho)$  is not uniform; it depends on the trader's racial identity and the level of social integration. Following the

---

<sup>48</sup> The existence and uniqueness of this equilibrium rely on the monotonic relationship between the fraction of informed traders and the value of private information. As the fraction of informed traders  $\phi_r$  increases, more private signals are aggregated into the price system, raising price informativeness  $\tau_\pi$ . A more informative price effectively "reveals" the private information to the uninformed traders, allowing them to free-ride on the costly acquisition of others. Consequently, as  $\phi_r$  rises, the information advantage of being informed shrinks. The welfare of uninformed traders  $CE_{UN,r}$  rises faster than the welfare of informed traders, causing the net benefit function  $\mathcal{B}_r(\phi_W, \phi_M; \rho)$  to be strictly decreasing in the aggregate amount of information (Grossman and Stiglitz 1980). The equilibrium is achieved at the unique point where this downward-sloping marginal benefit curve intersects the cost of acquisition  $c_r$ . If  $\mathcal{B}_r > c_r$ , more agents will become informed, driving  $\mathcal{B}_r$  down; if  $\mathcal{B}_r < c_r$ , agents will exit the information market, pushing  $\mathcal{B}_r$  back up until the indifference condition (34) is satisfied.

organizational finance literature, we decompose the total cost  $c_r$  into two components: a fixed cost for "hard" information analysis and a variable cost for "soft" information gathering (Stein, 2002; Liberti and Petersen, 2019). Conceptually, the total cost  $c_r$  can be expressed as:

$$c_r(\rho) = c_{\text{hard}} + c_{\text{soft}}(r, \rho),$$

where  $c_{\text{hard}}$  is the hard information cost and  $c_{\text{soft}}(r, \rho)$  is a soft information cost that depends on the racial group and racial integration.  $c_{\text{hard}}$  captures the resources required for proprietary data analysis, financial modeling, and algorithmic processing. We assume this cost is invariant to race or social structure, as "hard" information is by definition quantifiable, impersonal, and transferable without social context (Stein, 2002).

We interpret  $c_{\text{soft}}$  as the marginal cost of acquiring private information through "soft" channels such as conversations with management, industry contacts, or channel checks, all of which require social interaction.<sup>49</sup> Cohen et al. (2008) demonstrate that that social networks facilitate social networks significantly lower the cost of gathering such information, as connected agents are more forthcoming and accessible. However, if the primary sources of upstream information (e.g., corporate boards, insiders) are predominantly composed of the majority group, minority investors face higher search frictions in accessing these networks. We model this friction as a "communication surcharge" that scales with the degree of segregation. The cost for a trader to access these soft channels depends on the probability of needing to cross racial lines and the difficulty of that interaction:

$$\begin{cases} c_W(\rho) = c_{\text{hard}} + c_o [1 + (1 - \mu)(1 - \rho)] \\ c_M(\rho) = c_{\text{hard}} + c_o [1 + (\mu)(1 - \rho)] \end{cases} \quad (35)$$

Here,  $c_o$  is the baseline cost of in-group communication. The term  $(1 - \rho)$  represents the segregation friction, which will increase with a lower level of racial integration  $\rho$  and thus leads to

---

<sup>49</sup> The communication-based activities are needed in better understanding the future pay-off of assets, especially for acquiring soft, non-codified, or tacit knowledge.

an increase in total cost regardless of racial group. This formulation is motivated by Lazear (1999), who posits that cross-cultural interaction imposes a translation cost. When  $\rho < 1$ , the cost increases proportionally to the likelihood of interacting with the out-group. Since  $\mu > 0.5$  (majority is larger), the minority group faces a higher probability ( $\mu$ ) of needing to bridge this gap to access the aggregate information pool compared to the majority ( $1 - \mu$ ). Consequently, Minority investors face structurally higher information acquisition costs in segregated markets ( $c_M > c_W$ ). It is worth noting that the current linear specification for the communication surcharge, which scales with the probability of interacting with the out-group and  $(1 - \rho)$ , implicitly assumes that information is randomly distributed across racial groups. In practice, one might argue that the majority group holds an information set of greater quantity and quality, given its dominant presence in upstream information channels. If this were the case, the cost gap between majority and minority investors would only widen, implying that the results derived from our linear specification represent a lower bound on the impact of race-dependent acquisition costs on price efficiency. Any amplification of this gap would therefore strengthen our findings.

Consistent with survey evidence that interpersonal communication is a primary, low-cost driver of investor decisions (Shiller and Pound, 1989), we assume that soft information acquisition is generally less costly than hard information acquisition. Accordingly, we set the baseline soft information cost at one-tenth of the fixed analysis cost:  $c_o = \frac{1}{10} c_{hard}$ .

Finally, the equilibrium fractions of informed traders ( $\phi_W^*$ ,  $\phi_M^*$ ) are determined by the system of two equations derived from the indifference conditions by substituting the cost expression (35) into (34) and rearranging:

$$\left\{ \begin{array}{l} \frac{\tau_v + \tau_\pi + \phi_W^* \mu N \tau_y + \rho \phi_M^* (1 - \mu) N \tau_y + \tau_s - \tau_y}{\tau_v + \tau_\pi + \phi_W^* \mu N \tau_y + \rho \phi_M^* (1 - \mu) N \tau_y} = e^{2\gamma c_{hard} + 2\gamma c_o + 2\gamma c_o (1 - \mu)(1 - \rho)} \\ \frac{\tau_v + \tau_\pi + \rho \phi_W^* \mu N \tau_y + \phi_M^* (1 - \mu) N \tau_y + \tau_s - \tau_y}{\tau_v + \tau_\pi + \rho \phi_W^* \mu N \tau_y + \phi_M^* (1 - \mu) N \tau_y} = e^{2\gamma c_{hard} + 2\gamma c_o + 2\gamma c_o \mu (1 - \rho)} \\ \tau_\pi = \tau_x \frac{\tau_{signal}(\phi_W^*, \phi_M^*)^2}{\gamma^2}. \end{array} \right. \quad (36)$$

The equilibrium fractions of informed traders ( $\phi_W^*$  and  $\phi_M^*$ ) are determined by the system of indifference conditions where the marginal benefit of information equals the marginal cost for each group. After substituting the results from Proposition 1 into equation system (36), we can theoretically solve the equilibrium participation rates. However, given the complexity of the final expressions, involving highly non-linear (quadratic) interactions between aggregate signal precision, price coefficients, and logarithmic welfare terms, analytical discussion is almost impossible. Therefore, we rely on numerical simulation to show racial integration's impact on price informativeness, cost of equity, liquidity, welfare, and trading volumes.

We employ the same calibration parameters established in Section 3.4 to ensure consistency (e.g.,  $\tau_v = 25, \gamma = 2, \tau_x = 10$ ). Additionally, we set the fixed component of the information acquisition cost at  $c_{hard} = 0.01$ , which leaves the viable (soft) component of the cost to  $c_o = 0.001$ . We solve the fixed-point problem numerically for each level of integration  $\rho$ .

### 3.4.2 Price informativeness, cost of equity, and liquidity

Panel (b1) of Figure 3.2 shows that stock price informativeness is strictly increasing in racial integration. This result is nontrivial in an endogenous information environment, where two opposing forces are at play. On the one hand, greater integration improves the quality of social signals and overall price efficiency, which typically weakens incentives for private information acquisition through free-riding (Grossman and Stiglitz, 1980). On the other hand, integration reduces the cost of information acquisition for minority investors, as formalized in Section 3.4.2.

Our numerical results demonstrate that this cost-reduction effect dominates the free-riding channel. By lowering the frictional “communication surcharge” faced by minority traders, integration induces sufficient new information production to more than offset the disincentives created by more informative prices. As a result, the aggregate informational content embedded in prices increases.

[ Insert Figure 3.2 Here]

Consistent with the improvement in price efficiency, panel (b2) of Figure 3.2 shows that the cost of equity declines as integration increases. As the market aggregates more private signals (driven by lower acquisition costs for Minorities), the posterior uncertainty faced by the average trader diminishes. This reduction in estimation risk lowers the ex-ante risk premium required to clear the market (Easley and O'Hara, 2004). Thus, racial integration not only makes prices more accurate but also reduces the cost of capital for firms.

panel (b3) of Figure 3.2 demonstrates that market liquidity ( $1/\alpha_x$ ) improves with racial integration. As we discussed in Section 3.3.3, liquidity is driven by the aggressiveness of speculative trading, which increases with signal precision. In the endogenous setting, the net increase in total information production empowers rational traders to absorb noise shocks more effectively. Similarly to Section 3.3, a higher racial integration can improve liquidity if

and only if  $\frac{\alpha_v}{\alpha_x} = \frac{\tau_y[\rho N(\mu - \mu^2)(\phi_M^* + \phi_W^*) + (1 - \mu)(N - N\mu - 1)\phi_M^* + \mu(N\mu - 1)\phi_W^*] + \tau_s[\mu\phi_W^* + (1 - \mu)\phi_M^*]}{\gamma} > \sqrt{\frac{\tau_v}{\tau_x}} - \frac{\gamma}{\tau_x}$ . The

simulation confirms that the market remains in the "efficient regime" where the positive impact of reduced uncertainty on speculative depth dominates any adverse selection concerns (Han and Yang, 2013).

### 3.4.3 Trading volume and welfare analysis

We now examine the implications of endogenous information acquisition for trading volume and investor welfare. Unlike the exogenous setting, in which the supply of information is fixed, aggregate information quality here responds endogenously to changes in incentives. This feedback mechanism generates distinct equilibrium outcomes, particularly with respect to trading activity and the distribution of welfare across investor types.

#### 4.3.1 Trading Volume

Panel (b1) of Figure 1 shows that expected trading volume decreases with racial integration in the endogenous information setting. This finding stands in sharp contrast to the exogenous case (Section 3.3.4), where integration led to higher trading volume. The reversal is driven by a crowding-out effect: as integration improves price informativeness, incentives to acquire private signals decline for majority-group investors due to increased opportunities for free-riding. As a result, the proportion of informed majority traders falls.

Although minority participation rises in response to lower information-acquisition costs, the aggregate effect of free-riding dominates, leading to a reduction in overall speculative intensity relative to the exogenous benchmark. Consequently, while equilibrium prices become more informative, the volume of trade required to sustain those prices declines. A larger fraction of market participants remains uninformed and trades less aggressively, reflecting a substitution away from private information production toward reliance on public price signals.

#### 4.3.2 Welfare

The welfare consequences of racial integration are both substantial and highly asymmetric across demographic groups. Panel (b2) of Figure 3.1 shows that the welfare of noise traders improves (i.e., becomes less negative) as integration increases. This result is robust to endogenous information

acquisition. Because integration lowers the cost of equity and improves liquidity (Section 3.4.2), noise traders face a smaller implicit premium when satisfying liquidity demands. In effect, the “cost of liquidity” declines even as the market adjusts its information-production margin.

For rational but uninformed investors, the welfare effects diverge markedly from those in the exogenous setting. Panel (b3) of Figure 3.1 reveals that minority investors experience a substantial welfare gain. Lower information-acquisition costs ( $c_M$ ) allow minority traders to obtain private signals more cheaply, directly improving their Sharpe ratios. In addition, greater integration enables them to exploit the larger majority signal pool more effectively, further enhancing expected returns.

In contrast, the majority investors may experience stagnation or even a decline in welfare surplus. As integration erodes segmentation barriers, the structural advantage enjoyed by the dominant group, exclusive access to the largest and most informative signal pool, diminishes. Increased entry by minority traders, combined with a general reduction in risk premia (which lowers the returns to liquidity provision), compresses the excess returns previously captured by majority participants.

Crucially, panel (b4) of Figure 3.1 shows that, for a given increase in racial integration, the welfare gap closes more rapidly in the endogenous setting than in the exogenous benchmark. When minority traders can respond to declining information-acquisition costs by choosing to become informed, the redistributive effects of integration are amplified. By lowering entry barriers to participation in information-based trading, racial integration transforms a structural disadvantage into an opportunity for market participation and wealth accumulation.

### 3.5 Conclusion

This study integrates the sociology of networks with the microstructure of financial markets to examine a fundamental yet underexplored question: how does racial integration affect asset pricing efficiency and investor welfare? By embedding a racially stratified “islands” topology within a noisy rational expectations equilibrium framework, we demonstrate that social structure is not merely a backdrop for economic exchange but a determinant of market quality. We model racial segregation as a persistent information friction, a “trust tax”, that degrades the fidelity of signals transmitted across demographic boundaries.

Our analysis yields three primary insights into the mechanics of market efficiency. First, we show that racial integration functions as a technology shock that enhances stock price informativeness. In a segregated economy, valuable private information acquired by minority investors remains sequestered within local clusters due to homophily and communication costs (Lazear, 1999; McPherson et al., 2001). Integration lowers these barriers, effectively unlocking a latent pool of diverse signals. This aggregation effect reduces posterior uncertainty for all traders, leading to prices that more closely track fundamental values. As a result, we find that the cost of equity is strictly decreasing in integration, since reductions in estimation risk lower the ex-ante risk premium required to clear the market (Easley and O’Hara, 2004).

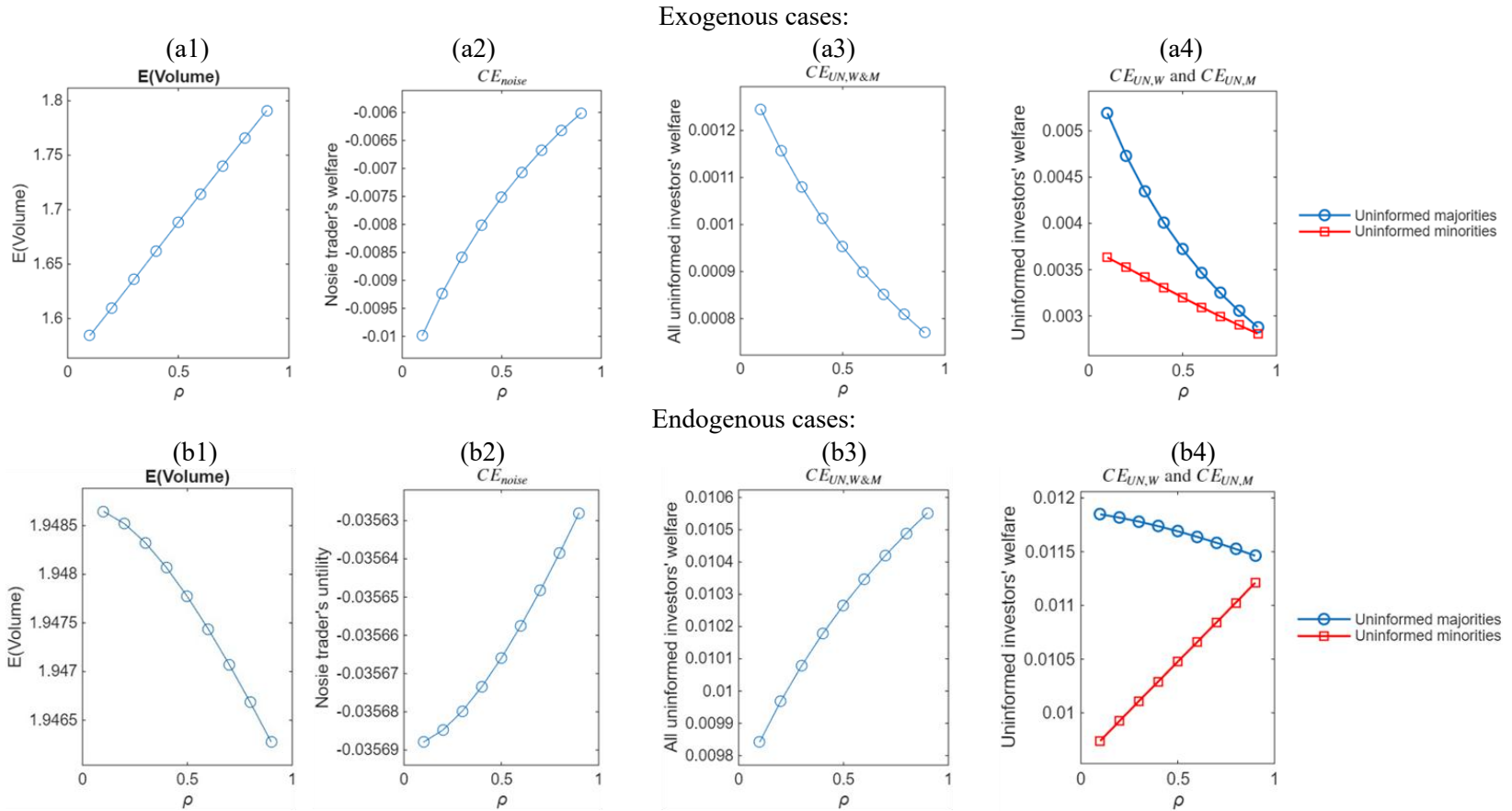
Second, we uncover a critical feedback loop between social structure and incentives for information production. When information acquisition is endogenous, integration plays a dual role: it increases incentives to free-ride on more informative prices, while simultaneously lowering the cost of acquiring “soft” information for minority investors. We show that this cost-reduction effect dominates the free-riding channel. By democratizing access to the aggregate information pool, integration stimulates higher participation rates among minority traders, thereby deepening market liquidity and mitigating the adverse selection typically associated with asymmetric information

(Han and Yang, 2013).

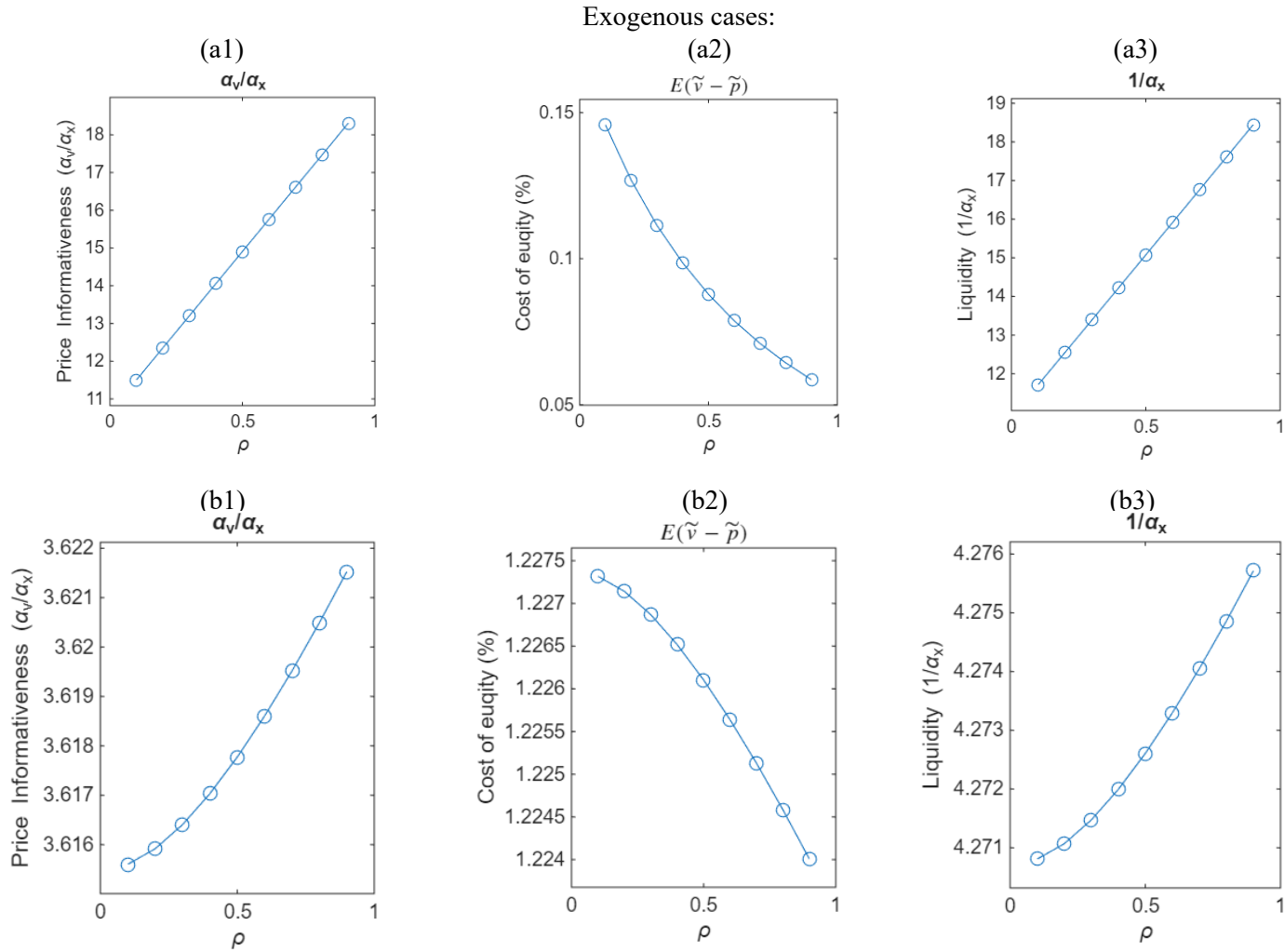
Third, our results highlight the redistributive implications of financial market integration. We document a structural welfare gap in segregated markets, in which minority investors experience systematic information disadvantage relative to the majority. Racial integration substantially narrows this gap. By eroding informational privilege among majority investors and reducing acquisition costs for minority participants, integration promotes a more equitable distribution of market surplus. These findings suggest that policies fostering social and racial integration may generate returns beyond social cohesion, serving as an indirect mechanism for reducing wealth inequality through financial markets.

