Essays on Retail Investor Behavior in Financial Markets

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Economics

have examined a dissertation entitled

"Essays on Retail Investor Behavior in Financial Markets"

presented by **Shushu Liang**

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature ______________________________

Typed name: Prof. John Campbell

Signature ______________________________

Typed name: Prof. Xavier Gabaix

Signature ______________________________

Typed name: Prof. Adi Sunderam

Signature ______________________________

Typed name: Prof. Mark Egan

Date: April 19, 2023

Essays on Retail Investor Behavior in Financial Markets

A dissertation presented

by

Shushu Liang

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2023

Dissertation Advisor:
Professor John Campbell

Author:
Shushu Liang

Essays on Retail Investor Behavior in Financial Markets

Abstract

This dissertation examines the trading behavior of retail investors using large datasets and both causal inference and structural estimation methods. The study has two main objectives: to understand the heterogeneity across different investors and to quantify the price impacts of various trading behaviors. Chapter 1 uses data on account-level stock holdings to quantify the importance of different mechanisms contributing to stock booms and busts during the 2015 Chinese stock market bubble. By estimating a structural model of heterogenous investor demand, I find that retail investors account for 78% of the variance in cross-sectional stock returns, but the contribution of different channels varies as the bubble evolves. Chapter 2 studies retail investors' return-chasing behavior leveraging the same data on 18 million individual equity accounts. I find that return chasing predicts both investor returns and stock returns more strongly than various other investor characteristics. Chapter 3 documents the decline in the average profitability of IPOs in the United States between 2010 and 2022. This work provides valuable insights into the drivers of stock price fluctuations and highlights the importance of understanding investor heterogeneity.

Contents

List of Tables

List of Figures

Acknowledgments

I am indebted to my advisors, John Campbell, Xavier Gabaix, Mark Egan, and Adi Sunderam, for their unwavering support, invaluable guidance, and constant encouragement throughout this journey. I am grateful to Sam Hanson, Malcolm Baker, and Jeremy Stein for their thoughtful advice. I would also like to thank Li-An Zhou and Gérard Roland for instilling in me a love for economics that brought me to graduate school.

I am grateful to all the friends who accompanied me these six years. In particular, I want to thank Ayushi Narayan for countless enlightening conversations and for always being a compassionate and supportive friend. I thank Romaine Campbell, Elisa Rubbo, Shuqi Dai, Ruxin Diao, Siqi Wang, and Miles Zheng for their companionship. Their presence was a constant source of comfort and inspiration during the challenging times. I also thank Zixuan Wang and Wentao Xiong for insightful conversations about career paths.

Above all, I owe an immeasurable debt of gratitude to my parents. None of my accomplishments would have been possible without their unfailing love and support. This dissertation stands as a testament to the curiosity that my parents ignited, fostered, and nurtured within me.

To my parents

Introduction

This dissertation focuses on studying the trading behavior of retail investors using large datasets and both causal inference and structural estimation methods. There are two main objectives of this dissertation. First, it highlights the importance of understanding the heterogeneity across different investors to gain insights into the entire distribution of trading patterns. Second, it aims to quantify the price impacts of various trading behaviors to provide a quantitative understanding of the importance of different types of investors.

I will now summarize the main findings of each chapter in sequence.

Chapter 1, "What Drives Stock Prices in a Bubble?" decomposes and quantifies the channels driving the boom and bust cycle of individual stocks. To shed light on the formation, expansion, and deflation of bubbles, I study how the cross section of stocks evolves during the 2015 Chinese stock market bubble. Using data on administrative account-level stock holdings covering a representative sample of 18 million retail investors and all institutional investors, I estimate a structural model of heterogeneous investor demand. The model allows me to attribute variation in stock returns to changing stock characteristics, changing investor preferences or beliefs, and the entrance of new investors. Improved stock fundamentals are initially key, accounting for 21% of the variance in cross-sectional stock returns during the formation of the bubble. In the expansion phase of the bubble, the entrance of new investors plays an important role, explaining 43% of the cross-sectional variance during the phase. Finally, the deflation phase is characterized by shifts in preferences or beliefs among existing retail investors, accounting for 25% of the cross-sectional variance in the period. I highlight the ways in which my structural model quantifies the forces in Kindleberger (1978)'s classic narrative.

In Chapter 2, "Who Chases Returns? Evidence from the Chinese Stock Market," I study retail investors' return-chasing behavior, leveraging the same data on 18 million individual equity accounts. This is the first paper to study the return chasing behavior at the individual investor level. I construct a new measure for an individual's return chasing propensity (RCP). Less sophisticated investors tend to exhibit higher RCP. Investors with a one standard deviation higher RCP earn 5.6% lower yearly returns on average. At the stock level, I construct a measure for a stock's return chasing ownership (RCO) by wealth-weighting its retail holders' RCP. Stocks with the highest RCO suffer 10% lower yearly returns on average. Return chasing predicts both investor returns and stock returns more strongly than various other investor characteristics.

Chapter 3, "Negative-Profit IPOs: Trends, Characteristics, and Investor Perceptions," documents a decline in the average profitability of IPOs in the United States between 2010 and 2022. Negative-profit IPOs are prevalent in the retail, technology, and pharmaceutical industries and are more likely to be backed by venture capital. However, negative-profit IPOs do not typically demonstrate significant improvement in fundamentals over the subsequent five years compared to positive-profit IPOs. While negative-profit IPOs tend to have higher first-day returns, they generally under-perform in the long run on average. Despite this under-performance, negative-profit IPOs have exhibited a fatter right tail in return distributions since 2010, suggesting that investors are becoming more tolerant of negative-profit IPOs potentially due to their increased likelihood of delivering superior returns.

Overall, this dissertation contributes to our understanding of the trading behavior of retail investors and the impact it has on stock prices. This work provides valuable insights into the drivers of stock price fluctuations and highlights the importance of understanding investor heterogeneity.

Chapter 1

What Drives Stock Prices in a Bubble?[1]

1.1 Introduction

Financial markets periodically display sharp price increases followed by sharp declines, which we refer to as "bubbles." Since the earliest recorded Dutch tulip bulb bubble in 1636, the drivers of these phenomena have been of great interest to researchers. The economics literature has hypothesized that various forces can contribute to bubbles: for example, interactions between sophisticated and unsophisticated investors (Abreu and Brunnermeier 2003; Greenwood and Nagel 2009), investors' time-varying beliefs or sentiments (Scheinkman and Xiong 2003; Baker and Wurgler 2006; Barberis et al. 2015; Nagel and Xu 2022), and wealth reallocation among heterogeneous investors (Atmaz and Basak 2018; Martin and Papadimitriou 2022a). There are also influential historical narratives about the drivers of bubbles. Notably, Kindleberger (1978) describes the phases of a typical bubble, which starts with good fundamental news (displacement), goes through a boom fueled by speculative trading and the entrance of new investors, and finally ends in a panic. While the literature and historical narratives have listed possible bubble mechanisms, a key remaining question is: how much do these various channels matter?

To understand the relative contribution of different channels, we focus on the cross

1. Co-authored with Weihua Chen and Donghui Shi

section of stocks for better identification and empirical implementation.[2] We ask: during an aggregate bubble episode, which types of stocks experience larger booms and busts? Which types of investors hold these stocks? Where do differences in cross-sectional stock returns come from? Even though the drivers of bubbles in the cross section need not be the same as those that lead to aggregate bubbles, cross-sectional evidence can build our understanding of aggregate macroeconomic events (Mian and Sufi 2009; Nakamura and Steinsson 2014; Pflueger, Siriwardane, and Sunderam 2020; Chodorow-Reich, Nenov, and Simsek 2021). We find that the drivers of aggregate bubbles hypothesized in the literature are also influential in generating cross-sectional stock boom-busts.[3]

We focus our analysis on the 2015 Chinese stock market bubble for several reasons. To begin with, we are able to obtain rich administrative account-level stock holdings and transactions records from the Shanghai Stock Exchange to facilitate our analysis. Our data cover a representative sample of 18 million retail investors and all institutional investors from 2011 to 2019 at a monthly frequency. These data enable us to systematically analyze every market player's behavior during an influential stock market bubble episode for the first time. Second, the 2015 Chinese stock market bubble features a rapid price increase and decrease, which is typical of stock price bubbles (the market price more than doubled and then crashed to half of its peak value within one and a half years). Third, this bubble is also considered to be a noteworthy event due to its scale and subsequent impact on global financial markets. At the time, the Shanghai Stock Exchange was the fifth largest stock exchange in the world. Finally, the Chinese stock market is a particularly good environment to study the price impacts of retail investors, who account for 84% of trading volume and hold 60% of floating market capitalization on average. The effect of retail investors may be of particular interest to other stock markets that have experienced an increase in retail involvement, such as the U.S. stock market since 2020.

2. Identifying causal links in the aggregate time series is challenging. Few historical bubble episodes have occurred, and it is hard to control for all relevant variables across different time periods.

3. To clarify the terminology, in this paper, "bubble, formation, expansion, and deflation" refer to phases of the aggregate bubble episode; "boom-bust, boom, and bust" refer to cross-sectional stock price paths.

We start with simple reduced-form regressions to understand which stock characteristics are associated with larger boom-busts in the cross-section. Of the 904 stocks, we classify 157 boom-bust stocks, which are defined as having both booms and busts above the median in the cross-section. We show that these boom-bust stocks are more popular among retail investors than among institutions. They have distinctive "boom-bust characteristics": they are less profitable, have a lower dividend, have a higher beta, are smaller, are younger, and are less likely to be state-owned enterprises (SOE).

Next, we show suggestive evidence that retail investors, especially new market entrants during the expansion phase of the bubble, can influence the price paths of boom-bust stocks. Retail net capital inflows during the expansion phase of the bubble account for 25% of pre-bubble market capitalization for boom-bust stocks. The number of new retail investors entering the market during the bubble peak was 7100% larger than the number of entering investors during the pre-bubble period.

To understand the quantitative significance of boom-bust characteristics and investor heterogeneity, we build a structural model that captures investors' heterogeneous demand for stocks. Following Koijen and Yogo (2019), we model investor demand for stocks with a factor structure, using stock characteristics as factors. We estimate the demand function for every investor type at each month, allowing for heterogeneity among different investors and over different time periods.

To identify investor demand elasticity, we leverage an institutional feature of the Chinese stock market: China is distinctive in having large institutional investors with extremely stable mandates (investment universe). We refer to these institutions as stable institutions, which are National Social Security Funds, Corporate Supplementary Pension Funds, and Qualified Foreign Institutional Investors. These institutions "aim for long-run stable growth, rather than short-term profits," as stated in their prospectus. We construct an instrument for a stock's price by considering whether the stock is held by stable institutions. Our instrument is in the spirit of the index inclusion literature (Harris and Gurel 1986; Shleifer 1986), which suggests that stocks included in the mandates of stable institutions have price pressure that is exogenous to retail investors' latent demand. This instrument builds and improves upon the instrument in Koijen and Yogo (2019) since stable institutions' mandates and flows are more likely to be exogenous to other investors' latent demand.

The demand estimation results show that retail investors have a high demand for boom-bust characteristics. Among retail investors, the ones with less wealth or the ones who enter the stock market during the bubble's expansion phase have an especially high demand for boom-bust characteristics. Institutions demand stocks with characteristics opposite of those demanded by retail investors, but are inelastic in their demand, which in turn limits the offsetting of retail demand.

Based on the estimated asset demand system, we decompose the variance in cross-sectional stock returns into changes stemming from stock fundamentals, investor types, and investor behaviors. Over the entire bubble period, retail investors contribute 78% of the variance in cross-sectional stock returns. Different elements play varying roles during different bubble phases. Improved stock fundamentals are initially key, accounting for 21% of the variance in cross-sectional stock returns during the formation of the bubble. This echoes Kindleberger's narrative that every bubble starts with some good fundamental news (displacement). However, stock fundamentals do not matter during the expansion and deflation phases of the bubble, when the price is decoupled from fundamentals. During the expansion phase, investors who newly enter the market play a significant role, explaining 43% of cross-sectional stock returns. The quote from Kindleberger (1978) depicts this period vividly:

> "*A follow-the-leader process develops as households see that others are profiting from speculative purchases. There is nothing as disturbing to one's well-being and judgment as to see a friend get rich.*"

Finally, during the deflation phase of the bubble, the shift in preferences or beliefs about stock characteristics among existing retail investors becomes more important, accounting for 25% of the stock return variance. We cannot disentangle investors' preferences or beliefs, so we refer to them interchangeably. One example of a shift in preferences/beliefs could be a reduced focus on profitability during the expansion of the bubble due to manias and speculative trading which then changes to a flight to quality when the panic hits.

After the variance decomposition, we conduct counterfactual analyses regarding the two most important channels contributing to stock boom-busts: the shift in preferences/beliefs among existing investors and new investor entry. In the first counterfactual exercise, we fix investors' preferences/beliefs over the bubble period. The counterfactual price path for boom-bust stocks is substantially flattened: the counterfactual cumulative return at the peak of the boom-bust cycle is reduced by 41%. We then carry out a second counterfactual exercise to exclude retail investors who entered during the expansion phase of the bubble. In this counterfactual scenario, prices for boom-bust stocks decrease during the expansion phase since existing investors think that prices are too high and decide to sell these stocks. In actuality, prices for boom-bust stocks kept increasing due to the massive inflows of new entrants, which overpowered the selling pressure from existing investors. If there had not been any late entrants, the counterfactual cumulative return at the peak of the boom-bust cycle would be reduced by 50%.

Related literature

There are many papers seeking to better understand bubbles. Historical narratives include Graham and Zweig (1973), Kindleberger (1978), and Brown (1991). Economists have tried to explain bubbles both empirically and theoretically (see a survey paper by Brunnermeier and Oehmke 2013). The contribution of this paper is twofold. First, methodologically, we focus on stock boom-busts in the cross section within one aggregate bubble for better identification and empirical implementation, while the previous literature mainly focuses on historical stock market bubbles. Our paper's focus on the cross section is similar in spirit to Chodorow-Reich, Guren, and McQuade (2022)'s study of the 2000s housing cycle at the city level.

Second, we decompose and quantify the various channels driving the boom and bust cycle of stocks during a market bubble. The channels we test and quantify include time-varying preferences or beliefs (Scheinkman and Xiong 2003; Baker and Wurgler 2006;

Barberis et al. 2015; Nagel and Xu 2022), the interplay among sophisticated and unsophisticated investors (Hong and Stein 1999; Abreu and Brunnermeier 2003; Brunnermeier and Nagel 2004; Greenwood and Nagel 2009), and wealth reallocation among heterogeneous investors (Atmaz and Basak 2018; Martin and Papadimitriou 2022a).[4] We find a major role for the entry of new investors, especially during the expansion phase of the bubble. While some earlier papers have studied herd behavior (Banerjee 1992; DeMarzo, Kaniel, and Kremer 2007), the role of entry has not received much attention in the recent literature. Our results suggest that models of bubbles should consider entry more seriously.

In estimating investors' asset demand system, the previous literature uses institutional stock holdings to estimate institutions' demand functions (Koijen, Richmond, and Yogo 2022; Haddad, Huebner, and Loualiche 2022; Gabaix and Koijen 2022). We bring in very detailed data on retail holdings that enable us to estimate heterogeneous retail demand functions. We also build upon the instrument for stock prices introduced in Koijen and Yogo (2019) by leveraging features of the Chinese institutional setting to make the exclusion restriction more likely to be satisfied.

By studying retail investors' trading behavior, this paper also contributes to the household finance literature (Calvet, Campbell, and Sodini 2009; Barber, Odean, and Zhu 2009; Barber and Odean 2013; Balasubramaniam et al. 2023) and the noise trader literature (De Long et al. 1990a; Shleifer and Summers 1990). We demonstrate the price impacts from retail investors and quantify the magnitudes of retail trading behavior contributing to stock booms and busts.

Finally, there are other papers studying bubbles in the Chinese stock market. Xiong and Yu (2011) study the Chinese warrants bubble. Bian et al. (2021), Liao, Peng, and Zhu (2022), and An, Lou, and Shi (2022) focus on the 2015 Chinese stock market bubble. While these studies describe the different forces driving bubbles, our contribution is to quantify the magnitudes of these forces.

4. We show that wealth reallocation among existing investors is relatively unimportant for cross-sectional stock returns, accounting for only 2% of return variance.

Organization of the paper

The remainder of the paper is organized as follows. In Section 3.2, we describe our data and provide more background information about the 2015 Chinese stock market bubble. In Section 1.3, we define boom-bust stocks and examine their differing characteristics compared to other stocks. In Section 1.4, we lay out a framework to estimate investors' heterogeneous asset demand functions. In Section 1.5, we quantify the different channels contributing to stock boom-busts through variance decomposition. In Section 1.6, we conduct counterfactual analyses. We conclude and discuss future research directions in Section 2.8.

1.2 Data and Background

In this section, we first introduce our data, which have significant advantages over alternative holdings data from other stock markets. We then take a closer look at the 2015 Chinese stock market bubble and provide an example of the cross-sectional stock boom-bust differences.

1.2.1 Data on Equity Holdings

We obtain monthly administrative account-level holdings and transactions records from the Shanghai Stock Exchange (SSE hereafter), from January 2011 through December 2019. There are two players in the market: institutional investors and retail investors. We obtain information on all institutional accounts (in total 42,974 institutional accounts) and a representative sample of 18 million retail accounts (about 20% of the entire retail population).

The Chinese stock market is a particularly good environment to study the price impacts of retail investors. Retail investors are considered to be more prone to behavioral mistakes, which the previous literature has emphasized to be an important driver of bubbles (Barberis, Shleifer, and Vishny 1998; Scheinkman and Xiong 2003; Baker and Wurgler 2006; Barber and Odean 2013). In the Chinese stock market, retail investors account for on average 84% of trading volume and hold 60% of floating stock market capitalization.[5] In addition to their

5. According to McCrank (2021), retail investors accounted for 17.1% of the trading volume in the U.S. stock

stock holdings, we also have investor demographics such as age, trading experience, and gender.

Our data on administrative account-level equity holdings have significant advantages over alternative datasets in other stock markets. In the U.S., there are data from the Securities and Exchange Commission Form 13F, which report quarterly holding filings for institutions whose market values exceed $100 million. Data on retail holdings in the U.S. are very limited: most research is based on the Odean data from a U.S. brokerage firm in the 1990s, with 78,000 accounts (Barber and Odean 2000). To study retail investors' trading behavior, researchers have also used Swedish data, which are annual snapshots of all Swedish residents' financial positions (Calvet, Campbell, and Sodini 2007; Calvet, Campbell, and Sodini 2009). The datasets closest to ours are the Taiwan Stock Exchange data used by Barber et al. (2009) and the Indian Stock Exchange data used by Campbell, Ramadorai, and Ranish (2014). Our dataset is advantageous because it covers both institutions and retail investors, has a representative and large-scale sample, updates at a monthly frequency, and experiences a noteworthy stock market bubble during the sample period. These granular data enable us to systematically examine every market player's behavior during a highly influential stock market bubble for the first time.

1.2.2 Data on Stock Characteristics

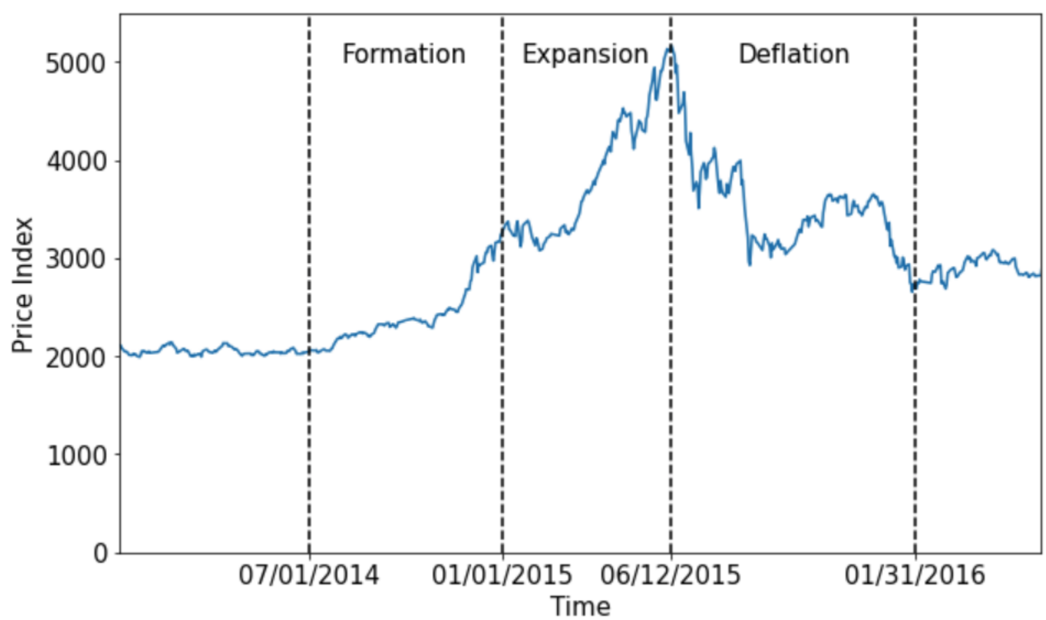
We use the Wind SSE Stock Database and China Stock Market & Accounting Research Database (CSMAR) to construct stock characteristics. These are the two standard stock databases for Chinese stock information. In cases where these two databases do not agree, we check financial statements to manually correct for errors. See Appendix A.2 for details of data construction.

market in January 2020.

Figure 1.1: *Time Series of Shanghai Stock Exchange Composite Index*

(a) *Price Index from 2011 through 2019*

(b) *Three Bubble Phases*



Notes: Panel (a) plots the time series of the price level for the Shanghai Stock Exchange composite index from 2011 through 2019. The price index excludes returns from cash dividends. The two dashed lines indicate the start (07/01/2014) and end (01/31/2016) of the bubble. Panel (b) zooms into the bubble period and separates the bubble into three phases according to the aggregate price trend: the formation (07/01/2014-12/31/2014), the expansion (01/01/2015-06/12/2015), and the deflation (06/13/2015-01/31/2016).

1.2.3 The 2015 Chinese Stock Market Bubble

Let us take a look at the 2015 Chinese stock market bubble. Panel (a) in Figure 1.1 plots the time series for the Shanghai Stock Exchange composite index. On July 1, 2014, the market price index was at 2,050. The market experienced a sharp price increase starting at the end of 2014, reaching the peak on June 12, 2015, when the market price index was at its historical high level of 5,166. This price level was more than twice its starting value less than a year earlier. Then the market went through a sharp decline: on January 31, 2016, the market price dropped to 2,738, a 47% decrease compared to 7 months prior.

Panel (b) of Figure 1.1 zooms into the bubble period and separates the bubble into three phases according to the aggregate price trend: the formation, expansion, and deflation of the bubble. From July 1, 2014 through June 12, 2015, the market experienced a price increase. To examine whether investors changed their behavior as the bubble evolved, we separate this price increase period into the formation and expansion phases, using the beginning of 2015 as a cutoff.[6] From June 13, 2015 through Jan 31, 2016, the market experienced a price decline, which we define as the deflation phase. Later, we will analyze stock performance and investors' behavior over these three bubble phases.

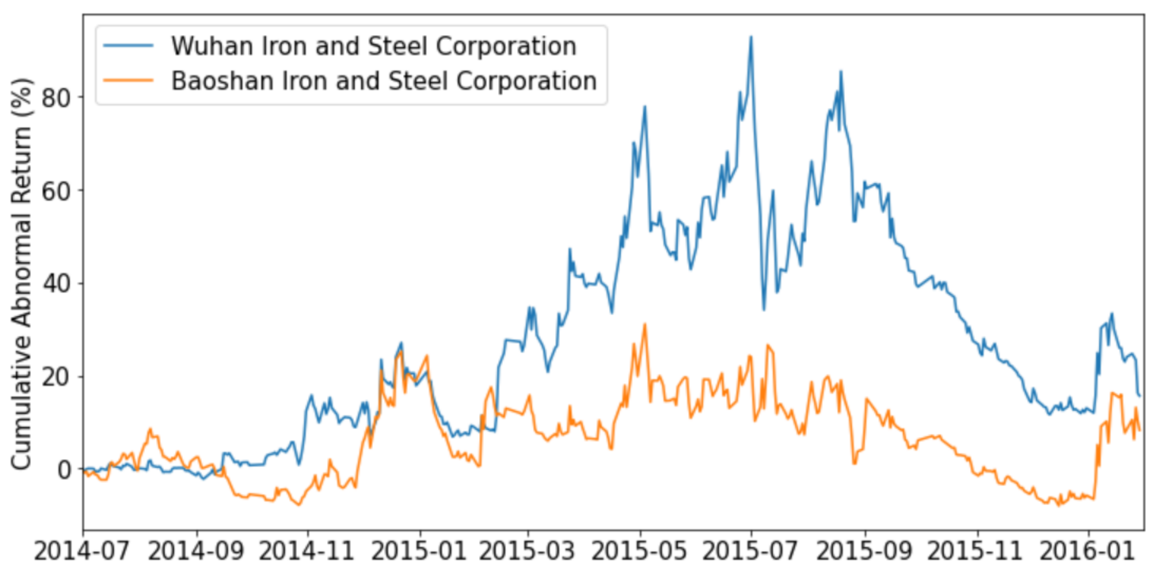
1.2.4 Cross-Sectional Boom-Bust Differences: An Example

This paper examines the cross-sectional differences in individual stock boom-busts during the 2015 Chinese stock market bubble. We ask: which stocks experience larger boom-busts? Who has a high demand for these stocks? To give an example of cross-sectional boom-bust differences, Figure 1.2 plots the cumulative returns for two stocks: "Wuhan Iron and Steel Corporation" and "Baoshan Iron and Steel Corporation." These two companies were in the same industry and operated in the same line of business: iron and steel. However, the price paths for the two stocks were very different. The returns plotted in Figure 1.2 are cumulative abnormal returns. The abnormal returns are calculated based on the CAPM, where the

6. Our results are robust if we use alternative cutoffs such as November 2014.

CAPM betas are calculated using 4-year daily returns prior to the bubble period. The reason we control for the CAPM beta here is that we want to examine other stock characteristics aside from the CAPM beta that may drive return differences. Appendix Table A.1 also plots the cumulative raw returns of these two stocks. The raw return differences are even more salient since the CAPM beta is basically an amplification of the market. Therefore, we are being conservative by focusing on abnormal returns. We can see that the Wuhan Iron and Steel Corporation went through a much larger boom-bust: during mid-2015, it had cumulative abnormal returns above 80%, and has almost dropped back to its starting value by the end of the bubble period. The Baoshan Iron and Steel Corporation, however, had its abnormal return path that is almost flat.

Figure 1.2: *Cumulative Abnormal Returns for Two Illustrative Stocks during the 2015 Chinese Stock Market Bubble*



	Retail Share	CAPM Beta	Book Value (CNY Billion)	Profitability-to-Book	Dividend-to-Book
Wuhan Iron and Steel Corporation	84.84%	0.87	36.14	1.20%	0.72%
Baoshan Iron and Steel Corporation	72.09%	0.79	114.13	5.66%	1.83%

Notes: This figure plots the cumulative abnormal returns for two illustrative stocks, “Wuhan Iron and Steel Corporation” and “Baoshan Iron and Steel Corporation,” during the 2015 Chinese stock market bubble period. The abnormal returns are calculated based on the CAPM, where the CAPM betas are calculated using 4-year daily returns prior to the bubble period. The table at the bottom of the figure reports average characteristics for these two stocks during the bubble period.

One may wonder how these two stocks differ in other characteristics. At the bottom of

Figure 1.2, we show average characteristics for these two stocks during the bubble period. Wuhan Iron and Steel Corporation—the one with a larger boom-bust—is owned by more retail investors, has a higher CAPM beta, is smaller in size, has a lower profitability-to-book ratio, and has a lower dividend-to-book ratio. The following sections of the paper first ask whether these differences in characteristics we see in this example can be generalized to other stocks. We then answer why some stocks experience larger boom-busts than other stocks.

1.3 Descriptive Evidence

In this section, we first define boom-bust stocks, and then show the differences in characteristics between boom-bust stocks and other stocks. Finally, we present suggestive evidence on which investors may contribute to the price paths of boom-bust stocks.

1.3.1 Boom-Bust Stocks

To examine cross-sectional differences in stock boom-busts, we need to first define which stocks are categorized as boom-bust. Focusing on the bubble period from 07/2014 through 01/2016, let us start by defining a stock's boom and bust sizes. Following the example in the previous section, we first calculate a stock's abnormal returns using the CAPM, where the CAPM betas are estimated using 4-year daily returns prior to the bubble. A stock's boom is defined as its highest cumulative abnormal returns since 07/2014, and the bust is defined as the percent difference between the stock's boom and its ensuing trough in cumulative abnormal returns.

Table 1.1 shows boom and bust size summary statistics for individual stocks. We have 904 stocks in the sample after restricting the data to stocks that had been in existence at least four years prior to the bubble.[7] Stocks have an average abnormal return boom of 43%, and

7. We exclude new stocks because of the inability to measure CAPM beta and their different price paths after IPO. We plan to study the role of IPO stocks during the bubble in future research.

an average bust of 34%. The standard deviations are quite high. The maximum of boom can go above 400%, and the maximum of bust can be as large as 76%.

Table 1.1: *Summary Statistics for Individual Stocks' Boom and Bust Sizes*

	Boom (%)	Bust (%)
	(1)	(2)
N	904	904
Mean	42.5	33.9
Std.Dev.	52.8	19.1
Min	-23.8	0
25th	9.0	19.6
50th	25.9	36.1
75th	61.0	48.4
Max	423.8	75.9

Notes: This table reports summary statistics for individual stocks' boom and bust sizes during the 2015 Chinese stock market bubble period: 07/2014-01/2016. We first calculate a stock's abnormal returns from CAPM, where the CAPM betas are estimated using 4-year daily returns prior to the bubble. A stock's boom is defined as its highest cumulative abnormal return since 07/2014, and the bust is defined as the percent difference between the stock's boom and its ensuing trough in cumulative abnormal returns.

After we calculate stocks' boom and bust sizes, we next define stocks as boom-bust stocks if both their boom and bust sizes are above the median. This way, we identify 157 boom-bust stocks.

Our results are robust to alternative ways of calculating a stock's boom and bust. We can use raw returns instead of abnormal returns, we can use a varying window to calculate the CAPM beta rather than a fixed window prior to the bubble period, and we can calculate the boom and bust based on a fixed aggregate time window rather than individual stock varying time windows. Furthermore, when identifying boom-bust stocks, we can use alternative cutoffs rather than the median threshold. All of the methods we explored to identify stocks that experienced sharp price increases and declines have yielded similar results.

1.3.2 Characteristics of Boom-Bust Stocks vs. Other Stocks

After identifying boom-bust stocks, we can now compare characteristics between boom-bust stocks and other stocks.

To clarify, our definition of boom-bust stocks relies on ex-post price paths, following Greenwood and Nagel (2009) and Chodorow-Reich, Guren, and McQuade (2022). There is no causal interpretation in this section. Rather, we use simple reduced-form regressions to understand which stock characteristics are associated with larger boom-busts in the cross-section.

We consider the following set of stock characteristics: CAPM beta, size, profitability, investment, dividend-to-book ratio, age, and an indicator for state-owned enterprise (SOE). This set of stock characteristics includes standard Fama-French factors and other stock characteristics that are commonly used in factor studies. CAPM beta is estimated using a rolling window of previous 4-year daily returns, restricting to stocks with at least 500 daily return observations. Size is the log value of a stock's book equity. Profitability is measured as the ratio of operating profits over book equity. Investment is the yearly growth rate in a stock's book equity. Dividend-to-book ratio is the ratio of annual dividends over book equity. Age is the number of years since a stock's IPO date. We also include the indicator for state-owned enterprises (SOE) since this firm type may affect stock performance in the Chinese stock market. Note that we deliberately avoid using stock prices to construct these factors (e.g., we use the log value of book equity rather than the log value of market capitalization to capture stock sizes), since by definition boom-bust stocks experience higher price increases and declines during the bubble period. This is also the reason why we do not include the value/growth factor here (price-to-book ratio).

Aside from standard stock characteristics, we additionally add retail share to capture ownership information. Retail share is defined as the ratio of stock value owned by retail investors to the total stock value owned either by retail investors or institutions.

With the set of stock characteristics ready, we can next compare characteristics between boom-bust stocks and other stocks. We estimate the following regression for each stock characteristic x_k:

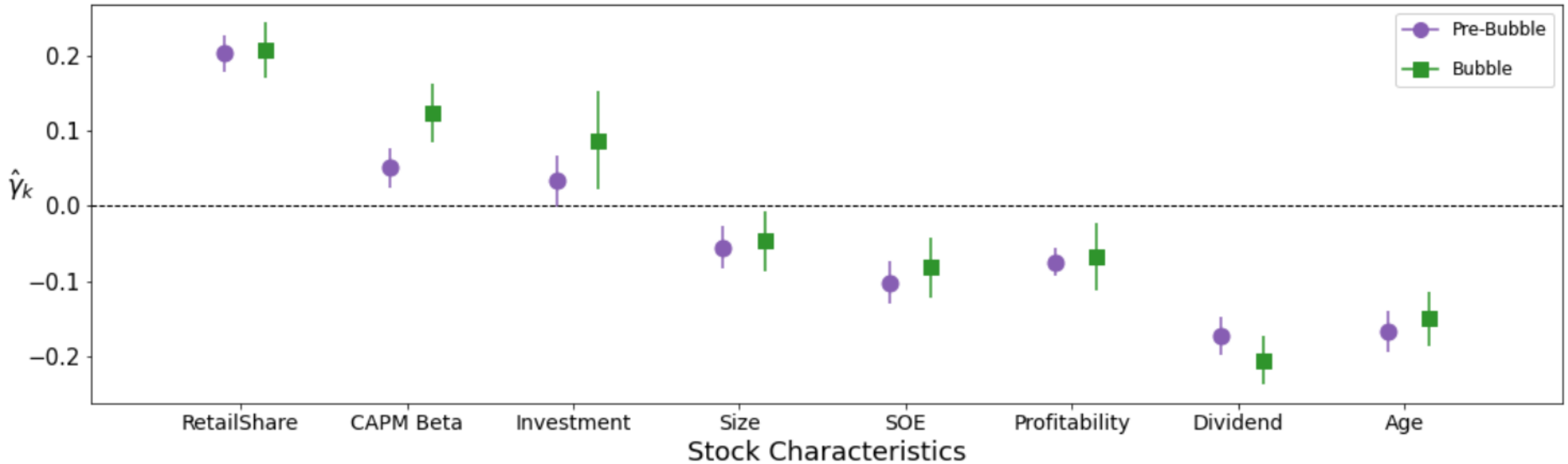
$$x_{k,t}(n) = \alpha_k + \gamma_k \mathbb{1}_{\text{Boom-Bust Stock}}(n) + \delta_{k,t} + \epsilon_{k,t}(n), \tag{1.1}$$

where $x_{k,t}(n)$ is stock n's characteristic x_k at time t. $\mathbb{1}_{\text{Boom-Bust Stock}}(n)$ indicates whether stock n is classified as a boom-bust stock. $\delta_{k,t}$ are time fixed effects. Observations are at the stock-by-month level. All stock characteristics x_k are standardized to cross-sectional z-scores so that we can compare the magnitudes of γ_k across different characteristics.

The estimated γ_k measures the average characteristic differences between boom-bust stocks and other stocks. In Figure 1.3, we plot the estimated γ_k for different stock characteristics, both during the bubble period (07/2014-01/2016) and prior to the bubble period (01/2011-06/2014). Boom-bust stocks are owned by more retail investors. They have distinctive characteristics, which we call "boom-bust characteristics": a higher CAPM beta, higher investment, smaller in size, less likely to be state-owned enterprises (SOE), less profitable, lower dividends, and younger. Going from the pre-bubble period to the bubble period, boom-bust stocks have a higher CAPM beta, slightly higher investment and a lower dividend, while the other characteristics remain relatively unchanged.

Figure 1.3 reveals that stocks that experienced larger boom-busts have distinctive "boom-bust characteristics." Natural questions to ask: why did these stocks experience larger boom-busts? Who were the investors creating a particularly high demand for these stocks? In the following subsection, we show suggestive evidence that retail investors, especially the ones who entered the stock market during the expansion phase of the bubble, may have contributed to the high prices for boom-bust stocks.

Figure 1.3: *Differences in Characteristics between Boom-Bust Stocks and Other Stocks*



Notes: This figure plots the differences in characteristics between boom-bust stocks and other stocks (the estimated γ_k), both during the bubble period (07/2014-01/2016) and the pre-bubble period (01/2011-06/2014). γ_k is estimated from the following regression:

$$x_{k,t}(n) = \alpha_k + \gamma_k \mathbb{1}_{\text{Boom-Bust Stock}}(n) + \delta_{k,t} + \epsilon_{k,t}(n),$$

where $x_{k,t}(n)$ is stock n's characteristic x_k at time t. $\mathbb{1}_{\text{Boom-Bust Stock}}(n)$ indicates whether stock n is classified as a boom-bust stock. $\delta_{k,t}$ are time fixed effects. Observations are at the stock-by-month level. Heteroskedasticity-robust standard errors are reported, and the error bars shown in the figure report the 95% confidence intervals. All stock characteristics x_k are standardized to cross-sectional z-scores so that we can compare the magnitudes of γ_k across different characteristics.

1.3.3 Suggestive Evidence on Who May Have Influenced Prices

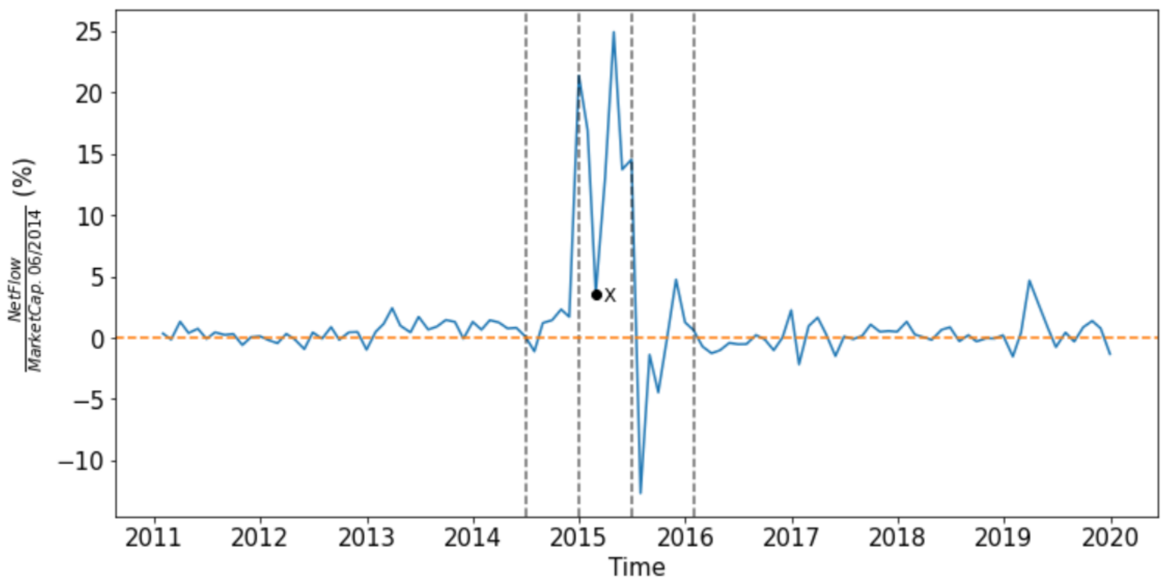
Figure 1.4 plots retail investors' net capital flows into boom-bust stocks. Net capital flows are calculated as the total value purchased minus the total value sold. Institutions serve as the counterpart. We divide the net capital flows by the stocks' market capitalization prior to the bubble (06/2014) to get the flow-to-market ratio. The grey dashed lines indicate different bubble phases: the formation, expansion, and deflation of the bubble. We can see that retail investors are net buyers of boom-bust stocks during the formation and expansion phases of the bubble, and are net sellers during the deflation phase. The retail net capital flows dropped at the beginning of 2015, however, were still positive and higher than the pre-bubble level. This drop in flows is likely a reaction to the flattening in market price at the time due to policy restrictions on leverage (Yu 2015; Chen and Qiu 2017). But this policy was soon reversed and the market price kept increasing. During the expansion phase of the bubble, the highest retail net flow reached 25% of pre-bubble market capitalization. Such huge flows suggest that retail investors may have enormous positive price pressure on boom-bust stocks during this time.

Figure 1.5 shows a time series of the number of new retail investors entering the Shanghai Stock Exchange. We can see that the number of new retail investors increased dramatically during the bubble's expansion phase. In June 2015, 4,116,485 new retail investors entered the stock market. This is 71 times larger compared to the number in June 2014. Such expansive entry into the stock market suggests that new entrants may have also contributed to price increases significantly.

1.4 Model

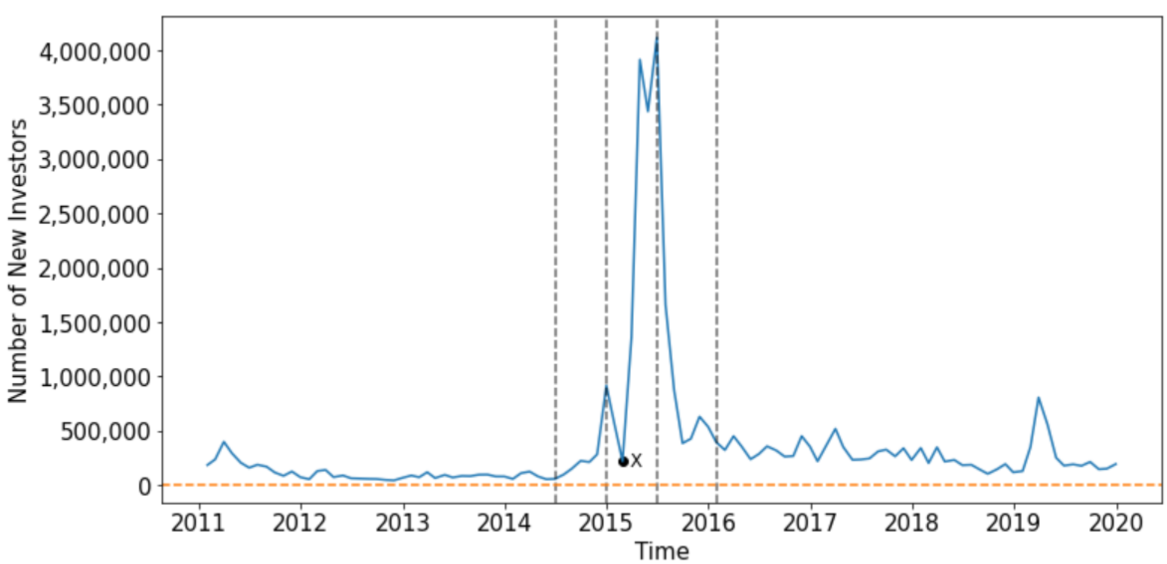
To understand the quantitative significance of boom-bust characteristics and investor heterogeneity, we build a structural model and estimate investors' heterogeneous demand functions for stocks. In this section, we set up the theoretical framework and lay out the procedure to estimate investors' asset demand functions.