In sum, our findings advance the “Social Finance” paradigm (Hirshleifer, 2020) by formalizing the link between demographic structure and financial equilibrium. We establish that racial segregation imposes a deadweight loss on the economy, manifesting in lower liquidity, higher costs of capital, and greater inequality. Conversely, our model predicts that integrated economies with strong cross-group connectivity exhibit superior price discovery and deeper liquidity, providing a theoretical foundation for the economic value of diversity within financial ecosystems.

**Figure 3.1: Volume, Welfare Effects of Racial Integration**



**Figure 3.2: Racial Integration and Price Efficiency, Liquidity, and Cost of Equity**



## Appendix to Chapter 3

### Appendix 3A.1: Derivation for Proposition 1

Recall that:

$$q_i = \frac{E_i[\tilde{v} | \mathcal{F}_{i,g}] - \tilde{p}}{\gamma \cdot \text{Var}_i(\tilde{v} | \mathcal{F}_{i,g})}, \quad (3A.1)$$

where  $q_i$  is the optimal demand, and  $\mathcal{F}_{i,g}$  is the information set of trader  $i$ , which depends on the trader's decision on being informative or not, and which group  $g$  they belong to. As all signals' distributions are normal, based on Bayesian updating, the posterior expectation and variance can

be expressed as posterior mean =  $\frac{\tau_{prior} \text{Prior-mean} + \sum \tau_{signal\ i} \text{signal\ } i}{\tau_{prior} + \sum \tau_{signal\ i}}$  and posterior variance =  $\frac{1}{\tau_{prior} + \sum \tau_{signal\ i}}$ . Therefore, based on equation (3A.1), after receiving signals,  $q_i$  should be:

$$q_i = \frac{\frac{\tau_{prior}(\text{Prior mean}) + \sum \tau_{signal\ i} \text{signal\ } i}{\tau_{prior} + \sum \tau_{signal\ i}} - \tilde{p}}{\gamma \cdot \frac{1}{\tau_{prior} + \sum \tau_{signal\ i}}}, \quad (3A.2)$$

Multiply the dominator and numerator of (3A.2) both by  $\tau_{prior} + \sum \tau_{signal\ i}$  and substitute  $\tau_{prior} = \tau_v$ , we have:

$$q_i = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \sum \tau_{signal\ i} \text{signal\ } i \right] - \frac{\tilde{p}}{\gamma} \left[ \tau_v + \sum \tau_{signal\ i} \right]. \quad (3A.3)$$

We can have the optimal demand for the majority and minority traders by substituting corresponding components from the main text into the equation (3A.3) :

$$q_{MA,i} = d_i(\tilde{p}; \mathcal{F}_{i,MA}) = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \tau_\pi \tilde{\xi} + \tau_y \sum_{j=1}^{\phi_{MA}\mu N} \tilde{y}_j + \rho \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} \tilde{y}_j' + \mathbf{1}_{\{i=IN\}} (\tau_s \tilde{s}_i - \tau_y \tilde{y}_j) \right] - \frac{\tilde{p}}{\gamma} \left[ \tau_v + \tau_\pi + \phi_{MA}\mu N \tau_y + \rho \phi_{MI}(1-\mu)N \tau_y + \mathbf{1}_{\{i=IN\}} (\tau_s - \tau_y) \right],$$

$$q_{MI,i} = d_i(\tilde{p}; \mathcal{F}_{i,MI}) = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \tau_\pi \tilde{\xi} + \rho \tau_y \sum_{j=1}^{\phi_{MA}\mu N} \tilde{y}'_j + \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} \tilde{y}_j + \mathbf{1}_{\{i=IN\}} (\tau_s \tilde{s}_i - \tau_y \tilde{y}_j) \right] - \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \rho \phi_{MA} \mu N \tau_y + \phi_{MI} (1 - \mu) N \tau_y + \mathbf{1}_{\{i=IN\}} (\tau_s - \tau_y)]. \quad (3A.4)$$

Here,  $\mathbf{1}_{\{i=IN\}}$  is equal to 1 if trader  $i$  is informed and equal to 0 otherwise, and the cross-group signal and within-group signal are defined as follows:

$$\tilde{y}'_j = \tilde{s}_j + \tilde{\eta}'_{lj}, \quad \tilde{\eta}'_{lj} \sim \mathcal{N}(0, 1/\rho\tau_\eta)$$

At the equilibrium price  $\tilde{p}$ , per-capita demand equals per-capita supply, meets the following equation:

$$\frac{1}{G} \sum_{g=1}^G \left[ \frac{1}{N} \sum_{i \in MA} d_i(\tilde{p}; \mathcal{F}_{i,MA}) + \frac{1}{N} \sum_{i \in MI} d_i(\tilde{p}; \mathcal{F}_{i,MI}) \right] = \tilde{x},$$

which is the same as

$$\frac{1}{G} \sum_{g=1}^G \left[ \frac{N\mu}{N} \bar{d}_i(\tilde{p}; \mathcal{F}_{i,MA}) + \frac{N(1-\mu)}{N} \bar{d}_i(\tilde{p}; \mathcal{F}_{i,MI}) \right] = \tilde{x}, \quad (3A.5)$$

We can have the following expression by substituting (3A.4) into (3A.5):

$$\begin{aligned} \tilde{x} = & \frac{1}{G} \sum_{g=1}^G \frac{1}{\gamma} [\tau_v \bar{v} + \tau_\pi \tilde{\xi} + \rho N(\mu - \mu^2)(\phi_{MI} + \phi_{MA})\tilde{y}'_j + (1 - \mu)(N - N\mu - 1)\phi_{MI}\tilde{y}_j \\ & + \mu(N\mu - 1)\phi_{MA}\tilde{y}_j + \tau_s[\mu\phi_{MA} + (1 - \mu)\phi_{MI}]\tilde{s}_i] \\ & + \frac{1}{G} \sum_{g=1}^G \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \tau_y[\rho N(\mu - \mu^2)(\phi_{MI} + \phi_{MA}) \\ & + (1 - \mu)(N - N\mu - 1)\phi_{MI} + \mu(N\mu - 1)\phi_{MA}] \\ & + \tau_s[\mu\phi_{MA} + (1 - \mu)\phi_{MI}]]. \end{aligned}$$

Substituting the signal definitions for  $\tilde{y}'_j$ ,  $\tilde{y}_j$ , and  $\tilde{s}_i$ , we can have:

$$\begin{aligned}
\tilde{x} &= \frac{1}{G} \sum_{g=1}^G \frac{1}{\gamma} [\tau_v \bar{v} + \tau_\pi \tilde{\xi} + \tau_{signal} \tilde{v}] + \frac{1}{G} \sum_{g=1}^G \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \tau_{signal}] + \\
&\frac{1}{G} \sum_{g=1}^G \frac{1}{\gamma} [\rho N(\mu - \mu^2)(\phi_{MI} + \phi_{MA})\tilde{\eta}_{ij}' + (1 - \mu)(N - N\mu - 1)\phi_{MI}\tilde{\eta}_{ij} + \mu(N\mu - 1)\phi_{MA}\tilde{\eta}_{ij} + \tau_{signal} \tilde{\varepsilon}_i], \\
\tau_{signal} &= \tau_y [\rho N(\mu - \mu^2)(\phi_{MI} + \phi_{MA}) + (1 - \mu)(N - N\mu - 1)\phi_{MI} + \mu(N\mu - 1)\phi_{MA}] + \\
&\tau_s [\mu\phi_{MA} + (1 - \mu)\phi_{MI}] + \tau_s [\mu\phi_{MA} + (1 - \mu)\phi_{MI}]. \tag{3A.6}
\end{aligned}$$

Building on Schneider (2009), we adopt the large economy assumption ( $G \rightarrow \infty$ ). By the law of large numbers, the last term becomes zero, as the means of all these noise terms are zero. So, we have:

$$\tilde{x} = \frac{1}{G} \sum_{g=1}^G \frac{1}{\gamma} [\tau_v \bar{v} + \tau_\pi \tilde{\xi} + \tau_{signal} \tilde{v}] + \frac{1}{G} \sum_{g=1}^G \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \tau_{signal}], \tag{3A.7}$$

Where  $\tau_{signal}$  is defined in equation (25). Next, we substitute  $\tilde{\xi} = \frac{\tilde{p} - \alpha_0 + \alpha_x \bar{x}}{\alpha_v}$  into and reorganize the formula to match  $\tilde{p} = \alpha_0 + \alpha_v \tilde{v} - \alpha_x \tilde{x}$ :

$$\begin{aligned}
\alpha_0 &= \frac{\tau_v \bar{v} + (\alpha_v/\alpha_x)\tau_x \bar{x}}{\tau_v + \tau_\pi + \tau_{signal}}, \\
\alpha_v &= \frac{\tau_\pi + \tau_{signal}}{\tau_v + \tau_\pi + \tau_{signal}}, \\
\alpha_x &= \frac{(\alpha_v/\alpha_x)\tau_x + \gamma}{\tau_v + \tau_\pi + \tau_{signal}},
\end{aligned}$$

where

$$\begin{aligned}
\tau_{signal} &= \tau_y [\rho N(\mu - \mu^2)(\phi_{MI} + \phi_{MA}) + (1 - \mu)(N - N\mu - 1)\phi_{MI} + \mu(N\mu - 1)\phi_{MA}] \\
&+ \tau_s [\mu\phi_{MA} + (1 - \mu)\phi_{MI}] \\
\frac{\alpha_v}{\alpha_x} &= \frac{\tau_{signal}}{\gamma} \text{ and } \tau_\pi = \tau_x \frac{\tau_{signal}^2}{\gamma^2}
\end{aligned}$$

## Appendix 3A.2: Derivation for trading volumes

(1) For the majority group, regarding the informed majority (IN, W):

$$q_{W,i} = d_i(\tilde{p}; \mathcal{F}_{i,W}) = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \tau_\pi \tilde{\xi} + \tau_y \sum_{j=1}^{\phi_{MA}\mu N} \tilde{y}_j + \rho \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} \tilde{y}_j' + (\tau_s \tilde{s}_i - \tau_y \tilde{y}_j) \right] - \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \phi_{MA}\mu N \tau_y + \rho \phi_{MI}(1-\mu)N \tau_y + (\tau_s - \tau_y)]. \quad (3A.8)$$

After substituting equations (4) to (6) into the demand function (3A.8) and reorganizing the expression, we can obtain:

$$d_i(\tilde{p}; \mathcal{F}_{i,W}^{IN}) = \frac{1}{\gamma} [h_{0,W}^{IN} + h_{v,W}^{IN} \tilde{v} + h_{x,W}^{IN} \tilde{x} + \tau_y \sum_{j=1}^{\phi_{MA}\mu N} (\tilde{\varepsilon}_i + \tilde{\eta}_{lj})] + \frac{1}{\gamma} \left[ \rho \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} (\tilde{\varepsilon}_i + \tilde{\eta}_{lj}') + \tau_s \tilde{\varepsilon}_i - \tau_y (\tilde{\varepsilon}_i + \tilde{\eta}_{lj}') \right], \quad (3A.9)$$

where

$$h_{0,W}^{IN} = \tau_v \bar{v} + \tau_\pi \frac{\alpha_x}{\alpha_v} \bar{x} - \alpha_o B_W^{IN},$$

$$h_{v,W}^{IN} = (1 - \alpha_v) B_W^{IN} - \alpha_v \tau_v,$$

$$h_{x,W}^{IN} = -\frac{\alpha_x}{\alpha_v} \tau_\pi + \alpha_x B_W^{IN},$$

$$B_W^{IN} = \tau_v + \tau_\pi + \phi_{MA}\mu N \tau_y + \rho \phi_{MI}(1-\mu)N \tau_y + (\tau_s - \tau_y).$$

Thus, we can see, the  $d_i(\tilde{p}; \mathcal{F}_{i,W}^{IN})$  follows a normal distribution given  $(\tilde{v}, \tilde{x})$ , and the conditional mean and standard deviation can be expressed as:

$$\mu_W^{IN}(\tilde{v}, \tilde{x}) = E[d_i(\tilde{p}; \mathcal{F}_{i,W}^{IN}) | \tilde{v}, \tilde{x}] = \frac{1}{\gamma} (h_{0,W}^{IN} + h_{v,W}^{IN} \tilde{v} + h_{x,W}^{IN} \tilde{x}), \quad (3A.10)$$

$$\sigma_W^{IN} = \sqrt{\text{Var}[d_i(\tilde{p}; \mathcal{F}_{i,W}^{IN}) | \tilde{v}, \tilde{x}]} = \frac{1}{\gamma} \sqrt{\phi_{MA}\mu N \tau_y + \rho \phi_{MI}(1-\mu)N \tau_y + (\tau_s - \tau_y)}, \quad (3A.11)$$

Conditional on  $(\tilde{v}, \tilde{x})$ , the expectation of the absolute value of trading volume (3A.9),  $|d_i(\tilde{p}; \mathcal{F}_{i,W}^{IN})|$ , follows a folded normal distribution, and its conditional mean is given by:

$$E[|d_i(\tilde{p}; \mathcal{F}_{i,W}^{IN})| | \tilde{v}, \tilde{x}] = \sigma_W^{IN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_W^{IN}(\tilde{v}, \tilde{x})]^2}{2\sigma_W^{IN}}} + \mu_W^{IN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_W^{IN}(\tilde{v}, \tilde{x})}{\sigma_W^{IN}}\right) \right], \quad (3A.12)$$

where  $\Phi(\star)$  is the cumulative distribution function of the standard normal distribution, and  $\mu_W^{IN}$ ,  $\sigma_W^{IN}$  are defined by equations (3A.10) and (3A.11).

**(2) For the majority group, regarding the uninformed majority (UN, W):**

$$\begin{aligned} d_i(\tilde{p}; \mathcal{F}_{i,W}^{UN}) &= \frac{1}{\gamma} [h_{0,W}^{UN} + h_{v,W}^{UN} \tilde{v} + h_{x,W}^{UN} \tilde{x} + \tau_y \sum_{j=1}^{\phi_{MA}\mu N} (\tilde{\varepsilon}_i + \tilde{\eta}_{ij})] \\ &\quad + \frac{1}{\gamma} [\rho \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} (\tilde{\varepsilon}_i + \tilde{\eta}_{ij}')], \end{aligned} \quad (3A.13)$$

where

$$h_{0,W}^{UN} = \tau_v \bar{v} + \tau_\pi \frac{\alpha_x}{\alpha_v} \bar{x} - \alpha_o B_W^{UN},$$

$$h_{v,W}^{UN} = (1 - \alpha_v) B_W^{UN} - \alpha_v \tau_v,$$

$$h_{x,W}^{UN} = -\frac{\alpha_x}{\alpha_v} \tau_\pi + \alpha_x B_W^{UN},$$

$$B_W^{UN} = \tau_v + \tau_\pi + \phi_{MA}\mu N \tau_y + \rho \phi_{MI}(1 - \mu) N \tau_y.$$

Similarity,

$$\mu_W^{UN}(\tilde{v}, \tilde{x}) = E[d_i(\tilde{p}; \mathcal{F}_{i,W}^{UN}) | \tilde{v}, \tilde{x}] = \frac{1}{\gamma} (h_{0,W}^{UN} + h_{v,W}^{UN} \tilde{v} + h_{x,W}^{UN} \tilde{x}), \quad (3A.14)$$

$$\sigma_W^{UN} = \sqrt{\text{Var}[d_i(\tilde{p}; \mathcal{F}_{i,W}^{UN}) | \tilde{v}, \tilde{x}]} = \frac{1}{\gamma} \sqrt{\phi_{MA}\mu N \tau_y + \rho \phi_{MI}(1 - \mu) N \tau_y}, \quad (3A.15)$$

Eventually, conditional on  $(\tilde{v}, \tilde{x})$ , the expectation of the absolute value of uninformed majority investor's trading volume (3A.13),  $|d_i(\tilde{p}; \mathcal{F}_{i,W}^{UN})|$ , follows a folded normal distribution, and its conditional mean is given by:

$$E[|d_i(\tilde{p}; \mathcal{F}_{i,W}^{UN})| | \tilde{v}, \tilde{x}] = \sigma_W^{UN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_W^{UN}(\tilde{v}, \tilde{x})]^2}{2\sigma_W^{UN}}} + \mu_W^{UN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_W^{UN}(\tilde{v}, \tilde{x})}{\sigma_W^{UN}}\right) \right], \quad (3A.16)$$

where  $\Phi(\star)$  is the cumulative distribution function of the standard normal distribution, and  $\mu_W^{UN}$ ,  $\sigma_W^{UN}$  are defined by equations (3A.14) and (3A.15).

**(3) For the minority group, regarding the informed minority (IN, M):**

$$q_{M,i} = d_i(\tilde{p}; \mathcal{F}_{i,M}) = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \tau_\pi \tilde{\xi} + \rho \tau_y \sum_{j=1}^{\phi_{MA}\mu N} \tilde{y}_j + \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} \tilde{y}_j' + (\tau_s \tilde{s}_i - \tau_y \tilde{y}_j) \right] - \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \rho \phi_{MA}\mu N \tau_y + \phi_{MI}(1-\mu)N \tau_y + (\tau_s - \tau_y)]. \quad (3A.17)$$

After substituting equations (4) to (6) into the demand function (3A.17) and reorganizing the expression, we can obtain:

$$d_i(\tilde{p}; \mathcal{F}_{i,M}^{IN}) = \frac{1}{\gamma} \left[ h_{0,M}^{IN} + h_{v,M}^{IN} \tilde{v} + h_{x,M}^{IN} \tilde{x} + \rho \tau_y \sum_{j=1}^{\phi_{MA}\mu N} (\tilde{\varepsilon}_i + \tilde{\eta}_{ij}') \right] + \frac{1}{\gamma} \left[ \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} (\tilde{\varepsilon}_i + \tilde{\eta}_{ij}') + \tau_s \tilde{\varepsilon}_i - \tau_y (\tilde{\varepsilon}_i + \tilde{\eta}_{ij}') \right], \quad (3A.18)$$

where

$$h_{0,M}^{IN} = \tau_v \bar{v} + \tau_\pi \frac{\alpha_x}{\alpha_v} \bar{x} - \alpha_o B_M^{IN},$$

$$h_{v,M}^{IN} = (1 - \alpha_v) B_M^{IN} - \alpha_v \tau_v,$$

$$h_{x,M}^{IN} = -\frac{\alpha_x}{\alpha_v} \tau_\pi + \alpha_x B_M^{IN},$$

$$B_M^{IN} = \tau_v + \tau_\pi + \rho \phi_{MA}\mu N \tau_y + \phi_{MI}(1-\mu)N \tau_y + (\tau_s - \tau_y).$$

So, we can eventually obtain,

$$\mu_M^{IN}(\tilde{v}, \tilde{x}) = E[d_i(\tilde{p}; \mathcal{F}_{i,M}^{IN}) | \tilde{v}, \tilde{x}] = \frac{1}{\gamma} (h_{0,M}^{IN} + h_{v,M}^{IN} \tilde{v} + h_{x,M}^{IN} \tilde{x}), \quad (3A.19)$$

$$\sigma_M^{IN} = \sqrt{\text{Var}[d_i(\tilde{p}; \mathcal{F}_{i,M}^{IN}) | \tilde{v}, \tilde{x}]} = \frac{1}{\gamma} \sqrt{\rho \phi_{MA}\mu N \tau_y + \phi_{MI}(1-\mu)N \tau_y + (\tau_s - \tau_y)}, \quad (3A.20)$$

Eventually, conditional on  $(\tilde{v}, \tilde{x})$ , the expectation of the absolute value of the informed minority investor's trading volume (3A.18),  $|d_i(\tilde{p}; \mathcal{F}_{i,M}^{IN})|$ , follows a folded normal distribution, and its conditional mean is given by:

$$E[|d_i(\tilde{p}; \mathcal{F}_{i,M}^{IN})| | \tilde{v}, \tilde{x}] = \sigma_M^{IN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_M^{IN}(\tilde{v}, \tilde{x})]^2}{2\sigma_M^{IN}}} + \mu_M^{IN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_M^{IN}(\tilde{v}, \tilde{x})}{\sigma_M^{IN}}\right) \right]. \quad (3A.21)$$

where  $\Phi(\star)$  is the cumulative distribution function of the standard normal distribution, and  $\mu_M^{IN}$ ,  $\sigma_M^{IN}$  are defined by equations (3A.19) and (3A.20).

**(4) For the minority group, regarding the uninformed minority (UN, M):**

$$q_{M,i} = d_i(\tilde{p}; \mathcal{F}_{i,M}) = \frac{1}{\gamma} \left[ \tau_v \bar{v} + \tau_\pi \tilde{\xi} + \rho \tau_y \sum_{j=1}^{\phi_{MA}\mu N} \tilde{y}_j + \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} \tilde{y}_j' \right] - \frac{\tilde{p}}{\gamma} [\tau_v + \tau_\pi + \rho \phi_{MA}\mu N \tau_y + \phi_{MI}(1-\mu)N \tau_y], \quad (3A.22)$$

After substituting equations (4) to (6) into the demand function (3A.22) and reorganizing the expression, we can obtain:

$$d_i(\tilde{p}; \mathcal{F}_{i,M}^{IN}) = \frac{1}{\gamma} \left[ h_{0,M}^{IN} + h_{v,M}^{IN} \tilde{v} + h_{x,M}^{IN} \tilde{x} + \rho \tau_y \sum_{j=1}^{\phi_{MA}\mu N} (\tilde{\varepsilon}_i + \tilde{\eta}_{ij}') \right] + \frac{1}{\gamma} \left[ \tau_y \sum_{j=1}^{\phi_{MI}(1-\mu)N} (\tilde{\varepsilon}_i + \tilde{\eta}_{ij}') + \tau_s \tilde{\varepsilon}_i - \tau_y (\tilde{\varepsilon}_i + \tilde{\eta}_{ij}') \right], \quad (3A.23)$$

where

$$h_{0,M}^{UN} = \tau_v \bar{v} + \tau_\pi \frac{\alpha_x}{\alpha_v} \bar{x} - \alpha_o B_M^{UN},$$

$$h_{v,M}^{UN} = (1 - \alpha_v) B_M^{UN} - \alpha_v \tau_v,$$

$$h_{x,M}^{UN} = -\frac{\alpha_x}{\alpha_v} \tau_\pi + \alpha_x B_M^{UN},$$

$$B_M^{UN} = \tau_v + \tau_\pi + \rho\phi_{MA}\mu N\tau_y + \phi_{MI}(1 - \mu)N\tau_y.$$

Similarly,

$$\mu_M^{UN}(\tilde{v}, \tilde{x}) = E[d_i(\tilde{p}; \mathcal{F}_{i,M}^{UN})|\tilde{v}, \tilde{x}] = \frac{1}{\gamma}(h_{0,M}^{UN} + h_{v,M}^{UN}\tilde{v} + h_{x,M}^{UN}\tilde{x}), \quad (3A.24)$$

$$\sigma_M^{UN} = \sqrt{Var[d_i(\tilde{p}; \mathcal{F}_{i,M}^{UN})|\tilde{v}, \tilde{x}]} = \frac{1}{\gamma}\sqrt{\rho\phi_{MA}\mu N\tau_y + \phi_{MI}(1 - \mu)N\tau_y}, \quad (3A.25)$$

$$E[d_i(\tilde{p}; \mathcal{F}_{i,M}^{UN})|\tilde{v}, \tilde{x}] = \sigma_M^{UN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_M^{UN}(\tilde{v}, \tilde{x})]^2}{2\sigma_M^{UN}}} + \mu_M^{UN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_M^{UN}(\tilde{v}, \tilde{x})}{\sigma_M^{UN}}\right) \right]. \quad (3A.26)$$

where  $\Phi(\star)$  is the cumulative distribution function of the standard normal distribution, and  $\mu_M^{IN}$ ,  $\sigma_M^{IN}$  are defined by equations (3A.24) and (3A.25).

Finally, the trading volume can be expressed by adding absolute trading volumes from different parties based on equations (3A.12), (3A.16), (3A.21), and (3A.26):

$$\tilde{q} = \frac{1}{2} \left\{ \begin{aligned} & \phi_{MA}\mu \left[ \sigma_W^{IN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_W^{IN}(\tilde{v}, \tilde{x})/\sigma_W^{IN}]^2}{2}} + \mu_W^{IN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_W^{IN}(\tilde{v}, \tilde{x})}{\sigma_W^{IN}}\right) \right] \right] \\ & + (1 - \phi_{MA})\mu \left[ \sigma_W^{UN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_W^{UN}(\tilde{v}, \tilde{x})/\sigma_W^{UN}]^2}{2}} + \mu_W^{UN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_W^{UN}(\tilde{v}, \tilde{x})}{\sigma_W^{UN}}\right) \right] \right] \\ & + \phi_{MI}(1 - \mu) \left[ \sigma_M^{IN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_M^{IN}(\tilde{v}, \tilde{x})/\sigma_M^{IN}]^2}{2}} + \mu_M^{IN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_M^{IN}(\tilde{v}, \tilde{x})}{\sigma_M^{IN}}\right) \right] \right] \\ & + (1 - \phi_{MI})(1 - \mu) \left[ \sigma_M^{UN} \sqrt{\frac{2}{\pi}} e^{-\frac{[\mu_M^{UN}(\tilde{v}, \tilde{x})/\sigma_M^{UN}]^2}{2}} + \mu_M^{UN}(\tilde{v}, \tilde{x}) \left[ 1 - 2\Phi\left(-\frac{\mu_M^{UN}(\tilde{v}, \tilde{x})}{\sigma_M^{UN}}\right) \right] \right] \end{aligned} \right\} + \frac{1}{2} |\tilde{x}|$$

## Chapter 4: Factor Timing: Evidence from Mutual Fund Skill

### 4.1 Introduction

Understanding time variations in the common factors of stock returns is central to asset pricing since these factors are tied to the stochastic discount factor (SDF), which reflects investor preferences (e.g., Lettau and Ludvigson, 2001; Haddad, Kozak, and Santosh, 2020). Meanwhile, mutual funds represent one of the most extensively studied investor types in finance as financial economists have spent decades debating whether mutual fund managers possess skill (Berk and Green, 2004; Barras, Scaillet, and Wermers, 2010; Berk and van Binsbergen, 2015; Barras, Gagliardini, and Scaillet, 2022).<sup>50</sup> This naturally raises the question: do mutual fund preferences, shaped by managerial skill, predict common factors through their connection to the SDF?

In this chapter, we build on the literature on mutual fund managerial skill to address this question. Specifically, if a skill measure can distinguish highly skilled funds from low-skilled funds and thereby reveal differences in their investment preferences, can these differences be used to forecast the future performance of common factors? To illustrate the intuition, consider a proposed measure of fund skill, where a high value indicates superior managerial ability. A highly skilled fund is expected to hold securities that perform well. If such securities are concentrated in value stocks, then value stocks should deliver superior future returns. Supporting this intuition, Bai, Hilscher, and Scherbina (2025) show that mutual funds exhibit identifiable factor tilt strategies and that exploiting such tilts can lead to performance improvement for skilled fund managers. Consequently, if highly skilled funds as a group collectively hold securities with attributes similar

---

<sup>50</sup> Numerous studies have proposed measures to identify skilled funds, and the consensus is that managerial skill is scarce, though some mutual fund managers do possess superior ability. Berk and Green (2004) provide a rational theory that explains the existence of such skill. Empirically, Barras, Scaillet, and Wermers (2010), Barras, Gagliardini, and Scaillet (2022) and Berk and van Binsbergen (2015) show that some fund managers indeed possess skill.

to value stocks, we would anticipate the value factor itself to earn high future returns. Building on this intuition, we introduce a new metric, the fund skill gap (FSG), to capture the differential forecasting ability of skilled versus unskilled funds. We define FSG as the difference in the attributes of stock holdings between highly skilled and low-skilled mutual funds. Conceptually, FSG is akin to the "issuer-repurchaser spread" of Greenwood and Hanson (2012), who use differences in corporate financing choices to predict factor returns, but in our study, the spread is based on investors' portfolio choices, specifically the informed demand of skilled vs. unskilled fund managers.

To construct the FSG metric, we employ several well-established measures of fund skill to distinguish highly skilled funds from their low-skilled counterparts.<sup>51</sup> We find that differences in holdings attributes of these two groups forecast future factor returns. The predictive power is most prominent for the value factor (HML), with weaker but still significant results for the investment (CMA) and profitability (RMW) factors. In contrast, we find little or mixed predictability for the size (SMB) and momentum (UMD) factors. Economically, skilled managers appear to excel at processing information about firms' cash flow fundamentals, rather than relying on short-lived technical signals. Within a conditional asset pricing framework, FSG can be interpreted as a summary measure of skilled managers' informative assessments of the state variables driving factor premia, providing a holding-based and implementable timing signal for fundamental-related discount rates.

To support the mechanism behind our results, we exploit systematic differences in managerial

---

<sup>51</sup> To this end, we proxy for managerial superior information using five standard skill measures: industry concentration (ICI), active share (AS), negative  $R^2$  ( $-R^2$ ), active weight (AW), and value added (VA). The key idea is that informed managers should (i) take discretionary, conviction driven positions that depart from mechanical benchmark exposures (conceptually captured by ICI, AS, and AW) (Kacperczyk, Sialm, and Zheng, 2005, Cremers and Petajisto, 2009, Doshi, Elkamhim, and Simutin, 2015), and (ii) generate value that is not mechanically explained by static factor loadings and is economically meaningful (captured by  $-R^2$  and VA) (Amihud and Goyenko, 2013, Berk and van Binsbergen, 2015). Because each proxy is noisy and reflects a different facet of information-based ability, we also use an average skill index that aggregates them to reduce measurement error when classifying high versus low skill funds.

discretion across fund mandates. The mechanism we propose is information-based portfolio tilts: if FSG reflects skilled managers' superior assessments of fundamentals and discount rates, then the signal should only be effective when managers are allowed to translate those assessments into active portfolio tilts. Mandates, therefore, provide a clean way to test the channel. When a mandate allows meaningful benchmark deviations, skilled managers can express time-varying views through holdings, widening the separation between high and low skill portfolios and increasing the signal-to-noise of FSG. When mandates bind tightly and compress feasible exposures, even informed managers have less room to act, so the same information should show up more weakly in FSG.

Consistent with these predictions, we find that FSG predictability is concentrated among growth-oriented funds, whose mandates permit larger deviations from benchmarks and more active style positioning; within this subsample, the FSG signal is especially strong for HML (and in some specifications extends to SMB). In contrast, capitalization targeted funds exhibit weak predictability across factors and even contrarian patterns for momentum, consistent with tighter constraints and capacity/turnover frictions limiting dynamic tilts. This mandate-based pattern directly supports our mechanism: when discretion is higher, skilled managers' portfolio tilts more cleanly embed forward-looking information about factor premia. In this sense, FSG can be interpreted as an aggregate information state variable in a conditional asset pricing framework, therefore summarizing skilled managers' assessments of the state variables driving expected factor returns and yielding a holdings-based, implementable timing signal for fundamental related factor premia.

We conduct a variety of analyses to ensure that the FSG predictive relationship is robust and not driven by other effects. First, we control for the factor's own recent performance by including the trailing 12-month factor return (to absorb momentum or seasonality) and time fixed effects in our predictive regressions. The FSG signal remains strongly significant for HML, suggesting it

contains information beyond well-known time series patterns (Ehsani and Linnainmaa, 2022; Arnott, Kalesnik, and Linnainmaa, 2023). Second, we rule out price pressure or mechanical effects well documented by the literature (e.g., Coval and Stafford, 2007; Lou, 2012; Li, 2022; Huang, Song, and Xiang, 2024). We find that excluding funds with extreme investor flows (top or bottom decile flows each quarter) does not diminish the results. Third, the predictability persists when we restrict the sample to funds that actively vary their factor exposures over time. Focusing on these actively managed funds helps us isolate sources of the predictive content of our FSG measure, which come from the managerial active change in factor exposure rather than passive style differences or pure stock picking.

We also show that the findings are not sensitive to how FSG is constructed: whether we compute FSG using portfolio factor exposures or directly from the characteristics of held stocks, and whether we define high/low skill cutoff by attribute rankings or raw values, the results for HML remain qualitatively similar. In fact, using techniques like principal component analysis to combine multiple skill measures further strengthens the predictive power, especially for the value premium. Overall, the evidence is consistent with FSG capturing a genuine informational edge possessed by skilled fund managers, rather than artifacts of data or trading frictions.

Our study makes three contributions to the literature. First, it advances research on mutual funds' ability to time common risk factors by showing that skill-sorted holdings differentials predict subsequent factor premia. Early holdings-based evidence casts doubt on mutual fund factor timing: Daniel, Grinblatt, Titman, and Wermers (1997) find little individual fund characteristic timing ability using holdings implied factor exposures. A central identification concern is that changes in a fund's holdings implied exposure are not uniquely informative about deliberate timing. They may instead reflect mechanical style drift and valuation effects, comovement with aggregate states that also shift expected factor returns, or passive drift rather than discretionary reallocation (Ferson and

Schadt, 1996; Brown and Goetzmann, 1997; Chan, Chen, and Lakonishok, 2002).<sup>52</sup> Recent work reinforces this point: Bai et al. (2025) separate active exposure shifts from passive drift and show that timing ability exists but is heterogeneous across funds. However, their approach relies on within-fund absolute average factor loading changes across four factors, which leaves open whether the mechanism operates through directional exposure tilts or the relative importance of different factors for timing. We take a complementary approach, identifying timing from cross-sectional skill heterogeneity in holdings. We construct FSG, a same-quarter high-skill minus low-skill holdings spread that captures skilled managers' directional factor positions while netting out date-level common exposure movements. Empirically, when high-skill funds tilt more toward value than low-skill funds, the HML premium subsequently rises. This cross-sectional design complements loading-change approaches (e.g., Bai et al., 2025) and turns dispersed portfolio positions into an implementable timing signal. Our analysis also helps interpret the informational content of different skill proxies.<sup>53</sup>

Second, our study contributes to the literature on factor timing. Prior studies have identified various signals that forecast factor returns, often related to valuation ratios or corporate actions. Greenwood and Hanson (2012) find that issuer repurchaser characteristic spread contains information about future factor returns, either due to time-varying risk premia or characteristic mispricings. Haddad et al. (2020) and Ehsani and Linnainmaa (2022) show more evidence on the predictability of principal component factors based on either portfolio valuation ratios or past returns. We add to this literature by introducing a fundamentally different kind of predictor: instead

---

<sup>52</sup> In the canonical timing framework, higher factor exposure ahead of strong factor performance is interpreted as timing. This interpretation is fragile if (i) exposures change mechanically (e.g., a growth portfolio appears more "value" after relative price declines), (ii) exposures move with the same state variables that shift expected factor returns (Ferson and Schadt, 1996), or (iii) exposure movements largely reflect drift rather than discretionary trading (Brown and Goetzmann, 1997; Chan et al., 2002).

<sup>53</sup> For example, the industry expertise based FSG predicts momentum in some specifications, consistent with Moskowitz and Grinblatt, 1999, showing that industry momentum explains a substantial share of conventional momentum returns, suggesting an industry information channel for momentum timing.

of coming from characteristics of the stocks or from ex post factor returns, our signal comes from the revealed preferences of sophisticated investors. This complements existing signals by drawing on the idea that certain investors may possess superior information or analytical foresight about factor payoffs.

Finally, we contribute to the demand-based asset pricing and return predictability literature that uses institutional holdings to study their price impact on asset prices.<sup>5</sup> Prior work attributes predictability to flow driven demand shocks and trading pressure arising from fire sales and rebalancing (Coval and Stafford, 2007; Frazzini and Lamont, 2008; Lou, 2012; Peng and Wang, 2023) and shows that these forces can impact characteristic based factor returns (Li, 2022; Huang et al., 2024; Peng and Wang, 2023). We instead isolate an information channel within mutual funds: skill-sorted holdings differences reflect managers' assessments of conditional expected factor premia, which is different from the uninformed flow, liquidity, and mechanical rebalancing effects emphasized in this literature (Coval and Stafford, 2007; Lou, 2012; Peng and Wang, 2023; Huang et al., 2024).

This chapter proceeds as follows. Section 4.2 describes the data. Section 4.3 defines our main metric, fund skill gap (FSG), and explains how we aggregate different skill measures using averaging or composite methods such as PCA and PLS. Section 4.4 presents the main results on predictability and evaluates robustness. Section 4.5 concludes.

## **4.2 Data**

### **4.2.1 Mutual fund sample**

Our mutual fund sample spans from January 1991 to December 2023 and focuses on U.S. actively managed equity funds. Monthly net returns and total net assets (TNA) are obtained from the CRSP Survivor Bias Free U.S. Mutual Fund database. Following Dou, Kogan, and Wu (2023),

we start our sample from 1991 to avoid coverage issues in earlier years. Next, we exclude index funds, exchange-traded funds (ETFs), sector funds, balanced funds, international funds, bond funds, and money market funds. This screening is based on CRSP investment objective codes and fund names, as in Doshi et al. (2015), and ensures that our sample is focused on funds with discretionary equity allocations within the U.S. market. Fund returns are reported net of management fees and transaction costs, excluding any front and back-end loads. We also obtain fund excess returns using the one-month Treasury Bill rate from Kenneth French's data library.<sup>54</sup>

Fund TNA is adjusted to be in 2023 U.S. dollars for inflation using the Consumer Price Index, as in Berk and van Binsbergen (2015), to ensure comparability of fund size over time. We remove observations where fund names are missing or the prior month's real TNA falls below \$10 million. This threshold follows Berk and van Binsbergen (2015) and helps mitigate concerns about artificial performance among very small or incubated funds. We also drop all records prior to a fund's official inception date, addressing the survival and incubation biases as in Evans (2010).

Holdings data are collected from Thomson Reuters S12 files and CRSP Mutual Fund Holdings. We use S12 from 1991 through 2008 Q3, and CRSP thereafter. This combination addresses data quality concerns documented in Zhu (2020), who finds that the S12 database omits many newly launched funds prior to 2008. A similar merging strategy is employed in Pástor, Stambaugh, and Taylor (2015) to improve the completeness of fund portfolio coverage. Share classes are aggregated to the fund level using the MFLINKS file by adding up total net assets from different share classes of the same fund. We also compute TNA weighted averages of fund characteristics from different share classes, consistent with the approach used in Jiang and Zheng (2018).

When processing our holdings data, we restrict the data to common stock shares only,

---

<sup>54</sup> [https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

identified by CRSP share codes 10 and 11. This restriction ensures that fund equity portfolio allocations reflect equity investment decisions by excluding preferred stocks, warrants, ADRs, and other hybrid securities that may not be actively chosen by fund managers. This approach follows standard practice in the mutual fund literature (Falkenstein, 1996; Daniel et al., 1997; Wermers, 2000; Cohen et al., 2008) and allows a consistent measure of equity exposure across funds. To ensure the quality of holdings-based measures we use later in our study, we impose two additional criteria. First, we restrict our sample to fund quarters with more than 10 distinct stock holdings. This requirement follows common practice in holdings-based studies (Kacperczyk et al., 2008; Jiang and Zheng, 2018) and avoids mechanically inflated signals from underdiversified or recently seeded funds. These filters collectively ensure that the calculation of our portfolio-level attributes is meaningful.

To ensure consistency in the timing of portfolio-level characteristic adjustments, we align all fund holdings on a common reporting schedule. In our main analysis, we retain only those holdings reported at calendar quarter ends, and the final sample has 144,347 fund quarter records with 3,636 funds and, on average, 1,173 funds in each quarter. Table 4.1 provides the breakdown of the number of funds in each quarter with different classifications using the CRSP investment objective code.

[Insert Table 4.1 Here]

#### **4.2.2 Mutual fund skill measures**

To evaluate mutual fund timing behavior and portfolio activeness, we employ a set of complementary measures that capture different dimensions of active management. These measures help us understand whether observed return patterns reflect genuine skill, dynamic exposure, or deviations from benchmark mandates as summarized in Table 4.2.

[Insert Table 4.2 Here]

We begin with the industry concentration index (ICI) of Kacperczyk et al. (2005), which

captures the extent to which managers specialize in particular sectors. Higher ICI reflects more focused bets, and we interpret it as evidence of selection skill when such specialization translates into persistent benchmark-adjusted outperformance (Kacperczyk et al., 2005). The second measure, active share (AS), introduced in Cremers and Petajisto (2009), measures the extent to which holdings deviate from benchmark allocations. Higher AS indicates managerial discretion and conviction, and we regard AS as informative of skill when it predicts subsequent abnormal returns rather than merely reflecting uncompensated tracking error or style drift (Cremers and Petajisto, 2009). We also examine  $R$  squared ( $R^2$ ) from rolling window regressions evaluating fund performance, first proposed in Amihud and Goyenko (2013). A lower  $R$  squared implies that more of a fund's return variance is unexplained by static factor exposures, signaling higher skill. This suggests that dynamic adjustments can potentially capture the factor timing behavior of mutual funds. In our tests, we use negative  $R^2$  for ranking consistency so that a higher value of the negative  $R^2$  indicates higher skill. Next, we calculate active weight (AW) from Doshi et al. (2015), which reflects a fund's aggregate deviations relative to value-weighted benchmarks. Similar to active share, high active weight can be related to style drift or tactical positioning in anticipation of factor movements. Finally, we include the value-added metric (VA) from Berk and van Binsbergen (2015), which captures how much gross alpha is translated into the dollar value for fund investors. This capital-weighted measure offers a direct economic assessment of overall managerial ability, and persistence in value added provides strong evidence for managerial skill.<sup>55</sup>

In summary, we consider five skill metrics to capture both return-based and holdings-based information. The value-added measure offers a capital-weighted outcome view of performance.