Figure 1.4: *Retail Investors' Net Capital Flows into Boom-Bust Stocks*



Notes: This figure plots retail investors' net capital flows into boom-bust stocks. Institutions serve as the counterpart. Net capital flows are calculated as the total value purchased minus the total value sold. We divide the net capital flows by the stocks' market capitalization prior to the bubble (06/2014) to get the flow-to-market ratio. The grey dashed lines indicate different bubble phases: the formation, the expansion, and the deflation of the bubble. During the beginning of 2015, there were shortly imposed policy restrictions on leverage that were soon lifted (indicated by point X).

Figure 1.5: *Number of New Retail Investors Entering the Shanghai Stock Exchange*



Notes: This figure shows a time series of the number of new retail investors entering the Shanghai Stock Exchange. The grey dashed lines indicate different bubble phases: the formation, the expansion, and the deflation of the bubble. During the beginning of 2015, there were shortly imposed policy restrictions on leverage that were soon lifted (indicated by point X).

1.4.1 Theoretical Framework

We follow Koijen and Yogo (2019) to model investors' asset demand functions. There are I investors, indexed by $i = 1, ..., I$. There are N stocks, indexed by $n = 0, ..., N$, where asset 0 is the outside asset. Each stock has K characteristics, indexed by $k = 1, ..., K$. We denote $w_{i,t}(n)$ as the portfolio weight that investor i allocates to asset n. We have $\sum_{n=0}^{N} w_{i,t}(n) = 1$. We assume investors' demand for stocks to follow a factor structure, with stock characteristics $x_{k,t}(n)$ as factors. That is, the demand function for investor i at month t is:

$$\ln\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = \alpha_{i,t} + \beta_{0,i,t}\left(\mathrm{me}_t(n) - \mathrm{be}_t(n)\right) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n) + \epsilon_{i,t}(n), \tag{1.2}$$

where $\mathrm{me}_t(n)$ denotes the log value of stock n's market equity, $\mathrm{be}_t(n)$ denotes the log value of stock n's book equity, and $x_{k,t}(n)$ denotes other stock characteristics that we will describe later in Section 1.4.2

The demand function models the relative weight that investor i allocates to stock n as a function of the stock's log market-to-book ratio and other stock characteristics $x_{k,t}(n)$. The log market-to-book ratio captures how expensive the stock is, and, in Section 1.4.2, we will construct an instrument for stock price to identify the demand elasticity (roughly $1 - \beta_{0,i,t}$). The demand function can be micro-founded based on a discrete choice model with i.i.d. Logit errors (see Koijen and Yogo (2019) for a detailed discussion on micro-foundation).

To complete the asset demand system, let $P_t(n)$ be stock n's price, $S_t(n)$ be stock n's shares outstanding, $\mathrm{ME}_t(n)$ be stock n's market equity, and $A_{i,t}$ be investor i's assets under management (AUM). We have the following market clearing condition for every stock n:

$$P_t(n) S_t(n) = \mathrm{ME}_t(n) = \sum_{i=1}^{I} w_{i,t}(n) A_{i,t}. \tag{1.3}$$

That is, a stock's total value should equal total investor demand for this stock.

Since the allocation weight $w_{i,t}(n)$ depends on the log price vector $\mathbf{p}$ according to (1.2), we can take log on both sides of the Equation (1.3) and rewrite the market clearing condition as:

$$\mathbf{p}_t = \mathbf{f}(\mathbf{p}_t) = \log\left(\sum_{i=1}^{I} A_{i,t} \mathbf{w}_{i,t}(\mathbf{p}_t)\right) - \mathbf{s}_t, \tag{1.4}$$

where the lowercase letters denote the logarithm of the corresponding uppercase variables and the bold letters are vector variables for all stocks.

Later, when we conduct counterfactual analyses, we can change conditions in the counterfactual scenario (e.g., change investors' $\beta_{i,t}$) and iterate to get the fixed point of the log price vector $\mathbf{p}_t^{new}$, which will be the new log price vector under the counterfactual scenario.

1.4.2 Demand Function Estimation

In this subsection, we illustrate the detailed procedure to estimate investors' demand functions.

Regression Specification

The demand function we need to estimate is Equation (1.2):

$$\ln\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = \alpha_{i,t} + \beta_{0,i,t}\left(\mathrm{me}_t(n) - \mathrm{be}_t(n)\right) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n) + \epsilon_{i,t}(n). \tag{1.2}$$

We estimate the demand function for every investor type i at each month t. Therefore, we allow the estimated coefficients $\beta_{k,i,t}$ to be fully flexible among different investors and over different time periods. Following Koijen and Yogo (2019), we restrict $\beta_{0,i,t} < 1$ to guarantee that there exists a unique equilibrium in counterfactual analyses.

We refer to the estimated $\beta_{k,i,t}$ as investor i's preferences/beliefs about stock characteristic k, and the estimated residual $\epsilon_{i,t}(n)$ as investor i's latent demand. To clarify the terminology, we do not attempt to separate investors' preferences for-or beliefs about-stock characteristics, so we refer to $\beta_{k,i,t}$ as preferences/beliefs interchangeably. $\beta_{k,i,t}$ incorporates anything that can affect investors' attitudes toward stock characteristics.

Equation (1.2) is estimated using an instrument for $\mathrm{me}_t(n)$, which we will describe in Section 1.4.2.

Data

For $w_{i,t}(n)$, the portfolio weight that investor i allocates to asset n, we directly compute it from our data on equity holdings. For $w_{i,t}(0)$, the portfolio weight on outside assets, we follow Koijen and Yogo (2019) to define outside assets as any stocks with missing stock characteristics.

For stock characteristics $x_{k,t}(n)$, we use the same set of variables in Section 1.3.2, but exclude retail share, since ownership information is endogenous to latent demand $\epsilon_{i,t}(n)$. We additionally add stocks' past one-year returns and 28 industry fixed effects into the stock characteristics set. Summary statistics for stock characteristics are reported in Table 1.2.

Parameterization

The demand function (1.2) is estimated for every investor type i at each month t. Theoretically, we can estimate the demand function for all individual investors. However, retail investors are highly under-diversified: most of them only hold a few stocks. Since the demand function is estimated through a cross-sectional regression, with only a few observations, there is not enough power to estimate the coefficients $\beta_{k,i,t}$ accurately. Therefore, we group retail investors into 750 types and estimate the demand function for every investor type at each month. That is, for investors within the same type, we parameterize their demand functions to be the same.

To classify retail investor types, we first label them into the following five dimensions independently:

1. Entry cohort: 5 groups
2. Wealth in the stock market: 5 groups
3. Investor tendency to buy stocks with high historic returns: 5 quintiles
4. Gender: 2 groups
5. Age: 3 terciles

Entry cohort captures when investors enter the stock market: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble phase. For stock market wealth, we calculate each investor's average stock market wealth in the preceding year and then label investors into 5 groups according to their average wealth. Investor tendency to buy stocks with high historic returns is measured by individual return chasing propensity (RCP) as constructed in Chen, Liang, and Shi (2022) (See Appendix A.2.3 on the construction of RCP). We also label investors by their gender and age.

The above independent sort gives us $5 \times 5 \times 5 \times 2 \times 3 = 750$ retail investor types. If certain retail types do not hold more than 500 stocks, we group them with the most similar types to reach a minimum of 500 holdings for each retail type. We set the threshold at 500 because this allows us to estimate the demand function more accurately. Our results are robust to alternative thresholds, for instance, 300.

For institutions, we pool them within the same institution category and with similar assets under management to reach a minimum of 500 holdings for each type. We have, on average, 246 institution types. Appendix A.2.7 and A.2.8 illustrate the procedures used to construct investor types in detail.

Instrumental Variable for Stock Price

To identify the demand elasticity, we need to find an instrumental variable for stock price, which enters the log value of market equity in Equation (1.2). We need an instrument for stock price because latent demand $\epsilon_{i,t}(n)$ is likely to be correlated with stock price. For example, there may be some good news for a stock that the demand function cannot capture and thus is absorbed by $\epsilon_{i,t}(n)$. In such an event, investors hold more of this stock due to the good news, and at the same time, the stock price increases because of this good news. If we run a simple OLS regression, we will get a biased estimate of $\beta_{0,i,t}$.

Table 1.2: *Summary Statistics for Stock Characteristics*

	2011-2013				2014-2016				2017-2019			
	N	Mean	Median	Std.Dev.	N	Mean	Median	Std.Dev.	N	Mean	Median	Std.Dev.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Market-to-Book	931	4.84	2.75	16.45	1064	13.10	3.46	103.88	1440	7.33	2.43	63.97
CAPM Beta	931	1.16	1.18	0.22	1064	1.05	1.07	0.21	1440	1.17	1.18	0.23
Book Value (CNY Billion)	931	12.42	1.94	69.40	1064	16.04	2.46	92.56	1440	17.14	2.68	104.67
Profitability-to-Book	931	0.07	0.08	0.13	1064	0.05	0.07	0.15	1440	0.08	0.09	0.11
Investment	931	0.19	0.11	0.33	1064	0.24	0.12	0.50	1440	0.15	0.10	0.23
Dividend-to-Book	931	0.02	0.01	0.02	1064	0.02	0.02	0.03	1440	0.03	0.02	0.03
Age	931	11.91	11.91	5.13	1064	13.32	14.44	6.53	1440	12.44	14.68	8.53
SOE	931	0.65	1.00	0.47	1064	0.57	1.00	0.49	1440	0.44	0.00	0.49
Past1YearReturn (%)	931	-0.19	-4.19	21.29	1064	26.29	25.31	22.25	1440	-5.86	-8.43	17.38

Notes: This table reports summary statistics of stock characteristics over the time. Characteristics for individual stocks are averaged over the given period, and summary statistics of those averaged stocks are reported. Market-to-Book is the ratio of a stock's market capitalization over its book equity. CAPM Beta is estimated using a rolling window of previous 4-year daily returns, restricting to stocks with at least 500 daily return observations. Book Value is in the unit of CNY billion (the exchange rate is roughly 6.5 CNY = 1 USD). Profitability-to-Book is measured as the ratio of operating profits over book equity. Investment is the yearly growth rate of a stock's book equity. Dividend-to-Book is the ratio of annual dividends over book equity. Age is the number of years since a stock's IPO date. SOE is a indicator for state-owned enterprises. Past1YearReturn is the stock return in the previous year.

We leverage the Chinese institutional setting to construct an instrument for prices, potentially improving upon the "mandate IV" in the literature (Koijen and Yogo 2019; Haddad, Huebner, and Loualiche 2022). To clarify the terminology, a mandate is a predetermined set of investable stocks for an institution.

China is distinctive in having large institutional investors with extremely stable mandates. We refer to these institutions as "stable institutions," which are the National Social Security Funds, Corporate Supplementary Pension Funds, and Qualified Foreign Institutional Investors. Empirically, we measure the mandate of an institution as all stocks that were held by this institution in the previous one year. Across stable institutions, a median of 97% of stocks held in a current month had been held at some point in the previous one year. Our instrument is in the same spirit as the "index inclusion" literature (Harris and Gurel 1986; Shleifer 1986): if a stock gets included in the mandate of a stable institution, say, a National Social Security Fund, it will have exogenous price pressure from that fund.

Formally, we construct the instrument in the following way: we calculate the counterfactual log value of the market equity for each stock n, in the hypothetical scenario that stable institutions hold a book equity-weighted portfolio within their mandate:

$$\widehat{\mathrm{me}}_t(n) = \log\left(\sum_{j \in Stable} A_{j,t} \frac{\mathbb{I}_{j,t}(n)\,\mathrm{BE}_t(n)}{\sum_{m=1}^{N} \mathbb{I}_{j,t}(m)\,\mathrm{BE}_t(m)}\right), \tag{1.5}$$

where $\widehat{\mathrm{me}}_t(n)$ is the counterfactual log market equity for stock n. Set *Stable* denotes stable institutions. $A_{j,t}$ denotes the AUM for institution j. $\mathbb{I}_{j,t}(n)$ is the indicator variable on whether stock n is in institution j's mandate. $\mathrm{BE}_t(n)$ denotes stock n's book equity.

The exclusion restriction for the instrument is:

$$A_{j,t}, \mathbb{I}_{j,t}(n) \perp \epsilon_{i,t}(n) \mid \boldsymbol{x}_t(n), \forall j \in \{\text{Stable institutions}\}. \tag{1.6}$$

That is, after controlling for the stock characteristic vector $\boldsymbol{x}_t(n)$, the AUM and mandate for stable institutions need to be exogenous to other investors' latent demand.

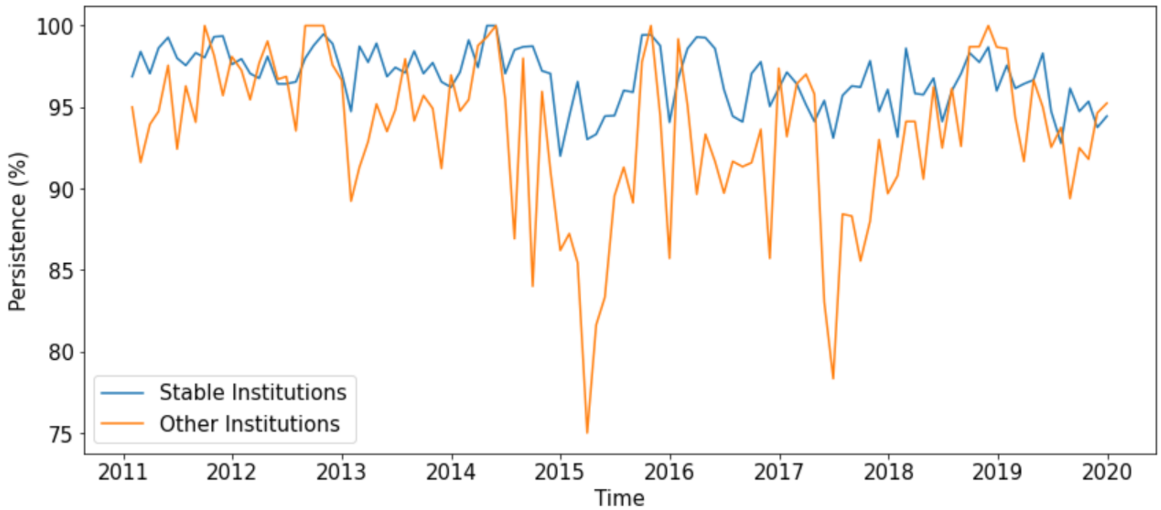
Our instrument based on the stable institutions has the following two advantages compared to the original instrument in Koijen and Yogo (2019), which uses all institutions to construct the mandate IV.

First, by conditioning on stable institutions, the mandate is more likely to be orthogonal to latent demand, i.e., $\mathbb{I}_{j,t}(n) \perp \epsilon_{i,t}(n) \mid \boldsymbol{x}_t(n), \forall j \in \{\text{Stable institutions}\}$. As mentioned above, across stable institutions, a median of 97% of stocks held in a current month had been held at some point in the previous one year: their mandates are extremely stable. What's more, in their prospectus, they stated that they "aim for long-run stable growth, rather than short-term profits."[8] Thus, a National Social Security Fund is unlikely to change its mandate $\mathbb{I}_{j,t}(n)$ due to retail latent demand $\epsilon_{i,t}(n)$. If we instead use all institutions to construct the mandate IV, some institutions, especially mutual funds, may alter their mandates to cater to retail latent demand. For example, if retail investors suddenly has enthusiasm for tech stocks ($\epsilon_{i,t}(n)$), a mutual fund may add tech stocks into its mandate to attract retail investors. In Figure 1.6, we plot the size-weighted median mandate persistence for stable vs. other institutions. Mandate persistence is defined as the percent of currently held stocks that were ever held in the previous year. We can see that stable institutions have a mandate persistence of around 97%. Other institutions have lower mandate persistence than stable institutions, with particularly low mandate persistence during the 2015 bubble period, which signifies that they altered their mandates during the bubble. Therefore, the exclusion restriction that mandates need to be independent of other investors' latent demand is more likely to hold when focusing on stable institutions only.

Second, AUM is also more likely to be orthogonal to latent demand when conditioning on stable institutions, i.e., $A_{j,t} \perp \epsilon_{i,t}(n) \mid \boldsymbol{x}_t(n), \forall j \in \{\text{Stable institutions}\}$. Flows of National Social Security Funds and Corporate Supplementary Pension Funds mainly come from Chinese residents' pension savings, which is a fixed ratio of personal income and cannot be easily changed by individuals. Similarly, Qualified Foreign Institutional Investors have their AUM coming from foreign investors and are unlikely to depend on Chinese retail investors' flows. Instead, if we used all institutions to construct the mandate IV, some institutions such

8. See annual reports of National Social Security Funds' at: http://www.ssf.gov.cn/portal/jjcw/sbjjndbg/webinfo/2016/06/1632636003321600.htm (2015, Chinese). See statements from Corporate Supplementary Pension Funds at : http://www.mohrss.gov.cn/xxgk2020/gzk/gz/202112/t20211228_431643.html (2011, Chinese).

Figure 1.6: *Size-Weighted Median Mandate Persistence for Stable vs. Other Institutions*



Notes: This figure plots the size-weighted median mandate persistence for stable vs. other institutions. Mandate persistence is defined as the percent of currently held stocks that were ever held in the previous year. Stable institutions are the National Social Security Funds, Corporate Supplementary Pension Funds, and Qualified Foreign Institutional Investors.

as mutual funds, may change their AUM based on retail inflows (Frazzini and Lamont 2008; Lou 2012), which would make their AUM $A_{j,t}$ correlated with retail latent demand $\epsilon_{i,t}(n)$.

Robustness

We did several robustness checks to make sure that our results are robust to the instrument we use. This includes using any two types of the three stable institutions to construct the instrument (as opposed to all of them), constructing the instrument using both equal-weighted and book-weighted counterfactual market equity, and including higher-orders of the instrument to capture first-stage non-linearity. To address the concern that other investors may want to hold some stocks because stable institutions hold them, we also constructed the instrument using only small-size stable institutions, whose holding behavior is less influential and observable to other investors. Our estimation results are robust to these alternatives.

1.4.3 Estimation Results

In this subsection, we report the demand estimation results. Summary statistics for the estimated demand function parameters are reported in Table 1.3. We find that retail investors prefer boom-bust characteristics compared to institutions. However, institutions have a low demand elasticity and therefore do not fully offset retail demand.

Institutions Have a Low Demand Elasticity

Price impacts from flows depend on the counter-party's demand elasticity. Based on our estimated demand functions, investor i's demand elasticity is roughly $1 - \hat{\beta}_{0,i,t}$.[9] To compare demand elasticities across different investors and across different time periods, we regress the estimated demand elasticities on the interaction between the investor's type (retail vs. institution) and time period indicators. Formally, we estimate the following regression:

$$1 - \hat{\beta}_{0,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1} \left[i \in I, t \in T \right], \tag{1.7}$$

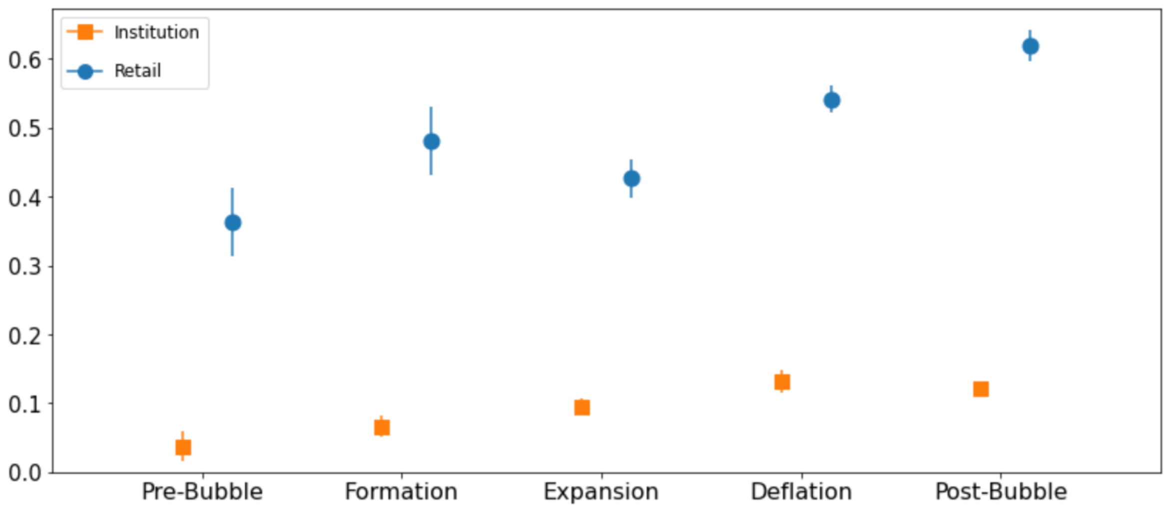
where $1 - \hat{\beta}_{0,i,t}$ is the previously estimated demand elasticity for investor i at time t, $I \in \{\text{Retail, Institution}\}$, $T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. Standard errors are clustered at the individual investor level. The estimated $\hat{\gamma}_{I,T}$ captures the average demand elasticities for different investor types over different time periods.

Figure 1.7 plots the estimated $\hat{\gamma}_{I,T}$ (coefficient estimates are reported in Column (1) of Table 1.4). We see that institutions have an extremely low demand elasticity of around 0.1 while retail investors have a higher demand elasticity compared to institutions at around 0.5. With such a low demand elasticity, institutions may not able to offset demand from retail investors. Gabaix and Koijen (2022) find that the stock market's extremely low macro demand elasticity can lead to large stock market fluctuations. In a similar light, our results

9. See detailed derivation in Appendix Section A.1.

show that the micro demand elasticity across stocks is low and, therefore, that flows in and out of an individual stock can generate large price fluctuations for that stock. Consider a back-of-the-envelope calculation: the AUM-weighted aggregate demand elasticity over the bubble period was roughly 0.25, which means that a 10% inflow into a stock would lead to a 40% return increase ($0.1/0.25 = 0.4$).

Figure 1.7: *Estimated Demand Elasticities of Retail Investors and Institutions over Different Periods*



Notes: This figure reports the average demand elasticities ($\hat{\gamma}_{I,T}$) of retail investors and institutions, over different time periods. $\hat{\gamma}_{I,T}$ is estimated by regressing the demand elasticity ($1 - \hat{\beta}_{0,i,t}$) on the interaction between the investor's type (retail vs. institution) and time period indicators:

$$1 - \hat{\beta}_{0,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}\left[i \in I, t \in T\right], \tag{1.8}$$

where $1 - \hat{\beta}_{0,i,t}$ is the previously estimated demand elasticity for investor i at time t, $I \in \{\text{Retail, Institution}\}$, $T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. The estimated $\hat{\gamma}_{I,T}$ captures the average demand elasticities for different investor types over different time periods. Standard errors are clustered at the individual investor level, and the error bars shown in the figure report the 95% confidence intervals.

Retail Investors Prefer Boom-Bust Characteristics

We now report retail investors' preferences/beliefs about various stock characteristics. We find that retail investors, especially those with less wealth and those who entered the market during the expansion phase of the bubble, prefer boom-bust characteristics as introduced in Section 1.3.2.

Table 1.3: *Summary Statistics for Demand Function Parameters*

	2011-2013				2014-2016				2017-2019			
	Retail		Institution		Retail		Institution		Retail		Institution	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\hat{\beta}_0$	0.67	0.29	0.98	0.04	0.44	0.26	0.86	0.13	0.39	0.31	0.92	0.09
$\hat{\beta}_{\text{CAPM Beta}}$	0.70	0.32	-0.18	0.21	0.25	0.38	-0.33	0.34	0.19	0.27	0.28	0.32
$\hat{\beta}_{\text{Size}}$	0.70	0.14	0.91	0.24	0.46	0.20	0.60	0.15	0.52	0.18	0.65	0.15
$\hat{\beta}_{\text{Profitability}}$	-0.17	0.62	1.87	1.03	-0.10	0.45	0.62	0.56	0.00	0.60	0.77	0.72
$\hat{\beta}_{\text{Investment}}$	-0.11	0.16	-0.08	0.19	-0.07	0.09	-0.10	0.09	-0.03	0.17	-0.06	0.23
$\hat{\beta}_{\text{Dividend}}$	-2.37	2.23	1.16	2.74	-1.85	2.00	-1.72	2.43	-2.08	1.48	-1.51	1.56
$\hat{\beta}_{\text{Age}}$	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
$\hat{\beta}_{\text{SOE}}$	-0.14	0.10	-0.09	0.10	-0.21	0.10	-0.11	0.13	-0.12	0.14	-0.19	0.12
$\hat{\beta}_{\text{Past1YearReturn}}$	-0.52	1.58	0.56	0.32	-0.38	1.74	0.42	0.38	-0.21	1.33	0.43	0.35

Notes: This table reports the summary statistics for the time-series mean of the estimated demand function parameters within the given period and investor type. The demand function for investor i at month t is:

$$\ln\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = \alpha_{i,t} + \beta_{0,i,t}\left(\mathrm{me}_t(n) - \mathrm{be}_t(n)\right) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n) + \epsilon_{i,t}(n),$$

where $\mathrm{me}_t(n)$ denotes the log value of stock n's market equity, $\mathrm{be}_t(n)$ denotes the log value of stock n's book equity, and $x_{k,t}(n)$ denotes other stock characteristics. We control for 28 industry fixed effects. Demand functions are estimated using an instrument for $\mathrm{me}_t(n)$, which we describe in Section 1.4.2.

Consider stock profitability and CAPM beta. Preferences/beliefs about other stock characteristics follow similar patterns and are reported in Appendix A.3. We first examine the preferences/beliefs differences between retail investors and institutions. As in Section 1.4.3, we regress the previously estimated characteristic preferences/beliefs on the interaction between the investor's type (retail vs. institution) and time period indicators:

$$\hat{\beta}_{k,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}\left[i \in I, t \in T\right], \tag{1.9}$$

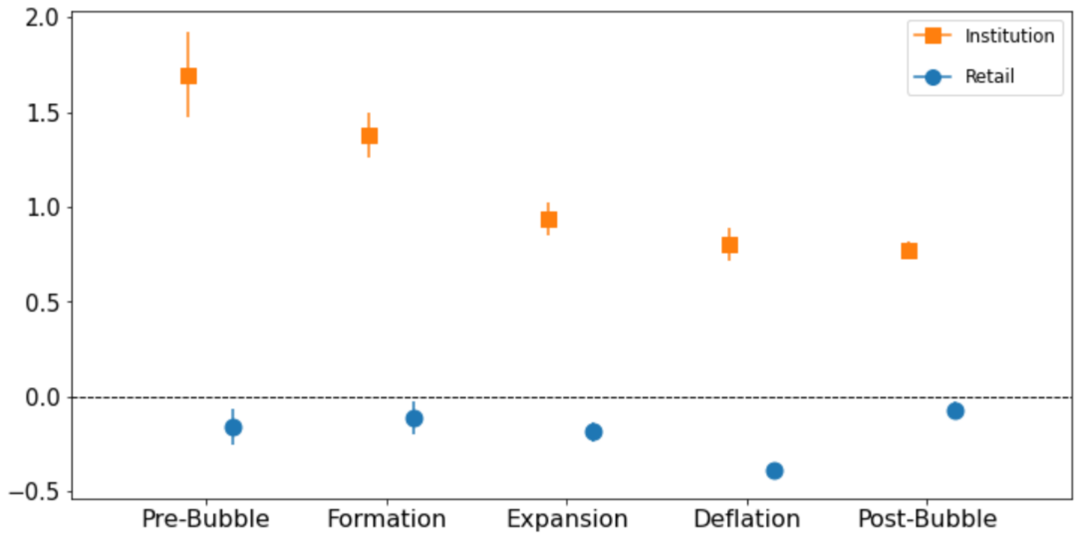
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $I \in \{\text{Retail, Institution}\}$, $T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T.

The estimated $\hat{\gamma}_{I,T}$ captures average preferences/beliefs about characteristic x_k of different investor types over different time periods. The results are plotted in Figure 1.8 (coefficient estimates are reported in Columns (2) and (3) of Table 1.4). Panel (a) reports investors' preferences/beliefs about profitability, and panel (b) reports investors' preferences/beliefs about CAPM beta. We see that retail investors are indifferent towards stock profitability on average, while institutions place great emphasis on profitability. A coefficient of 1.0 on profitability means that an investor will allocate an extra of 1% of wealth to a stock if the stock has a one percentage point increase in profitability-to-book ratio. As for CAPM beta, retail investors like higher beta stocks compared to institutions. A coefficient of 0.4 on CAPM beta means that an investor will allocate an extra of 4% of wealth to a stock if the stock's CAPM beta increases by 0.1. During the expansion and deflation phases of the bubble, institutions' average beta preferences went negative. Recall that low profitability and high beta are both boom-bust characteristics that we described in Section 1.3.2. It is also worth noting that similarly heightened retail demand can be seen for other boom-bust characteristics.

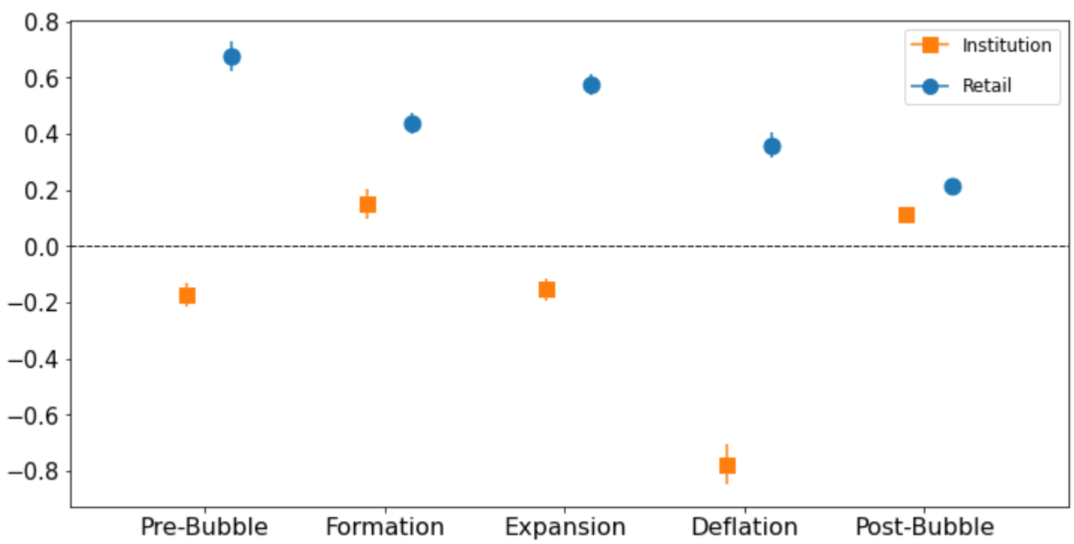
We can further investigate which retail investors have high demand for boom-bust characteristics. Figure 1.10 plots the preferences/beliefs of retail investors with different

Figure 1.8: *Preferences/Beliefs of Retail Investors and Institutions*

(a) *Preferences/Beliefs about Profitability*



(b) *Preferences/Beliefs about CAPM Beta*



Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_{I,T}$) of retail investors and institutions over different time periods. $\hat{\gamma}_{I,T}$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}\left[i \in I, t \in T\right],$$

where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $I \in \{\text{Retail, Institution}\}$, $T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. Panel (a) reports investors' preferences/beliefs about profitability, and panel (b) reports investors' preferences/beliefs about CAPM beta. Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Table 1.4: *Estimated Demand Function Coefficients of Retail Investors and Institutions*

Outcome Var:	$1-\hat{\beta}_0$	$\hat{\beta}_{\text{Profitability}}$	$\hat{\beta}_{\text{CAPM Beta}}$
	(1)	(2)	(3)
Institution: the Pre-Bubble Phase	0.037***	1.695***	-0.174***
	(0.011)	(0.113)	(0.022)
Institution: the Formation Phase	0.066***	1.380***	0.148***
	(0.008)	(0.060)	(0.027)
Institution: the Expansion Phase	0.095***	0.937***	-0.154***
	(0.007)	(0.045)	(0.021)
Institution: the Deflation Phase	0.132***	0.803***	-0.778***
	(0.008)	(0.045)	(0.036)
Institution: the Post-Bubble Phase	0.121***	0.775***	0.112***
	(0.003)	(0.021)	(0.008)
Retail: the Pre-Bubble Phase	0.364***	-0.157***	0.675***
	(0.025)	(0.048)	(0.026)
Retail: the Formation Phase	0.481***	-0.111**	0.438***
	(0.025)	(0.043)	(0.018)
Retail: the Expansion Phase	0.426***	-0.186***	0.574***
	(0.014)	(0.025)	(0.019)
Retail: the Deflation Phase	0.542***	-0.388***	0.359***
	(0.010)	(0.015)	(0.023)
Retail: the Post-Bubble Phase	0.619***	-0.068***	0.213***
	(0.012)	(0.022)	(0.009)
R-squared	0.392	0.204	0.194
Observations	59,494	59,494	59,494

Notes: This table reports $\hat{\gamma}_{I,T}$ from the following regression: $\hat{\beta}_{k,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}[i \in I, t \in T]$, where $\hat{\beta}_{k,i,t}$ is the previously estimated demand function coefficients for investor i at time t, $I \in$ {Retail, Institution}, $T \in$ {Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}, and $\mathbb{1}[i \in I, t \in T]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. The time period ranges from Jan 2011 to Dec 2019. Standard errors (in parentheses) are clustered at the investor level. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively.

levels of wealth (coefficient estimates are reported in Table 1.5). We classify retail investors into five groups according to their average stock market wealth in the previous year. We can see that low-wealth retail investors especially prefer boom-bust characteristics: they ignore profitability and take more risk by choosing high CAPM beta stocks. Interestingly, we see a monotone relationship between investors' preferences/beliefs and their wealth levels. The lower an investor's wealth level, the more likely that the investor will have a higher demand for boom-bust characteristics.

Figure 1.12 plots the average preferences/beliefs of retail investors entering the market at different time periods (coefficient estimates are reported in Table 1.6). We classify retail investors by their market entry cohort: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble phase. We focus on 01/2014 to 12/2016 and control for time fixed effects. Investors who entered the market during the expansion phase of the bubble have an especially high demand for boom-bust characteristics: they ignore stock profitability and take more risk by choosing high CAPM beta stocks.

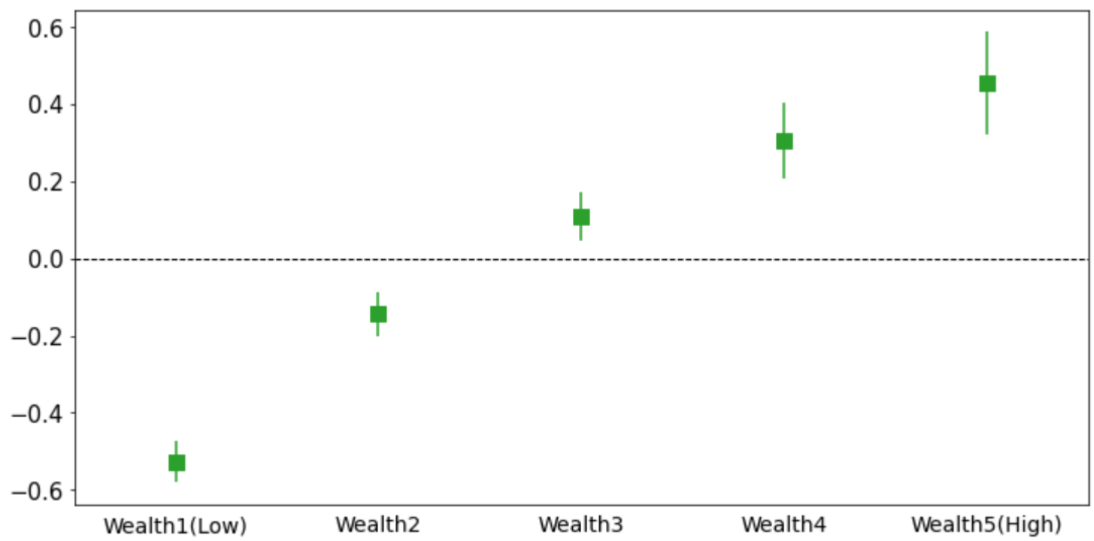
In sum, retail investors like boom-bust characteristics. Among retail investors, the ones who are likely to be unsophisticated (have less wealth or enter during the expansion phase of the bubble) especially prefer boom-bust characteristics. Coupled with the low demand elasticity from institutions, retail investors may be the main drivers of high prices for boom-bust stocks. We will quantify this contribution from retail investors in the next section.

1.5 Quantifying the Channels Driving Stock Boom-Busts

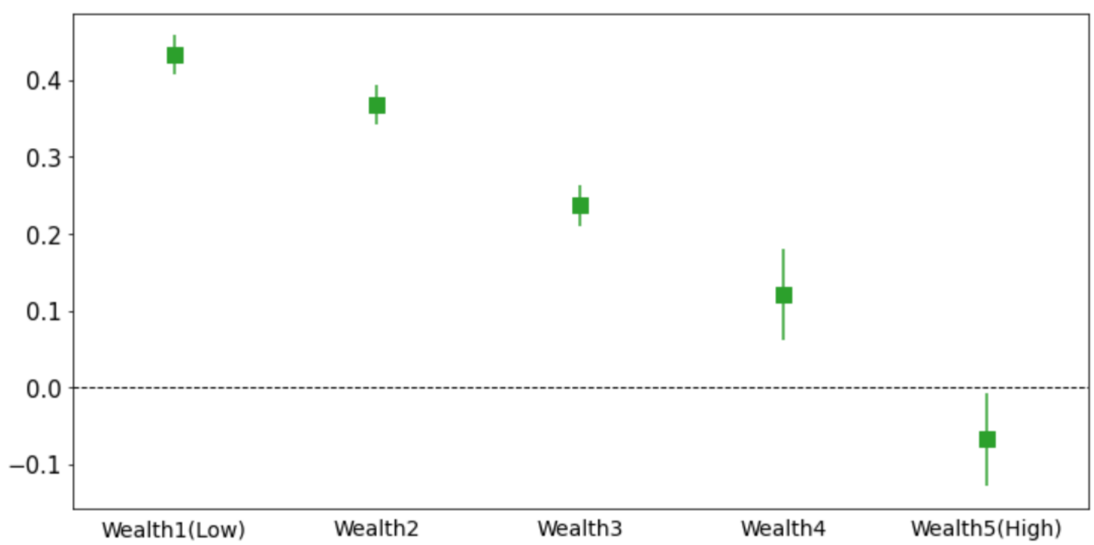
Having estimated heterogeneous demand functions, we can now implement the main task of this paper: to quantify the magnitudes by which various channels contribute to stock boom-busts. We decompose the contribution to cross-sectional stock return variance into three main components: retail investors, institutions, and stock supply. We change one element at a time in the asset demand system and recompute the counterfactual price vectors. In this section, we first lay out the procedure for return variance decomposition

Figure 1.10: *Preferences/Beliefs of Retail Investors with Different Levels of Wealth*

(a) *Preferences/Beliefs about Profitability*



(b) *Preferences/Beliefs about CAPM Beta*



Notes: This figure reports the preferences/beliefs ($\hat{\gamma}_W$) of retail investors with different levels of wealth. We classify retail investors into five groups according to their average stock market wealth in the previous year. $\hat{\gamma}_W$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_W \gamma_W \mathbb{1}\left[i \in W\right] + \delta_t,$$

where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $W \in$ {Wealth1 (Low), Wealth2, Wealth3, Wealth4, Wealth5 (High)}, and $\mathbb{1}\left[i \in W\right]$ is a dummy variable indicating whether investor i belongs to wealth group W, and δ_t are time fixed effects. Panel (a) reports investors' preferences/beliefs about profitability, and panel (b) reports investors' preferences/beliefs about CAPM beta. Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Table 1.5: *Estimated Preferences/Beliefs of Retail Investors with Different Levels of Wealth*

Outcome Var:	$\hat{\beta}_{\text{Profitability}}$	$\hat{\beta}_{\text{CAPM Beta}}$
	(1)	(2)
Wealth1 (Low)	-0.530***	0.462***
	(0.027)	(0.013)
Wealth2	-0.156***	0.401***
	(0.029)	(0.013)
Wealth3	0.087***	0.271***
	(0.033)	(0.014)
Wealth4	0.264***	0.152***
	(0.049)	(0.032)
Wealth5 (High)	0.416***	-0.036
	(0.068)	(0.032)
Time Fixed Effects	Y	Y
R-Squared	0.204	0.133
Observations	34,000	34,000

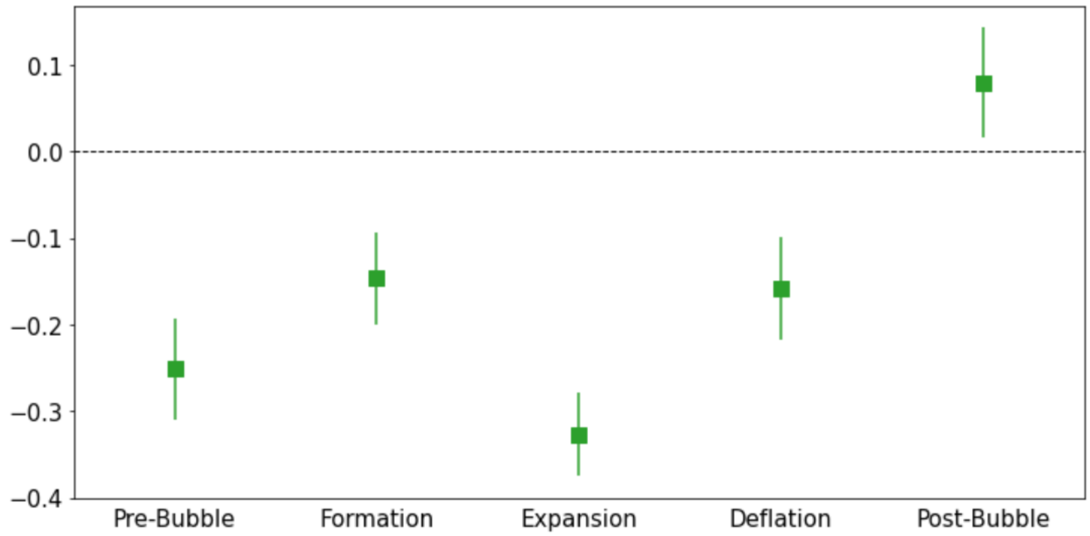
Notes: We classify retail investors into five groups according to their average stock market wealth in the previous year. This table reports $\hat{\gamma}_W$ from the following regression:

$$\hat{\beta}_{k,i,t} = \sum_W \gamma_W \mathbb{1}\left[i \in W\right] + \delta_t,$$

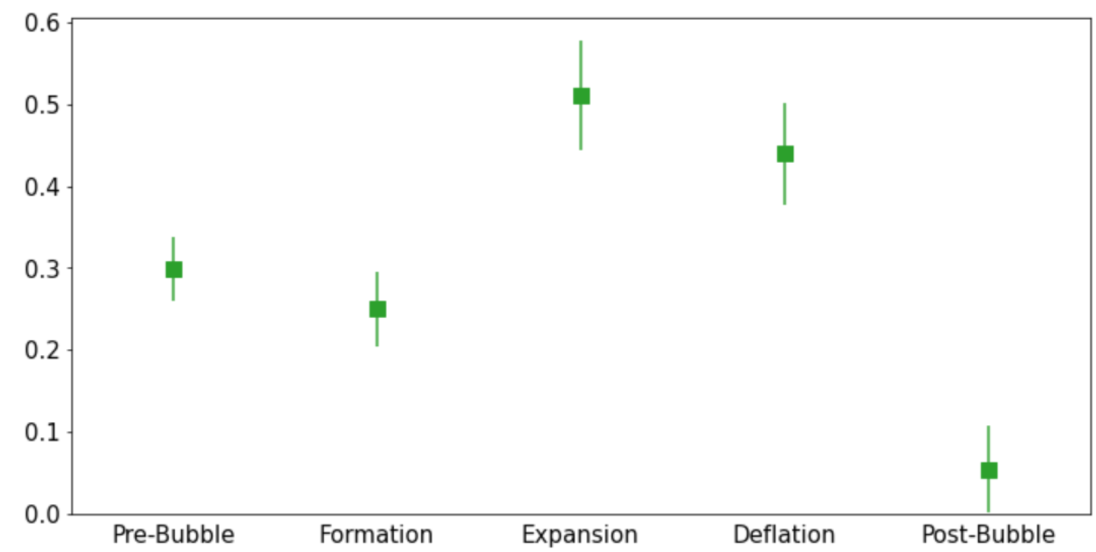
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $W \in \{\text{Wealth1 (Low), Wealth2, Wealth3, Wealth4, Wealth5 (High)}\}$, and $\mathbb{1}\left[i \in W\right]$ is a dummy variable indicating whether investor i belongs to wealth group W, and δ_t are time fixed effects. The time period ranges from Jan 2011 to Dec 2019. Standard errors (in parentheses) are clustered at the investor level. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively. The R-Squared values reported do not include the impact of fixed effects.