---

<sup>55</sup> The negative mean value-added in our sample is not inconsistent with Berk and van Binsbergen (2015), who document positive average value-added over 1977–2011 using a single-benchmark. The discrepancy likely reflects our use of a multi-factor alpha and an extended sample period (1993–2024) that captures an increasingly competitive environment for active management. Importantly, our analysis relies on the cross-sectional dispersion in skill rather than its average level, so the negative mean does not compromise the validity of the FSG measure.

Active share, active weight, and industry concentration capture the extent and direction of benchmark deviation, while signed  $R$  squared helps interpret whether deviations were value-enhancing. Used jointly, these measures give a richer picture of how funds actively respond to investment opportunities over time. Summary statistics of these measures are provided in panel A of Table 4.3.

[Insert Table 4.3 Here]

### 4.2.3 Factor returns

We obtain monthly factor returns from Kenneth French's data library. The factors we consider include size (SMB), value (HML), investment (CMA), and profitability (RMW) factors from Fama and French (2015), as well as the momentum factor (UMD) from Jegadeesh and Titman (1993) and Carhart (1997). All factors return series are aligned by the month-end with our fund sample. Summary statistics of factor returns are provided in panel B of Table 4.3.

## 4.3 Fund Skill Gap

### 4.3.1 Construct FSG using factor exposures and characteristics

Let  $\beta_{j,f,t-1}^{\text{decile}}$  denote fund  $j$ 's stock holdings' cross-sectional decile for factor exposure to a pricing factor  $f$  (e.g., HML factor) in quarter  $t - 1$ .<sup>56</sup> We define the fund skill gap (FSG) for factor  $f$  given a skill measure  $S$  as the average factor exposure decile of highly skilled funds' stock holdings minus the average factor exposure decile of low skill funds' stock holdings:

---

<sup>56</sup> To construct this, we: (i) estimate stock-level factor exposures using 36-month rolling regressions on the Fama–French five factors plus momentum, requiring at least 24 valid returns; (ii) assign each stock an ascending decile rank (i.e., higher exposure, higher rank) based on the breakpoints of common stocks listed on NYSE, NASDAQ, and AMEX (now NYSE American); and (iii) aggregate these ranks to the fund's portfolio level, using only long positions. Results based on raw factor exposures rather than deciles are similar.

$$FSG_{t-1}^{S,\beta_f} = \frac{\sum_{j \in \text{Highly skilled}} \beta_{j,f,t-1}^{\text{decile}}}{N_{t-1}^{\text{Highly skilled}}} - \frac{\sum_{j \in \text{Low skilled}} \beta_{j,f,t-1}^{\text{decile}}}{N_{t-1}^{\text{Low skilled}}}, \quad (37)$$

where  $N_{t-1}^{\text{Highly skilled}}$  ( $N_{t-1}^{\text{Low skilled}}$ ) is the number of funds in the top (bottom) decile of the skill measure  $S$ , and cross-sectional  $\beta_f$  Deciles for each quarter end are assigned using breakpoints of all common stocks listed on three major U.S. exchanges.<sup>57</sup> For instance, if we use value added (VA) as a measure for fund skill and consider value factor (HML), then  $FSG_{t-1}^{VA,\beta_{HML}} = 1$  indicates that highly skilled funds were on average one  $\beta_{HML}$  decile larger than low skill funds in quarter  $t - 1$ . In this case, a positive value of FSG suggests that, on average hold stocks with higher HML exposure than low skill funds. If fund skill is truly captured by some skill measures, such as value added, then FSG should positively predict the following monthly returns on the HML factor.

An alternative FSG measure is based on fund stock holdings' characteristics tied to a given pricing factor (e.g., book-to-market for HML). The caveat is that a single characteristic is an imperfect proxy for the corresponding factor exposure and may also load on stock level alpha (Kelly et al., 2019; Fama and French, 2020; Kozak and Nagel, 2024). For this reason, we use the rolling window factor exposures to construct our FSG measure in our main specification. However, estimating factor exposures requires specifying a factor model a priori, and model misspecification can induce additional measurement errors. Therefore, when assessing predictability, we also report results using a characteristic-based FSG measure constructed as follows.

Let  $X_{j,t-1}^{\text{decile}}$  denote fund  $j$ 's portfolio level (ascending) decile for characteristic  $X$  (e.g., book to market) in quarter  $t - 1$ .<sup>58</sup> We define the fund skill gap (FSG) for characteristic  $X$  given a skill measure  $S$  as the difference in average deciles between highly skilled and low skill funds:

---

<sup>57</sup> Greenwood and Hanson (2012) construct an issuer-repurchaser spread to measure firm's ability to forecast its own stock returns.

<sup>58</sup> The construction mirrors factor-exposure-based FSG except that ranking process is based on the characteristics instead of factor exposures. Results based on raw characteristics rather than deciles are also similar in this set-up.

$$FSG_{t-1}^{S,X} = \frac{\sum_{j \in \text{Highly skilled}} X_{j,t-1}^{\text{decile}}}{N_{t-1}^{\text{Highly skilled}}} - \frac{\sum_{j \in \text{Low skilled}} X_{j,t-1}^{\text{decile}}}{N_{t-1}^{\text{Low skilled}}}, \quad (38)$$

where  $N_{t-1}^{\text{Highly skilled}}$  ( $N_{t-1}^{\text{Low skill}}$ ) is the number of funds in the top (bottom) decile of the skill measure  $S$ , and  $X$ -deciles for each quarter end are assigned using breakpoints of all common stocks listed on three major U.S. exchanges. Interpretation mirrors the factor exposure-based FSG: a positive  $FSG_{t-1}^{S,X}$  indicates that highly skilled funds hold stocks with higher values of  $X$  (e.g., B/M). If  $S$  captures skill,  $FSG_{t-1}^{S,X}$  should positively predict subsequent returns on the factor sorted by  $X$  (e.g., HML).

### 4.3.2 Skill Aggregation

Since we use five distinct fund skill measures, combining them helps extract the common skill component among these measures and reduce bias from using a single skill proxy. We aggregate their information in two ways. The first approach is to construct a simple average skill index (ASI) by averaging the ranks from the five skill measures for each fund each quarter and then distinguishing highly skilled funds from low skill funds with the average skill index. The second approach is to conduct dimension reduction for the five FSG measures constructed with corresponding skill measures. We consider two dimension reduction approaches: principal component analysis (PCA) and partial least squares (PLS).

#### 4.3.2.1 Average skill index

We combine the five skill measures by averaging their within-quarter ranks. First, we orient each skill measure so that a higher value indicates greater skill. Next, within each quarter  $t - 1$ , we convert each skill measure  $k$  to a cross-sectional decile rank for every fund  $j$ :  $r_{j,k,t-1} = \mathcal{D}_{10}(S_{j,k,t-1}) \in \{1, \dots, 10\}$  ( $k = 1, \dots, 5$ ), where  $\mathcal{D}_{10}(\cdot)$  maps a value to its ascending decile in the

fund cross-section at  $t - 1$ . We then define the average skill index (ASI) for fund  $j$  in quarter  $t - 1$  as the simple average of the five decile ranks:

$$ASI_{j,t-1} = \frac{1}{5} \sum_{k=1}^5 r_{j,k,t-1} \in [1,10]. \quad (39)$$

Each quarter, funds in the top (bottom) ASI decile are classified as highly skilled (low-skilled). We then calculate  $FSG_{t-1}^{ASI,\beta_f}$  and  $FSG_{t-1}^{ASI,X}$  following the same procedure detailed in Section 4.3.1. Rank averaging makes different skill measures comparable and is robust to outliers.

#### 4.3.2.2 Use PCA and PLS to refine FSG measures

With different skill measures, we can construct corresponding FSG measures as proxies for the conditional attribute premia. However, each FSG time series may be correlated and act as a noisy proxy for the true attribute premia. Resorting to dimension reduction approaches to extract a common component across different FSG time series can help us resolve this potential issue. Specifically, we stack the five FSG series into a vector each quarter, standardize the inputs, and apply principal component analysis (PCA). We use the first principal component as an unsupervised composite (Connor and Korajczyk, 1993; Lettau and Pelger, 2020) to capture the variations shared across different measures, which after standardization, is invariant to scale, yielding a natural candidate to summarize potentially correlated but noisy FSG series. An alternative to PCA is partial least squares (PLS), which is tailored specifically to predict next month's factor returns.<sup>59</sup> We use PLS to form a latent combination of the five FSG series that maximizes its covariance with our forecasting target, therefore emphasizing the skill dimensions most predictive for future factor returns. As our focus is on economic interpretability, we estimate fixed (static) loadings using the full sample and apply these weights each quarter instead of estimating dynamic loadings using rolling windows.

---

<sup>59</sup> Kelly and Pruitt (2013) and Huang, Jiang, Tu, and Zhou (2015) use PLS to predict market returns.

## 4.4 Predictability of FSG on Factor Returns

In this section, we develop an empirical framework to motivate why FSG can contain information about time-varying factor premia. We then take this implication to the data by testing whether lagged FSG predicts subsequent factor returns using bivariate correlations and predictive regressions. We also aggregate multiple FSG signals using PCA and PLS based forecasters. Next, we examine how predictability varies across fund mandates to shed light on the information tilt mechanism. Finally, we address key robustness concerns, including flow-induced trading, factor turnover, and alternative FSG constructions, and show that our baseline results are robust to these variations.

### 4.4.1 Empirical framework

Consider a conditional attributes model of expected returns for individual stock  $i$ .<sup>60</sup>

$$\mathbb{E}_{t-1}[R_{i,t}] = \alpha_{t-1} + \beta_1 \cdot X_{i,t-1} + \beta_2 \cdot (\bar{T}_{t-1} \times X_{i,t-1}) + \mu_{i,t-1}. \quad (40)$$

Here attributes  $X_{i,t-1}$  may represent either firm-level characteristics (e.g., book-to-market ratio) or estimated factor exposures (e.g., rolling window beta estimates with respect to  $HML$ ), and  $\mu_{i,t-1}$  is the idiosyncratic part of the stock's expected return. This framework accommodates both characteristics-driven and factor exposure-driven sources of expected returns. Accordingly,  $\beta_1 + \beta_2 \cdot \bar{T}_{t-1}$  is the time varying attribute premium. Such an extension is valid as long as either characteristics or factor exposures explain the expected returns.

Assume fund  $s$  holds a weight of  $w_{i,s,t-1}(s)$  for stock  $i$  at time  $t - 1$ . Let  $\theta_{s,t-1}$  be a skill measure that serves as a proxy capturing a fund's ability to generate returns in the following period:

---

<sup>60</sup> Baker and Wurgler (2006) extend the characteristics model in Daniel and Titman (1997) with conditional terms and call the extended model as a conditional characteristics model. See also Greenwood and Hanson (2012) and Kelly et al. (2019) for similar specifications.

$$\theta_{s,t-1} = \mathbb{E}_{t-1}[R_{s,t}] + \epsilon_{s,t-1}, \quad (41)$$

where  $R_{s,t}$  is the fund portfolio return and is given as

$$R_{s,t} = \sum_i w_{i,st-1}(s) \cdot R_{i,t}, \quad (42)$$

and  $\epsilon_{s,t-1}$  is the unpredictable component of fund performance due to luck. Taking conditional expectations on both sides of (42) and substituting (40) for  $\mathbb{E}_{t-1}[R_{i,t}]$  gives the expected return for the fund portfolio (using the fact  $\sum_i w_{i,t-1}(s) = 1$ ):

$$\mathbb{E}_{t-1}[R_{s,t}] = \alpha_{t-1} + (\beta_1 + \beta_2 \cdot \bar{T}_{t-1}) \times X_{st-1} + \sum_i w_{i,t-1}(s) \cdot \mu_{i,t-1}, \quad (43)$$

where  $X_{st-1} = \sum_i w_{i,t-1}(s) \cdot X_{i,t-1}$ . Finally, substituting (43) into (41) for fund expected return yields:

$$\theta_{s,t-1} = \alpha_{t-1} + (\beta_1 + \beta_2 \cdot \bar{T}_{t-1}) \times X_{st-1} + \sum_i w_{i,t-1}(s) \cdot \mu_{i,t-1} + \epsilon_{s,t-1}. \quad (44)$$

where  $X_{st-1} = \sum_i w_{i,t-1}(s) \cdot X_{i,t-1}$  represents the fund's aggregate attributes of its stock holdings.<sup>61</sup>

A univariate cross-sectional regression of fund skill measure  $\theta_{s,t-1}$  at time  $t - 1$  on fund stock holdings attributes  $X_{s,t-1}$  therefore gives the time varying attribute premium as the cross-sectional regression slope indicated by equation (45) and taking the form of:

$$\delta_{t-1} = \beta_1 + \beta_2 \cdot \bar{T}_{t-1} \quad (45)$$

Suppose we want to forecast Nvidia's return in the coming year. The literature on the cross-section of expected stock returns suggests that we need to gather information on Nvidia's attributes and yield a forecast assuming that each attribute is associated with some average return in the cross-section. But our approach suggests that we can use the skill measure of mutual funds that hold

---

<sup>61</sup> Empirically, we specify  $X_{s,t-1}$  either as its portfolio-weighted characteristics or portfolio-weighted factor

stocks with the same attributes as Nvidia to recover the conditional expected return of Nvidia associated with these attributes. Consequently, we can back out the time-varying attribute premium  $\delta_{t-1}$  by comparing skill measures for funds with different aggregate attributes of their stock holdings in each cross-section.

We can simply implement this idea by calculating the difference in  $X_{s,t-1}$  between highly skilled and low-skilled funds instead of directly running cross-sectional regressions of  $\theta_{s,t-1}$  on  $X_{s,t-1}$ . The time series of these differences should positively predict returns associated with that characteristic. We adopt this implementation for our empirical tests.<sup>62</sup> The advantage is that sorting funds into skill-based groups improves the precision of the estimated attribute premium by averaging out idiosyncratic noise in individual fund skill measures.

#### 4.4.2 Long-short characteristics portfolio predictive regressions

We use the value factor as an example to demonstrate our approach. Figure 4.1 illustrates the timeline for our predictive regressions. At the end of each quarter, we compute the skill measure  $S$  for every fund, and then group funds into ascending deciles based on  $S$ , and use the top and bottom deciles to construct both factor exposure-based fund skill gap measure  $FSG^{S,\beta_{HML}}$  and the characteristics-based measure  $FSG^{S,BM}$ . We then examine both measures' abilities to forecast the HML return in the first month of the subsequent quarter. This timing yields four predictive observations per year and avoids look-ahead bias.

[Insert Figure 4.1 Here]

In general, we assess the predictability of FSG on its corresponding factor by estimating the

---

<sup>62</sup> To illustrate, if we treat each group of highly skilled and low skill funds as a single fund and run the cross sectional regression, then the expected return associated with the attribute is estimated as  $\hat{\delta}_{t-1} = \frac{\theta_{H,t-1} - \theta_{L,t-1}}{X_{H,t-1} - X_{L,t-1}}$ .  $\hat{\delta}_{t-1} > 0$  if highly skilled funds hold stocks with high values of the attribute, and  $\hat{\delta}_{t-1} < 0$  if highly skilled funds hold stocks with low values of the attribute.

following regression for a given skill measure-factor pair:

$$Ret_{f,t+1} = \beta_0 + \beta_{f,1}^S FSG_t^S + \bar{Ret}_{f,t}^{(12)} + FE_{quarter} + \epsilon_{t+1}, \quad (46)$$

where long short characteristics portfolio  $f \in \{HML, SMB, RMW, CMA, UMD\}$ ;  $S$  is one of the five skill measures (i.e.,  $ICI, AS, -R^2, AW, VA$ ) or the average skill index ( $ASI$ );  $FSG_t^S$  denotes either the factor exposure based  $FSG_t^{S,\beta_f}$  or the characteristics based  $FSG_t^{S,X_f}$ , where  $X_f$  is the sorting characteristic associated with factor  $f$  (e.g.,  $X_{HML}$  is the book-to-market ratio); and  $\bar{Ret}_{f,t}^{(12)}$  is the average return over the past 12 months, measured from  $t - 12$  through  $t$ . All regressions include quarter fixed effects to absorb seasonal patterns in fund behavior, and standard errors are Newey–West adjusted with four lags.

#### 4.4.2.1 Bivariate correlation coefficients

In this subsection, we report the Pearson bivariate correlations between factor return  $Ret_{f,t+1}$  and our skill gap measure  $FSG_t^S$ . Table 4.4 reports bivariate predictive correlations between the factor exposure-based skill gap (FSG) at  $t$  and next month factor returns at  $t + 1$  over 1991–2023. For each skill proxy (industry concentration, active share,  $-R^2$ , active weight, value added, and the average skill index), funds are sorted into ascending deciles to construct FSG signals. Panel A of Table 4.4 shows that information embedded in skilled managers' portfolio tilts is most useful for predicting fundamentals-related factor premia.

Using our average skill index as an overall measure for fund skill, we find that the factor exposure-based FSG displays economically and statistically meaningful correlations with subsequent returns on value (0.30,  $t = 3.49$ ), investment (0.18,  $t = 2.02$ ), and profitability (0.19,  $t = 2.11$ ) factors, while the links to size (0.09,  $t = 0.98$ ) and momentum (0.06,  $t = 0.69$ ) factors are weak. Quantitatively, a one decile increase in  $FSG_t^{ASI,\beta_{HML}}$  predicts a 30 bps increase in next month's HML returns, about 166.67% of the sample mean (0.18%). These patterns indicate

that the cross-section of managerial skill is expressed primarily through fund stock exposures to firms' fundamental valuation, asset growth, and profitability, rather than through trades from technical signals.

[Insert Table 4.4 Here]

Results for the Size factor are modest at best, which are consistent with institutional constraints. In reality, many funds face binding limits on holding micro and small-cap stocks, yielding narrow dispersion in the size tilts that FSG can capture. By contrast, tilts toward undervalued, profitable, or conservative investment firms are feasible within typical mandates and better reflect managers' information processing about fundamentals. In addition, the size premium bundles several effects (liquidity, visibility, or ownership constraints) and is difficult to forecast.

The joint dynamics between our average skill index, the factor exposure-based FSG, and next month's Value (HML) returns are illustrated in Figure 4.2. Corresponding plots for the other factors are reported in the Appendix. HML exhibits clear comovement with the lagged FSG, consistent with a strong predictive relation. During major downturns (e.g., 2000–2001 and 2008–2009), however, this comovement weakens as HML volatility rises sharply.<sup>63</sup>

[Insert Figure 4.2 Here]

In panel B of Table 4.4, using the characteristic-based FSG measures exhibits qualitatively similar results for the value factor, though for other factors, the predictive power of FSG becomes weaker. One explanation for the discrepancy in results from using factor exposure-based FSG and characteristics-based FSG is that using characteristics alone may induce bias in capturing stock tilts from factor timing. For instance, a highly skilled fund anticipating value factor to perform well may select stocks with high value factor exposure but not necessarily with a high book-to-market

---

<sup>63</sup> As shown in Appendix 4A.1, Momentum (UMD) displays rapid, high frequency swings that are difficult for mutual funds to exploit given mandate, liquidity, and turnover constraints. Effective momentum timing would require frequent position adjustments that are typically infeasible or suboptimal for mutual funds.

ratio. Constructing FSG from factor exposures aggregates information specifically related to factor timing, reducing potential bias introduced by characteristics per se. Overall, the value factor remains predictable even under the noisier version of our FSG measure (for our average skill index,  $t = 1.75$ ), reinforcing that skilled managers' information advantage is strongest in valuation-based factor timing. In summary, our preliminary correlation tests suggest that skill manifests in anticipatory tilts toward fundamental styles and away from short-lived technical signals.

#### 4.4.2.2 Regression results

Table 4.5 reports our regression results based on the setup detailed in Section 4.4.2. Panel A of Table 4.5 largely echoes the bivariate evidence for factor exposure based on FSG: predictability concentrates in the value factor (HML). The Average Skill FSG coefficient is large and precisely estimated (2.63, with a  $t$ -statistics of 3.99, significant at the 1% level). As shown in Table 4A.2 in the Internet Appendix, within  $R^2$  is the highest using Average Skill FSG to predict value factor returns, compared with using other single skill measures. One explanation is that the ASI aggregates information across several proxies (ICI, active share,  $-R^2$ , active weight, value added), which reduces noise in each individual skill measure when constructing our FSG measures. In summary, mutual funds appear to be skilled at timing factor premia associated with slow-moving fundamentals, therefore time variations in *HML* become more predictable from factor exposures of mutual fund stock holdings compared to other factors.

[Insert Table 4.5 Here]

The one departure in panel A of Table 4.5 from panel A of Table 4.4 is the result for the profitability factor (RMW): the Average Skill coefficient becomes insignificant once we include  $\bar{Ret}_{f,t}^{(12)}$  and quarter fixed effects. The reason may be that firm-level profitability is highly persistent and linked to a quarterly information cycle (Bernard and Thomas, 1989; Fama and French, 2015). The 12-month average return absorbs slow-moving persistence in RMW, while quarter fixed

effects soak up seasonality around earnings releases. When including lagged RMW returns, the residual month-to-month RMW return variation is smaller and less predictable by FSG. In contrast, the value added (VA) based FSG remains significant for RMW since VA measures skill from an outcome perspective and therefore captures managers' cross-sectional selection for profitable firms and is less collinear with trailing factor returns than the composite average skill index.

For the size factor (SMB), the factor exposure based on FSG using industry concentration as the skill measure is strongly predictive (1.08, with a *t*-statistics of 2.88, significant at the 1% level). Size premia bundle liquidity, financing conditions, and industry-specific demand-supply dynamics, which makes SMB relatively difficult to forecast using other generic skill proxies. Higher industry concentration plausibly reflects domain specialization: funds combine macro views (credit, growth, policy) with sector-specific knowledge about small-cap cyclical and capacity constraints, enabling sharper SMB timing than other skill measures. Momentum factor (UMD) remains hard to forecast—coefficients are sometimes negative and also imprecisely estimated—showing that managerial skill has little role in forecasting momentum returns. We also find that the investment factor (CMA) is predictable by our FSG measure using the average skill index (significant at the 5% level), providing additional evidence that managerial skill is more useful for forecasting fundamental-related factor premia.

Regression results using characteristics-based FSG are reported in panel B of Table 4.5. The overall patterns are similar for size, value, and investment factors. Specifically, the value factor predictability remains qualitatively unchanged, though predictability becomes weaker for other factors. This may be due to the fact that single characteristics are imperfect proxies when capturing fund stock tilts originated from factor timing. The detailed regression results can be found in 4A.2 in the Internet Appendix.

From a conditional asset pricing perspective, time variations in factor premia are related to investor preferences. The FSG constructed as the difference in portfolio tilts between highly skilled

and low skill funds can serve as an information state variable: when  $FSG_{t-1}$  for a given factor is high, skilled managers hold stocks with more factor exposure and therefore anticipate higher factor returns in the following period  $t$ . The strongest and most robust results lie in the value factor (HML), suggesting that valuation based factor is the most conditionally predictable factor from managerial information sets. Weaker results for momentum and (on average) size factors are consistent with the explanation that skilled managers rarely take advantage of transitory return dynamics upon forecasting returns and they are faced with tighter investment constraints that limit their portfolio tilts. On the other hand, the predictability of FSG using ICI as the skill measure for the size factor provides evidence that domain specialization is related to managers' ability to predict future performance of SMB.

#### 4.4.2.3 PCA and PLS for forecaster aggregation

In this section, we aggregate the information in the FSG time series using principal component analysis (PCA) and partial least squares (PLS). For each pricing factor  $f$ — *SMB*, *HML*, *UMD*, *CMA*, *RMW*— we collect the five skill-specific factor exposure based FSG signals (ICI, AS,  $-R^2$ , AW, VA) and construct (i) the first principal component,  $PC1_t^f$ , which captures the dominant common variation across FSG constructed with different skill measures, and (ii) a one-component PLS index,  $PLS1_t^f$ , chosen to maximize the covariance with the next month factor return  $Ret_{f,t+1}$ . Predictive content is evaluated by re-estimating the baseline specification in equation (46), replacing  $FSG_t^S$  with  $PC1_t^f$  or  $PLS1_t^f$ . Table 4.6 reports the regression coefficients together with the PCA and PLS loadings used to form the single predictive component. Through the loadings, we can see which skill-based FSG measures contribute the most to the single predictive component.

As shown in Table 4.6, using the first principal component or the one-component PLS index helps enhance the predictability of our FSG measures for factor returns, suggesting that the common component of our FSG measures constructed with five skill measures indeed contains

predictive information related to fund factor timing skill. Inspecting the loadings of each FSG series, we find that each FSG series contributes in the correct direction to either the first principal component or the one-component PLS index. For instance, FSG measures constructed with industry concentration, value added, and active share have the highest weights for the one-component PLS index in forecasting HML returns, consistent with the results in Table 4.5 showing these three FSG measures are the most significant ones in predicting HML returns.

[Insert Table 4.6 Here]

#### **4.4.3 Mechanism: predictability within different fund mandates**

To probe heterogeneity in mutual funds' factor timing ability, we partition the sample using CRSP objective codes and re-estimate the baseline factor exposure based FSG specifications for each of the three subsamples: (i) growth, growth and income, and income funds; (ii) growth only funds; and (iii) capitalization targeted funds (mid cap and small cap). The evidence indicates that predictability is concentrated among growth-oriented funds, with substantially weaker effects for income and cap-based categories. This pattern is consistent with growth mandates' emphasis on capital gains and greater willingness to take time-varying style exposures.

Table 4.7 reports results for the combined group of growth, growth and income, and income funds (2,650 funds). The estimates largely mirror the full sample results, with one notable exception: the FSG constructed from this group is predictive for the size (SMB) factor (0.856,  $t = 1.83$  and significant at 10% level) once controls are included. This suggests that, although size premia are generally difficult to time, the skill gap within this broad segment reflects managers' willingness to deviate from benchmarks and assume capacity or liquidity risk in smaller, less visible stocks when they possess superior information. In doing so, they can implement active, granular bets through which any size timing ability is most likely to manifest.

[Insert Table 4.7 Here]

Table 4.8 focuses on growth-only funds (1,990 funds). The results are similar to previous baseline results but show better SMB predictability in the regression specification (coefficient = 1.04,  $t = 2.48$  and significant at 5% level) compared to results in Table 4.8 (for SMB, 0.86,  $t = 1.83$ ). These results suggest that size factor timing, when present, is concentrated in mandates that explicitly pursue capital appreciation and allow greater benchmark deviations, which are more characteristic of growth-oriented funds.

[Insert Table 4.8 Here]

Table 4.9 examines capitalization targeted funds (mid-cap and small-cap; 1,066 funds).<sup>64</sup> Here, predictability is difficult to detect across the board: there exists little predictability for the size factor (by design), but also none for the value, investment, and profitability factors, and a negative association emerges between FSG and the momentum factor (coefficient =  $-5.24$ ,  $t = -2.19$ ). These patterns are consistent with the mandate-imposed constraints that compress cross-sectional dispersion in style exposures, leaving limited room to time factors. Additionally, higher trading costs and tighter capacity constraints of those funds can discourage dynamic tilts. Cap-based funds may trim recent winners to maintain capitalization targets and liquidity buffers, generating contrarian flow that works against return continuation. Overall, the cap-based subsample appears dominated by constraints rather than timing discretion, which attenuates their factor timing ability.

[Insert Table 4.9 Here]

---

<sup>64</sup> Since the cap based universe is thin, we sort funds into quintiles and use the top (bottom) group as the highly skilled (low skill) group. Using quintile sorts only starts from 1996 onward, which is the first year with at least 100 cap based funds to ensure sufficient cross sectional breadth.

#### 4.4.4 Robustness issues

This subsection mitigates concerns that could spuriously generate predictability in FSG. We first mitigate flow-induced trading pressure by filtering out funds experiencing extreme flows. We then account for mechanical exposure changes by restricting the sample to funds with elevated factor turnover over the past three years. Finally, we assess sensitivity to measurement choices by examining alternative FSG constructions. Together, these checks evaluate whether the baseline predictability reflects discretionary positioning rather than non-informational trading or measurement artifacts.

##### 4.4.4.1 Control for flows

If investors can detect managerial skills in financial markets, skilled funds will attract more flows. Skilled funds are therefore subject to flow-induced trading and impose temporary price pressure on stocks they hold, leading to return continuation for these stocks. On the other hand, if there exist more barriers to detecting skilled funds for investors, skilled funds are less subject to flow-induced trading, and the stocks they hold reflect more on their superior information or skills.

[Insert Table 4.10 Here]

To assess whether fund flows are associated with the skill measures, we sort funds each period into net flow terciles ( $f_1$  = low flow;  $f_3$  = high flow) and report in Table 4.10 the average rank for each skill proxy. Across terciles, ranks are relatively flat and centered around 5.5, indicating limited dispersion. Nevertheless, a mild monotonic pattern emerges from  $f_1$  to  $f_3$ : ICI,  $-R^2$  (lower benchmark fit), AW, VA, and the composite average all increase, while AS declines. These patterns suggest that higher flow funds score higher on most skill metrics, raising the concern that our previous results could partly reflect flow-driven price pressure rather than information-driven effects per se. To mitigate such concerns, we implement a flow filter. At the end of each quarter, we sort funds by the absolute value of net flow and drop the top and bottom deciles (i.e., funds with

the most extreme inflows and outflows). We then re-estimate all baseline specifications in Section 4.4.2 on this flow filtered subsample with the same controls used earlier. The results are shown in Table 4.11, which are qualitatively similar to our previous baseline results, indicating that the predictive relations we found earlier are not driven by contemporaneous flow-induced price pressure.

[Insert Table 4.11 Here]

#### 4.4.4.2 Mutual fund's factor turnover

Our skill proxies may capture both factor timing and pure stock picking, which generate fund-specific alpha. To sharpen identification on factor timing—distinct from stock picking—we focus our attention on funds that actively change their portfolio level factor exposures. For each fund–factor pair  $(j, f)$ , let  $\beta_{j,f,t}$  denote the estimated exposure in quarter  $t$ . We measure beta exposure activeness (BEA) (i.e., beta–turnover) over the past 36 months (12 quarters) as the standard deviation of symmetrized quarterly changes:

$$BEA_{j,t}^{\beta_f} = \text{sd} \left[ \frac{\beta_{j,f,h} - \beta_{j,f,h-1}}{0.5(|\beta_{j,f,h}| + |\beta_{j,f,h-1}|)} \right]_{h=t-11}^t,$$

requiring at least two observations in the window. The scaling renders the metric unit–free and self–benchmarked to the fund's recent exposure level. Within each factor  $f$ , we classify funds with  $BEA_{j,t}^{\beta_f}$  above the cross–sectional median as the active–beta group and re-estimate all baseline specifications on this subsample with the same controls. The results are directionally unchanged, and predictability is found concentrated in the value factor (see Table 4.12). These results are consistent with the view that when managers actively adjust factor exposures, our skill measures map more tightly into factor timing ability rather than stock picking.

[Insert Table 4.12 Here]

#### 4.4.4.3 Alternative FSG Constructions

**(1) FSG constructed with fund beta estimated by daily or monthly fund returns:** In our baseline test, we utilize quarterly holdings snapshots and monthly stock returns to construct the FSG. This exposure-based identification relies on the stability of underlying stock betas over a business cycle (approximately five years) to capture systematic risk timing. However, this approach may miss high-frequency timing executed via derivatives or intra-quarter adjustments not visible in holdings data. To address this, and to ensure our results are not driven by specific model identification choices, we include two alternative return-based specifications.

First, to capture high-frequency, short-term timing decisions, we construct the FSG using fund betas estimated from daily raw returns. At the end of each month  $t$ , we estimate factor loadings of mutual funds by regressing daily fund raw excess returns on the Fama-French five factors plus Momentum over a rolling 3-month window. This specification includes a Dimson (1979) correction for the market factor and requires a minimum of 50 valid daily observations. To ensure robust estimation, we apply standardized data filters and perform cross-sectional winsorization on the returns for each time period (see Appendix 4A.3 for detailed protocols). To mitigate the influence of outliers in the estimated coefficients, we convert the resulting betas into cross-sectional decile rankings at each month end. We then construct the FSG as the difference in the average beta rankings between high and low-skilled funds, as detailed in Section 4.3.1. This process generates a set of time series consisting of 306 monthly observations from January 1998 to December 2023.

Second, to align with the longer sample period of our baseline holdings-based test, we construct the FSG using fund betas estimated from monthly raw returns. At the end of each month  $t$ , we estimate factor exposures using a rolling 36-month window. Prior to estimation, we apply standardized filters and perform cross-sectional winsorization on the returns for each time period to ensure data quality (see Appendix 4A.3 for details). Consistent with the daily specification, we

convert the estimated betas into cross-sectional decile rankings at each month end rather than using raw values. The FSG is calculated as the difference in the average beta rankings between high and low-skilled funds. This specification yields a sample of 399 monthly observations ending in December 2023.

We examine the predictive power of these return-based FSG measures by estimating the baseline in-sample predictive regression for the five Fama-French factors and Momentum. Appendix 4A.3 reports the results. We find that the FSG constructed from both daily and monthly return-based betas positively and significantly predicts future Value and Size factor returns. These findings are quantitatively consistent with our baseline holdings-based estimates. This persistence provides extra support for our framework: it demonstrates that the link between managerial skill and factor timing is structural rather than an artifact of the holdings data or specific model assumptions. Whether estimated through high-frequency daily returns (capturing intra-quarter timing) or smoothed monthly returns, the skill gap consistently identifies an informational component for future factor returns from mutual fund managers.

**(2) FSG constructed with raw attributes:** Previously, we constructed factor exposure-based FSG measures using cross-sectional deciles. The decile transformation is standard in the literature and serves primarily to dampen the influence of outliers and to place estimates for factor exposures on a common, unit-free scale across time. Alternatively, we construct our FSG measures using raw attributes and conduct the same tests we did earlier with the alternative FSG measures. Specifically, for each factor  $f$  and skill proxy  $S$ , we construct the Raw FSG (RFSG) in quarter  $t$  as the difference in average factor exposures between the highly skilled and low skill groups:

$$RFSG_{t-1}^{S,\beta_f} = \frac{\sum_{j \in \text{Highly skilled}} \beta_{j,f,t-1}}{N_{t-1}^{\text{Highly skilled}}} - \frac{\sum_{j \in \text{Low skilled}} \beta_{j,f,t-1}}{N_{t-1}^{\text{Low skilled}}},$$

where groups are defined exactly as in earlier in Section 4.3.1; no re-scaling or ranking of  $\beta$  is

applied. We then re-estimate the bivariate correlations and the predictive regressions in equation (46), keeping all controls and fixed effects identical to the baseline.

Using  $RFSG_{t-1}^{S,\beta_f}$  yields estimates that are directionally unchanged relative to the decile-based benchmark. The strongest and most robust predictability continues to exist for HML; SMB and CMA exhibit predictability using the same skill proxies as before, while predictability for RMW remains weaker. Although the resulting magnitudes of estimated coefficients are naturally scaled in the units of  $\beta$ , the  $t$ -statistics and economic interpretations are similar to our previous results. In short, the ranking step in our baseline FSG construction mainly standardizes scale and limits outliers; it is not the driver for our findings. Detailed tables can be found in Appendix 4A.3.

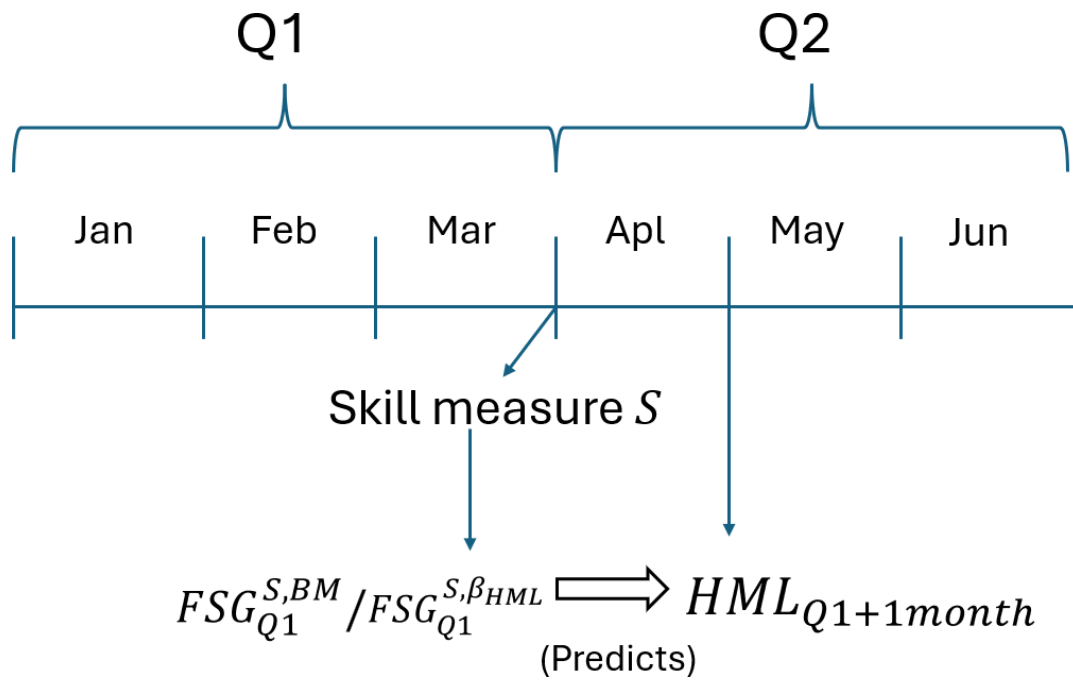
## 4.5 Conclusion

In this chapter, we introduce a simple, transparent Fund Skill Gap (FSG) measure that extracts mutual fund factor timing ability from the cross-section of mutual fund skills. Our FSG measure—constructed from fund portfolio holdings rather than returns—predicts future factor returns. Such predictability is most prominent for the value factor (HML) and relatively mild for the investment and profitability factors. We also find that fund managers do not exhibit superior factor timing skill for the momentum factor, with mixed evidence for the size factor. These patterns are robust after controlling for fund flows, using alternative signal constructions, with composite predictors, and standard controls. Further fund style analysis shows that factor timing skill mainly exists primarily among funds whose mandates allow greater discretion (i.e., growth funds), while it is negligible for funds with tight constraints (i.e., cap-based funds).

Two broader conclusions follow. First, active management skill manifests as an anticipatory style that tilts toward forecasting fundamental-related factor premia. In this sense, our FSG measure functions as a manager-information state variable that complements standard public signals and

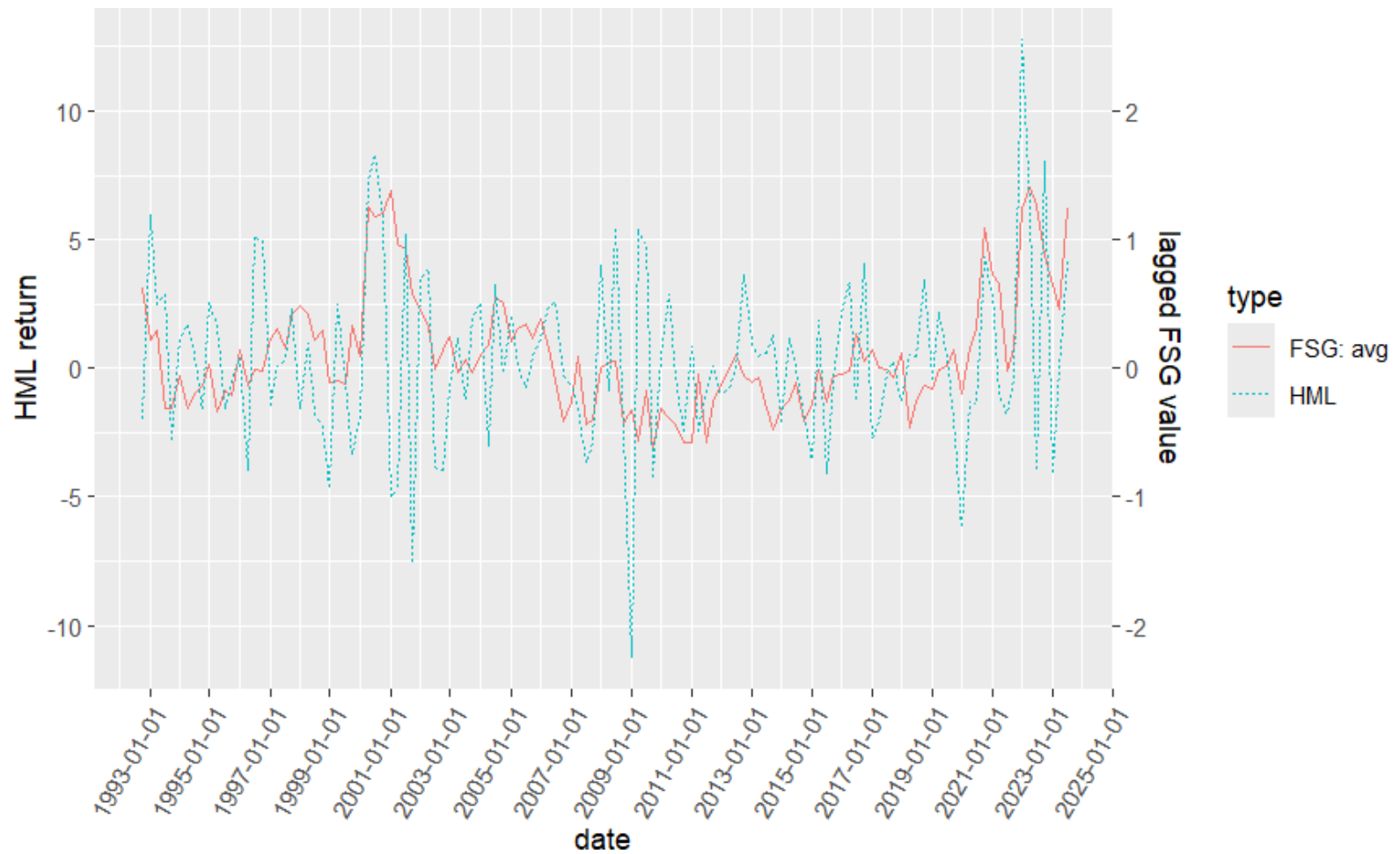
captures investor preferences that help explain time variations in factor premia. When skilled managers lean into value investing (or investing in stocks with conservative investment or high profitability), the corresponding factor premium is more likely to be realized. Second, our study shows holdings data contain forward-looking information about future factor returns, beyond what has been shown in the literature.