Figure 1.12: *Preference/Beliefs of Retail Investors in Different Cohorts*

(a) *Preferences/Beliefs about Profitability*



(b) *Preferences/Beliefs about CAPM Beta*



Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_C$) of retail investors entering the market at different time periods. We classify retail investors by their market entry cohort: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble phase. We focus on the time horizon from 01/2014 through 12/2016 and control for time fixed effects. $\hat{\gamma}_C$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_C \gamma_C \mathbb{1}\left[i \in C\right] + \delta_t,$$

where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $C \in$ {Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}, $\mathbb{1}\left[i \in C\right]$ is a dummy variable indicating whether investor i belongs to cohort C, and δ_t are time fixed effects. Panel (a) reports investors' preferences/beliefs about profitability, and panel (b) reports investors' preferences/beliefs about CAPM beta. Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Table 1.6: *Estimated Preferences/Beliefs of Retail Investors in Different Cohorts*

Outcome Var:	$\hat{\beta}_{\text{Profitability}}$	$\hat{\beta}_{\text{CAPM Beta}}$
	(1)	(2)
Cohort: Pre-Bubble	-0.251***	0.299***
	(0.030)	(0.020)
Cohort: Formation	-0.147***	0.250***
	(0.027)	(0.023)
Cohort: Expansion	-0.326***	0.510***
	(0.024)	(0.034)
Cohort: Deflation	-0.158***	0.439***
	(0.030)	(0.031)
Cohort: Post-Bubble	0.079**	0.054**
	(0.032)	(0.027)
Time Fixed Effects	Y	Y
R-Squared	0.030	0.067
Observations	10,887	10,887

Notes: We classify retail investors by their market entry cohort: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble phase. This table reports $\hat{\gamma}_C$ from the following regression:

$$\hat{\beta}_{k,i,t} = \sum_C \gamma_C \mathbb{1}\left[i \in C\right] + \delta_t,$$

where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $C \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, $\mathbb{1}\left[i \in C\right]$ is a dummy variable indicating whether investor i belongs to cohort C, and δ_t are time fixed effects. The time period ranges from Jan 2014 to Dec 2016. Standard errors (in parentheses) are clustered at the investor level. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively. The R-Squared values reported do not include the impact of fixed effects.

and then report decomposition results.

1.5.1 Return Variance Decomposition

According to the asset demand system in 1.4.1, we can write the price vector $\mathbf{p}_t$ as a function of shares outstanding $\mathbf{s}_t$, stock characteristics $\mathbf{x}_t$, investor AUM $\mathbf{A}_t$, investor preferences/beliefs $\boldsymbol{\beta}_t$, and latent demand $\boldsymbol{\epsilon}_t$:

$$\mathbf{p}_t = \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_t\right). \tag{1.10}$$

The log returns equal to changes in log prices. Therefore, we can decompose log returns as

$$\mathbf{r}_t = \mathbf{p}_t - \mathbf{p}_{t-1} = \Delta\mathbf{p}_t(\mathbf{s}) + \Delta\mathbf{p}_t(\mathbf{x}) + \sum_{i=1}^{I}\left(\Delta\mathbf{p}_t(\mathbf{A}_i) + \Delta\mathbf{p}_t(\boldsymbol{\beta}_i) + \Delta\mathbf{p}_t(\boldsymbol{\epsilon}_i)\right), \tag{1.11}$$

where

$$\begin{aligned}
\Delta\mathbf{p}_t(\mathbf{s}) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_{t-1}, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_{t-1}, \mathbf{x}_{t-1}, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\Delta\mathbf{p}_t(\mathbf{x}) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_{t-1}, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\sum_{i=1}^{I}\Delta\mathbf{p}_t(\mathbf{A}_i) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\sum_{i=1}^{I}\Delta\mathbf{p}_t(\boldsymbol{\beta}_i) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\sum_{i=1}^{I}\Delta\mathbf{p}_t(\boldsymbol{\epsilon}_i) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_t\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_{t-1}\right).
\end{aligned} \tag{1.12}$$

We change each element one at a time and recompute the counterfactual price vector $\mathbf{g}(\cdot)$. For investor-specific factors $\mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_t$, we update each investor-type i's change sequentially. For example, $\Delta\mathbf{p}_t(\mathbf{A}_i)$ denotes the log returns after we've changed investor i's AUM to the next period. $\sum_{i=1}^{I}\Delta\mathbf{p}_t(\mathbf{A}_i)$ denotes the log returns after changing everyone's AUM.

We compute the counterfactual price vector through an iterative process, using Newton's method based on Equation (1.4). The Jacobian matrix used in the iteration has the following analytical form:

$$\frac{\partial f(\mathbf{p};n)}{\partial p(m)} = \begin{cases} \frac{\sum_{i=1}^{I}\beta_{0,i}A_i w_i(\mathbf{p};n)(1-w_i(\mathbf{p};n))}{\sum_{i=1}^{I}A_i w_i(\mathbf{p};n)} & \text{if } m = n \\ \frac{\sum_{i=1}^{I}-\beta_{0,i}A_i w_i(\mathbf{p};n)w_i(\mathbf{p};m)}{\sum_{i=1}^{I}A_i w_i(\mathbf{p};n)} & \text{if } m \neq n \end{cases} \tag{1.13}$$

where $f(\mathbf{p}; n)$ denotes the n-th component of vector $\mathbf{f}(\mathbf{p})$, and $p(m)$ denotes the m-th component of vector $\mathbf{p}$.

Having computed the counterfactual price vectors, we then decompose the cross-sectional variance of log returns as

$$\begin{aligned}\text{Var}\left(\mathbf{r}_t\right) = & \text{Cov}\left(\Delta\mathbf{p}_t(\mathbf{s}), \mathbf{r}_t\right) + \text{Cov}\left(\Delta\mathbf{p}_t(\mathbf{x}), \mathbf{r}_t\right) \\ & + \sum_{i=1}^{I} \text{Cov}\left(\Delta\mathbf{p}_t(\mathbf{A}_i), \mathbf{r}_t\right) + \text{Cov}\left(\Delta\mathbf{p}_t(\boldsymbol{\beta}_i), \mathbf{r}_t\right) + \text{Cov}\left(\Delta\mathbf{p}_t(\boldsymbol{\epsilon}_i), \mathbf{r}_t\right).\end{aligned}$$

We then divide both sides of the equation by $\text{Var}\left(\mathbf{r}_t\right)$ to get shares of each element contributing to the cross-sectional return variance.

1.5.2 Decomposition Results

Decomposition results are reported in Table 1.7. We decompose the variance in cross-sectional stock returns into three main contributors: retail investors, institutions, and stock supply. Columns (1) to (3) report the variance decomposition during different bubble phases: the formation, expansion, and deflation. Column (4) reports the decomposition over the entire bubble period. Within retail investors, we further separate the new entrant contribution from that of existing investors. New entrants are defined as investors who had recently entered the stock market during the specific bubble phase. Among existing investors, we decompose returns into changes from investors' AUM, preferences/beliefs about stock characteristics, and latent demand.

The formation phase: booms start with improved stock fundamentals.

In Column (1) of Table 1.7, we can see that, during the formation phase of the bubble, 47% of cross-sectional return variance comes from retail investors, 30% from institutions, and 23% from stock supply. Focusing on stock supply, changes in stock characteristics explain 21% of return variance. That is, stocks can have higher returns due to good fundamental changes in their characteristics, e.g., higher profits. The results confirm the narrative that booms start with good fundamental changes (displacement).

The expansion phase: new entrants play a dominant role.

Table 1.7: *Cross-Sectional Return Variance Decomposition over Different Bubble Phases*

	% of Variance			
	Formation	Expansion	Deflation	Bubble Overall
Components	(1)	(2)	(3)	(4)
Retail Investors:	**46.6**	**81.2**	**91.1**	**78.1**
New entrants	20.8***	42.9***	19.0***	27.2***
	(7.3)	(8.1)	(5.7)	(4.4)
Existing:				
AUM (**A**)	-0.6	3.7***	2.9**	2.4***
	(1.3)	(1.2)	(1.2)	(0.7)
Preferences/Beliefs ($\boldsymbol{\beta}$)	4.2	11.6**	25.4***	16.3***
	(5.7)	(4.7)	(4.9)	(2.9)
Latent demand (ϵ)	22.2**	23.0**	43.9***	32.3***
	(9.6)	(9.7)	(6.8)	(5.0)
Institutions:	**30.3**	**20.4**	**14.8**	**20.0**
AUM (**A**)	14.6***	11.4***	8.7***	10.8***
	(4.5)	(3.9)	(3.1)	(2.3)
Preferences/Beliefs ($\boldsymbol{\beta}$)	10.4**	5.6**	5.8*	6.7***
	(4.8)	(2.5)	(3.2)	(2.1)
Latent demand (ϵ)	5.3	3.4	0.4	2.4
	(4.9)	(2.1)	(2.7)	(1.9)
Supply:	**23.1**	**-1.6**	**-5.8**	**1.9**
Stock characteristics (**x**)	21.2***	-7.1	1.5	3.0
	(5.9)	(7.2)	(3.4)	(3.2)
Shares outstanding (**s**)	2.0	5.5	-7.3	-1.1
	(8.9)	(8.7)	(8.4)	(5.3)
	100.0	**100.0**	**100.0**	**100.0**

Notes: This table reports the variance decomposition results for cross-sectional stock returns over different bubble phases. Columns (1) to (3) report the variance decomposition over the the formation, expansion, and deflation phases of the bubble. Column (4) reports the decomposition over the entire bubble period. We decompose the variance in cross-sectional stock returns into three main contributors: retail investors, institutions, and stock supply. The bolded numbers of the three main contributors sum up to 100. Within each main contributor, the unbolded numbers sum up to that contributor's bolded number. Within retail investors, we further separate the new entrant contribution from that of existing investors. New entrants are defined as investors who had recently entered the stock market during the specific bubble phase. Among existing investors, we decompose returns into changes from investors' AUM, preferences/beliefs about stock characteristics, and latent demand. The standard errors are clustered at the stock level.

Even though changes in stock characteristics play an important role during the formation phase, they do not matter during the expansion or deflation phase. Focusing on Column (2) of Table 1.7, the decomposition during the expansion of the bubble, we see a much larger role of retail investors, who contribute 81% of cross-sectional stock return variance. Within retail investors, new entrants explain more than half this contribution: investors who newly entered the stock market during the expansion phase contribute 43% of cross-sectional stock return variance.

Our results highlight the importance of late entrants in stock boom-busts: stocks held by late entrants experience much larger booms. This supports anecdotal evidence suggesting that an increasing number of entrants during a sharp price increase can be used as an indicator for widespread manias and to predict future price busts. The "shoeshine boy" indicator by Joseph Kennedy states: "*It is time to get out of the stock market when you got investment tips from a shoeshine boy*." Kindleberger (1978) also calls this period a "follow-the-leader process" where new investors come in after seeing others who profited from trading. According to our suggestive evidence in Section 1.3.3 (Figure 1.5), we also see that the number of new investors entering during the peak of the bubble was 71 times larger than the pre-bubble level. While some earlier papers have studied herd behavior (Banerjee 1992; DeMarzo, Kaniel, and Kremer 2007), the role of entry has not received much attention in the recent literature. Our results show that the number of new investors entering is explosive during the expansion phase of the bubble, and the contribution of late entrants to stock boom-bust is large. Consequently, models on bubbles may find it fruitful to consider entry more seriously.

The deflation phase: existing investors shift preferences/beliefs.

In Column (3) of Table 1.7, the decomposition during the deflation phase, we see that retail investors are still the dominant force. Existing retail investors contribute the most to return variance. Among existing retail investors, changes in investor preferences/beliefs about characteristics ($\boldsymbol{\beta}$) account for 25% of return variance. Investors have shifted their preferences/beliefs substantially during the bubble. For example, investors focus on stock fundamentals, like profitability, during normal market times. During the bubble expansion, however, they no longer focus on profitability due to increasing manias and, when the crash hits, there is a flight to quality: they suddenly change their preferences/beliefs and back in favor of profitability.

Another force that is salient among existing retail investors in Column (3) of Table 1.7 is latent demand. Latent demand accounts for 44% of return variance and is the residual that our demand function does not model. Latent demand likely captures the unexplained panic among retail investors during the bubble deflation.

Overall: retail investors dominate.

The Column (4) of Table 1.7 summarizes the overall contribution of each factor during the entire bubble. Overall, retail investors contribute 78% of cross-sectional return variance, institutions contribute the remaining 20% of variance, and the stock supply side is minimal. Among retail investors, new entrants are very influential by contributing 27% of return variance, while existing retail investors primarily influence stock returns by changing preferences/beliefs (16% of variance).[10]

Wealth reallocation among existing retail investors does not seem important.

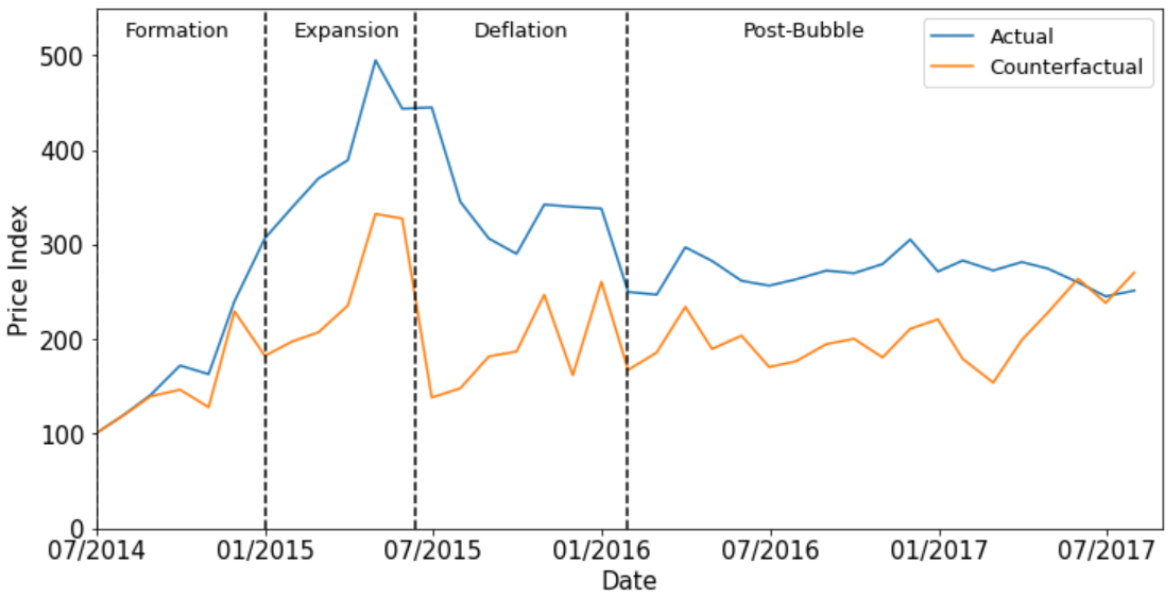
Table 1.7 also shows that the wealth reallocation channel (**A**) among existing retail investors does not seem important in generating cross-sectional differences in stock returns. Changes in retail investors' AUM only contribute 2.4% of return variance, on average. Some recent theoretical papers with heterogeneous investors have emphasized the importance of how wealth reallocation could generate high market volatility (Atmaz and Basak 2018; Martin and Papadimitriou 2022a). We show that this is not the case for cross-sectional stock returns.

10. It is worth noting that our results are about cross-sectional stock returns. The drivers of stock boom-busts in the cross section can be quantitatively different from those that lead to aggregate bubbles. For example, if institutions hold and trade a more diversified portfolio, they will mechanically have less influence on the cross section of stocks but may still be quite influential in driving aggregate stock market movements.

1.6 Counterfactuals

We have seen that existing investors shifting their preferences/beliefs and new investor entry are the two most important channels for determining differences in cross-sectional stock returns. We now conduct counterfactual exercises, computing counterfactual prices for boom-bust stocks if these two channels were to be removed.

Figure 1.14: *Actual vs. Counterfactual Price Index for Boom-Bust Stocks with Constant Retail Investor Preferences/Beliefs*



Notes: The blue line plots the actual price index for boom-bust stocks, and the orange line plots the counterfactual price index for boom-bust stocks where retail investors do not shift their preferences/beliefs. For the counterfactual implementation, we first compute each retail investor type's average pre-bubble preferences/beliefs ($\boldsymbol{\beta}_t$), and then hypothetically let all retail investors hold their average pre-bubble preferences/beliefs throughout the bubble. The other four elements ($\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \epsilon_t$) are unchanged. We compute the counterfactual price vector under this scenario.

1.6.1 What If Preferences/Beliefs Did Not Vary over Time?

In the previous section, investors shifting their preferences/beliefs is an important channel for stock boom-busts. Investors' attitudes towards stock characteristics change substantially over different bubble phases. This shift in preferences/beliefs may be due to a variety of reasons such as over-confidence, manias, panics, and time-varying sentiments. One may wonder, if we hold these time-varying preferences/beliefs constant over the bubble period, what would the prices for boom-bust stocks behave? We can experiment with such a hypothetical environment using our estimated asset demand system.

We first compute the average of each retail investor type's pre-bubble preferences/beliefs ($\boldsymbol{\beta}_t$), and then hypothetically make all retail investors hold their average pre-bubble preferences/beliefs throughout the bubble. The other four elements ($\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\epsilon}_t$) are unchanged. We then compute the counterfactual price vector under such scenario.

The counterfactual results are reported in Figure 1.14. The blue line plots the actual price index for boom-bust stocks, and the orange line plots the counterfactual price index where retail investors do not shift their preferences/beliefs. We see that the counterfactual price path flattens. The counterfactual cumulative return at the peak of the boom-bust cycle (05/2015) is reduced by 41%. The price path differences are concentrated over the expansion and deflation phases, which is consistent with our return decomposition results in Table 1.7.

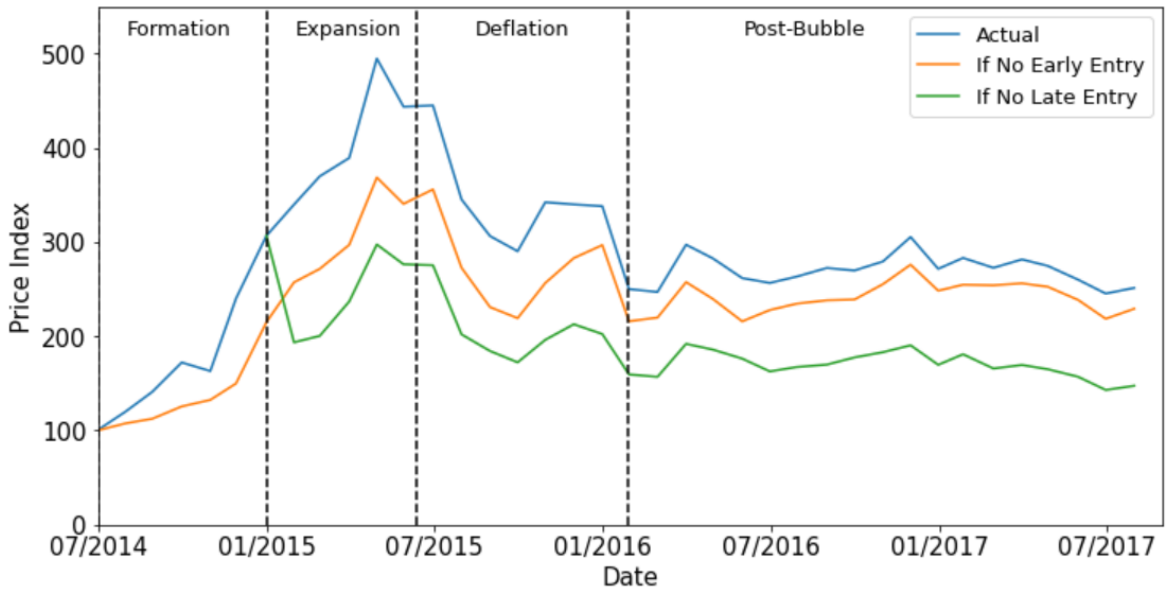
Our results on investor preferences/beliefs changes add to the literature focusing on time-varying investor beliefs or sentiments (Barberis, Shleifer, and Vishny 1998; Scheinkman and Xiong 2003; Baker and Wurgler 2006). We show that this channel is especially important during the expansion and deflation phases of the bubble.

1.6.2 What If No New Investors Entered?

Given that new entrants are influential in generating stock boom-busts, one may ask: if these new entrants did not enter the market, would there still be a boom-bust cycle? We can again carry out a counterfactual exercise where we hypothetically do not allow new entrants to participate in the market. Figure 1.15 reports the counterfactual price paths under such a scenario.

In Figure 1.15, the blue line plots the actual price index for boom-bust stocks. The orange line reports the counterfactual price index where we exclude investors who entered during the formation phase of the bubble. While we see that there is still a boom-bust, the size is noticeably smaller. The counterfactual cumulative return at the peak of the boom-bust cycle

Figure 1.15: *Actual vs. Counterfactual Price Index for Boom-Bust Stocks with No Retail Entry*



Notes: The blue line plots the actual price index for boom-bust stocks. The orange line plots the counterfactual price index where we exclude investors who entered during the formation phase of the bubble. The green line plots the counterfactual price index where we allow for early entrants but exclude investors who entered during the expansion phase of the bubble.

(05/2015) is reduced by 32%.

The green line in Figure 1.15 plots the counterfactual price index where we allow for early entrants but exclude investors who entered during the expansion phase of the bubble. Interestingly, we see that the price index falls at the beginning of 2015, suggesting that if late entrants had not been there, the price for boom-bust stocks would have decreased due to selling pressure from existing investors. In actuality, the price index for boom-bust stocks kept increasing because of the enormous inflow from late entrants. If there had not been any late entrants, the counterfactual cumulative return at the peak of the boom-bust cycle (05/2015) would be reduced by 50%.

1.7 Conclusion

This paper studies cross-sectional differences in stock boom-busts during the 2015 Chinese stock market bubble. We estimate heterogeneous investor demand and quantify the importance of various channels contributing to the variance in cross-sectional stock returns. We find that fundamental changes in stocks are important for explaining returns during the formation phase of the bubble, but not during the expansion and deflation phases. Instead, new investors (who entered during the expansion phase) significantly affect cross-sectional stock returns, accounting for 43% of return variance during the expansion phase. Existing investors influence stock returns mainly by changing their preferences or beliefs, contributing 25% to return variance over the deflation phase and an average of 16% over the entire bubble period. Overall, retail investors are the main drivers of cross-sectional stock returns (78% of variance). Institutions have a low demand elasticity and therefore do not offset the demand from retail investors.

Even though this paper focuses on explaining stock boom-busts in the cross-section, the lessons we learned have the potential to be extrapolated to aggregate bubbles as well. Our results indicating that booms start with good fundamental news, that price hikes are heavily influenced by new entrants, and that time-varying preferences or beliefs have important effects seem to be closely linked to influential accounts of aggregate bubbles by Kindleberger (1978) and others. We leave a formal mapping from the cross section to the aggregate for future research.

The two important channels, time of entry and time-varying preferences or beliefs, are taken as exogenous in the current paper. It will be interesting to endogenize these two forces to allow for more dynamics on the interplay among different forces. Identifying forces that affect investor entry and influence their preferences/beliefs may give us a dynamic model explaining the evolution of the cycle.

Finally, our results on how retail investors drive extreme stock price movements may be of particular interest for other stock markets that have experienced an increase in retail involvement, such as the U.S. stock market since 2020 (Bryzgalova, Pavlova, and Sikorskaya 2022). Understanding the price impacts of retail trading behavior is important for related policy discussions and market regulations.

Chapter 2

Who Chases Returns? Evidence from the Chinese Stock Market[1]

2.1 Introduction

How do retail investors trade? Is there heterogeneity in their trading behavior? Understanding retail investors' heterogeneous trading patterns is important for assessing households' financial welfare as well as for understanding price dynamics for stocks held mainly by retail investors. In this paper, we focus on return chasing, a trading behavior that conventional investing wisdom regards as a serious mistake.[2] Earlier studies have shown return chasing (alternatively, positive feedback trading or momentum trading) theoretically and empirically at the aggregate investor level (Hong and Stein 1999, De Long et al. 1990b, Grinblatt and Keloharju 2000, Barber, Odean, and Zhu 2009). However, little is known about the heterogeneity of return chasing among retail investors.

Ours is the first paper to study the heterogeneous return chasing propensities of individual investors. We use administrative investor account data from the Shanghai Stock

1. Co-authored with Weihua Chen and Donghui Shi

2. E.g., CNBC, Apr 7, 2022: "*7 Biggest Investing Mistakes, According to Experts*"; MarketRiders, Nov 30, 2021: "*Investment Mistakes to Avoid: Chasing Returns*"; MarketWatch, Aug 24, 2016: "*Here's How to Fix the Biggest Investing Mistake You've Been Making*".

Exchange. Our dataset covers a representative sample of 18 million retail investors from January 2011 through December 2019 at a monthly frequency. We obtain each account's holding and trading records, as well as investor demographics. The detailed micro data enable us to extend the existing literature by examining return chasing behavior at the individual investor level.

We ask three questions in the project: Who are return chasers? How do they perform? Does the entry and exit of return chasers impact future stock returns? We construct a new measure—Return Chasing Propensity (RCP hereafter)—to capture each investor's return chasing behavior based on her account trading and holding records.

To construct RCP, we look at each investor's stock holdings, match each stock to its returns prior to the time of purchase, and average these past returns across stocks to obtain an individual's RCP. Our RCP construct has the advantage that it is not contaminated by retail investors' inertia behavior. Retail investors are documented to exhibit high levels of inertia (Agnew, Balduzzi, and Sunden 2003; Calvet, Campbell, and Sodini 2009): they may not trade for several months purely because of inattention. If we simply look at the recent past returns of stocks held by retail investors, it is likely that we observe only random stock fluctuations at the moment of measurement rather than identifying the investor type. Our RCP changes only when there is active trading, and hence is not influenced by inertia.

RCP captures investors' tendency to buy stocks with recently high past returns. A natural question to ask is, who are return chasers? We find that RCP can be predicted by various investor characteristics that are associated with the investor being less sophisticated. Specifically, people who are younger, are less experienced in the stock market, have less wealth invested in the stock market, or have a stronger disposition effect are more likely to exhibit higher RCP. Investors who trade more frequently tend to have higher RCP as well. This pattern is consistent with previous findings that frequent retail traders are usually unsophisticated (Barber and Odean 2000; Barber and Odean 2001).

How do different retail investors perform? To answer this question, we use investor characteristics to predict each investor's future returns. We find that more sophisticated investors—investors who are older, are more experienced, hold a more diversified portfolio, or have a weaker disposition effect—are more likely to earn higher future returns. In particular, investors with higher RCP are significantly more likely to earn lower future returns, controlling for all the other investor characteristics. Among all the investor characteristics, RCP has the strongest predictive power for investors' future returns. A one standard deviation increase in RCP leads to 5.6% lower investor yearly returns on average.

One may wonder how RCP could play a role at the market aggregate level. We calculate market aggregate retail RCP, which is the wealth-weighted average RCP among all retail investors. We find that aggregate retail RCP closely tracks the market price index. Note that, with our RCP construct, we deliberately subtract market returns so that there is no mechanical relationship between aggregate retail RCP and the market index. Nonetheless, aggregate retail RCP co-moves closely with the market ups and downs.

We next shift our attention to the cross section of stocks. High RCP investors are attracted to stocks that have higher past returns, are larger in size, are more likely to be over-valued, and experience higher volatility and turnover. We construct a measure for stock-level return chasing ownership (RCO hereafter) by wealth-weighting RCP from retail investors who hold a particular stock (leaving this stock out of the investor RCP calculation to eliminate mechanical linkages between this stock's RCO and its past returns). To clarify the abbreviations we use: when we talk about RCP (return chasing propensity), we are talking about return chasing behavior at the **investor level**; when we talk about RCO (return chasing ownership), we are talking about the ownership structure, especially the composition of return chasers, at the **stock level**. We find that new investor entry is the most important driver of stock RCO changes. Stocks held by more return chasers (high RCO stocks) are more likely to earn lower future returns. We therefore implement a trading strategy that goes long stocks with low RCO and shorts stocks with high RCO. This strategy generates an annualized Sharpe ratio of 0.94. The RCO trading factor outperforms standard Fama/French factors in the Chinese stock market during our sample period.

We verify interaction effects between RCO and other stock attributes. Among stocks that have a higher retail share, the RCO trading strategy performs even better, with a more than doubled annualized Sharpe ratio of 1.94. We also find that the RCO trading strategy works better among stocks with higher past returns. We then show suggestive evidence that return chasers have price impacts on stocks they hold.

We next construct various other investor-characteristic ownership signals by aggregating investor characteristics at the stock level in the same way as for RCO. We find that RCO is one of the two strongest investor characteristic-based predictors of stock returns. RCO still significantly predicts future stock returns, after controlling for all the other ownership signals. The results show that RCO delivers distinctive information compared to other investor-characteristic ownership signals.

Finally, we decompose return chasing into industry return chasing and stock idiosyncratic return chasing. We find that more unsophisticated investors tend to chase stock idiosyncratic returns more often, which indeed hurts investor returns the most. Industries that are held by more industry return chasers under-perform. Within an industry, stocks that are held by more stock idiosyncratic return chasers also under-perform.

Related literature

This paper contributes to the household finance literature by being the first one to demonstrate heterogeneous return chasing tendencies at the individual investor level. Previous literature has shown, theoretically and empirically, positive feedback trading at the aggregate investor level (De Long et al. 1990b, Hong and Stein 1999, Hong, Lim, and Stein 2000, Barber, Odean, and Zhu 2009). We are able to investigate the heterogeneity of return chasing because of our very granular account-level data.[3]

There is a vast literature documenting heterogeneity in other investor trading patterns, although not for return chasing. Richer and better educated investors tend to have a weaker disposition effect (Dhar and Zhu 2006, Calvet, Campbell, and Sodini 2009), hold a more diversified portfolio (Calvet, Campbell, and Sodini 2007, Goetzmann and Kumar

3. Previous papers that also use the Shanghai Stock Exchange data include Bailey et al. (2009) and An, Lou, and Shi (2022).

2008, Gaudecker 2015), and display less inertia (Campbell 2006, Bilias, Georgarakos, and Haliassos 2010, Bianchi 2018). Other investor heterogeneities studied include differences in investor IQ (Grinblatt, Keloharju, and Linnainmaa 2012), ambiguity aversion (Dimmock et al. 2016), and individual experiences of macroeconomic shocks (Malmendier and Nagel 2011). Earlier papers often use survey data, nonrepresentative samples, or census data with yearly or even lower frequency. Our data has the advantage of its large scale, high frequency (monthly), and representativeness.

Return chasing behavior can result from investors' heterogeneous tendencies to extrapolate (Greenwood and Shleifer 2014; Barberis et al. 2015). We offer a framework to quantify this heterogeneity for retail investors and show that it has important impacts on both investor and stock returns. Liao, Peng, and Zhu (2021) use brokerage data to show that interaction between extrapolative beliefs and the disposition effect contributed to the high trading volume during the 2015 Chinese stock market bubble. While we do not focus on trading volume in this paper, we do confirm that investors with a stronger disposition effect are more likely to have higher RCP. And we show that this correlation exists not only during the bubble period but also during normal market periods.

Our paper also contributes to a small but growing literature that uses a stock's ownership structure to predict its future returns. Nagel (2005), Asquith, Pathak, and Ritter (2005) and Porras Prado, Saffi, and Sturgess (2016) find that stocks with low institutional ownership tend to under perform as a result of binding short-sale constraints. Morck, Shleifer, and Vishny (1988) and McConnell and Servaes (1990) study the impact of management ownership on a stock's market valuation from the corporate governance perspective. Choi, Jin, and Yan (2012) and Chen (2018) find that stocks with high ownership breadth of small investors under perform in the future. Our paper pushes the idea of stock ownership even further: we construct various kinds of investor-characteristic ownership signals, and find that RCO is one of the two strongest investor-characteristic ownership signals in predicting stock returns. We also find that a trading strategy that interacts RCO with a stock's retail share performs much better than using only retail share information or only RCO information.

Our result is consistent with previous findings that high retail share stocks are more likely to be over-valued, and also adds that heterogeneity within retail investors very much matters.

Organization of the paper

The remainder of our paper is organized as follows. In Section 2.2, we illustrate the measure for RCP and use Shanghai Stock Exchange data to calculate investor RCP. In Section 2.3, we examine the relationships between RCP and other investor characteristics, and further investigate the impact of RCP on future investor returns. In Section 2.4, we calculate stock level RCO and examine the stock attributes that attract return chasers. In Section 2.5, we construct various trading strategies based on stock RCO. In Section 2.6, we construct other investor-characteristic ownership signals and compare them with RCO. In Section 2.7, we decompose return chasing into industry return chasing and stock idiosyncratic return chasing. We conclude in Section 2.8.

2.2 Measure for Return Chasing Propensity (RCP)

In this section, we use investors' asset holding and trading data to construct an empirical measure for return chasing propensity (RCP).[4] We first illustrate our RCP construction, and then use the Shanghai Stock Exchange data to calculate investor RCP.

2.2.1 RCP Definition

We want to capture investors' heterogeneous propensities to chase returns. One challenge for measuring RCP is that retail investors tend to have high inertia (Agnew, Balduzzi, and Sunden 2003; Campbell 2006). It is quite common that a retail investor does not trade for several months purely because of inattention. If we simply look at the recent past returns on stocks that retail investors hold, it is likely that we capture random stock performance at the moment of measurement rather than identifying the investor type. To address this

4. In Appendix B.1, we use a simple model to motivate the idea that different propensities to chase returns can result from investors' heterogeneous tendencies to extrapolate. It is widely documented that people tend to extrapolate stocks' past performance to their future returns (Greenwood and Shleifer 2014; Barberis et al. 2015).

concern with inertia, we propose examining stock returns prior to the time of purchase. It is safe to assume that retail investors paid attention at least when they purchase a stock.

Formally, we construct the following individual level RCP for investor i at time t:

$$\mathrm{RCP}_{i,t} = \sum_{s \in \{S_{i,t}\}} \left[\gamma_s \left(R^s_{t(s)-K,t(s)} - R^M_{t(s)-K,t(s)} \right) \right], \tag{2.1}$$

where $S_{i,t}$ is the collection of stocks that investor i holds at time t. For each stock s that the investor holds at time t, we first identify the time when the investor purchased this stock: the purchase time is denoted as $t(s)$. We then calculate the K-period cumulative returns from time $t(s) - K$ to $t(s)$ for stock s, which is denoted as $R^s_{t(s)-K,t(s)}$. We subtract the corresponding market index returns $R^M_{t(s)-K,t(s)}$ during the same period from the cumulative stock returns.[5] Finally, we weight each stock s by weight γ_s.

For weight γ_s, we take two versions: the first is the simple arithmetic average so that $\gamma_s = 1/|S_{i,t}|$, where $|S_{i,t}|$ is the number of stocks that investor i holds at time t. The second is the weighted average so that $\gamma_s = v^s_{t(s)} / \sum_{\kappa \in \{S_{i,t}\}} v^\kappa_{t(\kappa)}$. $v^\kappa_{t(\kappa)}$ is stock κ's held value at the time of purchase $t(\kappa)$, again trying to eliminate the impact of retail investors' inertia.

Let us illustrate the past returns prior to the time of purchase $R^s_{t(s)-K,t(s)}$ with a simple example. Suppose that we try to calculate an investor's RCP on 12/31/2019. On 12/31/2019, the investor held Apple in her account. She purchased Apple on 9/20/2019. If $K = 12$ months, $R^s_{t(s)-K,t(s)}$ is the return on Apple from 9/20/2018 to 9/20/2019.

We note several advantages of measuring investors' RCP in this way. First, we use stock returns prior to the time of purchase, so an investor's RCP changes only when active trading occurs. Therefore our measure is not contaminated by inertia. Second, we subtract the corresponding market returns from stock returns. There is no mechanical correlation between the market performance and the RCP magnitude. Moreover, when we use value-weighted weights for RCP construction, a market index trader will always have zero RCP.

5. We have also tried to measure investor RCP using CAPM beta adjusted stock returns. We replicate results in Section 2.3 using this alternative RCP measure, as reported in Appendix B.2. The results are robust to this alternative RCP measure. We choose our current definition of RCP as in Equation (A.4) because we think this simple procedure is more likely to be aligned with retail investors' reasoning process.

Third, because we calculate RCP based on an investor's holding records, we observe RCP whenever the investor holds stocks in her account. In this way, we can think of RCP as capturing the information from a long-run average of past purchase actions. If we instead use only trading data, there will be a lot of missing values for the days when investors do not trade.

2.2.2 SSE Data

We obtain daily administrative investor account holding and trading records from the Shanghai Stock Exchange. There are two stock exchanges in China: the Shanghai Stock Exchange and the Shenzhen Stock Exchange. Stocks can be listed on only one of the two exchanges. The Shanghai Stock Exchange is the larger of the two, accounting for 62% of the market capitalization of these two exchanges between 2011 and 2019. We have complete information for all the stocks that are traded on the Shanghai Stock Exchange (SSE hereafter). That is, if we sum investor holdings for a particular stock, it is equal to the total shares that the firm issues. From 2011 through 2019, there are in total 87 million accounts on SSE. Retail investors account for 84% of the trading volume and hold 60% of the value of stocks in circulation. Table 2.1 states summary statistics for the entire investor population.

For Table 2.1, we separate the investors into three main groups: retail investors, institutions, and corporate. No. Accounts is the number of total accounts for each group during 2011-2019. We can see that retail investors hold the majority of accounts. Trading Volume is the daily average traded CNY value of all the stocks. Retail investors contribute 84% of the total trading volume. Size is the investors' average CNY value of all stock holdings in SSE. We can see that Corporate has the largest share of size, but those shares are seldom traded: they serve as either corporate assets or state-owned capital (this can be verified through the small corporate trading volume). As a result, actively traded shares are held mainly by retail investors and institutions. Retail investors hold about 50% more market capitalization than institutions.

In this paper, we focus on retail investors. Retail investors occupy a much larger share of

Table 2.1: *Summary Statistics for Investor Groups on SSE, 2011-2019*

	Retail	Institutional	Corporate	Total
	(1)	(2)	(3)	(4)
No. Accounts (Million)	87.39	0.08	0.11	87.58
Trading Volume (CNY Billion)	296.12	50.48	7.22	353.82
Trading Volume (%)	84%	14%	2%	100%
Size (CNY Trillion)	4.54	3.01	12.64	20.20
Size (%)	22%	15%	63%	100%

Notes: We separate the full investor population of the Shanghai Stock Exchange from Jan 2011 through Dec 2019 into three main groups: retail, institutional, and corporate investors. No. Accounts is the total number of accounts held by each group from 2011 through 2019. Trading Volume is the daily average traded CNY value of all stocks. Size is the average CNY value of all stock holdings on SSE. The exchange rate is roughly 6.5 CNY = 1 USD.

the Chinese stock market than the U.S. stock market, which makes it extremely important to understand retail traders' behavior in China.[6] All of the following analyses are based on a random sample of 18 million retail investor accounts (about 20% of the entire population), for a period of time that runs from January 2011 through December 2019. We analyze the data at a monthly frequency. We also obtain investor characteristics including age, trading experience, and gender.

2.2.3 Calculate Investor RCP

With our accounts' trading and holding data, we can now calculate the investor RCP according to Equation (A.4). For robustness, we let past returns length K vary over 1 month, 2 months, 3 months, 6 months, and 12 months. We also use the weight γ_s for both simple arithmetic average and weighted average. In this way, we obtain $5 \times 2 = 10$ different RCP measures.

We can assess the RCP measures by checking their predictive power for future return chasing purchases. To do that, we first identify each investor's stock purchase actions. For

6. According to McCrank (2021), retail investors accounted for 17.1% of the trading volume in the U.S. stock market in January 2020.

each stock purchase event, we use the investor's RCP before the purchase to predict the newly purchased stocks' RCP attribute. Formally, we run the following regression:

$$\text{RCP}_{i,t(p)}^{\text{New Purchased}} = \alpha + \rho \text{RCP}_{i,t(p)-1} + \epsilon_{i,t(p)}, \tag{2.2}$$

where $t(p)$ are the timestamps when investor i purchases stocks. $\text{RCP}_{i,t(p)}^{\text{New Purchased}}$ is investor i's newly purchased stocks' RCP at time $t(p)$, using only newly purchased stocks to calculate RCP according to Equation (A.4). $\text{RCP}_{i,t(p)-1}$ is investor i's RCP before the purchase date $t(p)$. We run this regression on different RCP measures. Appendix Table B.14 reports the regression results for all the RCP measures.

We find that $\text{RCP}_{Arithmetic}^{K=12}$ has the highest regression coefficient ρ of 0.221 and the highest R-squared of 3.7%, which indicates that RCP calculated using a simple arithmetic average of of the preceding 12 months' returns generates the least noise in capturing investors' return chasing behavior.[7] Therefore, we use $\text{RCP}_{Arithmetic}^{K=12}$ for all the main analyses of the paper.[8] The other 9 RCP measures are used for robustness checks. The results are very similar.

2.3 Investor Return Chasing Propensity (RCP)

In this section, we investigate investor RCP in several directions. We ask, who are high RCP investors? How do they perform? We also show the time series of investor RCP at the market aggregate level.

2.3.1 Who Are High RCP Investors?

Now that we have computed investor RCP, it is natural to ask which investors exhibit high RCP. To answer this question, we examine the relationships between RCP and other investor

7. The R-squared is small, which is not surprising because retail investors' purchasing behavior is known to be noisy and hard to predict (Merkle and Weber 2014).

8. We also carry out another test to regress each RCP measure on individual fixed effects, and check how much of the R-squared can be explained simply from these individual fixed effects. Table B.15 reports the R-squared for the 10 different RCP measures. The R-squared from these fixed effects regressions are stable at around 0.4–0.5, with the highest R-squared achieved by $\text{RCP}_{Arithmetic}^{K=12}$. The results again indicate that measure $\text{RCP}_{Arithmetic}^{K=12}$ captures the most persistent investor behavior.

characteristics.

Table 2.2: *Summary Statistics for Individual Investors*

	Panel (Investor-by-Month)					Cross-Sectional Variation		Time-Series Variation	
	Mean	Std.Dev.	25th	50th	75th	Mean	Std.Dev.	Mean	Std.Dev.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RCP (%)	32.26	80.77	-8.52	10.77	45.88	36.64	66.32	-0.76	10.51
Disposition Effect (%)	6.16	11.23	0.00	1.64	9.09	7.38	10.88	-0.02	0.90
Log(AvgInvest)	10.31	1.58	9.29	10.29	11.32	10.01	1.59	-0.00	0.00
Number Stocks	2.63	3.01	1	2	3	2.20	1.99	-0.00	0.16
HHI	0.74	0.29	0.50	0.90	1.00	0.80	0.20	-0.00	0.01
Trade Frequency	7.93	111.99	0	0	3	10.13	99.00	-0.15	4.36
VolatilityStockHeld (%)	3.38	3.93	2.37	2.93	3.74	3.37	2.76	-0.02	0.30
Age	45.96	12.52	36.58	45.00	54.08	42.18	12.54	-0.14	1.80
Experience	6.43	3.84	3.25	6.42	9.42	4.29	3.58	-0.14	1.79
Female	0.49	0.50	0	0	1	0.44	0.50	0.00	0.00

Notes: Columns (1) through (5) report summary statistics for the panel data, which contains investor-by-month observations from 18,147,831 investor accounts from Jan 2011 through Dec 2019. Columns (6) and (7) report cross-sectional moments of time-series means calculated for each account over its active life, giving equal weight to each account. Columns (8) and (9) report time-series moments of cross-sectional means (after demeaning each investor's data using the investor's specific time average), giving equal weight to each month. RCP is constructed by averaging stock returns prior to the time of purchase across all held stocks (Equation (A.4)). Disposition effect is defined as the difference between the proportion of gains realized and the proportion of losses realized over the past one year. Log(AvgInvest) is the log value of average stock market wealth that the investor holds in the past one year. Number Stocks is the number of stocks that the investor holds at the end of each month. HHI is the Herfindahl–Hirschman Index for an investor's portfolio, using the stock value as weight. Trade Frequency is how many times the investor trades in this month. VolatilityStockHeld is the average daily volatility (prior to the time of purchase) of stocks held by the investor. Age is the investor's age. Experience is the investor's investment age. Female is a dummy variable indicating the investor's gender.

Table 2.2 shows summary statistics for RCP and other investor characteristic variables. The first 5 columns report panel summary statistics (observations are investor-by-month). Columns (6) and (7) report cross-sectional moments of time-series means calculated for each account over its active life, giving equal weight to each account. Columns (8) and (9) report time-series moments of cross-sectional means (after demeaning each investor's data using the investor's specific time average), giving equal weight to each month. We can see how the variables vary within an investor over time from the time-series variation. Comparing Columns (2), (7) and (9), we can see that the variation in the panel data stem mainly from the cross-sectional variation: the standard deviation of the cross-sectional data is much larger than the standard deviation of the time-series data. This tells us that the differences across investors are much larger than the differences within an investor across time. Focusing on the panel summary statistics, our calculated RCP has an average of 32% and a standard deviation of 81%. This tells us that an average investor purchases stocks that have past yearly returns 32% higher than the market returns, indicating substantial average return chasing. We can also see that the RCP distribution is right-skewed, implying that some investors chase returns very aggressively. The disposition effect is defined as the difference between the proportion of gains realized and the proportion of losses realized (Odean 1998) over the previous one year. The disposition effect has an average of 6.2%, which is consistent the previous literature that retail investors are more likely to realize gains than losses (Odean 1998; Barber and Odean 2013). Log(AvgInvest) is the log value of average stock market wealth that the investor has held in the preceding year. Number Stocks is the number of stocks that the investor holds at the end of each month. The median of this number is 2, showing that the majority of retail investors are highly under-diversified. Similarly, HHI is the Herfindahl–Hirschman Index for an investor's portfolio, using the stock value as weight. The median HHI of 0.9 again confirms the under-diversification of retail investors. Trade Frequency is how many times the investor trades in a given month. The median trade frequency is 0, which verifies our previous concern that retail investors have high inertia. VolatilityStockHeld is the volatility of stocks held by the investor. It is measured as the average daily stock return volatility during the preceding year before the time of stock purchase, equal-weighted among held stocks. It captures whether the investor prefers buying volatile stocks. The average held stocks' daily volatility is at 3.4%. Age is the investor's age, with a mean at 46 years old. Experience is the investor's investment age, with mean of 6.4 years. Finally, Female is a dummy variable indicating the investor's gender.

One may be concerned that investors' portfolio diversification can influence RCP. Since we take either the arithmetic or weighted average of the stocks' past returns for the RCP calculation, the less diversified the investor's portfolio is, the more volatile the RCP measure will be. To address this concern, we control for both Number Stocks and HHI up to third-order polynomials in all our regressions.

Table 2.3: *Relationship Between RCP and Other Investor Characteristics*

Outcome Var:	RCP	RCP	$\text{RCP}_{t(p)}^{\text{New Purchased}}$
	(1)	(2)	(3)
RCP			0.155***
			(0.015)
Disposition Effect	0.108***	0.151***	0.002
	(0.023)	(0.012)	(0.017)
Trade Frequency	0.002**	0.002***	5.29e-04**
	(6.90e-04)	(6.26e-04)	(2.13e-04)
VolatilityStockHeld	0.473***	0.593***	0.084***
	(0.025)	(0.038)	(0.016)
Log(AvgInvest)	-0.634***	-0.777***	-0.257
	(0.108)	(0.088)	(0.209)
Age	-0.438***		-0.360***
	(0.044)		(0.057)
Age_t^2	0.0036***		0.003***
	(3.14e-04)		(4.09e-04)
Experience	-1.497***		-1.993***
	(0.422)		(0.264)
Experience_t^2	0.0907***		0.102***
	(0.026)		(0.017)
Female	1.734***		0.286
	(0.196)		(0.376)
Time Fixed Effects	Y	Y	Y
Investor Fixed Effects	N	Y	N
R-Squared	0.076	0.120	0.024
Observations.	662,470,542	663,944,347	110,548,377

Notes: Columns (1) and (2) report contemporary regression results: the unit of observation is individual-by-month. Column (3) report predictive regression results: the unit of observation is investor-by-purchase timestamp. Standard errors (in parentheses) are double-clustered at the investor and date levels. All regressions control for Number Stocks and HHI up to the 3rd polynomials. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively. The R-Squared values reported do not include the impact of fixed effects.

We next check the relationships between RCP and other investor characteristics. We run both contemporaneous regressions and predictive regressions. For the contemporaneous regressions, we regress RCP on the current-period investor characteristics:

$$\mathrm{RCP}_{i,t} = \alpha + \gamma X_{i,t} + \delta_t + \epsilon_{i,t}, \tag{2.3}$$

where $X_{i,t}$ are characteristics for investor i at time t, and δ_t are time fixed effects. The contemporaneous regressions help us examine the relationships between RCP and other investor characteristics in the same period. The results are reported in Columns (1) and (2) of Table 2.3.

For the predictive regressions, we want to ask whether investor characteristics $X_{i,t}$ are useful to predict future return chasing purchases. Because RCP changes only with active trading behavior, we use the next newly purchased stocks' RCP, i.e., $\mathrm{RCP}_{i,t(p)}^{\text{New Purchased}}$, as the dependent variable:

$$\mathrm{RCP}_{i,t(p)}^{\text{New Purchased}} = \alpha + \beta \mathrm{RCP}_{i,t(p)-1} + \gamma X_{i,t(p)-1} + \delta_{t(p)} + \epsilon_{i,t(p)}, \tag{2.4}$$

where the unit of observation is the investor-by-purchase timestamp. $t(p)$ are the timestamps when investor i purchases stocks. $\mathrm{RCP}_{i,t(p)}^{\text{New Purchased}}$ is investor i's newly purchased stocks' RCP at time $t(p)$, using only newly purchased stocks to calculate RCP (defined the same as in Regression (2.2)) . $\mathrm{RCP}_{i,t(p)-1}$ and $X_{i,t(p)-1}$ are investor i's RCP and other characteristics one month before the purchase date $t(p)$. $\delta_{t(p)}$ are again time fixed effects. The predictive regression results are reported in Column (3) of Table 2.3.

In Table 2.3, Column (1) reports the regression coefficients between RCP and other investor characteristics in the current period. We see that investors with a higher disposition effect are more likely to have higher RCP. A one standard deviation increase in the disposition effect (11.23%) leads to a 1.2% increase in RCP. Investors who trade more frequently also have higher RCP. This is consistent with findings from the previous literature (Barber and Odean 2000; Barber and Odean 2001) that frequent retail traders tend to be unsophisticated. Investors who hold more volatile stocks are more likely to have higher RCP. Investors with more stock market wealth tend to have lower RCP. Age has a negative impact on RCP, which means that older people tend to be more cautious and chase returns less. The effect on the quadratic term of Age is positive, which means that the effect of increasing age on RCP diminishes and changes sign at age 61 (about 12.5% of our observations involve investors who are above 61 years of age). Similar patterns exist for the stock market experience: more experienced investors are likely to have lower RCP but this sign also reverses at 8.3 years (about 33.6% of our observations involve investors with more than 8.3 years of investing experience). Finally, we see that females tend to have higher RCP. But the magnitude of the gender effect is very small. Females have 1.7% higher RCP than males, which accounts for only 2.1% of standard deviation of RCP.

Column (2) of Table 2.3 further adds investor fixed effects in the contemporaneous regression. We see that all of the time-varying variables have the exact same signs as those reported in Column (1), and magnitudes are similar. Column (3) shows the predictive regression results obtained using past investor characteristics to predict future return chasing behavior. Past RCP strongly predicts future purchased stocks' $\text{RCP}_{i,t(p)}^{\text{New Purchased}}$, which means that return chasing behavior is persistent, as shown in Section 2.2.3. For all other investor characteristics, the predictive regression results show the same patterns as in the contemporaneous regressions.

The results reported in Table 2.3 deliver a consistent message: investors with higher RCP are likely to be unsophisticated. They tend to be younger and less experienced, have a stronger disposition effect, trade more frequently, hold more volatile stocks, and have less wealth in the stock market.

2.3.2 How Do High RCP Investors Perform?

Given our descriptive portraits of high RCP investors, one may wonder, do high RCP investors lose or win money in the future? To answer the question, we predict investors' future returns based on various investor characteristics. We calculate investor returns as the ratio of an investor's current month's profits over her wealth in the stock market at the end of the preceding month. Formally,

$$r_{i,t} = \frac{\Pi_{i,t}}{W_{i,t-1}}, \tag{2.5}$$

where $r_{i,t}$ is the return for investor i at month t. $\Pi_{i,t}$ is the amount of profits that investor i earns in the stock market during month t. $W_{i,t-1}$ is investor i's stock market holding value at the end of the preceding month $t-1$.

This return calculation may not be accurate if there are large intra-month flows. For example, if an investor experiences a large capital inflow at the beginning of a given month, it is likely that we underestimate the cost of the investment by using "the previous month-end holding value $W_{i,t-1}$" as the denominator.[9] Especially, investor returns will have extreme values if the denominator—the previous month-end holding value $W_{i,t-1}$—is small. Therefore, we exclude observations that have the previous month-end holding value $W_{i,t-1}$ below the 5% percentile of the holding value distribution, which is 1,214 CNY (about 187 USD). In this way, monthly investor return has an average of -0.2% and a standard deviation of 197%. We can see that even after excluding observations that are likely to generate return outliers, investor return still has a large standard deviation, verifying that it is necessary to tease out outliers. We have tried other ways to deal with outliers like winsorize returns at different thresholds (1% or 5%) or drop return outliers. The results are robust.

We run the following regression:

$$r_{i,t+k} = \alpha + \beta r_{i,t} + \gamma X_{i,t} + \delta_t + \epsilon_{i,t+k}, \tag{2.6}$$

where $r_{i,t}$ is investor i's monthly return at time t. $r_{i,t+k}$ is the investor i's monthly return for k periods leading to time t. $X_{i,t}$ are investor i's characteristics, including $\text{RCP}_{i,t}$. δ_t are time fixed effects. The regression results are reported in Table 2.4. Standard errors are double-clustered at both the investor and time levels.

In Table 2.4, Columns (1) through (6) use future 1-month through 6-month returns,

9. Investor returns are hard to precisely calculate since there are constant inflows and outflows, and profits are made over different time horizons. We have tried other ways, such as the IRR returns used in Dichev (2007), to calculate the investor returns. The results are robust to alternative methods of calculating the returns.

Table 2.4: *The Impacts of Investor Characteristics on Future Returns*

Outcome Var:	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+4}	Ret_{t+5}	Ret_{t+6}
	(1)	(2)	(3)	(4)	(5)	(6)
Ret_t	0.022	-0.006	-0.022	-0.005	-0.005	-0.006
	(0.014)	(0.015)	(0.025)	(0.008)	(0.048)	(0.005)
RCP_t	-0.483***	-0.390***	-0.325***	-0.275***	-0.245***	-0.220***
	(0.139)	(0.092)	(0.062)	(0.055)	(0.061)	(0.053)
Disposition Effect_t	-0.221***	-0.213***	-0.226***	-0.207***	-0.201***	-0.176***
	(0.054)	(0.047)	(0.053)	(0.051)	(0.052)	(0.046)
$\text{VolatilityStockHeld}_t$	-0.067	-0.065	-0.065*	-0.054*	-0.046	-0.038
	(0.046)	(0.043)	(0.035)	(0.031)	(0.033)	(0.031)
Log(AvgInvest)_t	-0.219	-0.232	-0.221	-0.186	-0.187	-0.080
	(0.174)	(0.191)	(0.173)	(0.199)	(0.194)	(0.146)
HHI_t	-0.363***	-0.227***	-0.204**	-0.196**	-0.178**	-0.177**
	(0.132)	(0.087)	(0.086)	(0.081)	(0.081)	(0.081)
Trade Frequency_t	-0.289	-0.297*	-0.186	-0.222	-0.257**	-0.172
	(0.181)	(0.159)	(0.133)	(0.156)	(0.129)	(0.131)
Age_t	0.066**	0.070**	0.076**	0.074*	0.089**	0.076**
	(0.031)	(0.035)	(0.036)	(0.039)	(0.040)	(0.038)
Experience_t	0.296***	0.270***	0.258***	0.239***	0.225***	0.198***
	(0.053)	(0.049)	(0.050)	(0.047)	(0.055)	(0.040)
Female_t	0.128***	0.117***	0.114***	0.108***	0.103***	0.090***
	(0.029)	(0.028)	(0.028)	(0.028)	(0.032)	(0.025)
Time Fixed Effects	Y	Y	Y	Y	Y	Y
No. Observations	554,687,312	526,636,040	509,836,753	496,064,711	483,908,705	472,889,370
R-squared	6.95e-04	7.24e-05	6.39e-04	3.93e-05	3.68e-05	4.88e-05

Notes: Unit of observation is investor-by-month. Sample contains 18,147,831 investor accounts from the Shanghai Stock Exchange from Jan 2011 through Dec 2019. Standard errors (in parentheses) are double-clustered at the individual and time levels. All the variables except returns (Ret_{t+k}) are standardized to the number of standard deviations from the mean. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively. The R-Squared values reported do not include the impact of fixed effects.

respectively, as the dependent variable. We standardize each variable except returns (Ret_{t+k}) to have a mean of 0 and standard deviation of 1, so that the coefficients can be interpreted as the predicted return change for a one standard deviation increase in the variable. The coefficients on Ret_t are never significant, which indicates that investor returns do not have strong auto-correlations. The coefficients on RCP_t are always negative and highly significant: high RCP investors tend to earn lower returns in the future. A one standard deviation increase in current-month RCP will lead to 0.48% lower returns for the next month (an annualized return loss of 5.6%), and, even after six months, monthly returns are still 0.22% lower (an annualized return loss of 2.6%). The negative impact of RCP is quite strong and persistent.

We also see that a stronger disposition effect is associated with lower future returns. Investors who hold more volatile stocks tend to earn lower future returns, with marginal significance. Investor's wealth invested in the stock market (Log(AvgInvest)) does not have a significant relationship in predicting future returns. A higher HHI index, which implies a less diversified portfolio, is strongly associated with lower future returns. Investors who trade more frequently also earn lower future returns, with marginal significance. Note that the returns we consider here do not account for trading costs. The loss will be worse for frequent traders net of trading costs. This again confirms the findings of the previous literature that frequent retail traders tend to be unsophisticated and earn lower returns (Barber and Odean 2000; Barber and Odean 2001). People who are older and more experienced earn higher future returns. Finally, females are more likely to earn higher returns than males, which again is consistent with previous findings (Barber and Odean 2001).

The results indicate that sophisticated investors tend to earn higher average returns, while investors with behavioral biases do worse. Most interestingly, among the investor characteristics we consider, RCP has the strongest magnitude to predict future investor returns: a one standard deviation increase in RCP can lead to 5.6% lower yearly returns on average. This shows that RCP is a meaningful measure for capturing investor behavior, even after controlling for all the other investor characteristics. Later, in Section 2.6, we will show that this is also the case with stock return prediction.

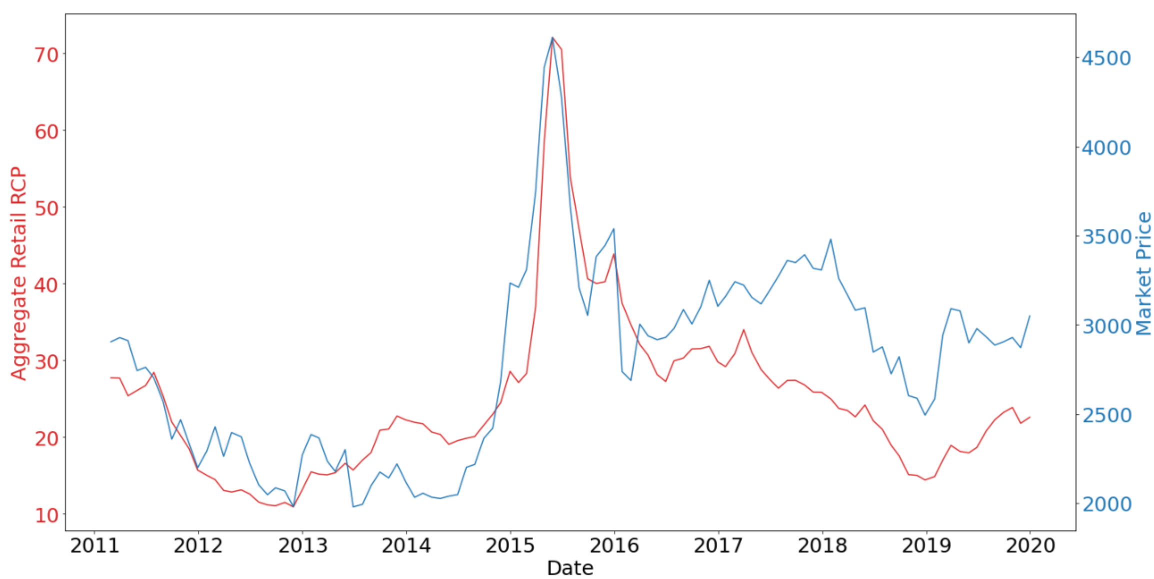
2.3.3 Aggregate Retail RCP

After we learnt about investor RCP, one may be curious about how RCP could play a role at the market aggregate level. We now aggregate investor RCP at the market level and demonstrate the time series of aggregate retail RCP. We define aggregate retail RCP as the wealth-weighted RCP from all retail investors:

$$\rho_t^M = \sum_i \rho_{i,t}\omega_{i,t}, \tag{2.7}$$

where ρ_t^M is market aggregate retail RCP, $\rho_{i,t}$ is investor i's RCP, and $\omega_{i,t}$ is investor i's wealth weight at time t.

Figure 2.1: *Relationship between Aggregate Retail RCP and Market Price*



Notes: This figure plots aggregate retail RCP and the market price index level. The red line represents aggregate retail RCP, while the blue line represents the market price index. Aggregate retail RCP is the wealth-weighted individual RCP from all retail investors. Individual RCP is constructed by averaging stock returns prior to the time of purchase across a investor's all held stocks (Equation (A.4)). Note that in our construction of RCP, we deliberately subtract market returns so that there is no mechanical relationship between aggregate retail RCP and the market index.

Figure 2.1 plots aggregate retail RCP and the market price index level. The red line represents aggregate retail RCP, while the blue line represents the market price index. We can see that aggregate retail RCP tracks the market price index quite well. There was a big stock market bubble in China from the end of 2014 to the middle of 2015. We see that aggregate retail RCP increases from 20% to over 70% during the run-up of the bubble, and drops to around 30% when the bubble bursts.

One may wonder, is there a mechanical relationship between aggregate retail RCP and the market performance? In constructing investor RCP (see Equation (A.4)), we deliberately subtract market returns, so there is no direct linkage between RCP and market returns. One possible mechanical effect is the "wealth reallocation" effect. That is, when the market exhibits momentum, say, during the run-up of the bubble, stocks with high past returns perform well so that high RCP investors earn money and acquire a higher wealth share in the market, which will increase aggregate retail RCP. In Appendix B.3.1, we decompose changes in aggregate retail RCP into different components and find that the "wealth reallocation" channel is tiny (contributes only -5%, even negatively, to the aggregate retail RCP volatility). Changes in aggregate retail RCP are driven mainly by within-investor RCP variations. We confirm that the close relationship between aggregate retail RCP and the market index is not mechanical. Investors do chase returns more during the market boom.

In Appendix B.3.2, we also show aggregate retail RCP for a range of wealth groups. We find that low-wealth groups have higher RCP on average and also react the most aggressively to market price changes, especially during the bubble period. This is consistent with our prior results that low-wealth investors tend to have higher RCP and earn lower average returns.

Overall, the close relationship between aggregate retail RCP and the market price index implies that RCP captures meaningful market dynamics, although it is hard to demonstrate a causal relationship in this context.

2.4 Stock Return Chasing Ownership (RCO)

We have shown that investor RCP is a useful measure for capturing retail investors' return chasing behavior and predicts future investor returns more strongly than any other investor characteristic we consider. When it comes to the cross section of stocks, can we still use RCP information? The answer is yes. In this section, we first introduce stock level Return Chasing Ownership (RCO hereafter), and then investigate stock characteristics that are attractive to return chasers.

Note that although RCP is an investor characteristic, RCO is a stock characteristic that captures the ownership structure of a given stock. A stock with a high RCO is one that is held disproportionately by aggressive return chasers. The analyses we conduct below are all at the stock level.

2.4.1 RCO Definition

RCO is an aggregation of RCP information from retail investors who hold a given stock. The higher RCO is, the more aggressively the stock's retail owners involve themselves in return chasing. Formally, we define a stock's RCO as the value-weighted RCP of its retail owners:

$$\text{RCO}_{s,t} = \sum_{i \in \mathbf{I}_{s,t}} w_{i,t} \text{RCP(-s)}_{i,t}, \tag{2.8}$$

where $\text{RCO}_{s,t}$ is RCO for stock s at time t. This is the weighted average of investor RCP across all retail investors who hold this stock at time t. The set $\mathbf{I}_{s,t}$ is the set of retail investors who hold stock s at time t. $\text{RCP(-s)}_{i,t}$ is investor i's calculated RCP, leaving this stock s out of the RCP calculation. In this way, we eliminate mechanical linkages between this stock's past returns and its RCO. Weight $w_{i,t}$ is retail investor i's share of the stock's total retail market value. We have $\sum_{i \in \mathbf{I}_{s,t}} w_{i,t} = 1$. Note that our focus in this paper is entirely on retail investors, so we do not consider shares held by institutions.

2.4.2 What Stocks Attract Return Chasers?

Stock RCO has an AR(1) coefficient of 0.88, implying that the half-life of shocks to stock RCO is about 5 months.[10] In Appendix Section B.4, we carry out a variance decomposition and find that changes in RCO mainly come from investor entries and exits (entries and exits together explain 71% of RCO variation). That is, when a stock's RCO increases, it is mainly because new investors with high RCP enter to hold this stock, or old investors with low RCP exit to completely sell this stock. One may wonder, what stocks will typically attract return chasers? We can answer this question by regress stock RCO on various stock attributes.

Table 2.5: *Summary Statistics for Stock Characteristics*

	Panel (Stock-by-Month)					Cross-Sectional Variation		Time-Series Variation	
	Mean	Std.Dev.	25th	50th	75th	Mean	Std.Dev.	Mean	Std.Dev.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RCO (%)	24.71	16.47	14.50	21.55	30.78	26.36	8.81	0.50	10.43
Past1YearReturn (%)	6.28	50.20	-24.04	-4.58	21.53	2.69	15.76	1.59	34.01
MktCap (CNY Billion)	26.06	109.70	3.56	6.66	14.96	21.94	90.32	-0.64	6.37
BM (Book-to-Market)	0.44	0.30	0.23	0.38	0.59	0.41	0.22	-0.01	0.10
β_{CAPM}	1.14	0.30	0.96	1.15	1.32	1.18	0.24	0.00	0.09
Volatility (%)	12.71	7.38	9.73	11.77	14.38	13.58	3.86	0.14	3.02
Turnover (%)	39.41	46.54	12.27	24.20	48.47	50.20	40.21	1.28	18.83

Notes: Columns (1) through (5) report summary statistics for the panel data, which contain 114,243 stock-by-month observations from 1,481 stocks on the Shanghai Stock Exchange from January 2011 to December 2019. Columns (6) and (7) report cross-sectional moments of time-series means calculated for each stock over the time period, giving equal weight to each stock. Columns (8) and (9) report time-series moments of cross-sectional means (after demeaning each stock's data using the stock's specific time average), giving equal weight to each month. RCO is the value-weighted RCP of a stock's retail holders. Past1YearReturn is stock return in the preceding year. MktCap is the value of the stock's market capitalization, in CNY billions (the exchange rate is roughly 6.5 CNY = 1 USD). BM is the book-to-market ratio. β_{CAPM} is the stock CAPM beta, which we estimate using a moving window of the preceding 2 years' daily stock returns (with at least 100 days of non-missing returns). Volatility is the standard deviation of monthly stock returns, and is calculated as the standard deviation of daily stock returns for the preceding 12 months, and multiplied by $\sqrt{22}$ to obtain the monthly return standard deviation (as there are on average 22 trading days in a month). Turnover is the ratio of the stock's traded value in the current month over the stock's market capitalization.

Table 2.5 reports summary statistics for various stock attributes. We limit our sample to stocks that are in the market for more than 6 months to exclude any special effects after IPOs. Given the panel structure of the data, we report both cross-sectional and time-series variations, apart from the panel summary statistics. Columns (1) through (5) report

10. We have tried multiple AR(k) regression specifications and find that the AR(1) model fits the RCO series the best. The regression results are in Appendix Table B.16.

summary statistics for the panel data, which contain 114,243 stock-by-month observations from 1,481 stocks on the Shanghai Stock Exchange from January 2011 to December 2019. Columns (6) and (7) report cross-sectional moments of time-series means calculated for each stock over the time period, giving equal weight to each stock. Columns (8) and (9) report time-series moments of cross-sectional means (after demeaning each stock's data using the stock's specific time average), giving equal weight to each month. Comparing Columns (7) and (9), we see that stock characteristics have comparable cross-sectional and time-series variations. Focusing on the panel summary statistics, we see that stock RCO has an average of 25% and a standard deviation of 16%. It tells us that an average stock is held by return chasers who have RCP of about 25%. Past1YearReturn is stock return in the preceding year, which averages 6%. MktCap is the value of the stock's market capitalization, in CNY billions (the exchange rate is roughly 6.5 CNY = 1 USD). The average stock has a MktCap of 26 CNY billion. BM is the book-to-market ratio, which averages 0.44. β_{CAPM} is the stock CAPM beta, which we estimate using a moving window of the preceding 2 years' daily stock returns (with at least 100 days of non-missing returns). The mean of β_{CAPM} is around 1. Volatility is the standard deviation of monthly stock returns, and is calculated as the standard deviation of daily stock returns for the preceding 12 months, and multiplied by $\sqrt{22}$ to obtain the monthly return standard deviation (as there are on average 22 trading days in a month). The average stock has a monthly volatility of 13%. Turnover is the ratio of the stock's traded value in the current month over the stock's market capitalization, which averages 39% monthly. Note that the Chinese stock market experiences much higher turnover than the U.S. stock market. This implies that ownership change on the Chinese stock market may be particularly useful for predicting future stock returns, which we verify in later sections.

We run both contemporaneous regressions and predictive regressions, as in Section 2.3.1, to check the relationships between stock RCO and other stock attributes. For the contemporaneous regressions, we regress the current period RCO on other stock attributes:

$$\text{RCO}_{s,t} = \alpha + \gamma X_{s,t} + \delta_t + \epsilon_{s,t}, \tag{2.9}$$

where $X_{s,t}$ are various stock attributes in the current period. δ_t are time fixed effects. The regression results are reported in Columns (1) and (2) of Table 2.6.

For the predictive regressions, we use all the current-period stock attributes to predict the next-period stock RCO:

$$\text{RCO}_{s,t+1} = \alpha + \beta \text{RCO}_{s,t} + \gamma X_{s,t} + \delta_t + \epsilon_{s,t+1}, \tag{2.10}$$

where $\text{RCO}_{s,t+1}$ is RCO for stock s at time $t+1$. We control for this period's $\text{RCO}_{s,t}$. The results are reported in Column (3) of Table 2.6.

Table 2.6: *The Relationship between Stock RCO and Other Stock Characteristics*

Outcome Var:	RCO_t	RCO_t	RCO_{t+1}	RCO_{t+1}
	(1)	(2)	(3)	(4)
RCO_t			0.850***	0.822***
			(0.011)	(0.014)
Past1YearReturn_t	3.349***	3.166***		0.445***
	(0.297)	(0.273)		(0.115)
$\beta_{\text{CAPM},t}$	0.329	0.124		0.007
	(0.278)	(0.280)		(0.082)
Log(MktValue)_t	1.804***	5.965***		0.185***
	(0.183)	(0.473)		(0.061)
BM_t	-3.004***	-1.830***		-0.500***
	(0.270)	(0.293)		(0.067)
Volatility_t	4.537***	3.678***		0.486**
	(1.410)	(1.262)		(0.200)
Turnover_t	2.977***	2.433***		-0.010
	(0.322)	(0.251)		(0.085)
Time Fixed Effects	Y	Y	Y	Y
Stock Fixed Effects	N	Y	N	N
R-squared	0.288	0.242	0.750	0.752
Observations	112,335	112,335	109,877	109,835

Notes: Unit of observation is stock-by-month. Sample contains 1,481 stocks at the Shanghai Stock Exchange from Jan 2011 to Dec 2019. All the variables except the stock RCO are standardized to the number of standard deviations from the mean. Standard errors (in parentheses) are double-clustered at the stock and time levels. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively. The R-Squared values reported do not include the impact of fixed effects.

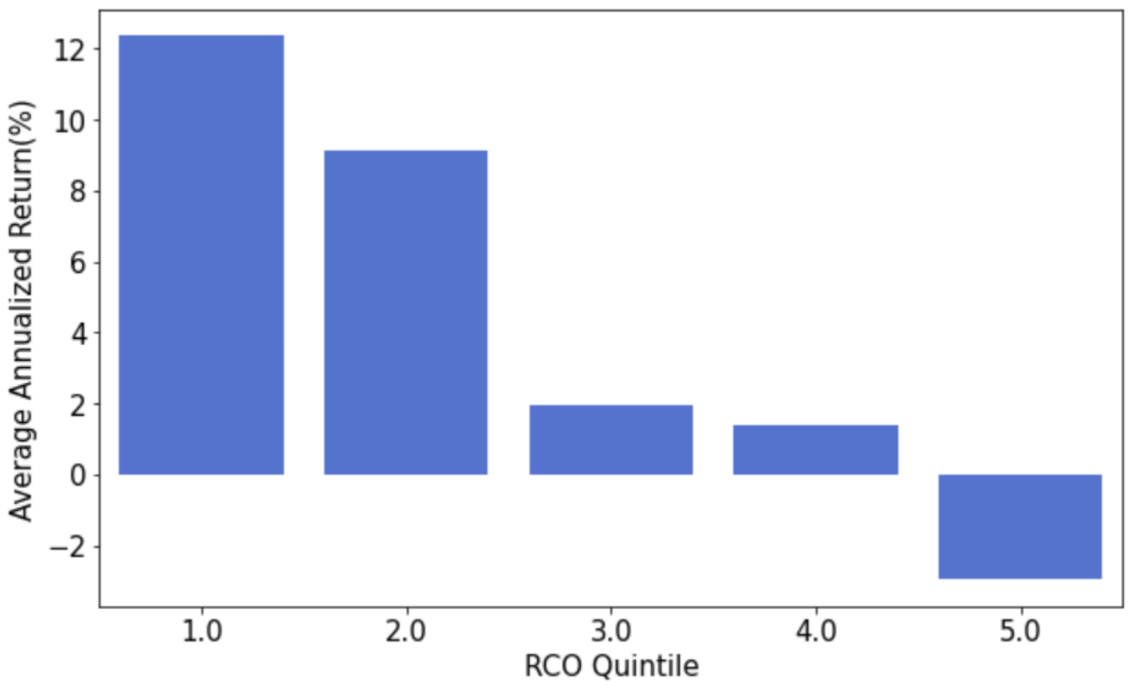
In Table 2.6, we standardize each variable except stock RCO to have a mean of 0 and standard deviation of 1, so that the coefficients can be interpreted as the predicted change for a one standard deviation increase in the variable. The standard errors are double-clustered at both the stock and time levels. In Column (1), we report the results obtained from the contemporaneous regression that regresses RCO on other stock attributes in the same period. For Column (2), we add stock fixed effects to the contemporaneous regression. We see that stocks with higher past returns, larger size, lower book-to-market ratios, higher volatility, and higher turnover are associated with higher RCO. A one standard deviation increase in stock past one-year returns will increase RCO by 3 percentage points. Size, book-to-market ratios, volatility, and turnover have comparable coefficient magnitudes. The coefficients on β_{CAPM} are not significant, indicating no clear relationship if return chasers are attracted to high- or low-beta stocks. In Column (3), we report the results obtained from the predictive regression with only the current period stock RCO as the regressor. The results simply indicate that stock RCO is persistent as we mentioned above. For Column (4), we add all the other stock attributes, to identify which stock attributes, conditioning on the current-period stock RCO, are useful for predicting the next period RCO. Given the high persistence of RCO, it is unsurprising that the magnitudes of the coefficients reported in Column (4) are smaller than those reported in Columns (1) and (2). Apart from the magnitudes, the signs of the three columns are consistent.

We are now ready to answer the question that we raised at the beginning of this section. Stocks that earn higher past returns, that have higher volatility and turnover, are larger in size, and are over-valued (with a lower book-to-market ratio) are more likely to attract return chasers. These stock features, except past returns, typically predict lower future risk-adjusted returns (Fama and French 1993; Ang et al. 2006). This is consistent with our previous findings that high RCP retail investors are more likely to earn lower average returns.

2.5 Trading Strategies Based on Stock RCO

We have learned that return chasers earn lower average returns. Does this also mean that high RCO stocks are more likely to under-perform? If so, can we use this information to design profitable trading strategies? In this section, we construct a stock RCO based trading factor and explore various trading strategies.

Figure 2.2: *Average Annualized Returns by Stock RCO Quintiles*



Notes: This figure shows the average future one-month returns for the 5 RCO quintiles (the numbers are annualized for easier interpretation). Quintile1 denotes the group of stocks with the lowest RCO, while quintile5 denotes the group of stocks with the highest RCO. RCO is constructed by wealth-weighting RCP from retail investors who hold this stock (leaving this particular stock out of the investor RCP calculation to eliminate mechanical linkages between this stock's RCO and its past returns).

2.5.1 The RCO Trading Factor

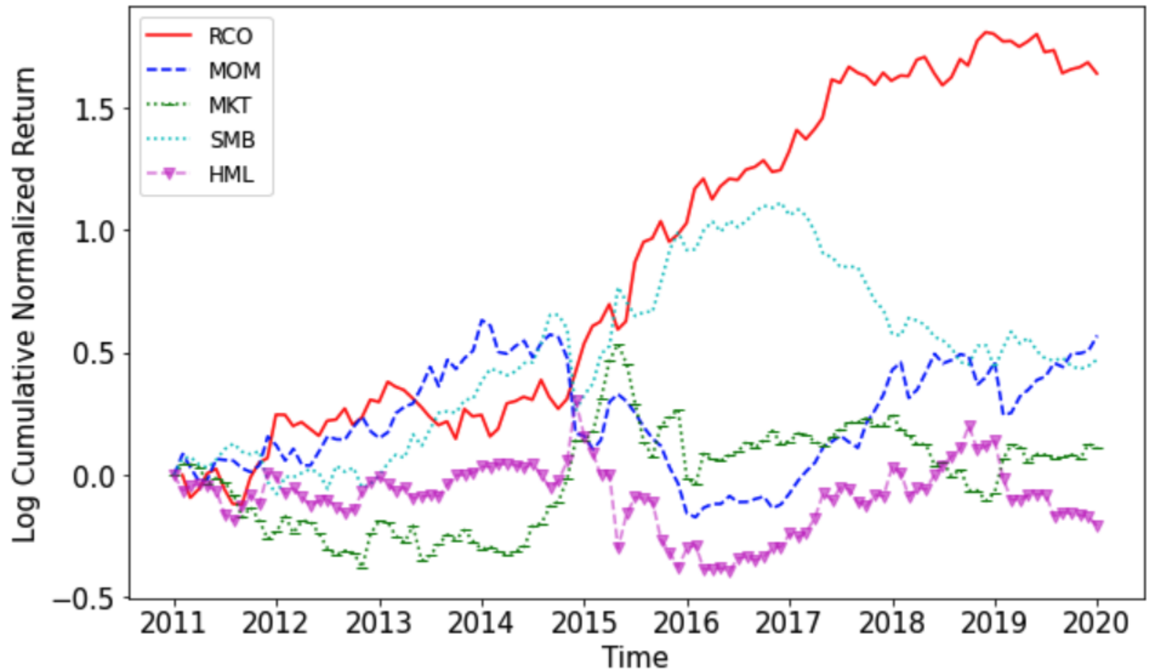
We first sort stocks into 5 quintiles by their RCO in the current month. The sorting is updated monthly because stock RCO changes every month. In Figure 2.2, we show the average future one-month returns for the 5 quintiles (the numbers are annualized for easier interpretation). Quintile1 is the group of stocks with the lowest RCO, while quintile5 is the group of stocks with the highest RCO. We observe a monotone relationship between stock future returns and RCO. The lowest RCO quintile stocks have on average 12.4% annualized future returns, while the highest RCO quintile stocks have instead -2.9% annualized returns. The 15.3% return difference is substantial.

The negative relationship between future stock returns and RCO indicates that stocks which are held by aggressive return chasers tend to under-perform. We next construct a trading strategy that goes long stocks in the lowest RCO quintile and shorts stocks in the highest RCO quintile. The portfolio returns are value-weighted by stocks' market capitalization. The long/short positions are updated monthly. This constructed RCO factor performs quite well in our sample period, with an annualized Sharpe ratio of 0.94 and a T-statistic of 2.75.[11]

Figure 2.3 plots the log cumulative normalized returns on RCO together with the standard Fama/French three factors and the momentum factor. All factors are rebalanced monthly. The return of the beginning month of our sample is set to zero. In this figure, the returns on SMB, HML, MOM, and RCO are all scaled to have the same standard deviation as MKT, so the height of each line at the right axis is proportional to the full-sample Sharpe ratio of each portfolio. From the figure we can see that RCO outperforms other standard factors during our sample period. It is worthwhile mentioning that our RCO factor is quite different from the traditional MOM factor. It is widely documented that there is missing momentum in the Chinese stock market (Liu, Stambaugh, and Yuan 2019, Haorui, Huihang, and Yan 2020), i.e., the classical momentum trading strategy does not work in China. This highlights the fact that stock RCO contains much broader signals than just stock past returns. The ownership structure that RCO captures delivers much richer information than what can be obtained using only stock characteristics.

11. One might worry about the cost of shorting individual stocks in China. While shorting individual stocks may be subject to some regulations, the market indices are widely available for shorting. In a robustness check, when we implement the RCO factor, instead of shorting stocks in the lowest RCO quintile we also try to short the available market index, e.g., the CSI 300 Index. This modified trading strategy has an annualized Sharpe ratio of 0.72 and a T-statistic of 2.13.

Figure 2.3: *Log Cumulative Normalized Returns on Different Factors*



Notes: This figure plots the log cumulative normalized returns on RCO together with the standard Fama/French three factors and the momentum factor. All the factors are rebalanced monthly. The returns on SMB, HML, MOM, and RCO are all scaled to have the same standard deviation as MKT, so the height of each line at the right axis is proportional to the full-sample Sharpe ratio of each portfolio. The return of the beginning month of our sample is set to zero. The RCO factor is constructed by going long stocks in the lowest RCO quintile and shorts stocks in the highest RCO quintile (portfolio returns are value-weighted by stocks' market capitalization).

2.5.2 Interaction between RCO and Stock Retail Shares

We have shown that the RCO trading factor is highly effective in the Chinese stock market. Our RCO measure is focused entirely on retail investors, so it is natural to conjecture that the effects may be more pronounced among stocks that have a higher retail share. We test this conjecture by double-sorting RCO and stock retail shares. We define a stock's retail share as the ratio of stock value owned by retail investors over the total stock value owned either by retail investors or institutions. We sort stocks into 3 terciles according to their retail shares in the current month. We then carry out an independent double-sort on RCO and retail share, and obtain $5 * 3 = 15$ groups.