Figure 4.1: Prediction Timeline



**Figure 4.2: HML Factor Returns and Lagged Average-Skill FSGs**

This figure plots Value (HML) factor returns and the lagged average skill fund skill gap (FSG). The blue dashed line shows the HML factor return (left y-axis, percent). The red solid line shows the average skill-based FSG,  $FSG_t^{ASI, \beta_{HML}}$ , constructed as in Section 4.3.2.1 (right y-axis) based on equation (39). The horizontal axis is calendar time; observations are quarterly and connected for visual continuity.



**Table 4.1: Number of Funds by Style (1991 – 2023)**

Mutual fund styles are assigned using the CRSP Objective Code (*crsp\_obj\_cd*). “Growth,” “Balance (Growth & Income),” and “Income” correspond to codes EDYG, EDYB, and EDYI. “Mid-cap,” “Small-cap,” and “Micro-cap” indicate funds whose CRSP objective explicitly targets those size segments (e.g., codes EDCM, EDCS, EDCI, or descriptions containing “Mid Cap,” “Small Cap,” “Micro Cap”). Counts are fund-year totals; 1991 and 2023 mark the beginning and end of the sample, respectively.

<b>Year</b>	<b>All Funds</b>	<b>Growth</b>	<b>Balance (G &amp; I)</b>	<b>Income</b>	<b>Mid-cap</b>	<b>Small-cap</b>	<b>Micro-cap</b>
1991	294	263	9	10	1	11	0
1992	358	309	18	13	1	17	0
1993	485	375	40	28	1	41	0
1994	497	364	42	29	7	55	0
1995	539	382	43	35	8	71	0
1996	734	478	74	47	20	115	0
1997	826	504	89	50	30	153	0
1998	1081	627	126	68	49	210	1
1999	1301	743	159	76	66	254	3
2000	1418	812	172	78	80	273	3
2001	1459	830	171	74	94	286	4
2002	1518	851	172	70	114	306	5
2003	1526	846	170	69	115	319	7
2004	1400	764	149	73	113	292	9
2005	1365	728	139	73	120	295	10
2006	1223	640	122	65	110	276	10
2007	1273	650	122	71	122	296	12
2008	1464	750	147	81	149	324	13
2009	1342	708	135	82	130	279	8
2010	2016	1038	234	127	208	399	10
2011	1986	1012	233	133	200	397	11
2012	1956	989	228	134	195	399	11
2013	1950	981	217	149	193	400	10
2014	1946	965	221	157	191	400	12
2015	1951	968	214	157	199	401	12
2016	1947	972	213	158	197	395	12
2017	1931	963	205	157	196	399	11
2018	1915	952	201	155	197	398	12
2019	1902	954	199	152	195	391	11
2020	1867	943	196	145	188	384	11
2021	1820	921	186	140	184	377	12
2022	1778	901	182	137	177	370	11
2023	1727	874	177	135	176	355	10

**Table 4.2: Summary of Mutual Fund Skill Measures**

Measure	Source	Core definition/formula	Data / Window	Interpretation
Industry concentration index (ICI)	Kacperczyk et al. (2005)	Industry focus via industry-level Herfindahl: $ICI_{j,t} = \sum_k (w_{j,t}^k - w_{mkt,t}^k)^2,$ where $w_{j,t}^k$ is the fund $j$ 's weight in industry $k$ at $t$ , and $w_{mkt,t}^k$ is the market portfolio's weight in industry $k$ at $t$ .	Holdings by 10 industry groups from Kacperczyk et al. (2005), quarterly.	Higher ICI $\Rightarrow$ more concentrated bets (specialization) and higher skills.
Active share (AS)	Cremers and Petajisto (2009)	Deviation from benchmark: $AS_{j,t} = \frac{1}{2} \sum_i  w_{j,t}^i - w_{b,t}^i ,$ where $w_{j,t}^i$ is the fund $j$ 's weight in stock $i$ at $t$ , and $w_{b,t}^i$ is the benchmark portfolio's weight in stock $i$ at $t$ .	Holdings vs. benchmark weights, quarterly.	Higher AS $\Rightarrow$ more benchmark deviation (stock selection) and higher skill.
$R^2$ (selectivity: $-R^2$ )	Amihud and Goyenko (2013)	From rolling factor regression $r_{p,t} = \alpha + \beta' f_t + \varepsilon_t$ , we use $-R^2$ as a selectivity proxy.	Monthly, 24m rolling window based on the Carhart 4-factor model.	Lower $R^2$ (higher $R^2$ ) $\Rightarrow$ more idiosyncratic selection/skill.
Active weight (AW)	Doshi et al. (2015)	Deviation from benchmark: $AW_{j,t} = \frac{1}{2} \sum_i  w_{j,t}^i - w_{vwj,t}^i ,$ where $w_{vwj,t}^i$ is the fund $j$ 's weight in stock $i$ at $t$ , and $w_{vwj,t}^i$ is the weight of stock $i$ from a value weighted index, merely based on all stocks included in the fund $j$ 's portfolio at $t$ .	Holdings vs. simple cap-weighted investing, quarterly.	Higher AW $\Rightarrow$ higher active tilts and higher skills.
Value-added (VA, dollars)	Berk and van Binsbergen (2015)	Dollars of skill: $VA_t = \alpha_t^{\text{gross}} \times AUM_{t-1},$ where $\alpha_t^{\text{gross}}$ is benchmark relative abnormal return $\times$ lagged total assets under management $AUM_{t-1}$ .	Monthly gross alpha is before fee alpha from the Carhart 4-factor model; lagged AUM is proxied by lagged TNA.	Direct dollar value from skill; higher VA $\Rightarrow$ more investor surplus created and higher skills.

**Table 4.3: Summary Statistics**

Panel A reports mutual fund characteristics and skill measures at the quarterly frequency; panel B reports factor return characteristics at the monthly frequency (1991–2023). Total net assets (TNA) are stated in 2023 dollars. Fund flow is computed from net returns as  $\text{Flow}_{j,t} = [\text{TNA}_{j,t} - \text{TNA}_{j,t-1}(1 + r_{j,t}^{\text{net}})] / \text{TNA}_{j,t-1}$ . Definitions and construction of the skill measures are summarized in Table 4.2. SMB (Small Minus Big) is the return spread between small-cap and large-cap portfolios. HML (High Minus Low) is the spread between value (high book to market) and growth (low book to market) firms. UMD (Up Minus Down) is the momentum factor, defined as the return spread between past winners and past losers. CMA (Conservative Minus Aggressive) is the investment factor, the spread between firms with conservative versus aggressive total asset growth. RMW (Robust Minus Weak) is the profitability factor, the return spread between firms with robust and weak operating profitability. All factors are from the Fama–French data library.

	N	Mean	SD	P25	Median	P75
<b>Panel A: Fund characteristics and skill measures (quarterly)</b>						
TNA (billions)	144,347	1.407	5.874	0.095	0.310	0.978
Flow	143,448	−0.001	0.193	−0.015	−0.006	0.005
Industry Concentration (ICI)	144,347	0.048	0.045	0.019	0.037	0.063
Active Share (AS)	119,284	0.775	0.153	0.680	0.806	0.897
Carhart-4 $R^2$	143,448	0.910	0.079	0.884	0.932	0.962
Active Weight (AW)	144,347	0.489	0.161	0.381	0.462	0.562
Value-added (VA, millions)	143,448	−1.969	59.125	−2.396	−0.107	1.373
<b>Panel B: Factors return (monthly, 1991–2023)</b>						
SMB (%)	396	0.162	3.073	−1.782	0.080	1.973
HML (%)	396	0.179	3.306	−1.637	−0.025	1.780
UMD (%)	396	0.391	4.742	−1.422	0.495	2.907
CMA (%)	396	0.208	2.194	−1.078	−0.010	1.380
RMW (%)	396	0.369	2.657	−0.923	0.380	1.422

**Table 4.4: FSG and Future Factor Return Correlations (Unconditional)**

This table reports Pearson correlation coefficients between the fund skill gap (FSG) signal at quarter  $t$  and monthly factor returns at  $t + 1$  from 1991 to 2023 with 132 data points. Funds are sorted by the skill measure listed in the first column with detailed definitions in Table 4.2; Average Skill follows the ASI construction in Section 4.3.2.1. Entries are correlations with  $t$ -statistics in parentheses. Panel A uses beta exposure-based FSG; Panel B uses characteristic-based FSG. Panel B signs Size and Investment to BMS (Big minus Small) and AMC (Aggressive minus Conservative) so that larger size and higher asset growth map to higher values. Factor returns are expressed in percent. No additional controls are included. Significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b><i>Panel A: Factor-Exposure-Based FSG (Bivariate Correlations)</i></b>					
Industry Concentration	0.16* (1.76)	0.26*** (2.92)	0.13 (1.42)	0.09 (1.00)	0.02 (0.26)
Active Share	0.08 (0.94)	0.24*** (2.74)	0.09 (0.97)	0.17 (1.89)	0.14 (1.57)
$R^2$ (Negative)	0.08 (0.91)	0.11 (1.19)	0.09 (1.01)	0.04 (0.46)	0.09 (1.03)
Active Weight	0.05 (0.52)	0.00 (−0.04)	−0.09 (−0.99)	0.12 (1.35)	0.05 (0.54)
Value-added	0.13 (1.50)	0.22** (2.44)	0.05 (0.57)	0.06 (0.68)	0.21** (2.32)
Average Skill	0.09 (0.98)	0.30*** (3.49)	0.06 (0.69)	0.18** (2.02)	0.19** (2.11)
	Size (BMS)	Value (HML)	Momentum (UMD)	Investment (AMC)	Profitability (RMW)
<b><i>Panel B: Characteristic-based FSG (bivariate correlations; Size signed to BMS, Investment to AMC)</i></b>					
Industry Concentration	0.12 (1.33)	0.14 (1.55)	0.10 (1.11)	0.09 (1.03)	−0.14 (−1.52)
Active Share	−0.19** (−2.15)	0.16* (1.75)	0.03 (0.38)	0.23*** (2.63)	−0.05 (−0.57)
$R^2$ (Negative)	−0.01 (−0.08)	0.02 (0.19)	−0.03 (−0.30)	0.00 (−0.05)	−0.10 (−1.09)
Active Weight	0.00 (0.01)	−0.14 (−1.57)	0.11 (1.22)	0.05 (0.59)	−0.03 (−0.28)
Value-added	0.08 (0.94)	0.20** (2.25)	−0.01 (−0.07)	0.05 (0.58)	0.13 (1.51)
Average Skill	0.07 (0.72)	0.16* (1.75)	−0.03 (−0.29)	0.11 (1.18)	−0.13 (−1.50)

**Table 4.5: FSG Predictive Regressions with Controls**

This table reports OLS regression coefficients relating monthly factor returns at  $t + 1$  to the fund skill gap (FSG) measured at quarter  $t$  over 1991–2023 (132 observations). For each factor  $f \in \{HML, SMB, RMW, CMA, UMD\}$  and skill proxy  $S$ , we estimate:  $Ret_{f,t+1} = \beta_0 + \beta_{f,1}^S FSG_t^S + \bar{Ret}_{f,t}^{(12)} + FE_{quarter} + \varepsilon_{t+1}$ , where  $FSG_t^S$  is either the factor exposure-based gap  $FSG_t^{S,\beta_f}$  or the characteristic-based gap  $FSG_t^{S,X_f}$  (with  $X_f$  the sorting characteristic for factor  $f$ , e.g.,  $X_{HML} = B/M$ ). Columns list factors; Funds are sorted by the skill measure listed in the first column, with detailed definitions in Table 4.2; Average Skill follows the ASI construction in Section 4.3.2.1.  $\bar{Ret}_{f,t}^{(12)}$  is the factor's 12-month average return from  $t - 12$  through  $t$ . All regressions include quarter fixed effects; parentheses report Newey–West  $t$ -statistics with four lags. Factor returns are in percent. Bivariate correlations are reported separately in Table 4.4. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b>Panel A: Factor-Exposure-Based FSG (coefficients with <math>t</math>-statistics)</b>					
Industry Concentration	1.08*** (2.88)	1.36** (2.31)	1.90* (1.74)	0.66 (0.82)	0.22 (0.22)
Active Share	0.66 (1.02)	1.54*** (2.74)	1.48 (0.83)	1.10 (1.37)	1.29 (1.10)
$R^2$ (Negative)	0.45 (1.36)	1.05 (1.26)	2.10 (0.95)	0.28 (0.40)	1.19 (0.72)
Active Weight	0.43 (0.92)	-0.03 (-0.02)	0.54 (0.16)	1.82 (1.36)	1.22 (0.83)
Value-Added	0.81* (1.71)	2.07** (2.47)	-0.10 (-0.06)	0.48 (0.80)	1.34** (2.05)
Average Skill	0.55* (1.92)	2.63*** (3.99)	1.44 (0.74)	1.40** (2.08)	1.93 (1.51)
	Size (BMS)	Value (HML)	Momentum (UMD)	Investment (AMC)	Profitability (RMW)
<b>Panel B: Characteristic-based FSG (coefficients with <math>t</math>-statistics; Size signed to BMS, Investment to AMC)</b>					
Industry Concentration	1.12* (1.79)	0.40 (1.03)	0.69 (1.08)	0.35 (1.01)	-0.55 (-1.34)
Active Share	-3.61*** (-3.67)	0.68 (1.45)	-0.04 (-0.06)	1.25*** (2.72)	0.40 (0.60)
$R^2$ (Negative)	0.02 (0.04)	0.34 (0.72)	-0.77 (-1.29)	0.21 (0.40)	-0.77 (-1.29)
Active Weight	0.18 (0.29)	-0.94 (-1.61)	1.43 (0.95)	0.49 (0.76)	0.09 (0.11)
Value-Added	0.65 (0.59)	0.89** (2.38)	0.03 (0.03)	0.27 (0.69)	0.42 (0.98)
Average Skill	0.83 (0.86)	0.74 (1.43)	-0.73 (-1.25)	0.65 (1.21)	-0.86 (-1.49)

**Table 4.6: PCA/PLS Components of FSG Signals and Factor Return Predictability**

This table aggregates the five skill-specific factor exposure based on FSG signals (ICI, AS,  $-R^2$ , AW, VA) for each factor  $f \in \{SMB, HML, UMD, CMA, RMW\}$  using (i) the first principal component,  $PC1_t^f$ , and (ii) a one-component partial least-squares index,  $PLS1_t^f$ , chosen to maximize covariance with next-month factor returns. Predictive content is evaluated by re-estimating equation (46), replacing  $FSG_t^S$  with  $PC1_t^f$  or  $PLS1_t^f$ . Reported are regression coefficients (with Newey–West  $t$ -statistics, lag 4, in parentheses). Rows labeled "Loading for ..." report the weights on each skill-based FSG in the composite. All specifications include  $R\bar{e}t_{f,t}^{(12)}$  and quarter fixed effects. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b>Panel A: PCA (first component)</b>					
PCA coefficient (t)	0.30* (1.91)	0.59*** (2.90)	0.52 (1.31)	0.16 (1.10)	0.48* (0.27)
Loading for PCA–FSG: ICI	0.49	0.61	0.57	0.57	0.56
Loading for PCA–FSG: AS	0.53	0.61	0.63	0.62	0.63
Loading for PCA–FSG: $-R^2$	0.44	0.24	0.37	0.52	0.49
Loading for PCA–FSG: AW	0.52	-0.35	0.35	0.10	-0.04
Loading for PCA–FSG: VA	0.14	0.27	-0.11	-0.06	0.23
<b>Panel B: PLS (one-component composite)</b>					
PLS coefficient (t)	0.45*** (2.65)	0.74*** (3.23)	0.63 (1.24)	0.30* (1.79)	0.52* (1.94)
Loading for PLS–FSG: ICI	0.60	0.61	0.50	0.61	0.29
Loading for PLS–FSG: AS	0.52	0.62	0.73	0.66	0.60
Loading for PLS–FSG: $-R^2$	0.43	0.28	0.53	0.43	0.42
Loading for PLS–FSG: AW	0.51	0.27	0.02	0.34	0.17
Loading for PLS–FSG: VA	0.35	0.41	0.17	0.18	0.66

**Table 4.7: Results for Style Funds with Factor-Exposure-Based FSG**

This table restricts the sample to "style funds"—growth, growth and income, and income funds—identified by CRSP objective codes (2,650 funds). Panel A reports Pearson correlations between the fund skill gap (FSG) measured at quarter  $t$  and monthly factor returns at  $t + 1$  over 1991–2023 (132 observations). Panel B reports OLS coefficients with the same specification as Table 4.5; all regressions include the factor's 12-month average return  $\bar{R}t_{f,t}^{(12)}$  and quarter fixed effects. Funds are sorted by the skill measure listed in the first column with definitions in Table 4.2; Average Skill follows the ASI construction in Section 4.3.2.1. Newey–West  $t$ -statistics (lag 4) are in parentheses for regression results. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b><i>Panel A: Bivariate correlations (style-fund subsample)</i></b>					
Industry Concentration (ICI)	0.17*(1.85)	0.26***(2.97)	0.07 (0.72)	0.14 (1.60)	−0.01 (−0.07)
Active Share (AS)	0.09 (0.96)	0.22**(2.49)	0.10 (1.06)	0.11 (1.19)	0.12 (1.31)
$R^2$ (Negative)	0.11 (1.24)	0.09 (1.05)	0.10 (1.12)	0.04 (0.39)	0.00 (0.03)
Active Weight (AW)	0.09 (0.99)	0.06 (0.70)	0.07 (0.72)	0.15*(1.68)	0.00 (0.02)
Value-Added (VA)	0.11 (1.19)	0.21**(2.42)	0.05 (0.55)	0.09 (1.02)	0.17*(1.89)
Average Skill (ASI)	0.10 (1.14)	0.26***(2.97)	0.05 (0.59)	0.21**(2.32)	0.09 (0.99)
<b><i>Panel B: Predictive regression coefficients (style-fund subsample)</i></b>					
Industry Concentration (ICI)	1.26***(2.62)	1.60**(2.58)	1.54 (1.36)	1.14 (1.36)	0.00 (0.00)
Active Share (AS)	0.73 (0.98)	1.44***(2.74)	1.66 (1.02)	0.77 (0.96)	1.15 (0.95)
$R^2$ (Negative)	0.90*(1.90)	1.23 (1.63)	2.17 (1.01)	0.31 (0.47)	0.21 (0.16)
Active Weight (AW)	0.79 (1.64)	1.15 (0.87)	4.03*(1.78)	2.24**(2.11)	0.07 (0.07)
Value-Added (VA)	0.72 (1.22)	1.85**(2.52)	0.27 (0.19)	0.55 (1.02)	1.08*(1.78)
Average Skill (ASI)	0.86*(1.83)	2.20***(3.70)	1.20 (0.69)	1.61**(2.50)	0.87 (0.80)

**Table 4.8: FSG Predictability Among Growth Funds**

This table restricts the sample to growth-only funds identified by CRSP objective codes (1,990 funds). Panel A reports Pearson correlations between the fund skill gap (FSG) measured at quarter  $t$  and monthly factor returns at  $t + 1$  over 1991–2023 (132 observations). Panel B reports OLS coefficients with the same specification as Table 4.5; all regressions include the factor's 12-month average return  $\bar{R}et_{f,t}^{(12)}$  and quarter fixed effects. Funds are sorted by the skill measure listed in the first column with definitions in Table 4.2; Average Skill follows the ASI construction in Section 4.3.2.1. Newey–West  $t$ -statistics (lag 4) are in parentheses for regression results. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b><i>Panel A: Bivariate correlations (growth-only subsample)</i></b>					
Industry Concentration (ICI)	0.19** (2.16)	0.27*** (3.12)	0.05 (0.51)	0.15* (1.67)	0.01 (0.06)
Active Share (AS)	0.08 (0.93)	0.21** (2.38)	0.08 (0.87)	0.13 (1.46)	0.10 (1.14)
$R^2$ (Negative)	0.12 (1.29)	0.11 (1.18)	0.10 (1.10)	0.05 (0.53)	−0.02 (−0.24)
Active Weight (AW)	0.10 (1.12)	0.03 (0.35)	0.02 (0.22)	0.18** (1.99)	0.01 (0.09)
Value-Added (VA)	0.12 (1.38)	0.20** (2.26)	0.07 (0.82)	0.12 (1.36)	0.17* (1.93)
Average Skill (ASI)	0.14 (1.60)	0.25*** (2.82)	0.06 (0.66)	0.20** (2.30)	0.10 (1.14)
<b><i>Panel B: Predictive regression coefficients (growth-only subsample)</i></b>					
Industry Concentration (ICI)	1.38*** (3.26)	1.62*** (2.76)	1.33 (1.07)	1.33 (1.46)	0.10 (0.11)
Active Share (AS)	0.67 (0.98)	1.39*** (2.70)	1.18 (0.77)	1.01 (1.26)	0.90 (0.86)
$R^2$ (Negative)	0.88* (1.86)	1.27* (1.77)	2.01 (1.04)	0.44 (0.66)	0.02 (0.02)
Active Weight (AW)	0.73* (1.80)	−0.32 (−0.46)	2.16 (1.12)	2.51** (2.24)	0.03 (0.03)
Value-Added (VA)	0.84 (1.28)	1.85** (2.16)	0.78 (0.45)	0.80 (1.49)	1.11* (1.80)
Average Skill (ASI)	1.04** (2.48)	2.03*** (3.49)	1.24 (0.76)	1.69*** (2.62)	0.99 (0.88)