Table 2.7: *Interactions Between RCO, Retail Share, and Past Returns*

Panel A: Returns on Double-Sort by RCO and Retail Share (%)

	RetailShare1 (Low)	RetailShare2	RetailShare3 (High)	RetailShare1 - RetailShare3
RCO1 (Low)	11.85	5.26	12.67	-0.82
RCO2	6.55	7.78	9.91	-3.36
RCO3	-1.54	-1.22	4.68	-6.22
RCO4	-0.15	-2.73	-3.40	3.25
RCO5 (High)	-2.13	-10.31	-18.32	16.19
RCO1 - RCO5	13.98	15.57	30.99	

Panel B: Returns on Double-Sort by RCO and Stock Past Returns (%)

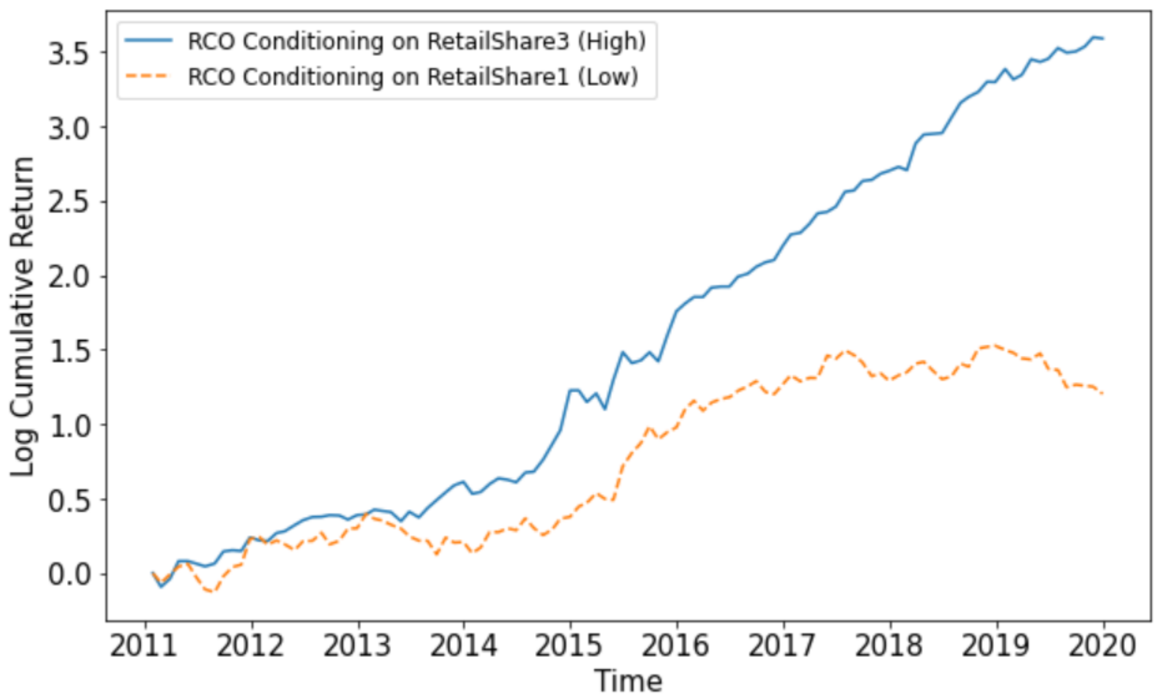
	PastReturn1 (Low)	PastReturn2	PastReturn3 (High)	PastReturn1 - PastReturn3
RCO1 (Low)	5.80	10.68	12.09	-6.29
RCO2	1.26	7.01	10.93	-9.67
RCO3	-5.07	-1.00	4.31	-9.38
RCO4	-3.22	-3.99	-1.12	-2.10
RCO5 (High)	-7.41	-6.15	-5.54	-1.87
RCO1 - RCO5	13.22	16.83	17.63	

Notes: Panel A reports the double-sort returns by stock RCO and retail share. Panel B reports the double sort returns by stock RCO and past returns. RCO1 is the group of stocks with the lowest RCO, while RCO5 is the group of stocks with the highest RCO. RetailShare1 is the group with the lowest retail share, while RetailShare3 is the group with the highest retail share. PastReturn1 is the group with the lowest past returns, while PastReturn3 is the group with the highest past returns. Each cell in the table reports the average annualized returns for the corresponding double-sort group. The last column and the last row show the long/short portfolio returns.

The average future one-month returns (annualized) for each group are reported in Panel A of Table 2.7. The last column and the last row show the long/short portfolio returns. RetailShare1 denotes the group with the lowest retail share, while RetailShare3 denotes the group with the highest retail share. Indeed, we see a clear pattern that the RCO trading strategy performs the best among stocks in the highest retail share tercile (Column RetailShare3), generating an average annualized return of 31%. We expect that stocks with the highest retail shares and the highest RCO will have the lowest future returns, which is exactly what we see in the table (the cell for RCO5 and RetailShare3 shows a future return of -18.32%, which is the lowest return in the table).

Instead, when we look at the returns from the long/short portfolios using retail shares only (the last column of Panel A, Table 2.7), the performance of the retail share factor by itself is mixed. Depending on which RCO quintile is in question, the retail share long/short portfolio returns can either be positive or negative. This finding suggests that using only aggregate information on retail shares may not be enough. Heterogeneity within retail investors matters. We obtain much richer information when open the black box of retail investors and identify who owns what.

Figure 2.4: *Log Cumulative RCO Returns Conditioning on Different Retail Share Terciles*



Notes: This figure plots the log cumulative returns for the RCO factor conditioning on different retail share terciles. The blue solid line refers to the RCO factor conditioning on stocks in the highest retail share tercile, while the orange dash line refers to the RCO factor conditioning on stocks in the lowest retail share tercile. The two portfolios are scaled to have the same standard deviation as MKT (market portfolio) as in Figure 2.3, so the height of each line at the right axis is proportional to the full-sample Sharpe ratio of each portfolio. The return of the beginning month of our sample is set to zero. The RCO factor is constructed by going long stocks in the lowest RCO quintile and shorts stocks in the highest RCO quintile (the portfolio returns are value-weighted by stocks' market capitalization, with long/short positions updated monthly). RCO and retail share are double sorted independently.

Figure 2.4 plots the log cumulative returns for the RCO factor conditioning on different retail share terciles. The blue solid line refers to the RCO factor conditioning on stocks in the highest retail share tercile, while the orange dash line refers to the RCO factor conditioning on stocks in the lowest retail share tercile. RCO and retail share are double sorted independently. The two portfolios are scaled to have the same standard deviation as MKT (market portfolio) as in Figure 2.3, so the height of each line at the right axis is proportional to the full-sample Sharpe ratio of each portfolio. We can see that RCO performs much better among high retail share stocks. The RCO factor conditioning on stocks in the highest retail share tercile generates an annualized Sharpe ratio of 1.94 (with T-statistic of 5.38).

2.5.3 Interaction between RCO and Past Stock Returns

RCO captures how aggressively a stock's retail owners chase past returns. We would expect the RCO trading factor to perform better for stocks which have higher past returns. As in the previous section, we double-sort stocks by its RCO and past returns. We first sort stocks into 3 terciles according to their preceding one-year returns. We then carry out an independent double-sort on RCO and past returns, which gives us $5 * 3 = 15$ groups.

The average future one-month returns (annualized) for each group are reported in Panel B of Table 2.7. PastReturn1 denotes the group with the lowest past returns, while PastReturn3 denotes the group with the highest past returns. We do see a monotone relationship between the RCO factor returns and past stock returns. For stocks in the highest past returns tercile, the RCO factor generates an average annualized return of 17.6%.[12]

2.5.4 Do Return Chasers Have Causal Effects on Stock Returns?

The strong relationship between RCO and stock returns, especially given the large magnitudes of return differences, may lead one to contemplate whether return chasers have causal effects on stock future returns. While it is generally difficult to draw causal conclusions regarding asset prices, we offer suggestive evidence that return chasers have price impacts on the stocks they hold.

Section 2.5.2 shows that the RCO trading factor performs better among stocks with a higher retail share. One may be tempted to conclude that return chasers must have price

12. In Appendix Table B.17, we also carry out an independent triple-sort on stock RCO, retail share, and past returns. The results get a bit noisy with the much increased 45 groups, but it is quite clear that the RCO factor works the best for stocks that have both the highest retail share and the highest past returns.

impacts, so that the more of them there are, the larger the price impacts they have on stocks. Retail share can, however, be influenced by various factors. To make the claim causal, we want variation in the retail share that is exogenous to other variables that may influence future stock returns. We propose that a stock's IPO price creates such exogenous variation in its retail share.

There is an institutional restriction in the Chinese stock market that every transaction has to be in contracts of 100 shares. For example, if a stock's price is \$100, the investor needs to have at least $100 * \$100 = \$10,000$ to trade this stock. As a result, a higher-priced stock captures a lower retail share because some low-wealth retail investors simply cannot afford it. In Appendix Table B.18, we show that a one standard deviation increase in log current price will make the retail share decrease by 7.3 percentage points. A stock's IPO price is strongly positively associated with its current price, and hence negatively associated with its current retail share. Insofar as factors that influenced a firm's IPO price many years in the past are likely to be uncorrelated with the stock's future performance, the IPO price generates potential exogenous variation in its retail share.

Nonetheless, we admit that there may be cases in which the exclusion restriction on IPO prices as an IV could be violated. For example, some firms might actively choose their IPO price to cater to retail investors. While in practice it is difficult for a firm to manipulate its IPO price,[13] we cannot completely rule out this possibility. Therefore, the results we show below are meant to be suggestive rather than conclusive.

As in the previous sections on sorting stocks, we first sort stocks into 2 equal-sized groups by their IPO prices. For Panel A of Table 2.8, we independently double-sort stocks

13. In the Chinese stock market, a firm needs to first file its IPO application, at which time the total shares for IPO issuance are determined (a resulting expected IPO price is also determined). The investment bank then needs to conduct its due diligence and asks for acceptable prices from main institutional investors. The final determined IPO price is based on various opinions of institutional investors and is usually different from the original "expected IPO price" indicated in the firm's IPO application. In this case, it is hard for this firm to split/merge shares before the IPO application to manipulate its final IPO price. (According to conversations with several investment bankers, it is almost never the case that the firm split/merge shares ahead to change its IPO price). What's more, according to regression results in Appendix Table B.18, a one standard deviation increase in Log(IPO Price) will decrease the retail share by 0.86 percentage points. The cross-sectional retail share has an average of 82% and a standard deviation of 16% (Table 2.9). Therefore there is also limited impact the firm could get from manipulating its IPO price.

by IPO price and RCO. We can see that indeed the RCO trading factor performs better in the Low IPO Price group, with an average annualized return of 21%, compared with 13% in the High IPO Price group. Because a low IPO price is associated with a high retail share, we have evidence that when we exogenously increase retail investors in some stocks, we can see higher price impacts imposed by return chasers. For Panel B of Table 2.8, we independently triple-sort stocks by IPO price, RCO, and past returns. We can see that, in the Low IPO Price group, the RCO factor performs better as past returns increase.

Table 2.8: *Interactions Between IPO Price, RCO, and Past Returns*

Panel A: Returns on Double-Sort by IPO Price and RCO (%)

	IPO Price Low	IPO Price High
RCO1 (Low)	11.23	10.22
RCO2	5.03	7.42
RCO3	0.48	-3.36
RCO4	-4.39	0.81
RCO5 (High)	-9.56	-2.48
RCO1 - RCO5	20.79	12.70

Panel B: Returns on Triple-Sort by IPO Price, RCO, and Past Returns (%)

	IPO Price Low			IPO Price High		
	PastRet1	PastRet2	PastRet3	PastRet1	PastRet2	PastRet3
RCO1 (Low)	5.50	9.23	15.86	6.65	12.97	8.99
RCO2	0.96	8.44	5.34	2.96	3.86	15.37
RCO3	0.13	-2.79	4.00	-7.05	1.38	1.16
RCO4	-2.62	-4.53	-3.50	-2.07	-3.28	2.27
RCO5 (High)	-0.14	0.65	-11.44	-6.35	-7.97	-1.27
RCO1-RCO5	5.64	8.57	27.29	13.00	20.94	10.26

Notes: Panel A reports the double-sort returns by stock IPO Price and RCO. Panel B reports the triple-sort returns by stock IPO Price, RCO, and past returns. Each cell reports the average annualized returns for the corresponding sorted group. The last row shows the long/short portfolio returns.

In general, as we use IPO price as a proxy for exogenous movements in retail shares, we obtain consistent results indicating that as the share of retail investors grows, the RCO factor performance strengthens. This finding suggests that return chasers do have some price impacts on the stocks they hold.

2.6 Other Investor-Characteristic Ownership Signals

The success of RCO may lead one to ask, is RCO a unique ownership characteristic? Can we get similar results if we use the ownership information of other investor characteristics? In this section, we construct various investor-characteristic ownership signals, and check whether RCO still predicts future stock returns after controlling for all the other ownership signals.

Table 2.9: *Summary Statistics for Stock Ownership Signals and General Characteristics*

	Panel (Stock-by-Month)					Cross-Sectional Variation		Time-Series Variation	
	Mean	Std.Dev.	25th	50th	75th	Mean	Std.Dev.	Mean	Std.Dev.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Investor-Characteristic Ownership Signals									
RCO (%)	24.71	16.47	14.50	21.55	30.78	26.36	8.81	0.50	10.43
Age	48.26	3.72	46.94	48.33	49.77	48.25	2.01	-0.11	1.98
Experience	7.20	1.68	6.06	7.28	8.36	7.26	0.91	-0.18	1.25
Female	0.43	0.08	0.40	0.44	0.47	0.42	0.06	0.00	0.02
Disposition Effect (%)	5.23	2.25	3.82	5.14	6.49	5.82	1.96	-0.06	0.67
Wealth (CNY Million)	161.51	2250.56	2.03	3.98	9.62	147.21	380.25	-39.25	907.53
Trade Frequency	163.92	515.98	36.64	72.08	153.02	165.22	115.80	1.76	132.04
Number Stocks	4.87	6.42	3.83	4.32	4.94	4.93	1.51	-0.16	3.04
Investor Return (%)	0.26	3.74	-1.72	-0.08	1.57	-0.02	1.11	0.09	2.76
Stock Characteristics									
Retail Share (%)	81.34	19.70	72.38	88.58	96.61	82.24	16.04	0.17	2.12
Past1YearReturn (%)	6.28	50.20	-24.04	-4.58	21.53	2.69	15.76	1.59	34.01
Volatility (%)	12.71	7.38	9.73	11.77	14.38	13.58	3.86	0.14	3.02
Turnover (%)	39.41	46.54	12.27	24.20	48.47	50.20	40.21	1.28	18.83
β_{Market}	0.99	0.26	0.84	1.00	1.14	1.01	0.22	0.01	0.04
β_{size}	0.60	0.54	0.31	0.64	0.92	0.72	0.51	0.00	0.05
β_{value}	-0.18	0.60	-0.49	-0.17	0.14	-0.15	0.47	-0.00	0.04

Notes: Columns (1) through (5) report summary statistics for the panel data, which contain 114,243 stock-by-month observations from 1,481 stocks on the Shanghai Stock Exchange from Jan 2011 to Dec 2019. Columns (6) and (7) report cross-sectional moments of time-series means calculated for each stock over the time period, giving equal weight to each stock. Columns (8) and (9) report time-series moments of cross-sectional means (after demeaning each stock's data using the stock's specific time average), giving equal weight to each month. Investor-characteristic ownership signals are constructed by Equation (2.11). For general stock characteristics, Retail Share is the ratio of stock value owned by retail investors over the total stock value owned either by retail investors or institutions. Volatility is the monthly standard deviation of stock returns. Past1YearReturn is stock return in the preceding year. Turnover is the ratio of the stock's traded value in the current month over the stock's market capitalization. $\beta_{Market,t}$, $\beta_{size,t}$ and $\beta_{value,t}$ are standard Fama/French three factor betas: the market beta, the size beta, and the value beta, respectively. We estimate the stock betas using the moving window of the preceding two years' daily returns (with at least 100 days of non-missing returns).

Note that we have various investor level characteristics. To construct ownership signals, we need to aggregate investor characteristics at the stock level. We construct these investor-characteristic ownership signals following the same procedure that we used for RCO. Formally, we have

$$\mathrm{X}_{s,t}^{\text{Ownership}} = \sum_{i \in \mathbf{I}_{s,t}} w_{i,t} \mathrm{X}_{i,t}, \tag{2.11}$$

where $\mathrm{X}_{s,t}^{\text{Ownership}}$ denotes the characteristic X ownership signal for stock s at time t. This variable is the weighted average of investor characteristics $\mathrm{X}_{i,t}$ across all retail investors who hold stock s at time t. Set $\mathbf{I}_{s,t}$ is the set of retail investors who hold stock s at time t. Weight $w_{i,t}$ is retail investor i's share of that stock's retail market value. We have $\sum_{i \in \mathbf{I}_{s,t}} w_{i,t} = 1$. Again our focus is only on retail investors.

The summary statistics for all our constructed investor-characteristic ownership signals are reported in Table 2.9. Given the panel structure of the data, we report both the cross-sectional and time-series variations for each variable in addition to the panel summary statistics. Columns (1) through (5) report summary statistics for the panel data, which contain 114,243 stock-by-month observations from 1,481 stocks on the Shanghai Stock Exchange from January 2011 to December 2019. Columns (6) and (7) report cross-sectional moments of time-series means calculated for each stock over the time period, giving equal weight to each stock. Columns (8) and (9) report time-series moments of cross-sectional means (after demeaning each stock's data using the stock's specific time average), giving equal weight to each month. Because we take the value-weighted investor characteristics to obtain the ownership signals, it may feel a bit difficult to interpret the numbers. Heuristically, one can think of these ownership signals as the characteristics of a "pseudo-representative" investor who holds the stock. For example, if a stock has an RCO of 25%, one can think of the representative stockholder as having an RCP of 25%. Similarly, a stock with age ownership of 48 means that investors who hold this stock are on average 48 years old. When we compare Columns (7) and (9), we can see that ownership signals exhibit comparable cross-sectional and time-series variations. Table 2.9 also lists summary statistics for other stock characteristics. We've discussed some of the characteristics in Table 1.2. The mean of Retail Share is 81%. We see quite a high average retail share here, mainly because small stocks have high retail shares, and this fact is highlighted when we give equal weight to each stock. $\beta_{Market,t}$, $\beta_{size,t}$ and $\beta_{value,t}$ are standard Fama/French three factor betas: the market beta, the size beta, and the value beta, respectively. We estimate the stock betas using the moving window of the preceding two years' daily returns (with at least 100 days of non-missing returns).

To be consistent with the trading strategies mentioned earlier, we rank each stock variable in each month cross-sectionally and convert them into quantiles. This rank transformation ensures that our results are robust to variables with different scales and distributions, which is exactly the case with our analysis, as can be seen in Table 2.9.[14]

We now run panel regressions to predict future stock returns by stock ownership signals:

$$r_{s,t+k} = \alpha + \gamma X_{s,t}^{\text{Ownership}} + Z_{s,t} + \delta_t + \epsilon_{i,t+k}, \tag{2.12}$$

where $r_{s,t+k}$ is stock s's monthly returns for k periods leading to time t. $X_{s,t}^{\text{Ownership}}$ are various ownership signals for stock s at time t. $Z_{s,t}$ are control variables for stock s. δ_t are time fixed effects. The regression results are reported in Table 2.10.

For Table 2.10, Columns (1) through (3) use future 1-month through 3-month returns, respectively, as dependent variables. Note that all the variables except returns (Ret_{t+k}) are transformed to their quantile ranks in the cross-section. Therefore we can easily compare the impact magnitudes of various ownership characteristics: the coefficient of each variable tells us the extent to which the future monthly return will be impacted if a variable rises from the lowest rank to the highest rank in the cross-section. Standard errors are double-clustered at the stock and time levels. We can see that there are 4 ownership signals that are significant at the 95% confidence level to predict the next 1-month return: stock RCO, Age, Experience, and Female. RCO has a coefficient of -0.847, which means that if a stock rises from the lowest RCO rank to the highest RCO rank in the cross-section, its return in the next month

14. This is a standard transformation to manage variables with varying scales and has also been used in Kelly, Pruitt, and Su (2019), Kozak, Nagel, and Santosh (2020), or Balasubramaniam et al. (2023), among others.

will decrease by 0.847% (about 10% annualized). Investor Experience also predicts future returns quite strongly: its impact magnitude is similar to that of RCO. We see that stocks owned by more experienced investors, by older investors, and by more females all tend to earn higher future returns. For other stock ownership signals, such as Disposition Effect, Wealth, and Trade Frequency, even though the coefficients may not be significant, the signs are as we would expect: stocks that are owned by investors who have a weaker disposition effect, are richer, and trade less frequently tend to earn higher returns in the following 3 months. The results again convey a consistent message: stocks held by more sophisticated investors are likely to outperform stocks held by less sophisticated investors. Note that we have controlled for stock retail share, volatility, past one-year returns, turnover, and the Fama/French three factors in all the regressions. Therefore our results are not driven by variations in these dimensions.

For Column (4) of Table 2.10, we add the interaction term between RCO and stock retail share to predict the following 1-month return, in the spirit of the double-sort trading strategy discussed in the previous section. We can see that the interaction term is especially large and highly significant, which is consistent with our double-sort trading strategy results that the RCO factor is extremely powerful among stocks with a high retail share. Column (5) of Table 2.10 further adds the triple interaction term between RCO, stock retail share, and stock past returns, and we obtain an even larger coefficient. This again is consistent with our previous sorting results that the RCO factor performs best among stocks with a higher retail share and higher past returns.

The most interesting result reported in Table 2.10 is that, after controlling a wide range of stock ownership signals, RCO remains quite significant. RCO and investor experience are the two strongest investor characteristic-based predictors of stock returns. The return impact is large: a stock with the highest RCO earns on average 10% lower yearly returns than the stock with the lowest RCO. We can see that RCO captures distinctive information about stockholders' behavior. In Appendix Table B.19, we also report the covariance matrix among ownership signals, which again verifies that RCO does not covary with other ownership signals much.

2.7 Chase Industry Returns or Idiosyncratic Returns?

So far we have shown that return chasing is a prevalent phenomenon in the Chinese stock market and it has substantial impacts on both investor returns and stock returns. It is worthwhile digging further to determine which types of returns investors chase. For example, an investor might pick a high past return stock because of the high returns on the industry in which the stock operates, or perhaps because of the stock's high idiosyncratic returns. Are there differences between these two versions of return chasing? In this section, we decompose return chasing into industry return chasing and stock idiosyncratic return chasing. Future research could take a similar approach to explore the mechanisms that drive return chasing more fundamentally.

According to the definition of RCP in Equation (A.4), we can decompose RCP into

$$\begin{aligned}
\mathrm{RCP}_{i,t} &= \sum_{s\in\{S_{i,t}\}}\left[\gamma_s\left(R^{s}_{t(s)-K,t(s)} - R^{M}_{t(s)-K,t(s)}\right)\right] \\
&= \sum_{s\in\{S_{i,t}\}}\left[\gamma_s\left(R^{s}_{t(s)-K,t(s)} - R^{I(s)}_{t(s)-K,t(s)} + R^{I(s)}_{t(s)-K,t(s)} - R^{M}_{t(s)-K,t(s)}\right)\right] \\
&= \underbrace{\sum_{s\in\{S_{i,t}\}}\left[\gamma_s\left(R^{s}_{t(s)-K,t(s)} - R^{I(s)}_{t(s)-K,t(s)}\right)\right]}_{\text{Stock idiosyncratic RCP (SRCP)}} + \underbrace{\sum_{s\in\{S_{i,t}\}}\left[\gamma_s\left(R^{I(s)}_{t(s)-K,t(s)} - R^{M}_{t(s)-K,t(s)}\right)\right]}_{\text{Industry RCP (IRCP)}},
\end{aligned} \tag{2.13}$$

where $R^{I(s)}_{t1,t2}$ is the return on industry $I(s)$ in which stock s operates during period $t1$ to $t2$.

In this way, investor RCP is decomposed as stock idiosyncratic RCP (SRCP) and industry RCP (IRCP). We use 28 industries based on the two-digit classification. With the decomposition ready, we can replicate all the results we have reported in this paper with the two newly constructed return chasing measures. Appendix B.5 lists the detailed analyses. Here we summarize the main findings:

1. Investors who have a stronger disposition effect, trade more frequently, hold more volatile stocks, are younger, are less experienced, or are male chase stock idiosyncratic returns to a greater extent than industry returns.

Table 2.10: *Predicting Future Stock Returns Using Ownership Signals*

Outcome Var:	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+1}	Ret_{t+1}
	(1)	(2)	(3)	(4)	(5)
Investor-Characteristic Ownership Signals					
RCO_t	-0.847***	-0.526*	-0.505	0.249	-0.787
	(0.296)	(0.299)	(0.337)	(0.443)	(0.762)
$\text{RCO}_t \times \text{Retail Share}_t$				-2.234***	1.178
				(0.662)	(1.197)
$\text{RCO}_t \times \text{Retail Share}_t \times \text{Past1YearReturn}_t$					-5.184**
					(2.099)
Age_t	0.626***	0.195	0.170	0.620***	0.619***
	(0.181)	(0.162)	(0.156)	(0.182)	(0.180)
Experience_t	0.855***	0.627***	0.520**	0.849***	0.815***
	(0.251)	(0.227)	(0.251)	(0.251)	(0.250)
Female_t	0.406**	-0.039	-0.063	0.377*	0.366*
	(0.197)	(0.170)	(0.174)	(0.197)	(0.195)
$\text{Disposition Effect}_t$	-0.638*	-0.534	-0.474	-0.591	-0.516
	(0.365)	(0.372)	(0.408)	(0.365)	(0.370)
Wealth_t	0.414*	0.144	0.269	0.418*	0.402*
	(0.222)	(0.209)	(0.194)	(0.222)	(0.222)
Trade Frequency_t	-0.142	-0.110	-0.182	-0.124	-0.070
	(0.220)	(0.235)	(0.220)	(0.221)	(0.217)
Number Stocks_t	-0.351	-0.309	-0.016	-0.360	-0.399
	(0.259)	(0.274)	(0.277)	(0.258)	(0.257)
Investor Return_t	-0.295	0.027	8.93e-05	-0.366	-0.464
	(0.532)	(0.495)	(0.538)	(0.530)	(0.529)
Controls					
Retail Share_t	-0.386	-0.063	-0.156	0.782	0.476
	(0.404)	(0.393)	(0.385)	(0.516)	(0.701)
Volatility_t	-0.143	-0.678*	-0.598	-0.154	-0.174
	(0.394)	(0.382)	(0.364)	(0.394)	(0.394)
Past1YearReturn_t	-0.513	-0.210	-0.129	-0.480	0.446
	(0.404)	(0.327)	(0.329)	(0.405)	(0.990)
Turnover_t	-0.827*	-0.511	-0.418	-0.807*	-0.792*
	(0.442)	(0.404)	(0.385)	(0.443)	(0.442)
$\beta_{Market,t}$	0.016	-0.212	-0.310	-0.006	0.001
	(0.323)	(0.321)	(0.317)	(0.325)	(0.325)
$\beta_{size,t}$	1.102	0.826	0.950	1.039	1.034
	(0.688)	(0.670)	(0.671)	(0.689)	(0.688)
$\beta_{value,t}$	-0.328	-0.128	-0.119	-0.257	-0.267
	(0.472)	(0.469)	(0.460)	(0.473)	(0.474)
Time fixed effects	Y	Y	Y	Y	Y
R-squared	0.006	0.003	0.003	0.006	0.007
No. Observations	110,640	108,357	106,347	110,640	110,640

Notes: Unit of observation is stock-by-month. Sample contains 1,481 stocks at the Shanghai Stock Exchange from Jan 2011 to Dec 2019. Standard errors (in parentheses) are double-clustered at stock and time levels. All variables are taken the quantile rank transformation. Column (5) controls for the interaction terms $\text{Past1YearReturn}_t \times \text{RetailShare}_t$ and $\text{Past1YearReturn}_t \times \text{RCO}_t$.

2. Both types of return chasing reduce future returns to investors. Chasing idiosyncratic returns exerts the most negative effects.
3. Industries with higher industry return chasing ownership (IRCO) under-perform.
4. Within an industry, stocks that exhibit higher stock idiosyncratic return chasing ownership (SRCO) also under-perform.

The results show that different styles of return chasing could be interesting too. We leave it to future research to further investigate the mechanisms that drive return chasing.

2.8 Discussion

This paper offers a novel measure to capture heterogeneous return chasing propensities at the individual investor level. We find that return chasing behavior has substantial impacts on both investors and stocks. High RCP investors are typically less sophisticated. They earn lower average future returns: a one standard deviation increase in RCP can lead to 5.6% lower yearly returns on average. Stocks that are owned by return chasers also have lower average future returns. The stock with the highest RCO earns 10% lower yearly returns than the stock with the lowest RCO. We are able to construct profitable trading strategies based on stocks' RCO, which outperform standard Fama/French factors in the Chinese stock market during our sample period. The RCO factor performs even better among stocks with higher retail share and higher past returns. Among all the investor characteristics we consider, return chasing has the strongest predictive power both for future investor returns and for stock returns.

Our results emphasize that heterogeneity in ownership structure can deliver rich information. Knowing "who owns what" helps us understand the cross-section of household positions as well as the performance of stocks that are held mainly by retail investors. This point is also stressed by Balasubramaniam et al. (2023). We highlight the importance of return chasing behavior; however, we are silent on the underlying reasons why investors differ in their propensities to chase returns. The heterogeneity may due to differences in beliefs or preferences. There may be different behavioral models they use, or different information sets they receive. Future research to quantify the contribution of various channels of return chasing can be very interesting.

One might worry that, in practice, investors may not have the administrative data we use, so the trading strategy in this project cannot be implemented. This is of course a valid concern. We demonstrate the profitable trading strategies for three reasons. First, we want to show the impacts of return chasing on stock performance with a rigorous analysis. Second, the strategy could be modified for other sources of ownership data; for example, brokerage data can also be used to compute the sample RCO (brokerage data are more available nowadays as in Liao et al. 2014 and Liao, Peng, and Zhu 2021). We conjecture that such a strategy could still work if a well-represented brokerage sample is used. Third, the existence of profitable trading strategies reveals the prevalence of return chasing behavior among retail investors, adding to the list of behavioral biases that affect such investors.

Finally, even though this paper uses Chinese stock market data, the large scale, high frequency, and representativeness of which are hard to match, we think similar patterns exist in the U.S. stock market as well, especially with the extensive retail involvement in that market since 2021. We are now experiencing a trend shifting towards direct investment. Our data on the Chinese stock market help us understand what happens when there is a large proportion of direct investing. It calls for attention to the roles that retail investors play in the stock market and how they influence stock performance.

Chapter 3

Negative-Profit IPOs: Trends, Characteristics, and Investor Perceptions

3.1 Introduction

Profitability plays a critical role in the IPO market, as companies with high profits are generally considered more attractive to investors due to their proven ability to generate income. However, in recent years, there has been a trend of more firms going public with negative profits. This shift challenges the traditional notion that profitability is a prerequisite for a successful IPO, and it begs the question of whether investors are adjusting their risk appetite in this changing market.

Our analysis of U.S. IPO data from 1980 to 2022 reveals a decrease in the average profitability of IPOs, with negative-profit IPOs experiencing a surge in popularity since 2010. What are the characteristics of these negative-profit IPOs? Do firms with negative profits at the time of issuance eventually catch up? What is the stock market's evaluation of these stocks? In this paper, we aim to shed light on the dynamics of negative-profit IPOs, exploring their prevalence, performance, and potential impact on the financial market.

Using data from CRSP, Compustat, and SDC Platinum, we identify two significant periods of increased prevalence of negative-profit IPOs in the U.S.—during the tech bubble period and from 2010 to 2022. We find that the decline in profitability is concentrated in the IPO market rather than among existing public firms. Negative-profit IPOs are more commonly found in the retail, technology, and pharmaceutical sectors and are often backed by venture capital.

Next, we compare both the fundamental and return performance of negative-profit IPOs to positive-profit IPOs. We find that negative-profit IPOs typically do not demonstrate significant improvement in fundamentals over the subsequent five years. In terms of stock returns, while negative-profit IPOs tend to have higher first-day returns, they generally under-perform in the long run on average.

Despite the worse average performance associated with negative-profit IPOs, they have a higher likelihood of delivering future superior returns. This potential for future success is reflected in the fatter right tail in return distributions that negative-profit IPOs have exhibited since 2010. This "higher risk, higher reward" pattern may explain investors' increased tolerance for negative profitability.

Related literature

The paper contributes to the IPO literature by systematically analyzing the trends and characteristics of IPO profitability. We document the prevalence of negative-profit IPOs since 2010 and contrast the performance of negative-profit and positive-profit IPOs both in terms of both future fundamental growth and market evaluation. Our findings are consistent with well-established phenomena in the IPO literature and provide new insights into the dimension of firm profitability. For example, the literature has extensively studied the phenomenon of "IPO underpricing," which refers to the high first-day returns for IPOs (Ritter 1991, Booth and Chua 1996, Purnanandam and Swaminathan 2004, Ljungqvist, Nanda, and Singh 2006, Ljungqvist 2007). Our paper demonstrates that this underpricing is particularly severe for negative-profit IPOs. In terms of long-term return performance, Loughran and Ritter (1995) documented the puzzle that new issuers tend to underperform in the long run, and we find that this underperformance is more common among negative-profit IPOs. The literature has also examined the role of venture capital in the IPO market (Lee and Wahal 2004, Kaplan and Lerner 2016, Hamao, Packer, and Ritter 2000), and we find that negative-profit IPOs are more likely to have venture capital backing.

There are different explanations for the increased acceptance of negative profitability among investors. Pástor and Veronesi (2003) demonstrate that market valuations can rise with greater uncertainty about future profitability, which is consistent with negative-profit IPOs having more dispersed future return distributions. Pflueger, Siriwardane, and Sunderam (2020) have shown that stocks with higher volatility can receive higher valuations when market risk is perceived low, which corresponds well to the observation that market risk has been low since 2010 and that negative-profit IPOs are often more volatile than positive-profit ones. Bali, Cakici, and Whitelaw (2011) find that investors have a high demand for lottery-like stocks, which may explain investors' interest in negative-profit IPOs in our case. Investors could also overestimate the possibility of negative-profit IPOs developing into future superstars (Kahneman and Tversky 1972; Gennaioli and Shleifer 2010; Bordalo et al. 2016). We leave it to future research to pinpoint the specific channels that contribute to the appeal of negative IPOs.

Organization of the paper

The remainder of the paper is organized as follows. In section 3.2, we will describe our dataset. Section 3.3 will present our findings on the decline in the average profitability of IPOs and the prevalence of negative-profit IPOs. Section 3.4 will compare the characteristics and performance of negative-profit IPOs and positive-profit IPOs. Section 3.5 will discuss the potential reasons for investors' increased tolerance of negative-profit IPOs. Finally, section 3.6 will conclude and discuss possible future research directions.

3.2 Data

We extracted monthly stock returns from the CRSP US Stock Database and identified a stock as an IPO when a new "permco" identifier appeared in the CRSP database. We obtained firm fundamental performance after IPO from the S&P Compustat Fundamentals database, which we updated annually. Firm fundamental performance prior to IPO was obtained from the Securities Data Company (SDC) new issues database. To ensure the accuracy of our IPO sample, we cross-referenced it with the list on Jay R. Ritter's website.[1] Industry classifications are obtained from Ken French's website.[2]

3.3 Rise in IPOs with Negative Profit since 2010

Let us begin by examining the historical trend in IPO profitability. We measure profitability as the ratio of operating income before depreciation over total assets. Figure 3.1 shows the time-series of average IPO profitability. The blue solid line plots the equal-weighted average profitability for all IPOs, while the red dashed line plots the share of IPOs with negative profits in each year. There are two periods when IPO profitability substantially decreased. One occurred during the tech bubble period, and the other from 2010 to 2022. The share of negative-profit IPOs increased from about 30% before 2010 (excluding the tech bubble) to over 60% in 2020.

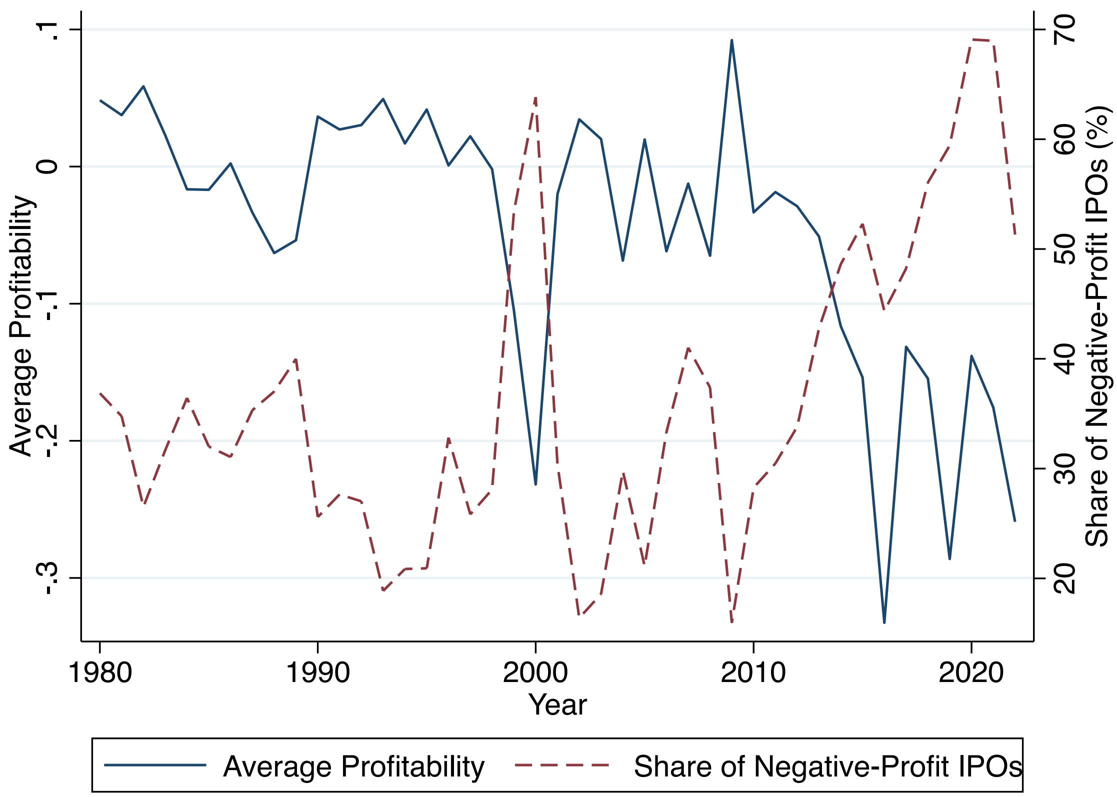
This pattern of declining profitability is robust to alternative profitability measures. In Figure 3.2, we plot the time series of other proxies for profitability. The dashed brown line measures profitability by the ratio of earnings over book equity, while the dash-dotted red line measures profitability by the ratio of free cash flow over book equity, where free cash flow is defined as net income plus depreciation and amortization, minus capital expenditures. Both earnings to book equity and free cash flow to book equity exhibit very similar trends to our baseline profitability measure. If anything, our baseline profitability measure serves as a lower bound for the declining trend since 2010.

One may wonder whether this declined profitability is an overall macroeconomic phenomenon or only restricted to the IPO sample. In Figure 3.3, we plot the time series of

1. See https://site.warrington.ufl.edu/ritter/ipo-data/.

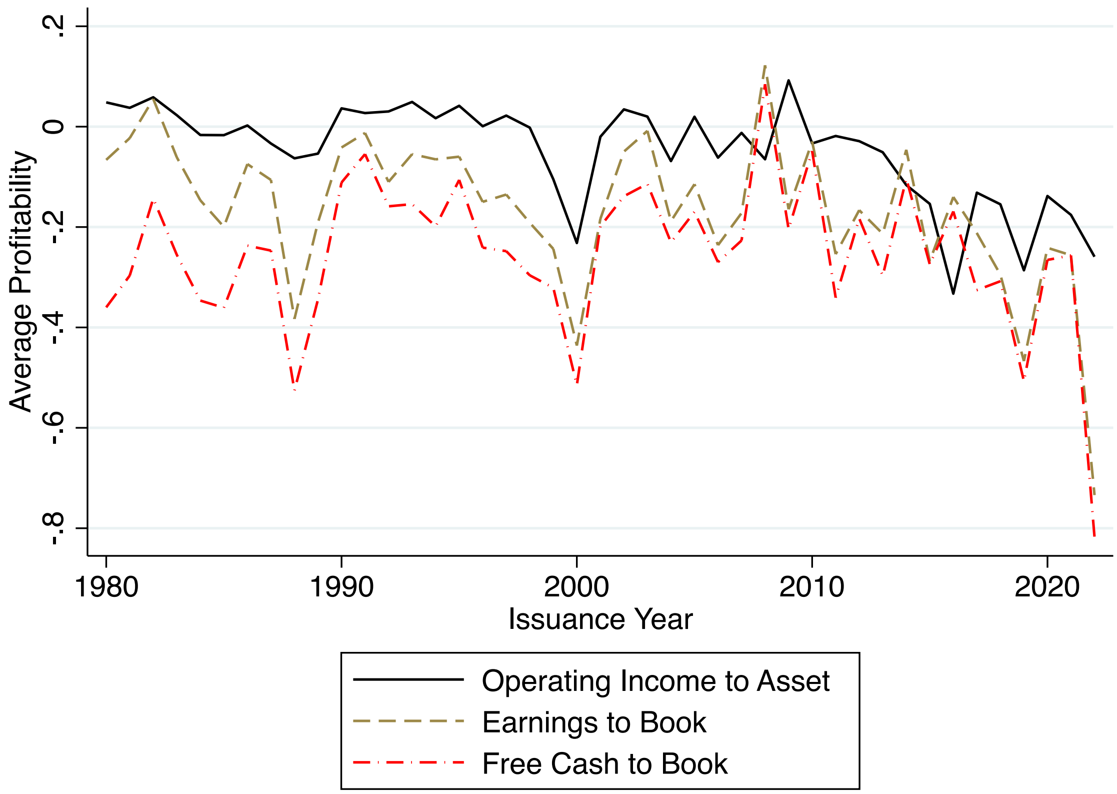
2. See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html on Industry Portfolios.

Figure 3.1: *Time Series of Profitability for IPOs*



Notes: Profitability is measured as the ratio of operating income before depreciation over total assets. The blue solid line plots the equal-weighted average profitability for all IPOs, while the red dashed line plots the share of IPOs with negative profits in each year.

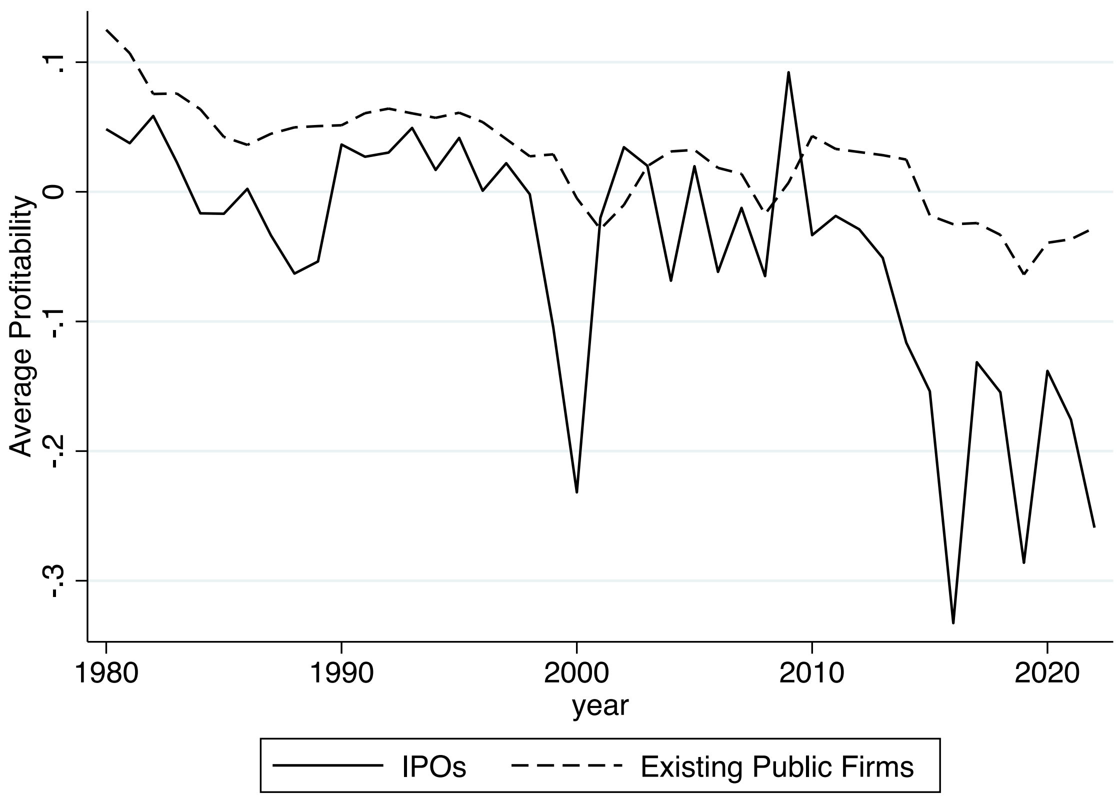
Figure 3.2: *Different Profitability Measures*



Notes: This figure plots the time series of profitability for IPOs, according to three different profitability measures. The solid black line refers to our baseline profitability measure: the ratio of operating income before depreciation over total asset. The dashed brown line measures profitability by the ratio of earnings over book equity. The dash-dotted red line measures profitability by the ratio of free cash flow over book equity, where free cash flow is defined as net income + depreciation & amortization − capital expenditures.

profitability for both the IPO sample and the existing public firm sample. We classify a stock as an IPO when a new "permco" identifier appears in the monthly CRSP database, while stocks with existing "permco" identifiers are classified as existing public firms. We observe that although existing public firms have slightly decreased their profitability over time, the sharp decline in profitability since 2010 is mostly concentrated in the IPO sample.

Figure 3.3: *Profitability of IPOs and Other Public Firms*



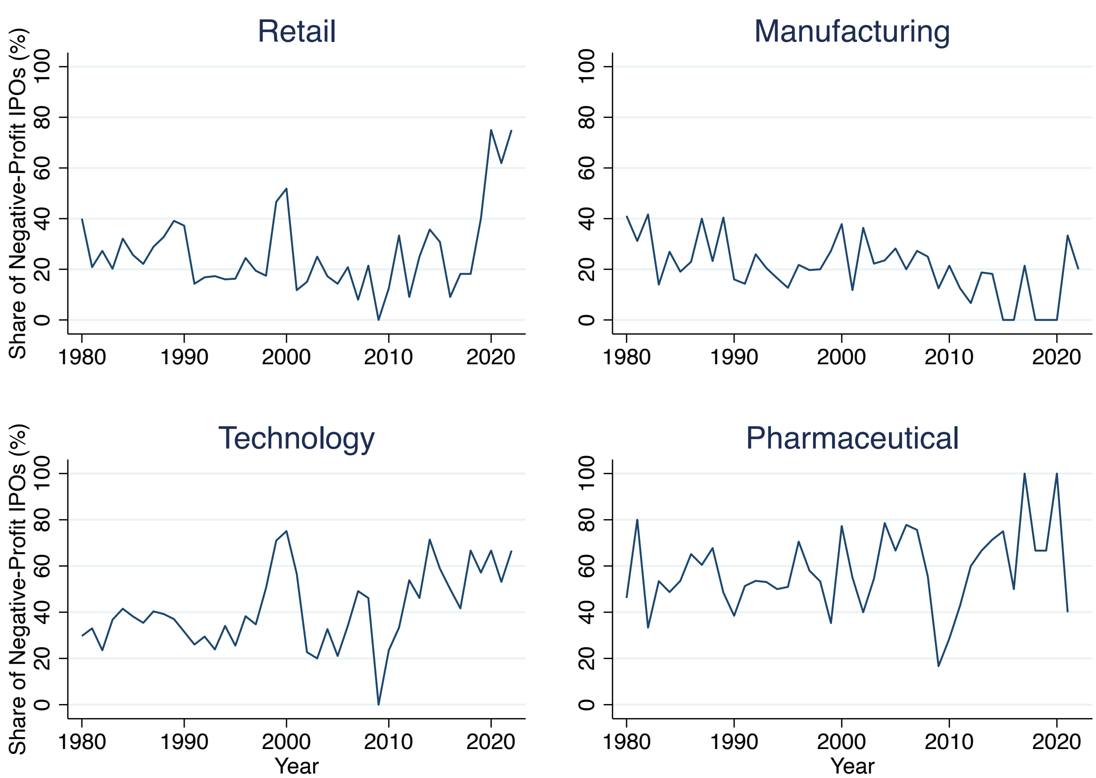
Notes: This figure plots the time series of profitability for both the IPO sample and the existing public firm sample. We classify a stock as an IPO when a new "permco" identifier appears in the monthly CRSP database. Stocks with existing "permco" are classified as existing public firms.

3.3.1 Industry Heterogeneity

Having demonstrated the overall declining trend of IPO profitability, Figure 3.4 zooms in on different industries to examine the industry heterogeneity of the increased prevalence of negative-profit IPOs. We classify industries according to the five-industry classification on

Ken French's website, omitting the fifth industry–"Others"–from this figure. We can see that the rise of negative-profit IPOs is concentrated in the retail, technology, and pharmaceutical industries, while the manufacturing industry does not exhibit an increase of negative-profit IPOs. Both retail and technology industries experienced an increase in negative-profit IPOs during the tech bubble. The retail industry began to experience a surge of negative-profit IPOs since 2020, while the technology industry had a higher prevalence of negative-profit IPOs starting much earlier, from 2010. The pharmaceutical industry has always had a higher ratio of IPOs with negative profits, and more firms went to IPO with negative-profit starting in 2017.

Figure 3.4: *Share of IPOs with Negative Profits in Different Industries*

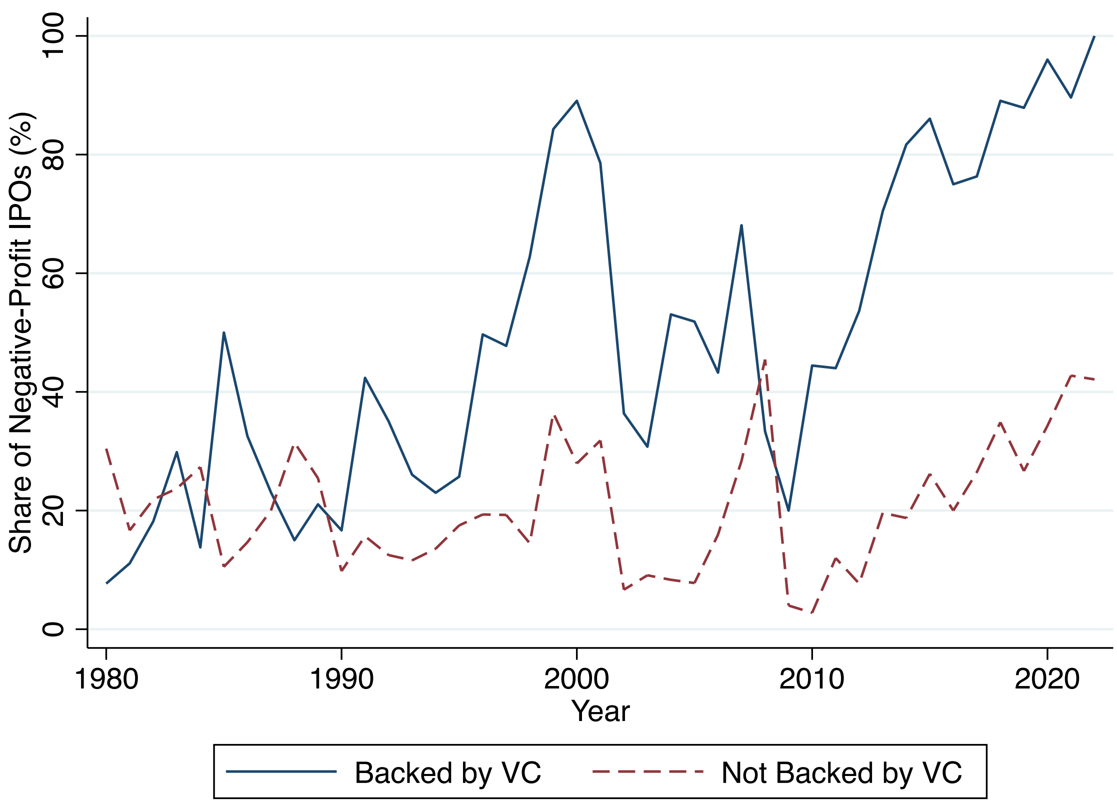


Notes: This figure plots the share of IPOs with negative profits in different industries. We classify industries according to the five-industry classification on Ken French's website, omitting the fifth industry–"Others"–from this figure.

3.3.2 The Role of Venture Capital

Venture capital is known for having a higher risk tolerance than other investors. With the increasing popularity of venture capital in recent decades, it is worth examining whether venture capitalists have also shifted their preferences towards firms that have yet to become profitable. Figure 3.5 plots the share of IPOs with negative profits according to whether the IPO is backed by venture capital or not. The overall trend of negative profitability is similar between IPOs backed by venture capital and those not backed by VC. However, the increase in negative profitability is much more pronounced for VC-backed IPOs, both during the tech bubble period and from 2010 through 2022. At the end of our sample period, all VC-backed IPOs have negative profits.

Figure 3.5: *Profitability for Venture Capital-Backed IPOs*



Notes: This figure plots the share of IPOs with negative profits according to whether the IPO is backed by venture capital or not.

Given the rapid increase in negative-profit IPOs backed by VC, it is reasonable to conjecture that venture capital has played a significant role in contributing to the declining trend of IPO profitability.

3.4 Negative-Profit vs. Positive-Profit IPOs

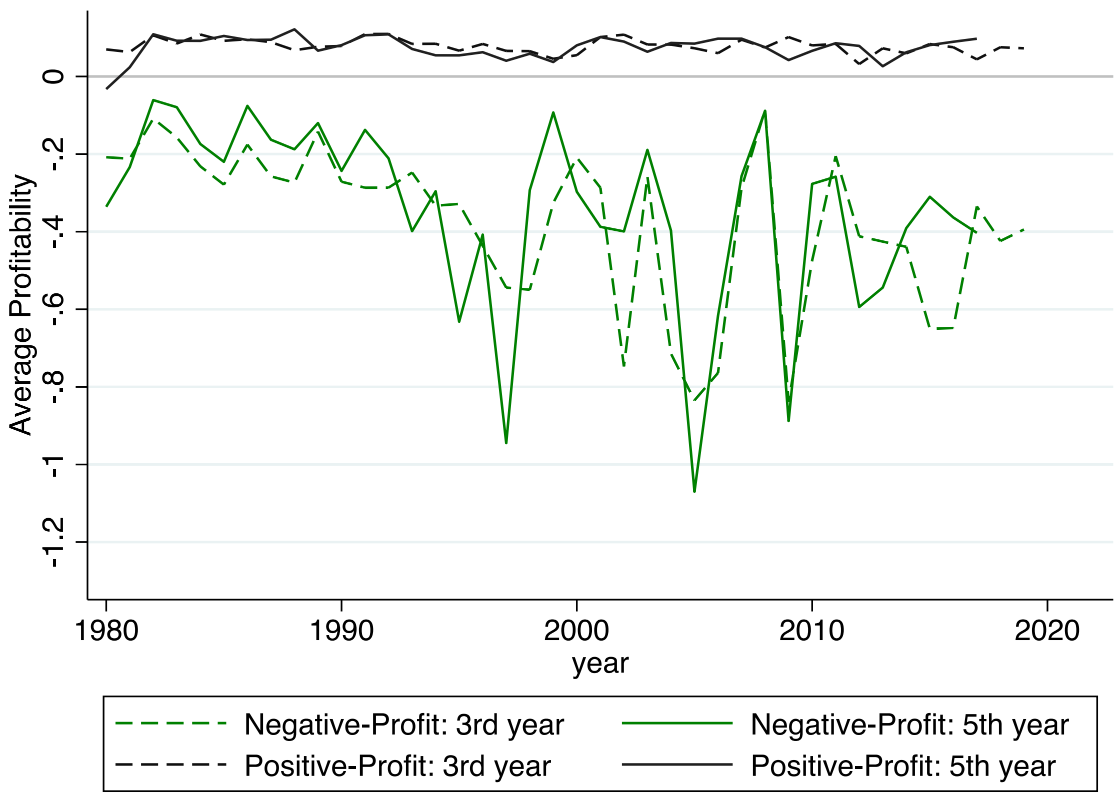
In this section, we will focus on the cross-sectional results by comparing the fundamental growth and return performance of negative-profit IPOs and positive-profit IPOs, building upon the time-series pattern of declined IPO profitability we have demonstrated.

3.4.1 Fundamental Growth

Firms that go to IPO with negative profits have worse fundamental situations to begin with. One may wonder whether these negative-profit IPOs catch up in the future. To explore this, we plot the future average profitability for negative-profit and positive-profit IPOs in Figure 3.6. The green lines represent negative-profit IPOs, while the black lines represent positive-profit IPOs. The dashed lines represent profitability performance in the third year after IPO, while the solid lines represent profitability performance in the fifth year after IPO. We observe that positive-profit IPOs remain to sustain positive profits, while negative-profit IPOs continue to lose money even after five years of IPO. Similar patterns can be seen in Figure 3.7, where we capture future performance with an alternative profitability measure—earnings per share (EPS). Furthermore, the EPS gaps between positive-profit and negative-profit IPOs widen over time.

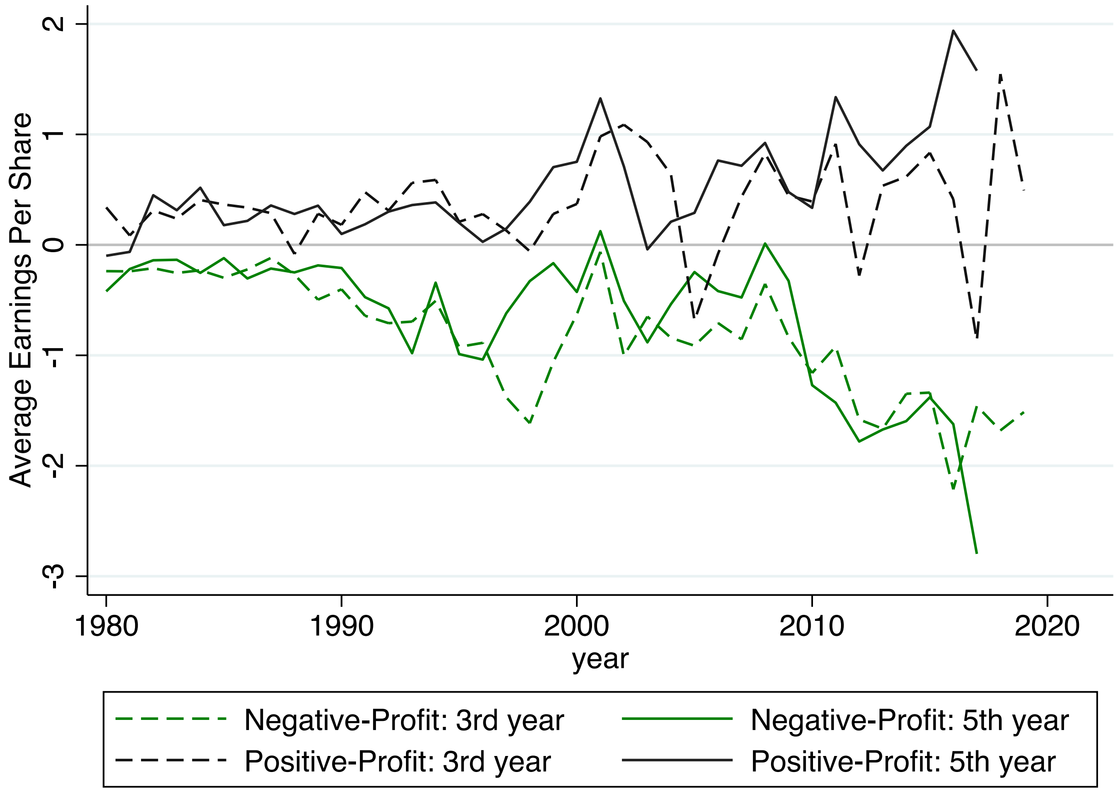
Admittedly, the sample period of this declining profitability phenomenon is relatively short, spanning only 12 years between 2010 and 2022. Thus, it is possible that better future performance could be realized beyond our sample period, say, in the 20th year. However, our results indicate that there is no evidence to suggest that negative-profit IPOs will catch up in terms of fundamentals, even after five years of the IPO.

Figure 3.6: *Future Profitability for Negative-Profit and Positive-Profit IPOs*



Notes: This figure plots future average profitability for negative-profit and positive-profit IPOs. The green lines represent negative-profit IPOs, while the black lines represent positive-profit IPOs. The dashed lines represent profitability performance in the third year after IPO, while the solid lines represent profitability performance in the fifth year after IPO.

Figure 3.7: *Future Earnings Per Share (EPS) for Negative-Profit and Positive-Profit IPOs*



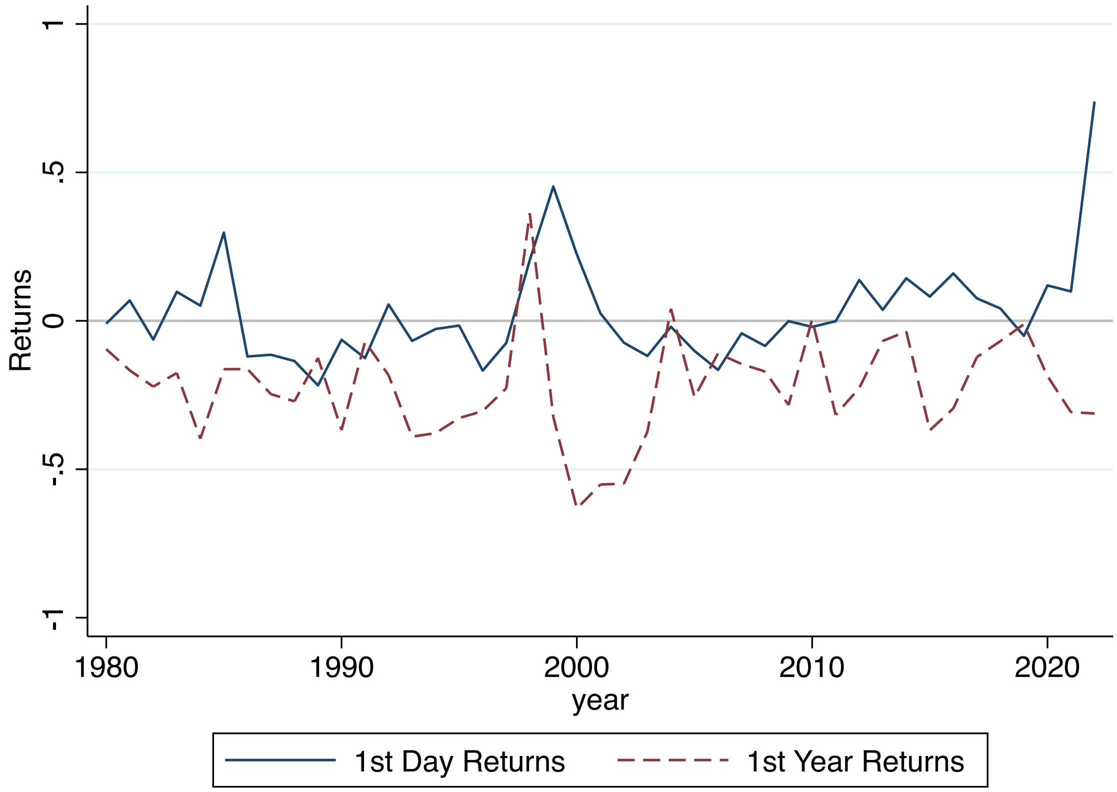
Notes: This figure plots future average earnings per share for negative-profit and positive-profit IPOs. The green lines represent negative-profit IPOs, while the black lines represent positive-profit IPOs. The dashed lines represent earnings per share in the third year after IPO, while the solid lines represent earnings per share in the fifth year after IPO.

3.4.2 Return Performance

Having demonstrated that negative-profit IPOs do not exhibit promising future fundamental growth, we move on to explore how the financial market evaluates these firms. We compare both the short-term and long-term return performance of IPOs in this subsection, using Figure 3.8 to plot return differences between negative-profit and positive-profit IPOs.

The blue solid line represents the equal-weighted average first day return difference when long negative-profit IPOs and short positive-profit IPOs. The dash red line represents the equal-weighted average return difference during the first year of IPO (including the

Figure 3.8: *Future Return Differences Between Negative-Profit and Positive-Profit IPOs*



Notes: This figure plots return differences between negative-profit and positive-profit IPOs. The blue solid line represents the equal-weighted average first day return difference when long negative-profit IPOs and short positive-profit IPOs. The dash red line represents the equal-weighted average return difference during the first year of IPO (including the first-day return) when long negative-profit IPOs and short positive-profit IPOs. If an IPO gets delisted, the return is imputed as zero.

first-day return) when long negative-profit IPOs and short positive-profit IPOs. If an IPO gets delisted, the return is imputed as zero. We see that the first-day return differences are generally positive, indicating that negative-profit IPOs obtain higher returns during the first day of IPO. However, the first-year return differences are mostly negative, suggesting that negative-profit IPOs earn lower returns during the first year of IPO. The results indicate that negative-profit IPOs are likely over-valued during the first day, and the returns revert back as time passes.

3.4.3 Return Distributions

While the long-run average returns for negative-profit IPOs are generally lower than those of positive-profit IPOs, investors with higher risk appetite may still prefer negative-profit IPOs if these stocks have a greater chance to achieve superior returns.

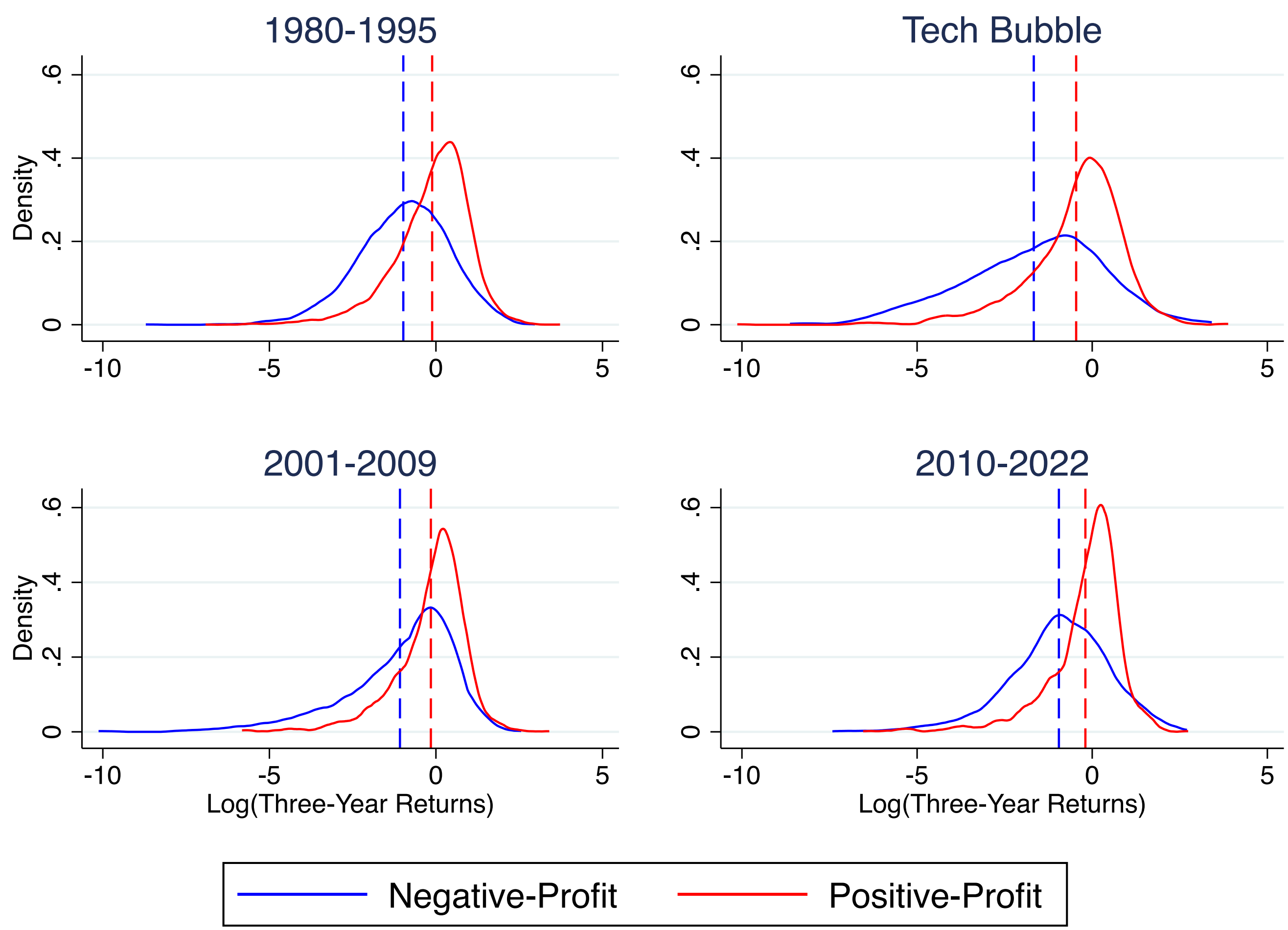
Figure 3.9 displays the density distributions for the logarithmic value of three-year returns during four periods: prior to the tech bubble (1980-1995), during the tech bubble (1996-2000), after the tech bubble until 2010 (2001-2009), and from 2010-2012. The vertical lines represent the average returns for negative-profit and positive-profit IPOs, respectively. The dashed lines depict equal-weighted returns, while the dotted lines represent proceeds-weighted returns. The figure reveals that during 1980-1995 and 2001-2009, the return distribution of negative-profit IPOs is completely dominated by the return distribution of positive-profit IPOs: negative-profit IPOs have lower average returns (both equal-weighted and proceeds-weighted) and do not exhibit fatter right tails. However, during the tech bubble and from 2010-2012, negative-profit IPOs still perform worse in terms of average returns, but they have a fatter right tail. This implies that negative-profit IPOs have a higher chance of achieving superior returns during these periods, making them more attractive to investors who pay attention to tail possibilities.

One might question whether there is a significant difference in tail fatness between negative-profit and positive-profit IPOs since 2010. To examine this, we can utilize quantile regressions. Specifically, we can run the following regression:

$$Q_\tau\left(r_i\right) = \alpha(\tau) + \beta(\tau)\mathbb{1}_{\text{Negative-Profit},i} + \epsilon_i. \tag{3.1}$$

Here, $Q_\tau\left(r_i\right)$ represents the τ-th quantile of the logarithmic value of three-year returns for IPO i, and $\mathbb{1}_{\text{Negative-Profit},i}$ is an indicator variable that denotes negative profits for IPOs. We report the regression findings in Table 3.1. The first panel covers observations from before 2010, excluding the tech bubble period of 1996-2000, while the second panel covers observations from after 2010. Columns (1), (2), and (3) report the 90th, 95th, and 99th quantile regression results, respectively.

Figure 3.9: *Return density distribution for IPOs with negative and positive profits*



Notes: This figure plots the density distributions for the logarithmic value of three-year returns across various periods. The vertical lines represent the average returns for negative-profit and positive-profit IPOs, respectively. The dashed lines depict equal-weighted returns, while the dotted lines represent proceeds-weighted returns.

Prior to 2010, negative-profit IPOs underperformed positive-profit IPOs across all quantiles. However, after 2010, negative-profit IPOs had higher three-year returns at the 95th and 99th percentiles. The differences in returns are statistically significant at a confidence level of 95%.

Table 3.1: *Quantile Regression Analysis of Three-Year Future Returns*

	Log(Three-Year Returns)		
	90th	95th	99th
	(1)	(2)	(3)
Before 2010			
Negative profit	-0.372***	-0.273***	-0.215**
	(0.0312)	(0.0420)	(0.0877)
Observations		9,979	
After 2010			
Negative profit	-0.00186	0.196**	0.424***
	(0.0520)	(0.0882)	(0.116)
Observations		2,462	

Notes: This table presents the results of a quantile regression analysis of the logarithmic value of three-year returns based on an indicator variable indicating negative profits for IPOs. The first panel includes observations prior to 2010, excluding the tech bubble period of 1996-2000, while the second panel includes observations after 2010. Columns (1), (2), and (3) report the 90th, 95th, and 99th quantile regression results, respectively. Standard errors are reported in parentheses, and statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% levels, respectively.

To formally examine the tail fatness, we estimate the Pareto coefficients for negative-profit IPOs. The Pareto coefficient is a summary statistic for tail fatness: the lower the Pareto coefficient, the fatter the right tail. We follow the estimation procedure in Gabaix and Ibragimov (2011) by estimating $\log(\text{Rank} - \frac{1}{2}) = a - \zeta \log(Z)$ to adjust for small sample bias. Table 3.2 reports the estimated Pareto coefficients ζ for IPOs with negative profits. We consider three return distributions: one-year, three-year, and five-year returns after IPO. Column (1) reports ζ before 2010 (excluding the tech bubble period 1996-2000), and column (2) reports ζ after 2010. Standard errors are reported in parentheses. For all three return measures, the estimated Pareto coefficients ζ become lower after 2010, indicating that the right tail for negative-profit IPOs becomes fatter since 2010.

Table 3.2: *Estimated Return Pareto Coefficient ζ for IPOs with Negative Profits*

	(1)	(2)
	$\zeta_{\text{before 2010}}$	$\zeta_{\text{after 2010}}$
1-year Return	2.52***	1.84***
	(0.50)	(0.37)
3-year Return	2.69***	2.43***
	(0.54)	(0.49)
5-year Return	1.91***	1.74***
	(0.38)	(0.35)

Notes: The table reports the estimated Pareto coefficients ζ for IPOs with negative profits. The Pareto coefficient is a summary statistic for tail fatness: the lower the Pareto coefficient, the fatter the right tail. We consider three return distributions: one-year, three-year, and five-year returns after IPO. Column (1) reports ζ before 2010 (excluding the tech bubble period 1996-2000), and column (2) reports ζ after 2010. Standard errors are reported in parentheses. We follow the estimation procedure in Gabaix and Ibragimov (2011) by estimating $\log(\text{Rank} - \frac{1}{2}) = a - \zeta \log(Z)$ to adjust for small sample bias. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively.

3.5 Possible Channels

There are several potential factors that could explain investors' increasing acceptance of negative profitability in IPOs. In this section, we discuss some possible explanations and leave further investigation to future research.

The growing popularity of negative-profit IPOs can be rational. Pástor and Veronesi (2003) found that higher uncertainty regarding future profitability can lead to increased market valuations. This idea is consistent with the notion that negative-profit IPOs have more diverse future return distributions since 2010. Additionally, Pflueger, Siriwardane, and Sunderam (2020) demonstrated that stocks with higher volatility can receive higher valuations when market risk is perceived to be low. This explanation fits well with the observation that market risk has been low since 2010 and that negative-profit IPOs are often more volatile than positive-profit ones. Finally, as previously noted, it is possible that the superior future performance of negative-profit IPOs has yet to be observed due to the limited sample period available.

Behavioral factors may also play a role in the rise of negative-profit IPOs. For example, Bali, Cakici, and Whitelaw (2011) found that some investors have a strong preference for lottery-like stocks, which could include negative-profit IPOs since 2010. Another possibility is that investors overestimate the likelihood of negative-profit IPOs becoming future superstars (Kahneman and Tversky 1972; Gennaioli and Shleifer 2010; Bordalo et al. 2016).

It is also possible that a combination of rational and behavioral investors is involved. Savvy investors with good stock-picking skills are more likely to identify the best-performing negative-profit IPOs, while less sophisticated investors may only select average negative-profit IPOs and suffer lower returns.

3.6 Conclusion

Our study sheds light on the changing dynamics of the IPO market, where profitability is no longer a critical determinant of a successful IPO. Negative-profit IPOs have become increasingly prevalent, particularly in the retail, technology, and pharmaceutical sectors, often with the backing of venture capital. However, our findings suggest that negative-profit IPOs do not generally exhibit significant improvement in fundamentals over the subsequent five years compared to positive-profit IPOs. While these negative-profit IPOs may have higher first-day returns, they tend to under-perform in the long run, albeit with a greater chance of delivering superior returns. Investors' growing acceptance of negative profitability could be due to factors such as higher uncertainty, lower perceived market risk, and the potential for superior future performance. Behavioral factors, such as investors' preference for lottery-like stocks and overestimating the likelihood of negative-profit IPOs becoming future superstars, may also play a role. Further research could be conducted to better understand the drivers of this phenomenon and its implications for investors and the financial market as a whole.

References

Abreu, Dilip, and Markus K. Brunnermeier. 2003. “Bubbles and crashes.” *Econometrica* 71 (1): 173–204.

Agnew, Julie, Pierluigi Balduzzi, and Annika Sunden. 2003. “Portfolio choice and trading in a large 401 (k) plan.” *American Economic Review* 93 (1): 193–215.

An, Li, Dong Lou, and Donghui Shi. 2022. “Wealth redistribution in bubbles and crashes.” *Journal of Monetary Economics* 126:134–153. ISSN: 0304-3932.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2006. “The cross-section of volatility and expected returns.” *The Journal of Finance* 61 (1): 259–299.

Asquith, Paul, Parag A Pathak, and Jay R Ritter. 2005. “Short interest, institutional ownership, and stock returns.” *Journal of Financial Economics* 78 (2): 243–276.

Atmaz, Adem, and Suleyman Basak. 2018. “Belief dispersion in the stock market.” *The Journal of Finance* 73 (3): 1225–1279.

Bailey, Warren, Jun Cai, Yan Leung Cheung, and Fenghua Wang. 2009. “Stock returns, order imbalances, and commonality: Evidence on individual, institutional, and proprietary investors in China.” *Journal of Banking & Finance* 33 (1): 9–19.

Baker, Malcolm, and Jeremy C Stein. 2004. “Market liquidity as a sentiment indicator.” *Journal of Financial Markets* 7 (3): 271–299.

Baker, Malcolm, and Jeffrey Wurgler. 2006. “Investor sentiment and the cross-section of stock returns.” *The Journal of Finance* 61 (4): 1645–1680.

Balasubramaniam, Vimal, Tarun Ramadorai, John Y Campbell, and Benjamin Ranish. 2023. “Who Owns What? A Factor Model for Direct Stockholding.” *The Journal of Finance.*

Bali, Turan G, Nusret Cakici, and Robert F Whitelaw. 2011. “Maxing out: Stocks as lotteries and the cross-section of expected returns.” *Journal of financial economics* 99 (2): 427–446.

Banerjee, Abhijit V. 1992. “A simple model of herd behavior.” *The Quarterly Journal of Economics* 107 (3): 797–817.

Barber, Brad M, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean. 2009. "Just how much do individual investors lose by trading?" *The Review of Financial Studies* 22 (2): 609–632.

Barber, Brad M, and Terrance Odean. 2001. "Boys will be boys: Gender, overconfidence, and common stock investment." *The Quarterly Journal of Economics* 116 (1): 261–292.

———. 2013. "The behavior of individual investors." In *Handbook of the Economics of Finance,* 2:1533–1570. Elsevier.

———. 2000. "Trading is hazardous to your wealth: The common stock investment performance of individual investors." *The Journal of Finance* 55 (2): 773–806.

Barber, Brad M, Terrance Odean, and Ning Zhu. 2009. "Systematic noise." *Journal of Financial Markets* 12 (4): 547–569.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015. "X-CAPM: An extrapolative capital asset pricing model." *Journal of Financial Economics* 115 (1): 1–24.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A model of investor sentiment." *Journal of Financial Economics* 49 (3): 307–343.

Bian, Jiangze, Zhi Da, Zhiguo He, Dong Lou, Kelly Shue, and Hao Zhou. 2021. "Margin trading and leverage management." *University of Chicago, Becker Friedman Institute for Economics Working Paper,* nos. 2021-29.

Bianchi, Milo. 2018. "Financial literacy and portfolio dynamics." *The Journal of Finance* 73 (2): 831–859.

Bilias, Yannis, Dimitris Georgarakos, and Michael Haliassos. 2010. "Portfolio inertia and stock market fluctuations." *Journal of Money, Credit and Banking* 42 (4): 715–742.

Booth, James R, and Lena Chua. 1996. "Ownership dispersion, costly information, and IPO underpricing." *Journal of Financial Economics* 41 (2): 291–310.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *The Quarterly Journal of Economics* 131 (4): 1753–1794.

Brown, John Dennis. 1991. *101 Years on Wall Street: An Investor's Almanac.* Prentice Hall Direct.

Brunnermeier, Markus K., and Stefan Nagel. 2004. "Hedge funds and the technology bubble." *The Journal of Finance* 59 (5): 2013–2040.

Brunnermeier, Markus K, and Martin Oehmke. 2013. "Bubbles, financial crises, and systemic risk." *Handbook of the Economics of Finance* 2:1221–1288.

Bryzgalova, Svetlana, Anna Pavlova, and Taisiya Sikorskaya. 2022. "Retail Trading in Options and the Rise of the Big Three Wholesalers." *Available at SSRN.*

Calvet, Laurent E, John Y Campbell, and Paolo Sodini. 2007. "Down or out: Assessing the welfare costs of household investment mistakes." *Journal of Political Economy* 115 (5): 707–747.

———. 2009. "Fight or flight? Portfolio rebalancing by individual investors." *The Quarterly Journal of Economics* 124 (1): 301–348.

Campbell, John Y. 2006. "Household Finance." *The Journal of Finance* 61 (4): 1553–1604.

Campbell, John Y, Tarun Ramadorai, and Benjamin Ranish. 2014. "Getting better or feeling better? How equity investors respond to investment experience." *NBER Working Paper.*

Chen, Jiahui, and Qipu Qiu. 2017. "The 2015 Chinese stock market crash (in Chinese)." *Caixin,* https://special.caixin.com/event_0619_1/.

Chen, Weihua, Shushu Liang, and Donghui Shi. 2022. "Who chases returns? Evidence from the Chinese stock market." *SSRN (June 30, 2022).*

Chen, Yin. 2018. "Retail Investors Ownership Breadth and Stock Returns."

Chodorow-Reich, Gabriel, Adam M Guren, and Timothy J McQuade. 2022. "The 2000s housing cycle with 2020 hindsight: A Neo-Kindlebergerian view."

Chodorow-Reich, Gabriel, Plamen T Nenov, and Alp Simsek. 2021. "Stock market wealth and the real economy: A local labor market approach." *American Economic Review* 111 (5): 1613–57.

Choi, James J., Li Jin, and Hongjun Yan. 2012. "What Does Stock Ownership Breadth Measure?" *Review of Finance* 17, no. 4 (October): 1239–1278. ISSN: 1572-3097.

De Long, J Bradford, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. 1990a. "Noise trader risk in financial markets." *Journal of political Economy* 98 (4): 703–738.

———. 1990b. "Positive feedback investment strategies and destabilizing rational speculation." *the Journal of Finance* 45 (2): 379–395.

DeMarzo, Peter, Ron Kaniel, and Ilan Kremer. 2007. "Technological innovation and real investment booms and busts." *Journal of Financial Economics* 85 (3): 735–754.

Dhar, Ravi, and Ning Zhu. 2006. "Up close and personal: Investor sophistication and the disposition effect." *Management Science* 52 (5): 726–740.

Dichev, Ilia D. 2007. "What are stock investors' actual historical returns? Evidence from dollar-weighted returns." *American Economic Review* 97 (1): 386–401.

Dimmock, Stephen G, Roy Kouwenberg, Olivia S Mitchell, and Kim Peijnenburg. 2016. "Ambiguity aversion and household portfolio choice puzzles: Empirical evidence." *Journal of Financial Economics* 119 (3): 559–577.

Fama, Eugene F., and Kenneth R. French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33 (1): 3–56. ISSN: 0304-405X.

Frazzini, Andrea, and Owen A Lamont. 2008. "Dumb money: Mutual fund flows and the cross-section of stock returns." *Journal of Financial Economics* 88 (2): 299–322.

Gabaix, Xavier. 2009. "Power laws in economics and finance." *Annual Review of Economics* 1 (1): 255–294.

Gabaix, Xavier, and Rustam Ibragimov. 2011. "Rank - 1 / 2: A Simple Way to Improve the OLS Estimation of Tail Exponents." *Journal of Business & Economic Statistics* 29 (1): 24–39.

Gabaix, Xavier, and Ralph SJ Koijen. 2022. "In search of the origins of financial fluctuations: The inelastic markets hypothesis." *Available at SSRN.*

Gaudecker, Hans-Martin Von. 2015. "How does household portfolio diversification vary with financial literacy and financial advice?" *The Journal of Finance* 70 (2): 489–507.

Gennaioli, Nicola, and Andrei Shleifer. 2010. "What comes to mind." *The Quarterly Journal of Economics* 125 (4): 1399–1433.

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

Graham, Benjamin, and Jason Zweig. 1973. *The intelligent investor.* HarperBusiness Essentials New York.

Greenwood, Robin, and Stefan Nagel. 2009. "Inexperienced investors and bubbles." *Journal of Financial Economics* 93 (2): 239–258.

Greenwood, Robin, and Andrei Shleifer. 2014. "Expectations of returns and expected returns." *The Review of Financial Studies* 27 (3): 714–746.