**Table 4.9: FSG Predictability Among Cap-Targeted Funds**

This table restricts the sample to capitalization-targeted funds identified by CRSP objective codes (1,066 funds). Panel A reports Pearson correlations between the fund skill gap (FSG) measured at quarter  $t$  and monthly factor returns at  $t + 1$  over 1996–2023 (110 observations). Panel B reports OLS coefficients with the same specification as Table 4.5; all regressions include the factor's 12-month average return  $\bar{R}et_{f,t}^{(12)}$  and quarter fixed effects. Funds are sorted by the skill measure listed in the first column with definitions in Table 4.2; Average Skill follows the ASI construction in Section 4.3.2.1. Newey–West  $t$ -statistics (lag 4) are in parentheses for regression results. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b><i>Panel A: Bivariate correlations (cap-based subsample)</i></b>					
Industry Concentration (ICI)	0.13 (1.34)	0.15 (1.59)	−0.04 (−0.45)	0.01 (0.12)	−0.04 (−0.38)
Active Share (AS)	−0.01 (−0.10)	0.23** (2.49)	−0.04 (−0.39)	0.05 (0.54)	−0.01 (−0.06)
$R^2$ (Negative)	0.13 (1.39)	0.00 (−0.03)	0.16 (1.65)	−0.05 (−0.55)	0.15 (1.53)
Active Weight (AW)	−0.04 (−0.40)	−0.04 (−0.47)	−0.13 (−1.32)	−0.09 (−0.98)	0.01 (0.08)
Value-Added (VA)	−0.07 (−0.70)	−0.03 (−0.29)	−0.04 (−0.42)	0.08 (0.83)	0.16* (1.70)
Average Skill (ASI)	0.06 (0.59)	0.02 (0.18)	−0.17* (−1.79)	0.03 (0.27)	0.10 (1.07)
<b><i>Panel B: Predictive regression coefficients (cap-based subsample)</i></b>					
Industry Concentration (ICI)	1.47* (1.75)	0.54 (1.05)	−0.10 (−0.07)	−0.09 (−0.11)	−0.42 (−0.57)
Active Share (AS)	−0.18 (−0.21)	3.33* (1.96)	−2.35 (−0.79)	0.10 (0.06)	−0.56 (−0.41)
$R^2$ (Negative)	1.63* (1.68)	0.23 (0.21)	2.58 (1.22)	−0.70 (−0.70)	1.83 (1.39)
Active Weight (AW)	−0.09 (−0.07)	−0.40 (−0.41)	−4.95 (−1.16)	−1.38 (−0.70)	0.02 (0.01)
Value-Added (VA)	−0.10 (−0.14)	−0.03 (−0.03)	−1.75 (−0.89)	0.91 (0.88)	1.32* (1.73)
Average Skill (ASI)	0.90 (0.82)	−0.14 (−0.14)	−5.24** (−2.19)	−0.02 (−0.01)	1.07 (0.82)

**Table 4.10: Average Skill Ranks by Fund-Flow Terciles**

Funds are sorted into terciles by net flow each period (f1 = low flow, f3 = high flow). For each tercile, we report the average rank for each skill proxy—Industry Concentration (*ICI*), Active Share (*AS*),  $R^2$ , Active Weight (*AW*), Value-Added (*VA*)—and the composite *average skill rank*. Higher values indicate higher ranks. Definitions of the skill measures are in Table 4.2.

	ICI	AS	$-R^2$	AW	VA	Average skill rank
f1 (Low flow)	5.343	5.726	5.210	5.436	5.482	5.444
f2 (Mid flow)	5.425	5.628	5.357	5.473	5.477	5.475
f3 (High flow)	5.702	5.078	5.885	5.885	5.553	5.549

**Table 4.11: FSG Predictability Excluding Extreme Flow Funds**

This table uses a flow filtered sample that excludes, each quarter, funds in the top quintile of the absolute net flow distribution (most extreme in and outflows). Panel A reports Pearson correlations between the fund skill gap (FSG) measured at quarter  $t$  and monthly factor returns at  $t + 1$  over 1991–2023 (132 observations). Panel B reports OLS coefficients with the same specifications as Table 4.5. Funds are sorted by the skill measure listed in the first column with detailed definitions in Table 4.2; Average Skill follows the ASI construction in Section 4.3.2.1. All regressions include quarter fixed effects. Newey–West  $t$ -statistics (lag 4) are in parentheses for regression results. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b><i>Panel A: Bivariate correlations (flow-filtered sample)</i></b>					
Industry Concentration (ICI)	0.16* (1.76)	0.26*** (2.92)	0.13 (1.42)	0.09 (1.00)	0.02 (0.26)
Active Share (AS)	0.08 (0.94)	0.24*** (2.74)	0.09 (0.97)	0.17 (1.89)	0.14 (1.57)
$R^2$ (Negative)	0.08 (0.91)	0.11 (1.19)	0.09 (1.01)	0.04 (0.46)	0.09 (1.03)
Active Weight (AW)	0.05 (0.52)	0.00 (-0.04)	-0.09 (-0.99)	0.12 (1.35)	0.05 (0.54)
Value-Added (VA)	0.13 (1.50)	0.22** (2.44)	0.05 (0.57)	0.06 (0.68)	0.21** (2.32)
Average Skill (ASI)	0.09 (0.98)	0.30*** (3.49)	0.06 (0.69)	0.18** (2.02)	0.19** (2.11)
<b><i>Panel B: Predictive regression coefficients (flow-filtered sample)</i></b>					
Industry Concentration (ICI)	0.97** (2.50)	1.42** (2.49)	2.60** (2.17)	0.67 (0.80)	0.18 (0.17)
Active Share (AS)	0.66 (1.04)	1.67*** (2.92)	1.48 (0.91)	1.43* (1.88)	1.36 (1.19)
$R^2$ (Negative)	0.48 (1.47)	1.26 (1.63)	1.57 (0.75)	0.33 (0.48)	1.52 (0.90)
Active Weight (AW)	0.43 (0.98)	0.23 (0.18)	-1.76 (-0.50)	2.22 (1.60)	0.90 (0.57)
Value-Added (VA)	0.81 (1.52)	1.96** (2.29)	0.32 (0.19)	0.41 (0.70)	1.36** (2.13)
Average Skill (ASI)	0.45 (1.44)	2.53*** (4.01)	1.29 (0.66)	1.43** (2.12)	1.96 (1.52)

**Table 4.12: Results of the Exposure Turnover-Filtered Subsample**

This table restricts the sample to funds that actively adjust portfolio-level factor exposures. For each fund-factor pair  $(j, f)$ , Beta-Exposure Activeness (BEA) over the past 36 months (12 quarters) is  $\text{sd} \left[ \frac{\beta_{j,f,h} - \beta_{j,f,h-1}}{0.5(|\beta_{j,f,h}| + |\beta_{j,f,h-1}|)} \right]_{h=t-11}^t$ , requiring at least two observations. Within each factor, funds above the cross-sectional median of BEA form the active-beta group. Panel A reports Pearson correlations between the fund skill gap (FSG) at quarter  $t$  and monthly factor returns at  $t + 1$ . Panel B reports OLS coefficients with the same specification as Table 4.5; all regressions include the factor's 12-month average return  $\bar{R}_{f,t}^{(12)}$  and quarter fixed effects. Columns list factors; Size and Investment are signed to BMS and AMC. Newey–West  $t$ -statistics (lag 4) are in parentheses for regression results. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

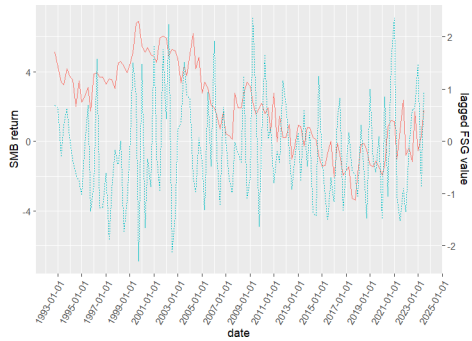
	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b>Panel A: Bivariate correlations (active-beta subsample)</b>					
Industry Concentration (ICI)	0.12 (1.36)	0.23*** (2.65)	0.00 (0.03)	0.08 (0.84)	-0.01 (-0.07)
Active Share (AS)	0.07 (0.75)	0.19** (2.14)	0.02 (0.23)	0.08 (0.89)	0.09 (1.00)
$R^2$ (Negative)	0.13 (1.45)	0.14 (1.52)	0.07 (0.78)	0.12 (1.37)	0.08 (0.87)
Active Weight (AW)	0.09 (0.97)	0.03 (0.31)	0.10 (1.06)	-0.01 (-0.16)	0.05 (0.53)
Value-Added (VA)	0.12 (1.32)	0.20** (2.23)	0.03 (0.36)	0.09 (1.01)	0.23*** (2.66)
Average Skill (ASI)	0.11 (1.22)	0.25*** (2.88)	0.07 (0.79)	0.13 (1.48)	0.17* (1.90)
<b>Panel B: Predictive regression coefficients (active-beta subsample)</b>					
Industry Concentration (ICI)	0.93 (1.65)	1.74** (2.41)	0.99 (0.53)	0.53 (0.66)	-0.11 (-0.09)
Active Share (AS)	0.49 (0.76)	1.80** (2.15)	0.27 (0.21)	0.62 (0.71)	0.75 (0.58)
$R^2$ (Negative)	1.26* (1.90)	1.94** (2.02)	1.90 (0.75)	1.19* (1.69)	1.17 (0.67)
Active Weight (AW)	1.12 (1.41)	0.70 (0.55)	4.11 (1.58)	0.05 (0.04)	0.82 (0.62)
Value-Added (VA)	1.28 (1.61)	2.73*** (2.03)	0.04 (0.02)	0.81 (1.14)	2.13** (2.42)
Average Skill (ASI)	0.99* (1.66)	2.67*** (2.66)	2.10 (1.00)	1.11 (1.54)	2.04 (1.27)

## Appendix to Chapter 4

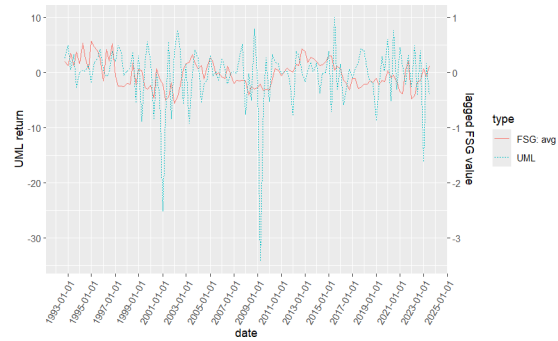
### Appendix 4A.1: Time-series comovement between factor returns and FSG

**Figure 4A.1: Comovement of Factor Returns with Lagged Average-Skill FSG**

Panels (a)–(d) correspond to Size (SMB), Momentum (UMD), Investment (CMA), and Profitability (RMW), respectively. The blue dashed series plots the factor return  $Ret_{f,t}$  (left y-axis, percent). The red solid series plots the average-skill-based FSG,  $FSG_t^{ASI, \beta_f}$ , constructed as in Section 4.3.2.1 (right y-axis). The horizontal axis is calendar time; observations are quarterly and are connected by lines for visual continuity.



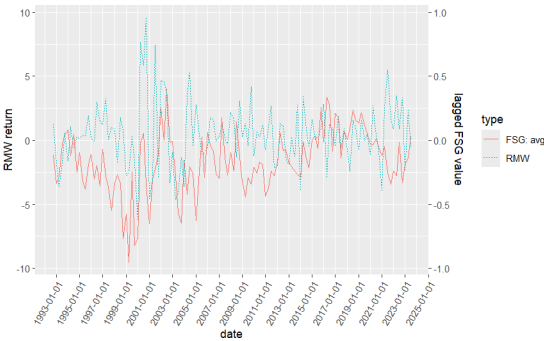
(a) SMB and average skill-based FSG



(b) UMD and average skill-based FSG



(c) CMA and average skill-based FSG



(d) RMW and average skill-based FSG

## Appendix 4A.2: Detailed regression results of baseline results

**Table 4A.1: Factor-Exposure-Based FSG Regression**

This table reports OLS regressions of next month factor returns (in percent) on the fund skill gap (FSG) at  $t - 1$ , where FSG is constructed from characteristics (rather than factor exposures) by sorting funds on the skill proxy indicated by each column: (1) Industry Concentration (*ICI*), (2) Active Share (*AS*), (3)  $-R^2$ , (4) Active Weight (*AW*), (5) Value-Added (*VA*), and (6) the average skill index (*ASI*). Factor returns are: Size (BMS), Value (HML), Momentum (UMD), Investment (AMC), and Profitability (RMW). All specifications control for the past 12-month average return of the corresponding factor,  $\bar{Ret}_{f,t}^{(12)}$  (measured from  $t - 13$  through  $t - 1$ ) and include quarter fixed effects. Newey–West standard errors with four lags are reported in parentheses. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	(1)	(2)	(3)	(4)	(5)	(6)
	ICI	AS	$-R^2$	AW	VA	ASI
<b>Panel A: Dependent variable = Size (SMB)</b>						
$FSG_{t-1}$	1.08*** (0.38)	0.66 (0.64)	0.45 (0.33)	0.43 (0.47)	0.83 (0.53)	0.49 (0.31)
Within $R^2$	0.025	-0.003	-0.003	-0.004	0.008	0.005
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	132	132	132	132	132	132
<b>Panel B: Dependent variable = Value (HML)</b>						
$FSG_{t-1}$	1.36** (0.59)	1.54*** (0.56)	1.05 (0.84)	-0.03 (1.40)	2.07** (0.84)	2.63*** (0.66)
Within $R^2$	0.010	0.054	0.011	-0.006	0.054	0.116
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	132	132	132	132	132	132
<b>Panel C: Dependent variable = Momentum (UMD)</b>						
$FSG_{t-1}$	1.90* (1.09)	1.48 (1.78)	2.10 (2.22)	0.54 (3.51)	-0.10 (1.67)	1.44 (1.94)
Within $R^2$	0.040	0.031	0.035	0.027	0.027	0.031
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	132	132	132	132	132	132
<b>Panel D: Dependent variable = Investment (CMA)</b>						
$FSG_{t-1}$	0.66 (0.80)	1.10 (0.80)	0.28 (0.69)	1.82 (1.34)	0.48 (0.59)	1.40** (0.67)
Within $R^2$	-0.023	-0.016	-0.028	-0.012	-0.024	0.006
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	132	132	132	132	132	132
<b>Panel E: Dependent variable = Profitability (RMW)</b>						
$FSG_{t-1}$	0.22 (1.00)	1.29 (1.18)	1.19 (1.65)	1.22 (1.47)	1.34** (0.65)	1.93 (1.28)
Within $R^2$	0.046	0.060	0.057	0.052	0.075	0.078
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	132	132	132	132	132	132

### Appendix 4A.3: Alternative FSG constructions

**(a) Daily funds return beta estimation:** To estimate factor exposures using daily fund returns, we employ a rolling window of three months (approximately 63 trading days) ending at month end  $t$ . To ensure data quality, we apply several filters. First, to mitigate incubation bias, we follow Evans (2010) and remove the first 18 months of return history for each fund, as funds with poor private performance are often closed without reporting. Second, we exclude fund month observations in which total net assets at the end of the previous month fall below \$5 million, in order to avoid biases arising from liquidity constraints. Third, to eliminate data entry errors, we set daily returns to missing if their absolute value exceeds 300 percent. Finally, to reduce the influence of extreme outliers without discarding valid observations, we apply cross-sectional winsorization at each time point following Bali et al. (2016), setting the lower and upper bounds at the 0.5th and 99.5th percentiles, respectively.

To mitigate the downward bias in beta estimates caused by non-synchronous trading in the underlying holdings of mutual funds, we follow Dimson (1979) and include a lagged term for the market factor. We strictly restrict the Dimson correction to the market factor only, as applying lags to all factors would substantially inflate the parameter space and sharply reduce degrees of freedom in short rolling windows. Therefore, the final regression specification for day  $d$  is given by:  $R_{i,d} - R_{f,d} = \alpha_{i,t} + \beta_{MKT}^0 MKT_d + \beta_{MKT}^1 MKT_{d-1} + \sum_{k \in F} \beta_{i,k} F_{k,d} + \epsilon_{i,d}$ , where  $\mathbf{F} = \{SMB, HML, RMW, CMA, UMD\}$  represents the set of synchronous factors. The total market beta is defined as the sum of the synchronous and lagged coefficients ( $\beta_{MKT}^0 + \beta_{MKT}^1$ ), while other factor betas are the synchronous coefficients. We require a minimum of 50 valid daily return observations within the window ( $N \geq 50$ ) to ensure statistical stability.

**(b) Monthly funds return beta estimation:** For the monthly specification, we estimate betas using a 36-month rolling window. To ensure that our analysis is robust to data errors and biases

inherent in mutual fund databases, we apply the same filtering protocol used in the daily specification. After data processing, we estimate factor betas using a procedure consistent with our baseline stock level beta estimation, as described in Section 4.3.1. We require at least 24 valid monthly return observations within the 36-month window to ensure reliable estimation. Funds that do not meet this threshold are treated as missing for month  $t$ .

**Table 4A.2: Factor Return Predictability Using Daily-Return Beta-Based FSG**

This table constructs the fund skill gap (FSG) using fund betas estimated from daily raw returns (rolling 3-month window; Dimson correction for the market; minimum 50 valid trading days per window), then tests predictability for monthly factor returns. The sample spans January 1998 through December 2023 and comprises 306 monthly observations. Panel A reports univariate predictive regression (no controls). Panel B adds the factor's 12-month average return ( $\bar{R}et_{f,t}^{(12)}$ ) and quarter fixed effects. Newey–West  $t$ -statistics (lag 4) are shown in parentheses. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b><i>Panel A: Predictive regressions without controls (daily-beta FSG)</i></b>					
Industry Concentration (ICI)	0.12** (2.04)	0.10* (1.80)	0.08 (1.45)	0.01 (0.14)	0.03 (0.45)
Active Share (AS)	0.10* (1.69)	0.09* (1.66)	0.06 (1.04)	0.01 (0.13)	-0.04 (-0.65)
$R^2$ (Negative)	0.12** (2.10)	0.12** (2.19)	0.05 (0.86)	0.05 (0.86)	0.11** (2.00)
Active Weight (AW)	0.14** (2.43)	0.01 (0.18)	0.09 (1.55)	0.05 (0.83)	-0.08 (-1.31)
Value-Added (VA)	0.00 (0.06)	0.11** (1.99)	0.03 (0.54)	0.11* (1.94)	0.15*** (2.58)
Average Skill (ASI)	0.16*** (2.76)	0.15*** (2.65)	0.09 (1.60)	0.09* (1.65)	0.11** (1.97)
<b><i>Panel B: Predictive regressions with controls (daily-beta FSG)</i></b>					
Industry Concentration (ICI)	0.31** (2.19)	0.18 (1.29)	0.32 (1.51)	0.01 (0.05)	0.05 (0.33)
Active Share (AS)	0.57* (1.80)	0.18 (1.17)	0.22 (0.84)	0.00 (0.01)	-0.05 (-0.32)
$R^2$ (Negative)	0.30** (2.55)	0.20 (1.27)	0.18 (0.66)	0.01 (0.09)	0.25* (1.73)
Active Weight (AW)	0.36** (2.49)	-0.05 (-0.24)	0.43 (1.16)	0.09 (0.65)	-0.22 (-1.00)
Value-Added (VA)	0.01 (0.04)	0.16 (1.58)	-0.01 (-0.10)	0.08 (1.31)	0.19** (2.51)
Average Skill (ASI)	0.51*** (3.30)	0.24* (1.74)	0.25 (1.07)	0.08 (0.87)	0.19 (1.38)

**Table 4A.3: Factor Return Predictability Using Monthly-Return Beta-Based FSG**

This table constructs the fund skill gap (FSG) using fund betas estimated from monthly raw returns (rolling 36-month window, minimum 24 valid records per window), then tests predictability for monthly factor returns. The dataset provides 399 monthly observations ending in December 2023. Panel A reports univariate predictive regressions (no controls). Panel B adds the factor's 12-month average return ( $\bar{R}et_{f,t}^{(12)}$ ) and quarter fixed effects. Newey–West  $t$ -statistics (lag 4) are shown in parentheses. Factor returns are in percent. Significance: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .10$ .