Grinblatt, Mark, and Matti Keloharju. 2000. "The investment behavior and performance of various investor types: a study of Finland's unique data set." *Journal of Financial Economics* 55 (1): 43–67.

Grinblatt, Mark, Matti Keloharju, and Juhani T Linnainmaa. 2012. "IQ, trading behavior, and performance." *Journal of Financial Economics* 104 (2): 339–362.

Haddad, Valentin, Paul Huebner, and Erik Loualiche. 2022. "How competitive is the stock market? Theory, evidence from portfolios, and implications for the rise of passive investing."

Hamao, Yasushi, Frank Packer, and Jay R Ritter. 2000. "Institutional affiliation and the role of venture capital: Evidence from initial public offerings in Japan." *Pacific-Basin Finance Journal* 8 (5): 529–558.

Haorui, Bai, Wu Huihang, and Ke Yan. 2020. "On the Reason for "Weak Monthly Momentum Effect" in the Chinese Stock Market: Evidence from T+ 1 Overnight Discount." *Journal of Finance and Economics* 46 (04): 140–154.

Harris, Lawrence, and Eitan Gurel. 1986. "Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures." *the Journal of Finance* 41 (4): 815–829.

Hong, Harrison, Terence Lim, and Jeremy C Stein. 2000. "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies." *The Journal of Finance* 55 (1): 265–295.

Hong, Harrison, and Jeremy C Stein. 1999. "A unified theory of underreaction, momentum trading, and overreaction in asset markets." *The Journal of finance* 54 (6): 2143–2184.

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective probability: A judgment of representativeness." *Cognitive Psychology* 3 (3): 430–454.

Kaplan, Steven N, and Josh Lerner. 2016. *Venture capital data: Opportunities and challenges.* Technical report. National Bureau of Economic Research.

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

Kindleberger, Charles P. 1978. "Manias, panics and crashes: a history of financial crises." *The American Historical Review* 84, no. 2 (April): 416–417.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo. 2022. "Which Investors Matter for Equity Valuations and Expected Returns?"

Koijen, Ralph SJ, and Motohiro Yogo. 2019. "A demand system approach to asset pricing." *Journal of Political Economy* 127 (4): 1475–1515.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2020. "Shrinking the cross-section." *Journal of Financial Economics* 135 (2): 271–292.

Lee, Peggy M, and Sunil Wahal. 2004. "Grandstanding, certification and the underpricing of venture capital backed IPOs." *Journal of Financial Economics* 73 (2): 375–407.

Liao, Jingchi, Cameron Peng, and Ning N Zhu. 2021. "Extrapolative bubbles and trading volume." *Review of Financial Studies.*

Liao, Jingchi, Cameron Peng, and Ning Zhu. 2022. "Extrapolative bubbles and trading volume." *The Review of Financial Studies* 35 (4): 1682–1722.

Liao, Li, Zhisheng Li, Weiqiang Zhang, and Ning Zhu. 2014. "Exercise to lose money? Irrational exercise behavior from the Chinese warrants market." *Journal of Futures Markets* 34 (5): 399–419.

Liu, Jianan, Robert F Stambaugh, and Yu Yuan. 2019. "Size and value in China." *Journal of Financial Economics* 134 (1): 48–69.

Ljungqvist, Alexander. 2007. "IPO underpricing." In *Handbook of Empirical Corporate Finance,* 375–422. Elsevier.

Ljungqvist, Alexander, Vikram Nanda, and Rajdeep Singh. 2006. "Hot markets, investor sentiment, and IPO pricing." *the Journal of Business* 79 (4): 1667–1702.

Lou, Dong. 2012. "A flow-based explanation for return predictability." *The Review of Financial Studies* 25 (12): 3457–3489.

Loughran, Tim, and Jay R Ritter. 1995. "The new issues puzzle." *The Journal of Finance* 50 (1): 23–51.

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression babies: do macroeconomic experiences affect risk taking?" *The Quarterly Journal of Economics* 126 (1): 373–416.

Martin, Ian W. R., and Dimitris Papadimitriou. 2022a. "Sentiment and Speculation in a Market with Heterogeneous Beliefs." *American Economic Review* 112 (8): 2465–2517.

Martin, Ian, and Dimitris Papadimitriou. 2022b. "Sentiment and speculation in a market with heterogeneous beliefs." *American Economic Review.*

McConnell, John J., and Henri Servaes. 1990. "Additional evidence on equity ownership and corporate value." *Journal of Financial Economics* 27 (2): 595–612. ISSN: 0304-405X.

McCrank, John. 2021. "Factbox: The U.S. retail trading frenzy in numbers." *Reuters* (January 29, 2021). https://www.reuters.com/article/us-retail-trading-numbers-idUSKBN29Y2PW.

Merkle, Christoph, and Martin Weber. 2014. "Do investors put their money where their mouth is? Stock market expectations and investing behavior." *Journal of Banking & Finance* 46:372–386.

Mian, Atif, and Amir Sufi. 2009. "The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis." *The Quarterly Journal of Economics* 124 (4): 1449–1496.

Morck, Randall, Andrei Shleifer, and Robert W. Vishny. 1988. "Management ownership and market valuation: An empirical analysis." *Journal of Financial Economics* 20:293–315. ISSN: 0304-405X.

Nagel, Stefan. 2005. "Short sales, institutional investors and the cross-section of stock returns." *Journal of Financial Economics* 78 (2): 277–309.

Nagel, Stefan, and Zhengyang Xu. 2022. "Asset pricing with fading memory." *The Review of Financial Studies* 35 (5): 2190–2245.

Nakamura, Emi, and Jon Steinsson. 2014. "Fiscal stimulus in a monetary union: Evidence from US regions." *American Economic Review* 104 (3): 753–92.

Odean, Terrance. 1998. "Are investors reluctant to realize their losses?" *The Journal of Finance* 53 (5): 1775–1798.

Pástor, L'uboš, and Pietro Veronesi. 2003. "Stock valuation and learning about profitability." *The Journal of Finance* 58 (5): 1749–1789.

Pflueger, Carolin, Emil Siriwardane, and Adi Sunderam. 2020. "Financial market risk perceptions and the macroeconomy." *The Quarterly Journal of Economics* 135 (3): 1443–1491.

Pohl, Walt, Karl Schmedders, and Ole Wilms. 2020. "Asset pricing with heterogeneous agents and long-run risk." *Journal of Financial Economics.*

Porras Prado, Melissa, Pedro A. C. Saffi, and Jason Sturgess. 2016. "Ownership Structure, Limits to Arbitrage, and Stock Returns: Evidence from Equity Lending Markets." *The Review of Financial Studies* 29, no. 12 (July): 3211–3244. ISSN: 0893-9454.

Purnanandam, Amiyatosh K, and Bhaskaran Swaminathan. 2004. "Are IPOs really underpriced?" *The Review of Financial Studies* 17 (3): 811–848.

Ritter, Jay R. 1991. "The long-run performance of initial public offerings." *The Journal of Finance* 46 (1): 3–27.

Scheinkman, Jose A, and Wei Xiong. 2003. "Overconfidence and speculative bubbles." *Journal of political Economy* 111 (6): 1183–1220.

Shleifer, Andrei. 1986. "Do demand curves for stocks slope down?" *The Journal of Finance* 41 (3): 579–590.

Shleifer, Andrei, and Lawrence H Summers. 1990. "The noise trader approach to finance." *Journal of Economic perspectives* 4 (2): 19–33.

Xiong, Wei, and Jialin Yu. 2011. "The Chinese warrants bubble." *American Economic Review* 101 (6): 2723–53.

Yu, Fenghui. 2015. "Who causes the stock market skyrocketing? (in Chinese)." *The Beijing News,* http://epaper.bjnews.com.cn/html/2015-01/21/content_558501.htm?.

Appendix A

Appendix to Chapter 1

A.1 Omitted Derivations

A.1.1 Individual Demand Elasticity

Denote $\mathbf{q}_{i,t} = \log(A_{i,t}\mathbf{w}_{i,t}) - \mathbf{p}$ the vector of log shares held by investor i.

The elasticity of individual demand is:

$$-\frac{\partial \mathbf{q}_{i,t}}{\partial \mathbf{p}_t} = \begin{bmatrix} -\frac{\partial \mathbf{q}_{i,t}}{\partial p_t(1)}, & \cdots, & -\frac{\partial \mathbf{q}_{i,t}}{\partial p_t(n)} \end{bmatrix}$$
$$= \mathbf{I} - \beta_{0,i,t} \begin{bmatrix} 1 - w_{i,t}(1) & -w_{i,t}(2) & \cdots & -w_{i,t}(n) \\ -w_{i,t}(1) & 1 - w_{i,t}(2) & \cdots & -w_{i,t}(n) \\ \vdots & \vdots & \ddots & \vdots \\ -w_{i,t}(1) & -w_{i,t}(2) & \cdots & 1 - w_{i,t}(n) \end{bmatrix}, \tag{A.1}$$

where the diagonal terms are stock own-demand elasticities, and the off-diagnoal terms are cross-demand elasticities.

For investor i, a stock's own-demand elasticity (the diagonal term) depends on investor i's portfolio weight for the particular stock. For investor i at time t, the average own-demand elasticity across all the stocks held is:

$$\frac{1}{|\mathcal{N}_{i,t}|} \sum_{m \in \mathcal{N}_{i,t}} [1 - \beta_{0,i,t}(1 - w_{i,t}(m))] = 1 - \beta_{0,i,t} + \beta_{0,i,t} \frac{1 - w_{i,t}(0)}{|\mathcal{N}_{i,t}|}. \tag{A.2}$$

Since we have imposed that $|\mathcal{N}_{i,t}| \geq 500$, we estimate the average demand elasticity to be approximately as $1 - \beta_{0,i,t}$.

A.1.2 Aggregate Demand Elasticity

Denote $\mathbf{q}_t = \log\left(\sum_{i=1}^{I} A_{i,t}\mathbf{w}_{i,t}\right) - \mathbf{p}_t$ as the vector of log shares held across all investors.

The elasticity of aggregate demand is

$$
\begin{aligned}
-\frac{\partial \mathbf{q}_t}{\partial \mathbf{p}_t} &= \begin{bmatrix} -\frac{\partial \mathbf{q}_t}{\partial p_t(1)} & \cdots & -\frac{\partial \mathbf{q}_t}{\partial p_t(n)} \end{bmatrix} \\
&= \begin{bmatrix}
1 - \frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(1)(1-w_{i,t}(1))}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(1)} & \frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(1) w_{i,t}(2)}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(1)} & \cdots & \frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(1) w_{i,t}(n)}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(1)} \\
\frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(2) w_{i,t}(1)}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(2)} & 1 - \frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(2)(1-w_{i,t}(2))}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(2)} & \cdots & \frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(2) w_{i,t}(n)}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(2)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(n) w_{i,t}(1)}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(n)} & \frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(n) w_{i,t}(2)}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(n)} & \cdots & 1 - \frac{\sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} w_{i,t}(n)(1-w_{i,t}(n))}{\sum_{i=1}^{I} A_{i,t} w_{i,t}(n)}
\end{bmatrix}
\end{aligned}
\tag{A.3}
$$

A.2 Details on Data Construction

A.2.1 Industry Classification

We employ SWS industry classification, a commonly used industry classification standard for companies listed on the Chinese stock market. The 28 industries in SWS are: Agriculture, Banking, Building Materials, Chemical, Commerce & Trade, Communication, Computer, Conglomerate, Construction, Defense Military, Electrical Equipment, Electronics, Equipment, Food & Beverage, Household Appliances, Leisure Service, Light Manufacturing, Media, Medical Biology, Mining, Non-Bank Finance, Non-Ferrous Metal, Public Utility, Real Estate, Steel, Textile & Apparel, Transportation, and Vehicle.

A.2.2 Data from Financial Statements

For each stock at a given time, we obtain fundamentals data from the most recently available financial statement.

For context, the China Securities Regulatory Commission requires all listed companies to disclose their financial statements to the public within strict time windows. Annual reports are prepared and disclosed to the public within 4 months following the end of a fiscal year, semi-annual reports within 2 months following a sixth-month fiscal period, and quarterly reports within 1 month following a fiscal quarter. Annual reports and first quarter reports are both released by April.

According to the statement release schedule, we use the following algorithm to construct stock characteristics from financial statements:

1. For January, February, March, and April: we use third-quarter financial statements from the prior year (because first quarter and annual reports haven't been released yet).

2. For May, June, July, and August: we use first quarter reports in the current year. For some variables missing in the quarterly reports, we use information from the annual report in the previous year.

3. For September and October: we use semi-annual reports in the current year.

4. For November and December: we use third quarter reports in the current year.

A.2.3 Constructing Investor Return Chasing Propensity (RCP)

We construct the measure for an investor's return chasing propensity (RCP) following Chen, Liang, and Shi (2022). Investor RCP captures how aggressively a retail investor chases returns. To construct RCP, we look at each investor's stock holdings, match each stock to its returns prior to the time of purchase, and average these past returns across stocks to obtain an individual's RCP. We can formally denote the RCP of investor i at time t as:

$$\mathrm{RCP}_{i,t} = \sum_{s \in \{S_{i,t}\}} \left[\gamma_s \left(R^{s}_{t(s)-K,t(s)} - R^{M}_{t(s)-K,t(s)} \right) \right], \tag{A.4}$$

where $S_{i,t}$ is the collection of stocks that investor i holds at time t. For each stock s that the investor holds at time t, we first identify the time at which the investor purchased this stock: the purchase time is denoted as $t(s)$. We then calculate the K-period cumulative returns from time $t(s)-K$ to $t(s)$ for stock s, which is denoted as $R^s_{t(s)-K,t(s)}$. We subtract the corresponding market index returns $R^M_{t(s)-K,t(s)}$ during the same period from the cumulative stock returns. Finally, we weight each stock s by weight γ_s.

We follow Chen, Liang, and Shi (2022) and use $K = 12$ months and γ_s as equal weights.

A.2.4 Obtain a Representative Sample of Retail Accounts

To obtain a representative sample of retail accounts at the Shanghai Stock Exchange, we first randomly generate an anonymous ID for all retail accounts at the SSE. This account ID has trailing digit that randomly ranges from 0 to 9. We then select accounts with IDs ending with 1 or 6, which yields 20% of the entire retail population. This way, we obtain a random and representative sample of 18 million retail accounts.

A.2.5 Reweight Sample Retail Holdings to Align with the Size of the Full Retail Population

Our sample covers about 20% of the entire retail population. To assess the price impacts of retail investors, we are interested in obtaining an estimate of the account-level holdings of the entire retail population. One possible way to reweight the sample is to multiple the holdings of every retail account in our sample by 5, so that each account in our sample represents about 5 accounts in the entire retail population. We adopt a similar reweighting scheme but with better accuracy. Because we obtain the number of total shares held by all retail investors for each stock in each month, we can reweight sample holdings using a stock-month specific multiplier.

For each stock n at month t, we denote the number of total shares held by the entire retail population as $s^p_t(n)$, and the number of shares held by our retail sample as $s^s_t(n)$. We can compute the multiplier $x_t(n)$ for each stock n at month t:

$$x_t(n) = \frac{s^p_t(n)}{s^s_t(n)}. \tag{A.5}$$

The multiplier $x_t(n)$ is close to 5, confirming our random sample generation scheme above.

To get an estimate of the account-level holdings of the entire retail population, we multiply our sample account-level holdings by the calculated multiplier for each stock in each month. We denote the sample holdings for account i at month t as $s^s_{i,t}(n)$. We can therefore reweight to get the population holdings $s^p_{i,t}(n)$ for account i at month t:

$$s^p_{i,t}(n) = s^s_{i,t}(n)x_t(n), \forall i. \tag{A.6}$$

We use these reweighted holdings to assess price impacts of retail investors.

A.2.6 Wealth Group Classification

To classify retail investors into different wealth groups, we first calculate a retail investor's average wealth in the stock market in the prior year and then classify investors by their average wealth.

One problem with classifying investors into wealth groups is that most accounts are owned by investors with low wealth (lower than 100,000 CNY, approximately 15,000 USD); grouping retail investors into 5 quintiles would yield 4 groups that are all dominated by low-wealth retail investors. This phenomenon is also known as the power law of the wealth distribution (Gabaix 2009).

China Securities Regulatory Commission has classified retail investors into five groups according to their average stock market wealth. Their classification, however, has only existed since 2017. In their classification, the wealth ranges for the five groups are: less than 100K, 100K-500K, 500K-5M, 5M-10M, and above 10M (CNY). The five groups account for 71.7%, 19%, 8.6%, 0.4%, and 0.3% of accounts, respectively.

While this classification only appears in the data beginning in 2017, we extend it to prior years by ranking accounts by their average wealth in the previous year, and then assigning investors to five wealth groups using the aforementioned ranks and account number percentage ratios.

A.2.7 Classify Types for Retail Investors

To classify retail investor types, we first label them into the following five dimensions independently:

1. Entry cohort: 5 groups
2. Wealth in the stock market: 5 groups
3. Investors' tendency to buy high past return stocks (return chasing propensity: RCP): 5 quintiles
4. Gender: 2 groups
5. Age: 3 terciles

To make sure that each retail investor type holds more than 500 stocks in the cross-section, we use the following algorithm in the classification:

1. Classify investors by cohort according to their entry time.
2. Split each cohort into 5 wealth groups. If each subgroup still hold more than 500 stocks, we proceed with the subgroup classification; otherwise, we end the classification process for this month.
3. Split each subgroup into 5 RCP subgroups. If each subgroup still holds more than 500 stocks, we proceed with the subgroup classification; otherwise, we end the classification process for this month.
4. Split each entry time-wealth-RCP group into 2 gender groups. If each subgroup still holds more than 500 stocks, we proceed with the subgroup classification; otherwise, we end the classification process for this month.
5. Split each subgroup into 3 age groups. If each subgroup still holds more than 500 stocks, we proceed with the subgroup classification; otherwise, we end the classification process for this month.

This algorithm makes sure that each retail type holds more than 500 stocks.[1]

A.2.8 Classify Types for Institutions

We group institutions by institution category and AUM range to reach a minimum of 500 holdings for each institution type. We have, on average, 246 institution types.

Below are the five institution category:

1. Mutual funds
2. Hedge funds
3. Brokerage associated funds
4. Trusts
5. Others

For stable institutions (the National Social Security Funds, Corporate Supplementary Pension Funds, and Qualified Foreign Institutional Investors), we do not classify them since they are used to construct the instrument. Their holdings are taken as exogenously given and we do not estimate their demands.

We use the following algorithm to pool institutions within the same category and with similar assets under management:

1. We rank institutions within the same category by their AUM.
2. Institutions with more than 500 stocks are classified as a single type. Institutions with fewer than 500 stocks are grouped consecutively (by AUM rank) until the group collectively holds more than 500 stocks.

1. We set the threshold at 500 since with at least 500 observations, we can estimate the demand function accurately. Results are robust to thresholds, for instance, 300.

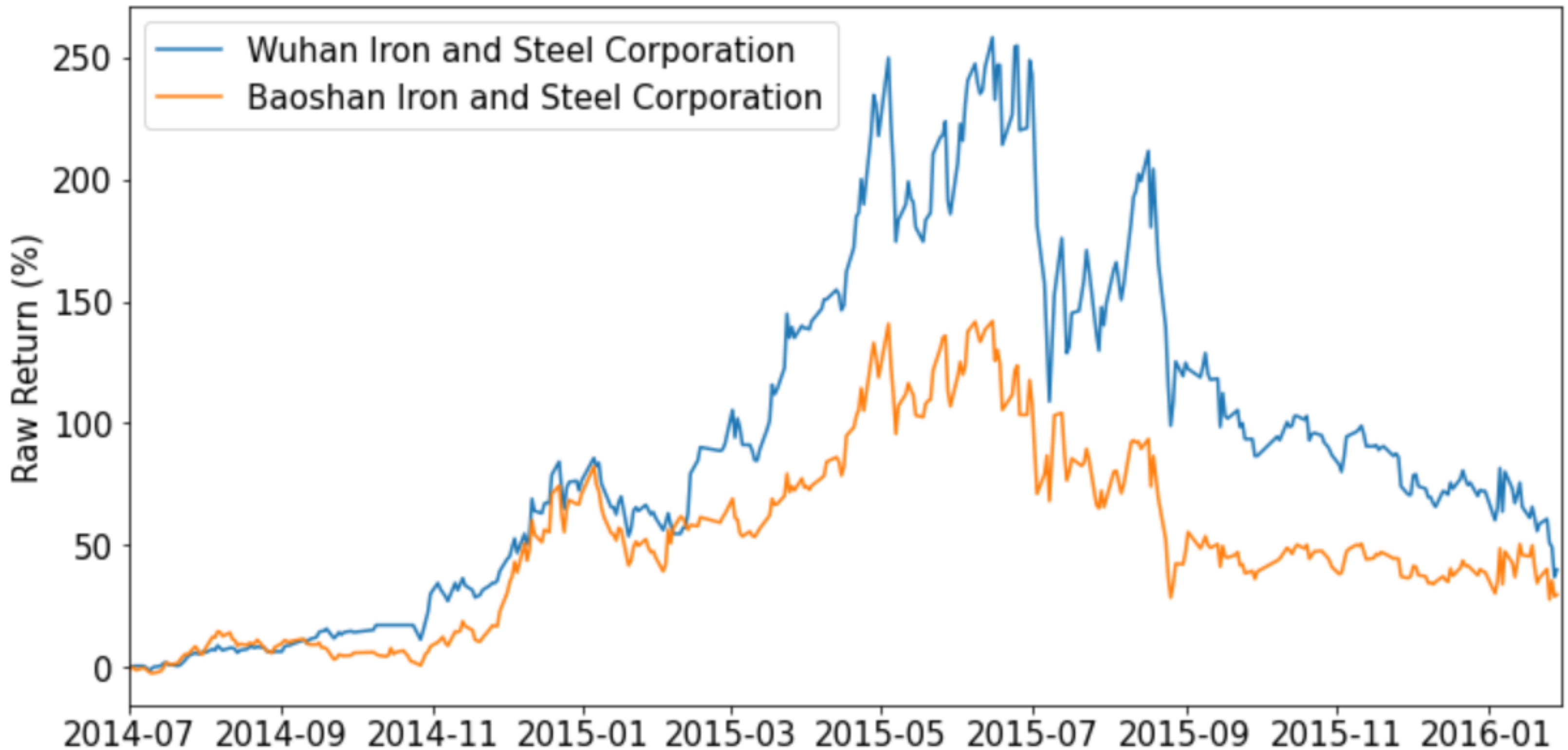
A.3 Preferences/Beliefs about Stock Characteristics

We additionally report investors' estimated preferences/beliefs about dividends, size, age, and the construction industry in Appendix Figures A.2 through A.7. The construction industry went through a sharp price increase and a sharp decline during the bubble, and serves as an example for the estimated industry fixed effects. In Appendix Figure A.2, we see that retail investors prefer stocks that have lower dividends and are smaller. In Appendix Figure A.3, we see that retail investors who had preferred older stocks before the bubble shifted to relatively younger stocks than institutions during the expansion and deflation phases of the bubble, and maintained their preferences for younger stocks during the post bubble period. Retail investors also have more enthusiasm on the construction industry than institutions.

In Appendix Figures A.4 and A.5, we show the heterogeneity of preferences/beliefs among retail investors by investor wealth. We see that low-wealth retail investors prefer boom-bust characteristics consistently. They like stocks that have lower dividends, are smaller, are younger, and are in the construction industry. The trends are almost always monotone: the less wealth an investor has, the more likely she will prefer boom-bust characteristics.

In Appendix Figures A.6 and A.7, we examine the heterogeneity among retail investors by market entry cohort. We see that investors who entered during the formation or expansion phase of the bubble prefer stocks that have lower dividends, are smaller, are younger, and are in the construction industry than the ones who entered prior to the bubble.

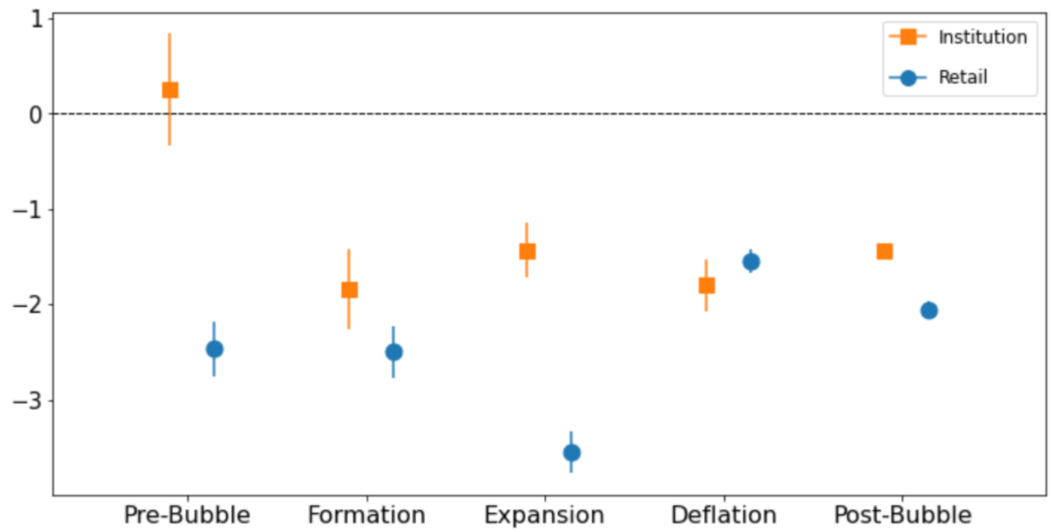
Figure A.1: *Cumulative Raw Returns for Two Illustrative Stocks during the 2015 Chinese Stock Market Bubble*



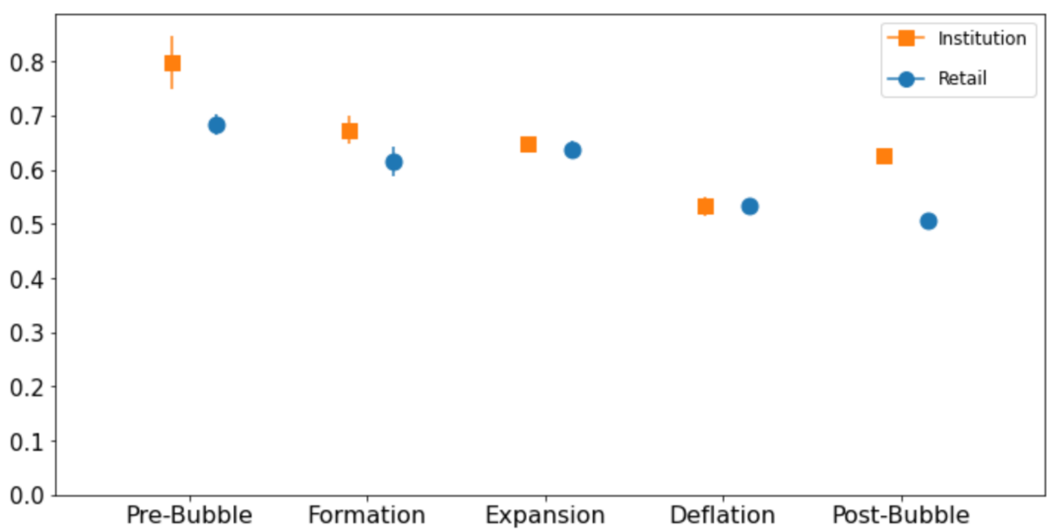
Notes: This figure plots the cumulative raw returns for two illustrative stocks, "Wuhan Iron and Steel Corporation" and "Baoshan Iron and Steel Corporation," during the bubble period.

Figure A.2: *Preferences/Beliefs of Retail Investors and Institutions (Dividend/Size)*

(a) *Preferences/Beliefs about Dividend*



(b) *Preferences/Beliefs about Size*



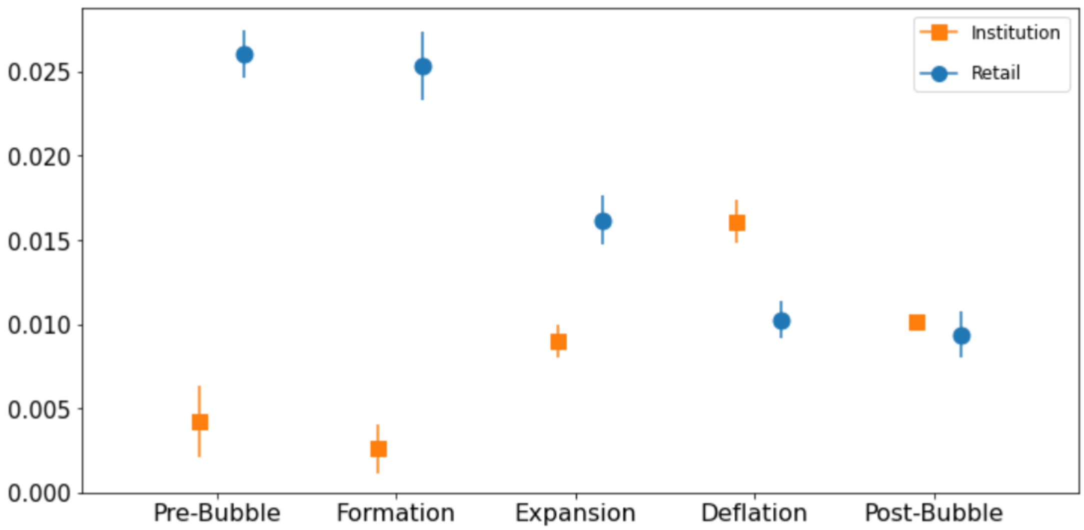
Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_{I,T}$) of retail investors and institutions over different time periods. $\hat{\gamma}_{I,T}$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}\left[i \in I, t \in T\right],$$

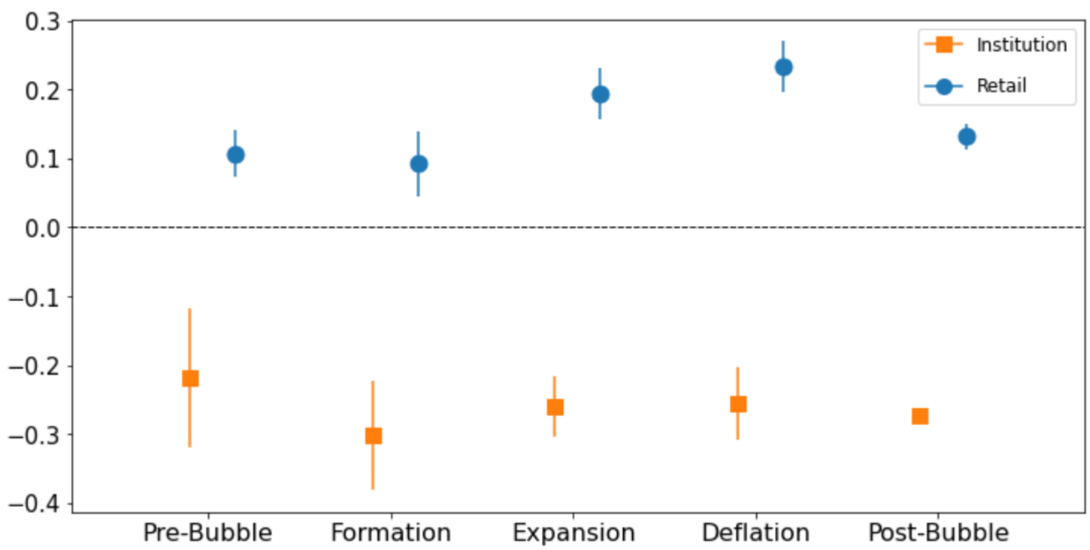
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $I \in \{\text{Retail, Institution}\}$, $T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}[i \in I, t \in T]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. Panel (a) reports investors' preferences/beliefs about dividend, and panel (b) reports investors' preferences/beliefs about size (measured by log book equity). Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Figure A.3: *Preferences/Beliefs of Retail Investors and Institutions (Age/Industry)*

(a) *Preferences/Beliefs about Age*



(b) *Preferences/Beliefs about Construction Industry*



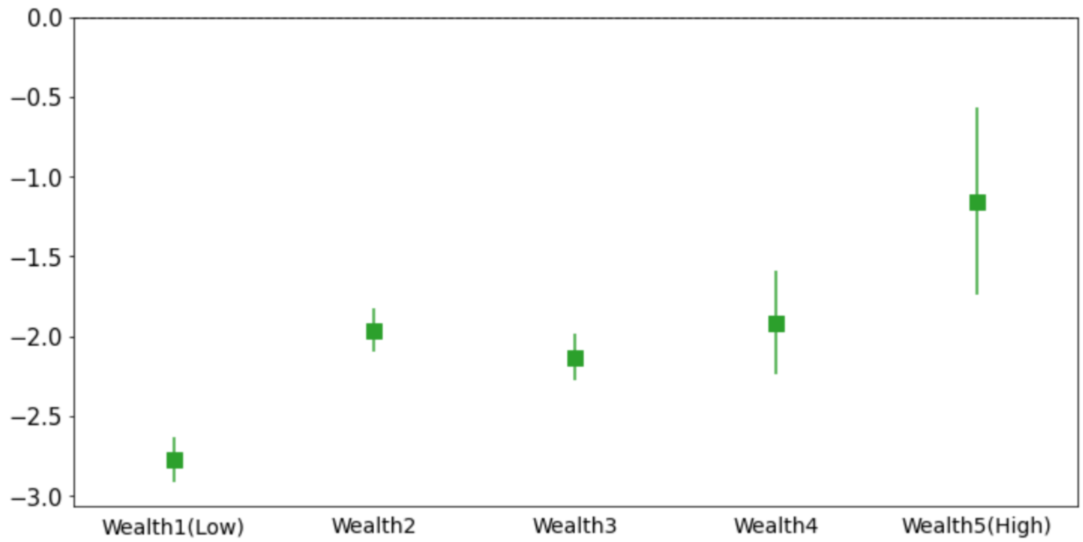
Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_{I,T}$) of retail investors and institutions over different time periods. $\hat{\gamma}_{I,T}$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}\left[i \in I, t \in T\right],$$

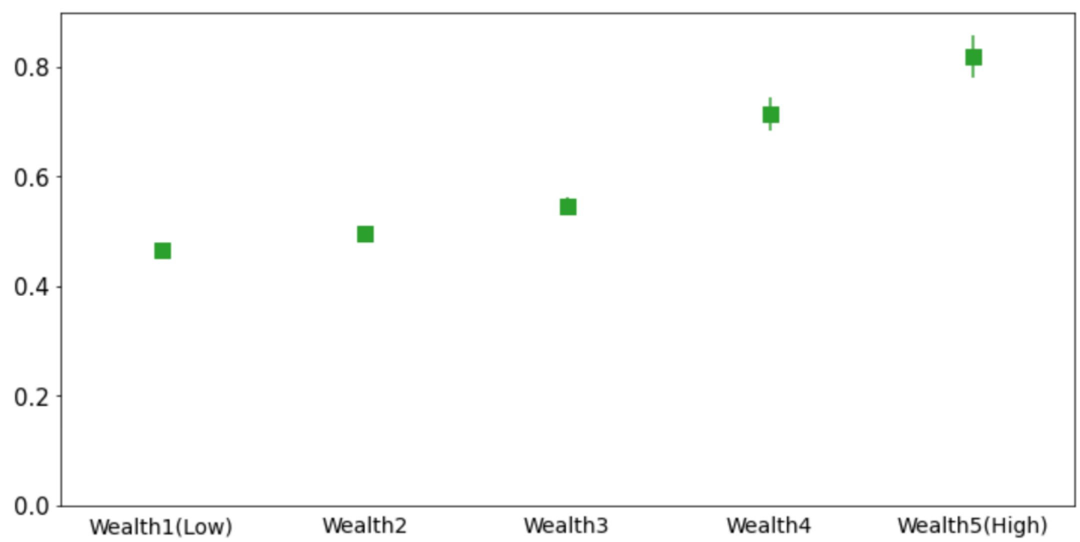
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $I \in \{\text{Retail, Institution}\}, T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. Panel (a) reports investors' preferences/beliefs about age, and panel (b) reports investors' preferences/beliefs about the construction industry. Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Figure A.4: *Preferences/Beliefs of Retail Investors with Different Wealth (Dividend/Size)*

(a) *Preferences/Beliefs about Dividend*



(b) *Preferences/Beliefs about Size*



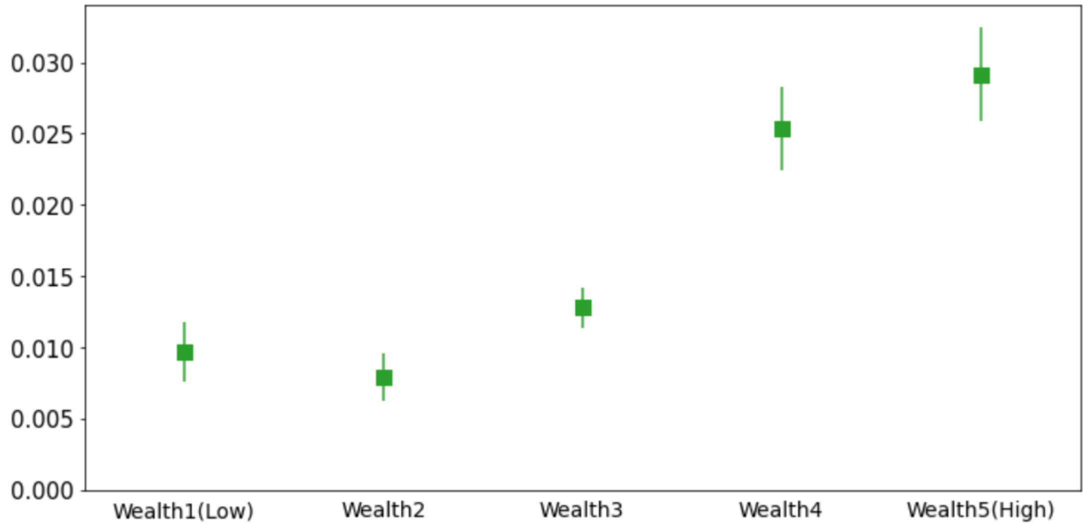
Notes: This figure reports the preferences/beliefs ($\hat{\gamma}_W$) of retail investors with different levels of wealth. We classify retail investors into five groups according to their average stock market wealth in the previous year. $\hat{\gamma}_W$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_{W} \gamma_W \mathbb{1}\left[i \in W\right] + \delta_t,$$

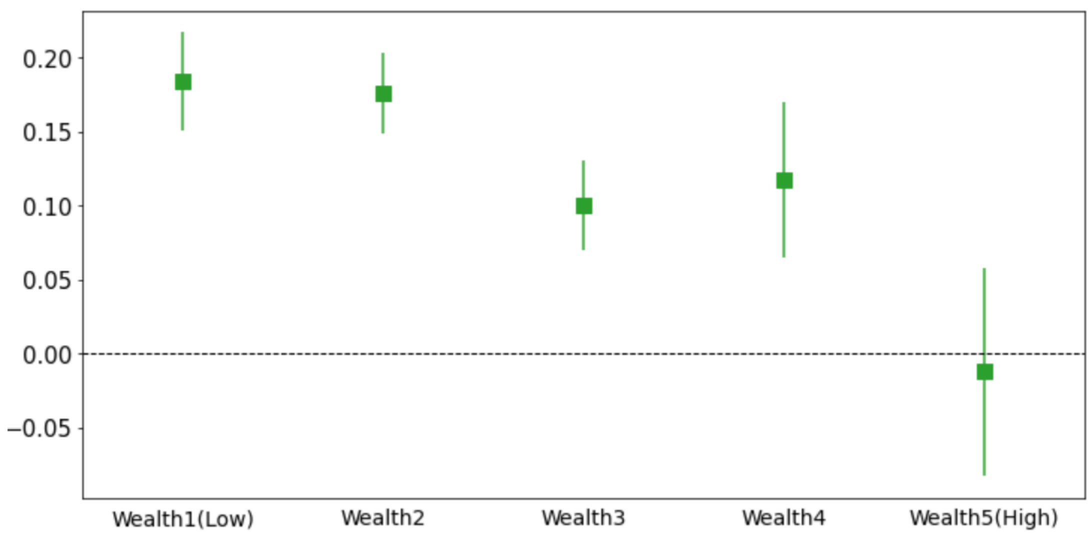
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $W \in$ {Wealth1 (Low), Wealth2, Wealth3, Wealth4, Wealth5 (High)}, and $\mathbb{1}\left[i \in W\right]$ is a dummy variable indicating whether investor i belongs to wealth group W, and δ_t are time fixed effects. Panel (a) reports investors' preferences/beliefs about dividend, and panel (b) reports investors' preferences/beliefs about size (measured by log book equity). Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Figure A.5: *Preferences/Beliefs of Retail Investors with Different Wealth (Age/Industry)*

(a) *Preferences/Beliefs about Age*



(b) *Preferences/Beliefs about Construction Industry*



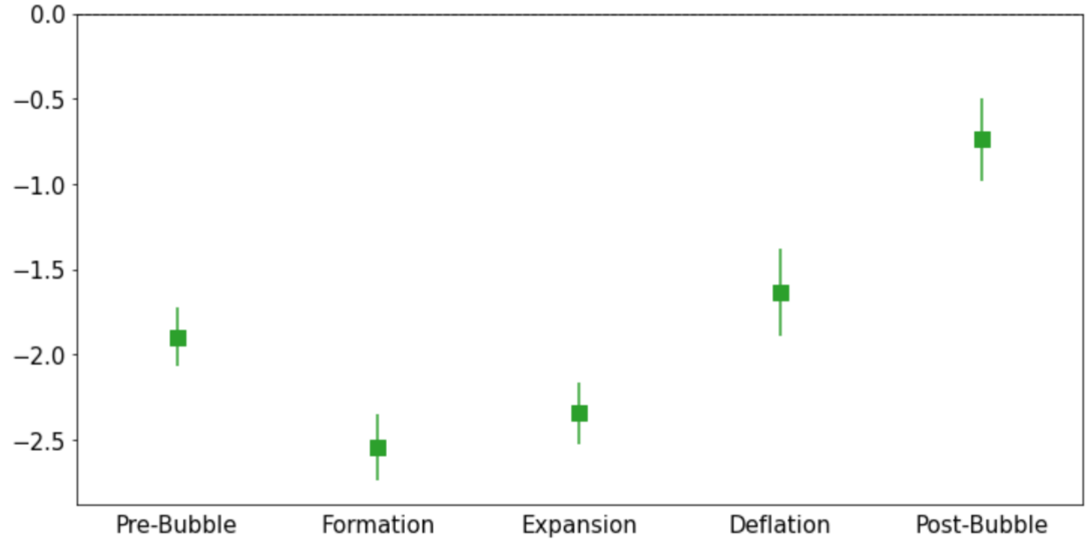
Notes: This figure reports the preferences/beliefs ($\hat{\gamma}_W$) of retail investors with different levels of wealth. We classify retail investors into five groups according to their average stock market wealth in the previous year. $\hat{\gamma}_W$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_W \gamma_W \mathbb{1}\left[i \in W\right] + \delta_t,$$