	Size (SMB)	Value (HML)	Momentum (UMD)	Investment (CMA)	Profitability (RMW)
<b><i>Panel A: Predictive regressions without controls (monthly-beta FSG)</i></b>					
Industry Concentration (ICI)	0.05 (0.91)	0.09* (1.90)	0.08 (1.52)	0.01 (0.14)	0.00 (0.04)
Active Share (AS)	0.04 (0.84)	0.13*** (2.61)	0.05 (0.92)	0.00 (−0.06)	0.00 (−0.08)
$R^2$ (Negative)	0.08 (1.53)	0.18*** (3.62)	−0.08 (−1.69)	0.01 (0.15)	0.12** (2.41)
Active Weight (AW)	0.13** (2.54)	−0.04 (−0.89)	0.09* (1.77)	0.06 (1.16)	0.04 (0.80)
Value-Added (VA)	0.00 (−0.05)	0.11** (2.24)	0.04 (0.90)	0.04 (0.74)	0.16*** (3.14)
Average Skill (ASI)	0.10* (1.92)	0.18*** (3.68)	0.05 (0.96)	0.05 (1.06)	0.19*** (3.85)
<b><i>Panel B: Predictive regressions with controls (monthly-beta FSG)</i></b>					
Industry Concentration (ICI)	0.13 (1.04)	0.16 (1.39)	0.30** (2.35)	−0.01 (−0.08)	0.00 (0.02)
Active Share (AS)	0.16 (0.76)	0.23* (1.92)	0.27 (1.27)	−0.03 (−0.30)	0.01 (0.11)
$R^2$ (Negative)	0.17* (1.89)	0.29** (2.44)	−0.33* (−1.90)	−0.02 (−0.18)	0.24 (1.33)
Active Weight (AW)	0.37** (2.46)	−0.20 (−0.97)	0.54* (1.82)	0.08 (0.58)	0.12 (0.77)
Value-Added (VA)	0.06 (0.46)	0.17* (1.77)	0.07 (0.48)	0.00 (0.01)	0.18*** (3.06)
Average Skill (ASI)	0.29** (2.30)	0.32** (2.56)	0.17 (0.97)	0.02 (0.25)	0.38*** (2.98)

## Chapter 5: General Conclusion

This dissertation examines how information produced and transmitted by people and institutions embedded in real communities and markets is reflected in asset prices. The first two essays establish that racial residential integration fundamentally alters the information environment of financial markets. Empirically, I show that firms located in more racially integrated communities exhibit greater stock price informativeness and lower information asymmetry, driven by improved local information networks that facilitate broader and faster information diffusion. Theoretically, I demonstrate that racial integration functions as a technology shock that unlocks diverse signals, reduces the cost of equity, and promotes a more equitable distribution of market surplus by narrowing the structural welfare gap between majority and minority investors.

The third essay shifts focus to institutional investors and shows that skilled mutual fund managers' portfolio tilts contain forward-looking information about systematic risk factor returns. The fund skill gap measure captures skilled managers' informative assessments of conditional expected factor premia, providing a holdings-based timing signal that complements standard public signals.

Collectively, these findings highlight the importance of social structure in facilitating information flow, the role of institutional investors as information intermediaries, and the distributional consequences of informational frictions in financial markets. They suggest that policies fostering social integration may generate returns beyond social cohesion by improving market efficiency and reducing wealth inequality through financial channels. Future research could explore how other dimensions of social diversity affect information diffusion, how technological changes reshape information networks, and how regulatory interventions can promote more inclusive and efficient financial markets.

## Reference List

- Agarwal, S., Choi, H.-S., He, J., & Sing, T. F. (2019). Matching in housing markets: The role of ethnic social networks. *The Review of Financial Studies*, 32, 3958-4004.
- Agarwal, V., Jiang, W., Luo, Y., & Zou, H. (2025). The real effect of sociopolitical racial/ethnic animus: Mutual fund manager performance during AAPI hate. *Available at SSRN*.
- Ahern, K. R. (2017). Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics*, 125, 26-47.
- Alesina, A., Baqir, R., & Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114, 1243-1284.
- Alesina, A., & La Ferrara, E. (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115, 847-904.
- Ambrose, B. W., Conklin, J. N., & Lopez, L. A. (2021). Does borrower and broker race affect the cost of mortgage credit? *The Review of Financial Studies*, 34, 790-826.
- Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5, 31-56.
- Amihud, Y., & Goyenko, R. (2013). Mutual fund's R2 as predictor of performance. *The Review of Financial Studies*, 26(3), 667-694.

- Amiram, D., Owens, E., & Rozenbaum, O. (2016). Do information releases increase or decrease information asymmetry? New evidence from analyst forecast announcements. *Journal of Accounting and Economics*, 62, 121-138.
- Ananat, E. O. (2011). The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3, 34-66.
- Arnott, R. D., Kalesnik, V., & Linnainmaa, J. T. (2023). Factor momentum. *The Review of Financial Studies*, 36(8), 3034-3070.
- Bai, T., Hilscher, J., & Scherbina, A. (2025). Unencumbered by style: Why do funds change factor loadings, and does it help? *Journal of Banking and Finance*, 181, 107544.
- Baik, B., Kang, J.-K., & Kim, J.-M. (2010). Local institutional investors, information asymmetries, and equity returns. *Journal of Financial Economics*, 97, 81-106.
- Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32, 259-280.
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), 1645-1680.
- Bali, T. G., Engle, R. F., & Murray, S. (2016). Empirical asset pricing: The cross section of stock returns. *John Wiley & Sons*.

- Bali, T. G., Hirshleifer, D., Peng, L., Tang, Y., & Wang, Q. (2021). Social interactions and lottery stock mania. *NBER Working Paper 29543*.
- Barras, L., Scaillet, O., & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance*, *65*(1), 179-216.
- Barras, L., Gagliardini, P., & Scaillet, O. (2022). Skill, scale, and value creation in the mutual fund industry. *The Journal of Finance*, *77*(1), 601-638.
- Berk, J. B., & Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, *112*(6), 1269-1295.
- Berk, J. B., & van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, *118*(1), 1-20.
- Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research*, *27*, 1-36.
- Bernile, G., Korniotis, G., Kumar, A., & Wang, Q. (2015). Local business cycles and local liquidity. *Journal of Financial and Quantitative Analysis*, *50*, 987-1010.
- Bernile, G., Bhagwat, V., & Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of Financial Economics*, *127*, 588-612.
- Bernile, G., Kumar, A., Sulaeman, J., & Wang, Q. (2019). Has local informational advantage disappeared? *The Review of Financial Economics*, *37*, 38-60.

- Bertrand, M., Luttmer, E. F. P., & Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, *115*, 1019-1055.
- Bhattacharya, U., & Daouk, H. (2002). The world price of insider trading. *The Journal of Finance*, *57*, 75-108.
- Bird, A., Karolyi, G. A., & Ruchti, J. (2023). How do firms respond to political uncertainty? Evidence from U.S. gubernatorial elections. *Journal of Accounting Research*, *61*, 1185-1230.
- Boehmer, E., & Kelley, E. K. (2009). Institutional investors and the informational efficiency of prices. *The Review of Financial Studies*, *22*, 3563-3594.
- Boehmer, E., Jones, C. M., & Zhang, X. (2013). Shackling short sellers: The 2008 shorting ban. *The Review of Financial Studies*, *26*, 1363-1400.
- Boone, A. L., & White, J. T. (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, *117*, 508-533.
- Brown, S. J., & Goetzmann, W. N. (1997). Mutual fund styles. *Journal of Financial Economics*, *43*, 373-399.
- Brown, S. J., Sotes-Paladino, J., Wang, J., & Yao, Y. (2017). Starting on the wrong foot: Seasonality in mutual fund performance. *Journal of Banking and Finance*, *82*, 133-150.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57-82.

- Chan, L. K. C., Chen, H.-L., & Lakonishok, J. (2002). On mutual fund investment styles. *The Review of Financial Studies*, 15, 1407-1437.
- Charles, C. Z. (2003). The dynamics of racial residential segregation. *Annual Review of Sociology*, 29, 167-207.
- Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129, 1553-1623.
- Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). The opportunity atlas: Mapping the childhood roots of social mobility. *Dataset, Opportunity Insights*.
- Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., & The Opportunity Insights Team. (2022a). Social capital I: Measurement and associations with economic mobility. *Nature*, 608, 108-121.
- Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., & The Opportunity Insights Team. (2022b). Social capital II: Determinants of economic connectedness. *Nature*, 608, 122-134.
- Chetty, R., Dobbie, W., Goldman, B., Porter, S. R., & Yang, C. S. (2024). Changing opportunity: Sociological mechanisms underlying growing class gaps and shrinking race gaps in economic mobility. *NBER Working Paper No. 32697*.
- Cho, S., & Lucio, J. (2025). Vouchers welcome here: Source of income antidiscrimination laws in American cities. *Housing Policy Debate*.

- Christoffersen, S. E. K., & Sarkissian, S. (2009). City size and fund performance. *Journal of Financial Economics*, 92, 252-275.
- Chuprinin, O., & Sosyura, D. (2018). Family descent as a signal of managerial quality: Evidence from mutual funds. *The Review of Financial Studies*, 31, 3756-3820.
- Chyn, E., Haggag, K., & Stuart, B. A. (2022). The effects of racial segregation on intergenerational mobility: Evidence from historical railroad placement. *American Economic Journal: Applied Economics*, Forthcoming.
- Cohen, L., Frazzini, A., & Malloy, C. (2008). The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy*, 116(5), 951-979.
- Connor, G., & Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *The Journal of Finance*, 48(4), 1263-1291.
- Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54, 2045-2073.
- Coval, J. D., & Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, 109, 811-841.
- Coval, J., & Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2), 479-512.

- Cremers, K. M., & Petajisto, A. (2009). How active is your fund manager? A new measure that predicts performance. *The Review of Financial Studies*, 22(9), 3329-3365.
- Csaszar, F. A. (2012). Organizational structure as a determinant of performance: Evidence from mutual funds. *Strategic Management Journal*, 33, 611-632.
- Cui, J., Jo, H., & Na, H. (2018). Does corporate social responsibility affect information asymmetry? *Journal of Business Ethics*, 148, 549-572.
- Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, 52(3), 1035-1058.
- Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance*, 52(1), 1-33.
- Davis, J. C., & Henderson, J. V. (2008). The agglomeration of headquarters. *Regional Science and Urban Economics*, 38, 445-460.
- Derenoncourt, E. (2022). Can you move to opportunity? Evidence from the Great Migration. *The American Economic Review*, 112, 369-408.
- Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, liquidity, and the cost of capital. *The Journal of Finance*, 46, 1325-1359.
- Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7(2), 197-226.

- Doshi, H., Elkamhi, R., & Simutin, M. (2015). Managerial activeness and mutual fund performance. *The Review of Asset Pricing Studies*, 5(2), 156-184.
- Dou, W., Kogan, L., & Wu, W. (2023). Common fund flows: Flow hedging and factor pricing. *The Journal of Finance*, Forthcoming.
- Durbin, J. (1954). Errors in variables. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 22, 23-32.
- Dyer, T., Köchling, G., & Limbach, P. (2024). Traditional investment research and social networks: Evidence from Facebook connections. *Available at SSRN*.
- Ehsani, S., & Linnainmaa, J. T. (2022). Factor momentum and the momentum factor. *The Journal of Finance*, 77(3), 1877-1919.
- Ellis, J., Smith, J., & White, R. (2020). Corruption and corporate innovation. *Journal of Financial and Quantitative Analysis*, 55, 2124-2149.
- Evans, R. B. (2010). Mutual fund incubation. *The Journal of Finance*, 65(4), 1581-1611.
- Falkenstein, E. G. (1996). Preferences for stock characteristics as revealed by mutual fund portfolio holdings. *The Journal of Finance*, 51(1), 111-135.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22.
- Fama, E. F., & French, K. R. (2020). Comparing cross-section and time-series factor models. *The Review of Financial Studies*, *33*(5), 1891-1926.
- Fernandez, R. M., & Fernandez-Mateo, I. (2006). Networks, race, and hiring. *American Sociological Review*, *71*, 42-71.
- Ferson, W. E., & Schadt, R. W. (1996). Measuring fund strategy and performance in changing economic conditions. *The Journal of Finance*, *51*(2), 425-461.
- Frazzini, A., & Lamont, O. A. (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics*, *88*(2), 299-322.
- Furutani, S., Shibahara, T., Akiyama, M., & Aida, M. (2023). Analysis of homophily effects on information diffusion on social networks. *IEEE Access*, *11*, 79974-79983.
- Gabaix, X., & Koijen, R. S. (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis. *Technical Report, National Bureau of Economic Research*.
- Galvez, M. (2010). What do we know about housing choice voucher program location outcomes? A review of recent literature. *Urban Institute*.
- George, T. J., Kaul, G., & Nimalendran, M. (1991). Estimation of the bid-ask spread and its components: A new approach. *The Review of Financial Studies*, *4*, 623-656.

- Golub, B., & Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, *127*, 1287-1338.
- Greenwood, R., & Hanson, S. G. (2012). Share issuance and factor timing. *The Journal of Finance*, *67*(2), 761-798.
- Grossman, S. J., & Stiglitz, J. E. (1976). Information and competitive price systems. *The American Economic Review*, *66*, 246-253.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*, 393-408.
- Haddad, V., Kozak, S., & Santosh, S. (2020). Factor timing. *The Review of Financial Studies*, *33*(5), 1980-2018.
- Halberstam, Y., & Knight, B. G. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, *143*, 73-88.
- Han, L., Huang, X., Kadan, O., & Wu, J. (2021). The true colors of money: Racial diversity and asset management. *Available at SSRN*.
- Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, *46*, 1251-1271.
- Hinz, O., & Spann, M. (2008). The impact of information diffusion on bidding behavior in secret reserve price auctions. *Information Systems Research*, *19*, 351-368.

- Hirshleifer, D., Peng, L., & Wang, Q. (2025). News diffusion in social networks and stock market reactions. *The Review of Financial Studies*, 38, 883-937.
- Hong, H., Kubik, J. D., & Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance*, 60, 2801-2824.
- Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, 28(3), 791-837.
- Huang, S., Song, Y., & Xiang, H. (2024). Noise trading and asset pricing factors. *Management Science*.
- Jackson, M. O., & Rogers, B. W. (2007). Relating network structure to diffusion properties through stochastic dominance. *B.E. Journal of Theoretical Economics*, 7, Article 6.
- Jahn, J., Schmid, C. F., & Schrag, C. (1947). The measurement of ecological segregation. *American Sociological Review*, 12, 293-303.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65-91.
- Jiang, H., & Sun, Z. (2014). Dispersion in beliefs among active mutual funds and the cross-section of stock returns. *Journal of Financial Economics*, 114, 341-365.
- Jiang, H., & Zheng, L. (2018). Active fundamental performance. *The Review of Financial Studies*, 31(12), 4688-4719.

- Kacperczyk, M., Sialm, C., & Zheng, L. (2005). On the industry concentration of actively managed equity mutual funds. *The Journal of Finance*, *60*(4), 1983-2011.
- Kacperczyk, M., Sialm, C., & Zheng, L. (2008). Unobserved actions of mutual funds. *The Review of Financial Studies*, *21*(6), 2379-2416.
- Kacperczyk, M., Nieuwerburgh, S. V., & Veldkamp, L. (2014). Time-varying fund manager skill. *The Journal of Finance*, *69*, 1455-1484.
- Kacperczyk, M., Sundaresan, S., & Wang, T. (2021). Do foreign institutional investors improve price efficiency? *The Review of Financial Studies*, *34*, 1317-1367.
- Kain, J. F. (2004). A pioneer's perspective on the spatial mismatch literature. *Urban Studies*, *41*, 7-32.
- Kelly, B. T., & Ljungqvist, A. (2007). The value of research. *EFA 2008 Athens Meetings Paper*.
- Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, *68*(5), 1721-1756.
- Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*.
- Kim, D., Wang, Q., & Wang, X. (2022). Geographic clustering of institutional investors. *Journal of Financial Economics*, *144*, 547-570.

- Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*, 97-126.
- Kozak, S., & Nagel, S. (2024). When do cross-sectional asset pricing factors span the stochastic discount factor? *Available at SSRN 4443643*.
- Kuchler, T., Li, Y., Peng, L., Stroebel, J., & Zhou, D. (2022). Social proximity to capital: Implications for investors and firms. *The Review of Financial Studies*, *35*, 2743-2789.
- Kumar, A., Niessen-Ruenzi, A., & Spalt, O. G. (2015). What's in a name? Mutual fund flows when managers have foreign-sounding names. *The Review of Financial Studies*, *28*, 2281-2321.
- Lettau, M., & Ludvigson, S. (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, *109*(6), 1238-1287.
- Lettau, M., & Pelger, M. (2020). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies*, *33*(5), 2274-2325.
- Leuz, C., & Verrecchia, R. E. (2000). The economic consequences of increased disclosure. *Journal of Accounting Research*, *38*, 91-124.
- Li, J. (2022). What drives the size and value factors? *Review of Asset Pricing Studies*, *12*(4), 845-885.
- Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *Review of Corporate Finance Studies*, *8*, 1-41.

- Lichter, D. T., Parisi, D., & Taquino, M. C. (2015). Toward a new macro-segregation? Decomposing segregation within and between metropolitan cities and suburbs. *American Sociological Review*, *80*, 843-873.
- Lin, N. (2002). Social capital: A theory of social structure and action. *Cambridge University Press*.
- Lof, M., & van Bommel, J. (2023). Asymmetric information and the distribution of trading volume. *Journal of Corporate Finance*, *82*.
- Lou, D. (2012). A flow-based explanation for return predictability. *The Review of Financial Studies*, *25*(12), 3457-3489.
- Lu, Y., Naik, N. Y., & Teo, M. (2022). Race, discrimination, and hedge funds. *Available at SSRN*.
- Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, *67*, 281-315.
- McClure, K. (2010). Housing choice voucher marketing opportunity index: Analysis of data at the tract and block group level. *Working Paper, Urban Institute*.
- McDonald, S., Lin, N., & Ao, D. (2009). Networks of opportunity: Gender, race, and job leads. *Social Problems*, *56*, 385-402.
- McDonald, S., & Day, J. C. (2010). Race, gender, and the invisible hand of social capital. *Sociology Compass*, *4*, 532-543.

- McDonald, S. (2011). What's in the 'old boys' network? Accessing social capital in gendered and racialized networks. *Social Networks*, 33, 317-330.
- McLean, R. D., Pontiff, J., & Reilly, C. (2025). Taking sides on return predictability. *Journal of Financial Economics*, *Forthcoming*.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415-444.
- Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4), 1249-1290.
- Moskowitz, T. J. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses: Discussion. *The Journal of Finance*, 55, 1695-1703.
- Mouw, T., & Entwisle, B. (2006). Residential segregation and interracial friendship in schools. *American Journal of Sociology*, 112, 394-441.
- Nguyen, G., & Nguyen, T. H. (2025). Social networks and venture capital investments around the world. *The British Accounting Review*, 101628.
- Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, 116(1), 23-45.

- Pedulla, D. S., & Pager, D. (2019). Race and networks in the job search process. *American Sociological Review*, 84, 983-1012.
- Peng, C., & Wang, C. (2023). Factor rebalancing. Available at SSRN 3327849.
- Pool, V. K., Stoffman, N., & Yonker, S. E. (2015). The people in your neighborhood: Social interactions and mutual fund portfolios. *The Journal of Finance*, 70, 2679-2732.
- Rajkumar, K., Saint-Jacques, G., Bojinov, I., Brynjolfsson, E., & Aral, S. (2022). A causal test of the strength of weak ties. *Science*, 377, 1304-1310.
- Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, 48, 240-267.
- Saffi, P. A., & Sigurdsson, K. (2011). Price efficiency and short selling. *The Review of Financial Studies*, 24, 821-852.
- Sard, B. (2008). A research note: The effects of housing vouchers on residential integration.
- Shu, T., Sulaeman, J., & Yeung, P. E. (2012). Local religious beliefs and mutual fund risk-taking behaviors. *Management Science*, 58, 1779-1796.
- Sigelman, L., Bledsoe, T., Welch, S., & Combs, M. W. (1996). Making contact? Black-white social interaction in an urban setting. *American Journal of Sociology*, 101, 1306-1332.
- Teles, D., & Su, Y. (2022). Source of income protections and access to low-poverty neighborhoods. *Urban Institute*.

- Theil, H. (1972). *Statistical decomposition analysis. Amsterdam: North Holland.*
- Tolbert, C. M., & Sizer, M. (1996). U.S. commuting zones and labor market areas: A 1990 update. *Staff Paper AGES-9614, U.S. Department of Agriculture Economic Research Service.*
- Trounstein, J. (2016). Segregation and inequality in public goods. *American Journal of Political Science, 60*, 709-725.
- Turner, M. A. (2003). Strengths and weaknesses of the housing voucher program. *Urban Institute.*
- U.S. Census Bureau. (2024). About the topic of race. <https://www.census.gov/topics/population/race/about.html>. Accessed December 20, 2024.
- Wardlaw, M. (2020). Measuring mutual fund flow pressure as a shock to stock returns. *The Journal of Finance, 75*, 3221-3243.
- Wei, C., & Zhang, L. (2020). Trust and local bias. *Journal of Financial and Quantitative Analysis, 55*, 2211-2245.
- Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance, 55(4)*, 1655-1695.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data. Cambridge: MIT Press.*
- Wu, D. M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica, 41*, 733-750.

Zhu, Q. (2020). The missing new funds. *Management Science*, 66(3), 1193-1204.