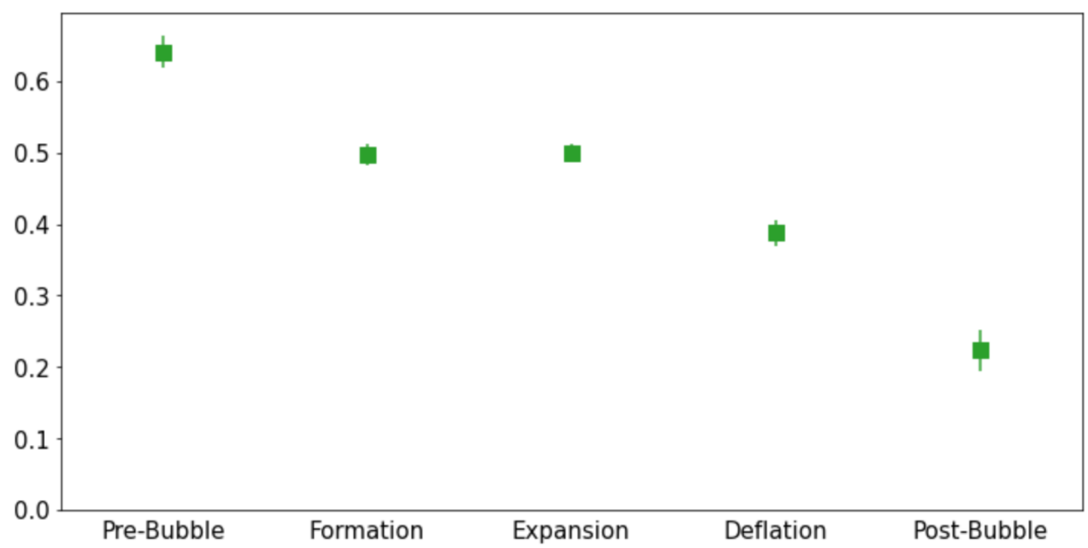
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $W \in$ {Wealth1 (Low), Wealth2, Wealth3, Wealth4, Wealth5 (High)}, and $\mathbb{1}\left[i \in W\right]$ is a dummy variable indicating whether investor i belongs to wealth group W, and δ_t are time fixed effects. Panel (a) reports investors' preferences/beliefs about age, and panel (b) reports investors' preferences/beliefs about the construction industry. Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Figure A.6: *Preferences/Beliefs of Retail Investors in Different Cohorts (Dividend/Size)*

(a) *Preferences/Beliefs about Dividend*



(b) *Preferences/Beliefs about Size*



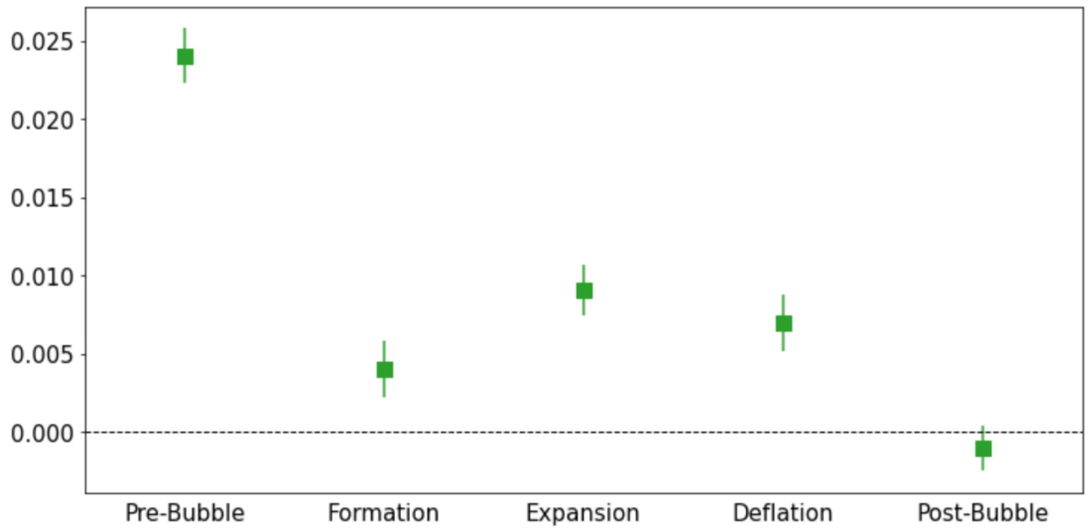
Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_C$) of retail investors entering the market at different time periods. We classify retail investors by their market entry cohort: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble phase. We focus on the time horizon from 01/2014 through 12/2016 and control for time fixed effects. $\hat{\gamma}_C$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_C \gamma_C \mathbb{1}\left[i \in C\right] + \delta_t,$$

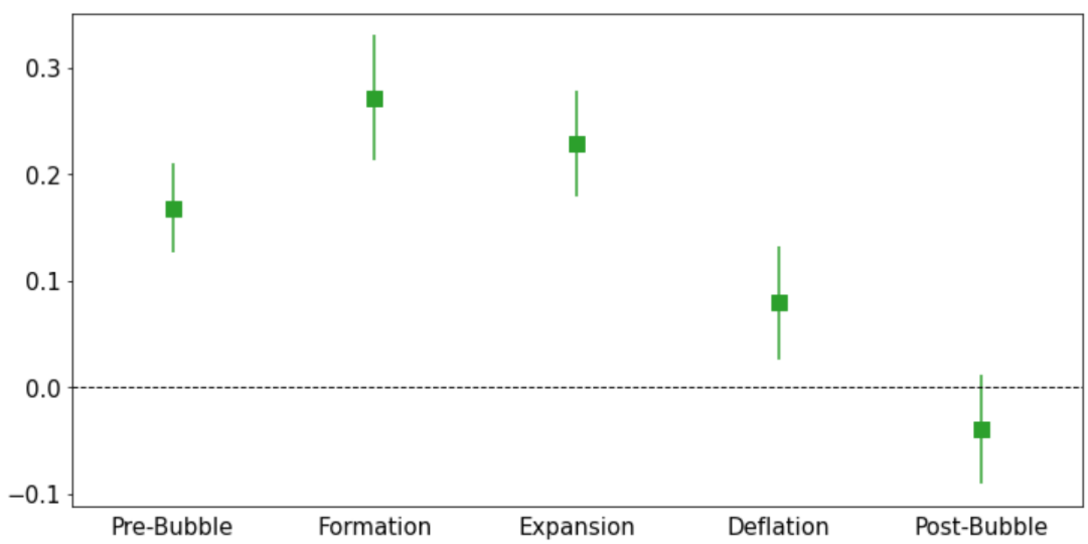
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $C \in$ {Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}, $\mathbb{1}\left[i \in C\right]$ is a dummy variable indicating whether investor i belongs to cohort C, and δ_t are time fixed effects. Panel (a) reports investors' preferences/beliefs about dividend, and panel (b) reports investors' preferences/beliefs about size (measured by log book equity). Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Figure A.7: *Preferences/Beliefs of Retail Investors in Different Cohorts (Age/Industry)*

(a) *Preferences/Beliefs about Age*



(b) *Preferences/Beliefs about Construction Industry*



Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_C$) of retail investors entering the market at different time periods. We classify retail investors by their market entry cohort: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble phase. We focus on the time horizon from 01/2014 through 12/2016 and control for time fixed effects. $\hat{\gamma}_C$ is estimated by the following regression:

$$\hat{\beta}_{k,i,t} = \sum_C \gamma_C \mathbb{1}\left[i \in C\right] + \delta_t,$$

where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $C \in$ {Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}, $\mathbb{1}\left[i \in C\right]$ is a dummy variable indicating whether investor i belongs to cohort C, and δ_t are time fixed effects. Panel (a) reports investors' preferences/beliefs about age, and panel (b) reports investors' preferences/beliefs about the construction industry. Standard errors are clustered at the investor level, and the error bars shown in the figure report the 95% confidence intervals.

Appendix B

Appendix to Chapter 2

B.1 A Model with Heterogeneous Extrapolative Beliefs

We use this model to motivate the idea that different propensities to chase returns can result from investors' heterogeneous tendencies to extrapolate.

There are 3 periods: $t = 0$, $t = 1$ and $t = 2$. Past returns are realized at $t = 0$. The investors choose their portfolios at $t = 1$ and consume at $t = 2$.

There are 2 risky assets, indexed by h and l, each with fixed supply of 1. Asset h has a high past return $R_{h,0}$, and asset l has a low past return $R_{l,0}$, i.e., $R_{h,0} > R_{l,0}$. There is also a risk-free asset in elastic supply with risk-free rate R_f.

We have heterogeneous investors that are indexed by i. Investors observe the risky assets' past return $R_{k,0}, k \in \{h, l\}$, and extrapolate on the difference between $R_{k,0}$ and the market benchmark return $R_{M,0}$, i.e., $R_{k,0} - R_{M,0}$. The extrapolative tendency is defined by a parameter ρ_i. We have

$$E_1^i \left[R_{k,2} - R_{M,2}\right] = \rho_i \left[R_{k,0} - R_{M,0}\right], \quad k \in \{h, l\}, \tag{B.1}$$

where E_1^i means investor i's expectation at time $t = 1$. The market benchmark return R_M is the default benchmark that investors compare their investment with. Here we simply assume that R_M is the arithmetic average of the two assets, i.e., $R_{M,t} = (R_{h,t} + R_{l,t})/2$. The results hold more generally for other forms of R_M with fixed weights on individual assets.

Equation (B.1) means that at time $t = 1$, investor i expects the future gap between the risky asset return $R_{k,2}$ and the market benchmark return $R_{M,2}$ to correlate with the past gap $R_{k,0} - R_{M,0}$, with extrapolation parameter ρ_i. If $\rho_i > 0$, the investor believes in momentum, and if $\rho_i < 0$, the investor believes in mean-reversion [1].

Returns of risky assets are lognormally distributed, with the covariance matrix of log returns as Σ. For the following analysis, we will simply assume the two risky assets have zero covariance so that $\Sigma = \begin{bmatrix} \sigma_h^2 & 0 \\ 0 & \sigma_l^2 \end{bmatrix}$. The same results hold for any covariance structure.

Investors have CRRA utility. At time $t = 1$, investors choose the wealth share α_h^i, α_l^i to invest in the two risky assets, trying to maximize their expected utility at $t = 2$. We denote W_t^i as individual i's wealth at time t. Investors have the objective function:

$$\max_{\alpha_h^i, \alpha_l^i} E_1^i \left[\frac{(W_2^i)^{1-\gamma}}{1-\gamma} \right] \tag{B.2}$$

s.t.

$$W_2^i = R_f(1 - \alpha_h^i - \alpha_l^i)W_1^i + \alpha_h^i W_1^i R_{h,2} + \alpha_l^i W_1^i R_{l,2}.$$

We can solve the optimal portfolio allocation as

$$\begin{pmatrix} \alpha_h^i \\ \alpha_l^i \end{pmatrix} \approx \begin{pmatrix} \frac{1}{\gamma\sigma_h^2} \left[E_1^i \left[R_{M,2} \right] + \rho_i (R_{h,0} - R_{M,0}) - R_f \right] \\ \frac{1}{\gamma\sigma_l^2} \left[E_1^i \left[R_{M,2} \right] + \rho_i (R_{l,0} - R_{M,0}) - R_f \right] \end{pmatrix}, \tag{B.3}$$

where the approximation becomes more accurate as the time interval shrinks, and is exact in continuous time when asset prices follow diffusions. We use the discrete time set up just to keep the model simple.

Given that $R_{h,0} - R_{M,0} > 0$, we have $\partial\alpha_h^i / \partial\rho_i > 0$. That is, investors' wealth share in the high past return stock α_h^i is increasing in ρ_i. This is intuitive: if an investor is more extrapolative, she is more likely to buy high past return stocks. This simple model motivates

1. Equation (B.1) assumes that investors extrapolate returns at $t = 0$ rather than at $t = 1$. This is just for simplicity so that investors extrapolate past realized returns. We can get rid of this time gap if we adopt a continuous time framework.

that heterogeneous extrapolative beliefs lead to different portfolio choices.

Finally, we close the model by setting the aggregate asset demand equals to the supply. We can solve the prices for the two stocks P_h and P_l as:

$$\begin{pmatrix} P_h \\ P_l \end{pmatrix} = \begin{pmatrix} \int_0^1 W_1^i \alpha_h^i di \\ \int_0^1 W_1^i \alpha_l^i di \end{pmatrix} = \begin{pmatrix} \frac{1}{\gamma\sigma_h^2}\left[\int_0^1 E_1^i \left[R_{M,2} - R_f\right] W_1^i di + (R_{h,0}^e - R_{M,0}^e) \int_0^1 W_1^i \rho_i di\right] \\ \frac{1}{\gamma\sigma_l^2}\left[\int_0^1 E_1^i \left[R_{M,2} - R_f\right] W_1^i di + (R_{l,0}^e - R_{M,0}^e) \int_0^1 W_1^i \rho_i di\right] \end{pmatrix}. \tag{B.4}$$

We can see that the market price is influenced by the wealth weighted extrapolation parameter $\int_0^1 W_1^i \rho_i di$. Let us define $\rho = \int_0^1 W_1^i \rho_i di$ as the market extrapolation degree.

The higher the aggregate extrapolation parameter $\rho = \int_0^1 W_1^i \rho_i di$, the higher is the high-past return stock price P_h, and the lower is the low-past return stock price P_l, given that $R_{l,0}^e < R_{M,0}^e < R_{h,0}^e$. This is also intuitive: a higher aggregate extrapolation parameter means higher aggregate demand for the high past return stock, and therefore pushes the price P_h higher. It suggests that if aggregate return chasing tendency is high, stocks with high past return are more likely to be over-valued.

B.2 Alternative CAPM Beta Adjusted RCP

We can define the CAPM beta adjusted RCP measure as:

$$\text{RCP}_{i,t}^{\text{adj}} = \sum_{s \in \{S_{i,t}\}} \left[\gamma_s \left(\sum_{k=1}^{K} \text{R-adj}_{t(s)-k}^s\right)\right], \tag{B.5}$$

where all the notations are the same as in Equation (A.4), except for the CAPM beta adjusted returns R-adj_t^s. R-adj_t^s is defined as the sum of the intercept and residual from the regression:

$$R_t^s - R_t^f = \alpha_0^s + \beta^s (R_t^M - R_t^f) + \epsilon_t^s, \tag{B.6}$$

where R_t^f is the risk free rate at time t, R_t^M is the market index return at time t. The estimated coefficient $\hat{\beta}^s$ is stock s's estimated CAPM beta. We estimate this regression using the moving window of past 2 years' stock daily returns (with at least 100 days of non-missing returns). We have

$$\text{R-adj}_t^s = \alpha_0^s + \epsilon_t^s = (R_t^s - R_t^f) - \hat{\beta}^s(R_t^M - R_t^f). \tag{B.7}$$

We now replicate results in Section 2.3 using this alternative RCP measure. The patterns are robust. Replication results are reported in Table B.1 and Table B.2.

B.3 Aggregate Retail RCP

B.3.1 Decompose Aggregate Retail RCP Volatility

Given that aggregate retail RCP co-moves with the market ups and downs, it is natural to ask what are the sources for this volatility. To answer this question, we carry out the following decomposition:

$$\begin{aligned}\Delta\rho_t^M = & \underbrace{\sum_{i\in Both}\omega_{i,t-1}\Delta\rho_{i,t}}_{\Delta\text{within}} + \underbrace{\sum_{i\in Both}\rho_{i,t-1}\Delta\omega_{i,t}}_{\Delta\text{between: wealth reallocation}} + \underbrace{\sum_{i\in Both}\Delta\rho_{i,t}\Delta\omega_{i,t}}_{\Delta\text{cross}} \\ & + \sum_{i\in Entry}\rho_{i,t}\omega_{i,t} - \sum_{i\in Exit}\rho_{i,t-1}\omega_{i,t-1}.\end{aligned} \tag{B.8}$$

This decomposition is an accounting identity according to the definition of aggregate retail RCP ρ_t^M in Equation (2.7). We decompose the changes in aggregate retail RCP into 5 components. The first one is the "within" term: holding the last period wealth weights constant, what is the impact from changing RCP within each investor. The second one is the "between" term: holding the last period individual RCP constant, what is the impact from changing wealth weights among investors. This is also referred as the "wealth reallocation" term. The third one is the "cross" term that is the cross product of the first 2 terms: whether changes in wealth weights and changes in individual RCP move in the same direction. We also have the "entry" and "exit" terms. They capture investors who newly enter the market and newly exit the market respectively.

This decomposition is used to answer the following question: when aggregate retail RCP increases, is it because every investor in the economy changes to a higher individual RCP ("within")? Or is it because high RCP investors acquire a higher wealth share ("between")? Or is it because high RCP investors enter the market ("entry")? We compute the 5 components, and calculate each component's contribution to the aggregate retail RCP changes through a variance decomposition:

$$\text{Contribution}_i = \frac{Cov(\Delta\rho_t^M, \text{Component}_i)}{Var(\Delta\rho_t^M)}. \tag{B.9}$$

Given that we have $\Delta\rho_t^M = \sum_i \text{Component}_i$, we guarantee that the sum of contributions equals to 1. We have the variance decomposition results that the "within" term is the most important: it contributes 76% of aggregate retail RCP volatility. The "entry" term also has a sizeable share that it accounts for 17% of aggregate retail RCP volatility. For the remaining terms, "cross" accounts for 10%, "exit" accounts for 2%, and "between" accounts for -5%.

In Figure B.1, we also plot the time-series of the 5 components. The figure shows the importance of each component visually: the "within" term track the aggregate RCP changes the most closely.

The reason for the "between" term being so small is that the individual wealth weights do not change much, i.e., $\Delta\omega_{i,t}$ is small. One simple test we can do is to regress the investors' wealth weights on investor fixed effects. The R^2 from this fixed effects regression is high at 0.71, which confirms that individual wealth weights are quite stable over time.

It is important to make sure that our decomposition results are robust. We carry out various robustness checks. Table B.3 and Table B.4 reports the decomposition results using different RCP measures. The decomposition results are quite stable across different RCP measures. One might worry that the wealth weights being stable because we look at monthly changes. In Table B.5 and Table B.6, we show that even if we look at 6-month aggregate retail RCP changes, the between effects still have a negligible role. We do see bigger roles for entry and exit terms. This is expected since we count some of the within component as entry and exit when we look at the 6-month horizon compared to the 1-month horizon case.

Our results that aggregate retail RCP volatility comes mainly from the "within" and "entry" terms shed light on the recent theoretical debate on the sources of aggregate market volatility. Recent papers like Atmaz and Basak (2018), Pohl, Schmedders, and Wilms

(2020), and Martin and Papadimitriou (2022b) rely heavily on the "between" term (wealth reallocation) to generate the observed volatility of market prices and returns. They need the individual wealth weights to be volatile enough so that the wealth weighted outcome will be volatile too. We show that this is not the case in the data. Instead, previous literature that uses a representative agent, or using CARA utility functions so that there is no wealth effects, implicitly assumes that there is only the "within" term (Barberis et al. 2015; Nagel and Xu 2022). We show that this is actually not a bad assumption since the "within" term accounts for 76% of market volatility. Papers that emphasize disagreements and short-sale constraints regard the "entry" and "exit" terms as the most important driver of market volatility (Scheinkman and Xiong 2003; Baker and Stein 2004). As heterogeneous agent models are of increasing interest, our results offer important moments for these models to match. To the best of our knowledge, this is the first empirical study that quantifies the relative magnitude of each component. Admittedly, our aggregate retail RCP volatility is not the same object as the volatility of market returns. We leave it to future work to connect these two objects more carefully.

B.3.2 Aggregate Retail RCP for Different Wealth Groups

In this appendix section, we look at RCP among different wealth groups to give us a better idea how RCP differs across investors. We separate the sample into 5 wealth groups according to investors' wealth in the stock market: less than CNY 100K, between 100K and 500K, between 500K and 5M, between 5M and 10M, and more than 10M. The exchange rate is roughly 6.5 CNY = 1 USD. Table B.7 reports summary statistics for the 5 groups during Jan 2011 to Dec 2019. We can see that most accounts are in the low wealth group. Trading Volume is the average daily traded CNY value of all the stocks. Size is the investors' average CNY value of all holdings of stocks in SSE. We can see that each group's trading volume share is similar to their market size share. The lowest wealth group has a relatively larger share of the trading volume, while the highest wealth group has a relatively lower trading volume share compared to its market size.

Figure B.2 plots the wealth weights of the 5 groups over time. Note that group 1 is the group with the lowest wealth, and group 5 has the highest wealth. We can see that groups 2, 3, and 4 have relatively stable wealth shares, while group 1's wealth share decreases and group 5's wealth share increases. It is quite surprising that the richest group gets a higher wealth share purely at the expense of the lowest wealth group. There can be multiple reasons for this pattern. For example, maybe rich people gain a lot of money from the stock market and hence their stock value grows. It can also be that rich people start to invest a higher percentage of their total wealth into the stock market. Or maybe it is just rich households' outside wealth grow a lot and even if they keep the risky asset share constant, their stock market wealth needs to increase correspondingly. Since we don't have investors' wealth data outside of the stock market, it is beyond the scope of this paper to draw a conclusion. However, we do see from previous analyses that poorer investors are more likely to earn lower average returns.

In Figure B.3, we plot the calculated RCP in each group (we use the same method to aggregate individual RCP into the group aggregate RCP as before). We can see that before the end of 2014, the 5 groups have quite similar RCP. However, since the start of the stock market bubble, group 1 increases its RCP most dramatically. In Figure B.4, we plot the market price index level on the top of Figure B.3. It is interesting to observe that the responsiveness of RCP towards market price increase ranks exactly with the wealth level: group 1 increases its RCP the most, while group 5 increases its RCP the least. We see that low wealth groups have both a higher slope of RCP increase and a higher level of RCP compared to high wealth groups. This is consistent with our prior results that low wealth investors are more likely to have higher RCP and earn lower average returns.

B.4 Why Does Stock RCO Change?

One may wonder what leads to changes in stock RCO? Is it because new investors who have different RCP enter to buy the stock? Or old investors sell the stock? Or is it because investors who already hold the stock change their individual RCP? To answer the question, we decompose changes in stock RCO into the following five components:

$$\Delta\text{RCO}_{s,t} = \underbrace{\sum_{i\in Both} \omega_{i,s,t-1}\Delta\rho_{i,s,t}}_{\Delta\text{within}} + \underbrace{\sum_{i\in Both} \rho_{i,s,t-1}\Delta\omega_{i,s,t}}_{\Delta\text{between: wealth reallocation}} + \underbrace{\sum_{i\in Both} \Delta\rho_{i,s,t}\Delta\omega_{i,s,t}}_{\Delta\text{cross}} + \sum_{i\in Entry} \rho_{i,s,t}\omega_{i,s,t} - \sum_{i\in Exit} \rho_{i,s,t-1}\omega_{i,s,t-1}, \tag{B.10}$$

where $\Delta\text{RCO}_{s,t} = \text{RCO}_{s,t} - \text{RCO}_{s,t-1}$. $\rho_{i,s,t}$ denotes RCP(-s) for investor i who holds stock s at time t. Weight $w_{i,t}$ is retail investor i's share of that stock's total retail market value. $i \in Both$ indicates investors who hold stock s both in the previous period and the current period.

This decomposition is an accounting identity according to the definition of RCO in Equation (2.8). We decompose the changes in RCO into five components. The "within" term means that holding the last period stock holders' wealth weights in this stock constant, what is the impact from changing RCP within each investor. The "between" term means that holding the last period stock holders' individual RCP constant, what is the impact from changing wealth weights across different stock holders. The "cross" term is the cross product of the first two terms. Finally, the "entry" and "exit" terms capture the impact from investors who newly enter to buy this stock and newly exit to completely sell this stock, respectively. If there are partial sales or purchases made by existing shareholders, the impact will be classified as the combination of the "within", "between", and "cross" terms.

We compute the 5 components, and calculate each component's contribution to the aggregate stock RCO changes through a variance decomposition:

$$\text{Contribution}_i = \frac{Cov(\Delta\text{RCO}_{s,t}, \text{Component}_i)}{Var(\Delta\text{RCO}_{s,t})}. \tag{B.11}$$

Given that we have $\Delta\text{RCO}_{s,t} = \sum_i \text{Component}_i$, we guarantee that the sum of contributions equals to 1.

We carry out this variance decomposition for each stock's RCO. Since we have 1,481 stocks in our sample, we have 1,481 decomposition results. On average, "within", "between", "cross", "entry", and "exit" account for 28%, 1%, -0.2%, 47%, and 24% of RCO changes respectively.[2] The "entry" term plays a particularly important role. That is, when a stock's RCO increases, it is mainly because high RCP investors enter and start to hold this stock. Some variations are also explained by the "within" term (investors who hold this stock increase their individual RCP) and the "exit" term (low RCP investors sell this stock and exit).

Overall, when we talk about changes in stock RCO, we can roughly think of these changes as driven by different investors entering or exiting this stock ("entry" and "exit" together explain 71% of the total variation).

B.5 Chase Industry Returns or Idiosyncratic Returns?

We decompose investor RCP into stock idiosyncratic RCP (SRCP) and industry RCP (IRCP) as in Equation (2.13).

We then replicate Table 2.3 using both SRCP and IRCP as dependent variables. In Table B.10, we check the relationship between SRCP/IRCP and investor characteristics. Column (1) uses SRCP as the dependent variable, while Column (2) uses IRCP as the dependent variable. Column (3) uses the difference between SRCP and IRCP as the dependent variable. We can see that investor characteristics that are associated with the investor being more unsophisticated are correlated with higher SRCP and lower IRCP, except for Log(AvgInvest). That is, investors who have a stronger disposition effect, trade more frequently, hold more volatile stocks, are younger, are less experienced, or are male chase stock idiosyncratic returns to a greater extent than industry returns.

In terms of investor return prediction, we replicate Table 2.4 using both SRCP and IRCP as predictors. Results are reported in Table B.11. We can see that a one standard deviation increase in SRCP leads to 0.42% lower next month return for the investor; while a one standard deviation increase in IRCP leads to 0.26% lower next month return. Both types of

2. While we talk about the average of 1,481 decomposition results here, the full summary statistics for the 1,481 decomposition results can be seen in Internet Appendix Table B.8. The robustness of the decomposition results using different RCP measures is demonstrated in Internet Appendix Table B.9.

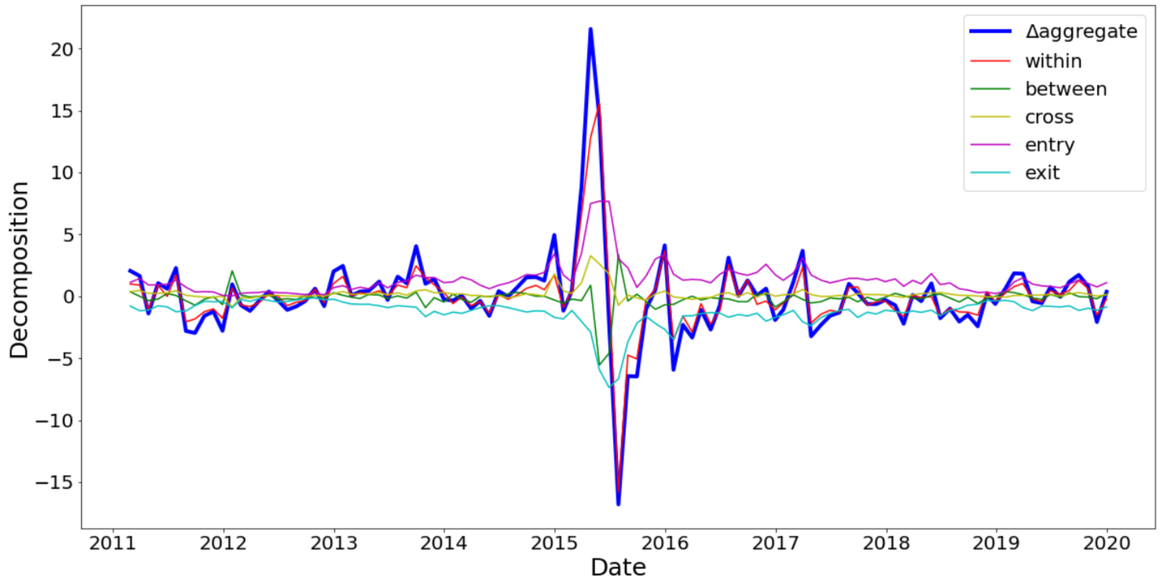
return chasing reduce future returns to investors. Chasing idiosyncratic returns exerts the most negative effects.

At the stock level, we construct stock idiosyncratic return chasing ownership (SRCO) and industry return chasing ownership (IRCO) following the same procedure we used for RCO.

Table B.12 uses industry return chasing ownership (IRCO) together with other ownership variables to predict future industry returns. We have 28 industries each period. We can see that industries with high IRCO under-perform: an industry rises from the lowest IRCO to the highest IRCO can lead to 1.4% next-month return loss.

Table B.13 uses stock idiosyncratic return chasing ownership (SRCO) together with other ownership variables to predict future stock returns, after controlling for industry fixed effects. We can see that within an industry, stocks with high SRCO under-perform: a stock rises from the lowest SRCO to the highest SRCO can lead to 0.5% next-month return loss.

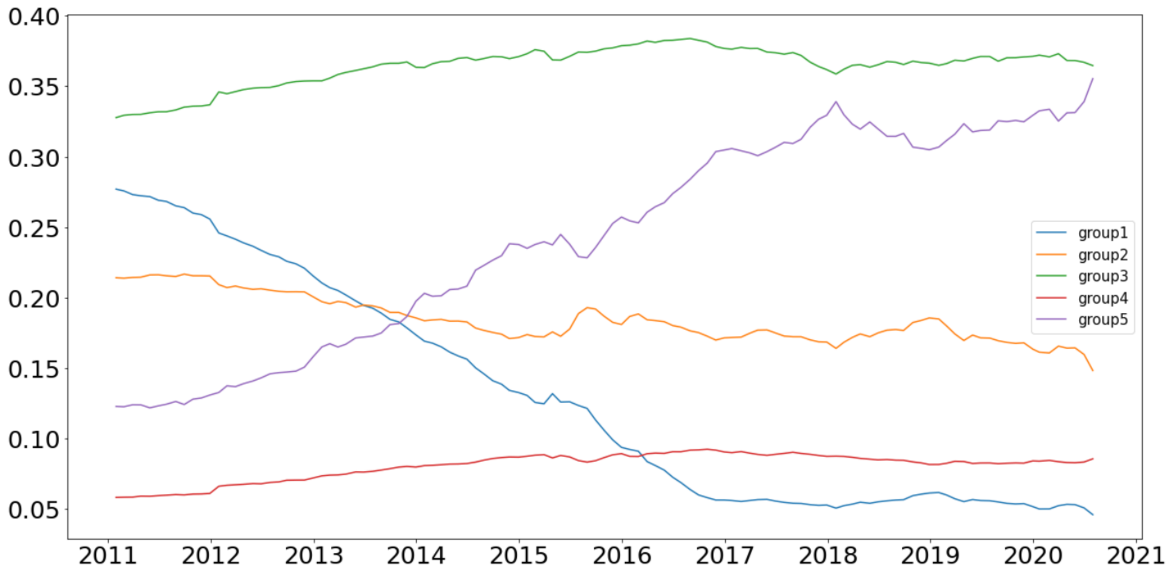
Figure B.1: *The Time-Series of the Aggregate Retail RCP Change Decomposition*



Notes: This figure plots the time-series of the aggregate retail RCP change and its 5 components. The decomposition is as follows:

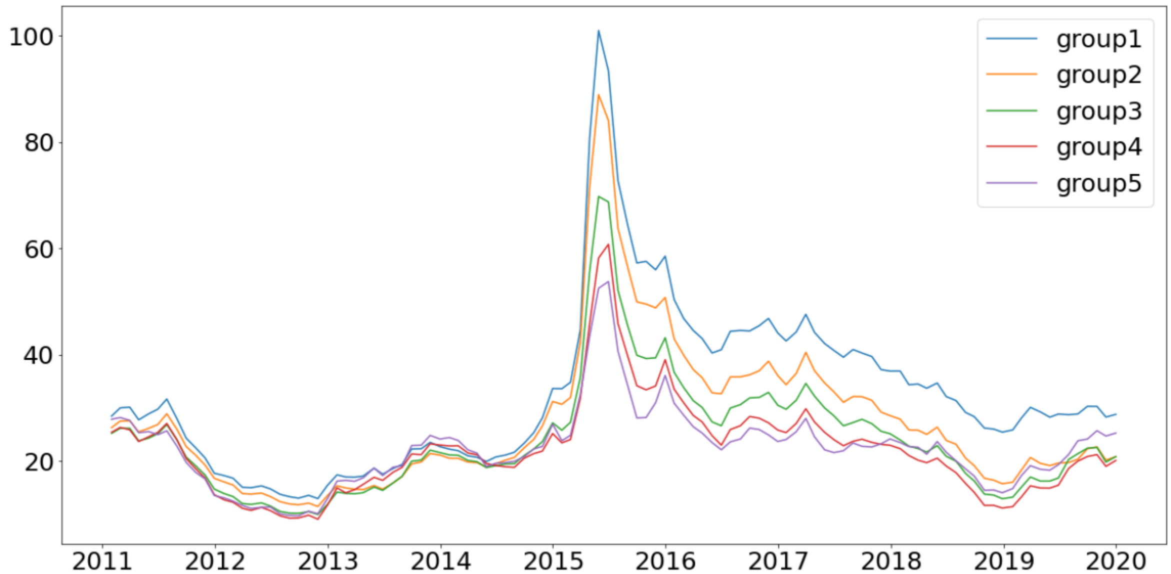
$$\Delta\rho_t^M = \underbrace{\sum_{i\in Both} \omega_{i,t-1}\Delta\rho_{i,t}}_{\Delta\text{within}} + \underbrace{\sum_{i\in Both} \rho_{i,t-1}\Delta\omega_{i,t}}_{\Delta\text{between: wealth change}} + \underbrace{\sum_{i\in Both} \Delta\rho_{i,t}\Delta\omega_{i,t}}_{\Delta\text{cross}} + \sum_{i\in Entry} \rho_{i,t}\omega_{i,t} - \sum_{i\in Exit} \rho_{i,t-1}\omega_{i,t-1}. \tag{B.8}$$

Figure B.2: *Wealth Weights of Different Wealth Groups*



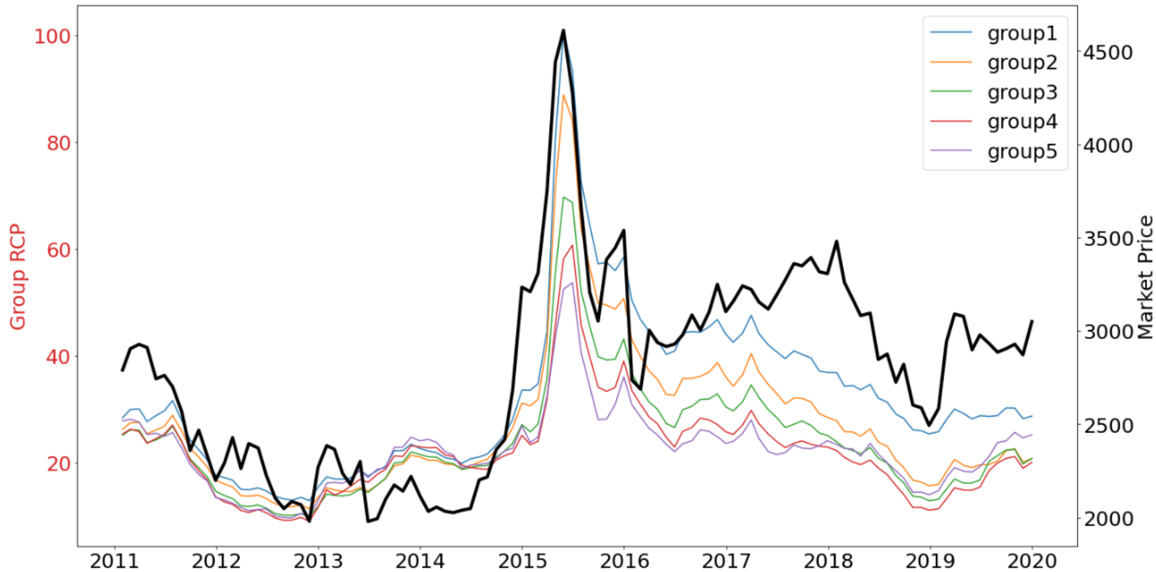
Notes: This figure plots the wealth weights of the 5 groups along the time. The 5 wealth groups are defined according to investors' wealth in the stock market: less than CNY 100K, between 100K and 500K, between 500K and 5M, between 5M and 10M, and more than 10M. Roughly 6.5 CNY equal to 1 dollar. Group 1 is the group with the lowest wealth, and group 5 is the one with the highest wealth.

Figure B.3: *RCP of Different Wealth Groups*



Notes: This figure plots the calculated RCP in each wealth group. The 5 wealth groups are defined according to investors' wealth in the stock market: less than CNY 100K, between 100K and 500K, between 500K and 5M, between 5M and 10M, and more than 10M. The exchange rate is roughly 6.5 CNY = 1 USD. Group 1 is the group with the lowest wealth, and group 5 is the one with the highest wealth.

Figure B.4: *RCP of Different Wealth Groups and Market Price*



Notes: This figure plots the market price index level (thick black line) on the top of Figure B.3. The 5 wealth groups are defined according to investors' wealth in the stock market: less than CNY 100K, between 100K and 500K, between 500K and 5M, between 5M and 10M, and more than 10M. The exchange rate is roughly 6.5 CNY = 1 USD. Group 1 is the group with the lowest wealth, and group 5 is the one with the highest wealth.

Table B.1: *Relationship RCP and Other Investor Characteristics*

Replicate Table 2.3 Using CAPM Beta Adjusted RCP

Outcome Var:	RCP	RCP	$\text{RCP}^{\text{New Purchased}}_{i,t(p)}$
	(1)	(2)	(3)
RCP			0.180***
			(0.007)
Disposition Effect	0.036***	0.032***	-0.008
	(0.014)	(0.004)	(0.006)
Trade Frequency	0.001***	9.81e-04***	1.26e-04*
	(2.75e-04)	(2.73e-04)	(6.5e-05)
VolatilityStockHeld	0.469***	0.508***	0.057***
	(0.018)	(0.018)	(0.009)
Log(AvgInvest)	0.716***	0.724***	-0.009
	(0.100)	(0.046)	(0.105)
Age	-0.329***		-0.179***
	(0.014)		(0.014)
Age^2_t	0.003***		0.002***
	(1.22e-04)		(1.36e-04)
Experience	-1.692***		-0.825***
	(0.116)		(0.073)
Experience^2_t	0.108***		0.046***
	(0.008)		(0.005)
Female	1.008***		1.287***
	(0.088)		(0.110)
Time Fixed Effects	Y	Y	Y
Investor Fixed Effects	N	Y	N
R-Squared	0.057	0.107	0.031
Observations.	650,653,925	652,095,551	104,601,293

Notes: Columns (1) and (2) report contemporary regression results: unit of observation is individual-by-month. Column (3) report predictive regression results: unit of observation is investor-by-purchase timestamp. Standard errors (in parentheses) are double clustered at investor and date level. All regressions controlled Number Stocks and HHI up to 3rd polynomials. The ***, **, * denote significance at the 1%, 5%, 10% level correspondingly. The R-Squared reported does not include the impact from fixed effects.

Table B.2: *The Impacts of Investor Characteristics on Future Returns*

Replicate Table 2.4 Using CAPM Beta Adjusted RCP

Outcome Var:	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+4}	Ret_{t+5}	Ret_{t+6}
	(1)	(2)	(3)	(4)	(5)	(6)
Ret_t	-0.002	-0.003	-0.004	0.001	0.001	0.006**
	(0.009)	(0.009)	(0.006)	(0.005)	(0.005)	(0.003)
RCP_t^{adj}	-0.296***	-0.253***	-0.229***	-0.208***	-0.197***	-0.181**
	(0.103)	(0.089)	(0.077)	(0.074)	(0.072)	(0.071)
Disposition $Effect_t$	-0.099**	-0.104**	-0.111***	-0.109***	-0.107**	-0.103**
	(0.045)	(0.043)	(0.042)	(0.042)	(0.042)	(0.042)
$VolatilityStockHeld_t$	-0.133**	-0.126**	-0.125**	-0.117**	-0.109**	-0.096**
	(0.060)	(0.054)	(0.049)	(0.046)	(0.043)	(0.040)
$Log(AvgInvest)_t$	0.079	0.096*	0.098*	0.099*	0.101*	0.097*
	(0.068)	(0.057)	(0.055)	(0.054)	(0.053)	(0.053)
HHI_t	-0.135	-0.034	-0.033	-0.027	-0.027	-0.029
	(0.086)	(0.037)	(0.033)	(0.032)	(0.030)	(0.030)
Trade $Frequency_t$	-0.058***	-0.045**	-0.030	-0.025	-0.036*	-0.035**
	(0.022)	(0.022)	(0.023)	(0.022)	(0.020)	(0.017)
Age_t	0.042*	0.042*	0.041*	0.040	0.039	0.038
	(0.024)	(0.025)	(0.024)	(0.025)	(0.025)	(0.024)
$Experience_t$	0.217***	0.179***	0.169***	0.163***	0.155***	0.146***
	(0.044)	(0.033)	(0.029)	(0.028)	(0.026)	(0.024)
$Female_t$	0.079***	0.068***	0.065***	0.063***	0.060***	0.058***
	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Time Fixed Effects	Y	Y	Y	Y	Y	Y
No. Observations	545,405,372	518,296,772	501,873,295	488,384,652	476,456,560	465,631,673
R-squared	9.97e-04	7.46e-04	6.92e-04	6.13e-04	5.72e-04	5.56e-04

Notes: Unit of observation is investor-by-month. Sample contains 18,147,831 investor accounts from the Shanghai Stock Exchange from Jan 2011 to Dec 2019. Standard errors (in parentheses) are double clustered at the individual and time level. All the variables except returns (Ret_{t+k}) are standardized to number of standard deviations from the mean. The ***, **, * denote significance at the 1%, 5%, 10% level correspondingly. The R-Squared reported does not include the impact from fixed effects.

Table B.3: *Decomposition Results For Different RCP Measures*

RCP measure	Within(%)	Between(%)	Cross(%)	Entry(%)	Exit(%)
RCP calculated using simple arithmetic average					
$RCP^{K=1}_{Arithmetic}$	77.96	1.19	4.65	9.56	6.64
$RCP^{K=2}_{Arithmetic}$	74.72	2.37	5.86	11.50	5.55
$RCP^{K=3}_{Arithmetic}$	74.36	2.13	5.83	13.32	4.37
$RCP^{K=6}_{Arithmetic}$	66.54	1.70	12.13	17.19	2.45
$RCP^{K=12}_{Arithmetic}$	75.94	-5.24	10.16	16.64	2.51
RCP calculated using weighted average					
$RCP^{K=1}_{Weighted}$	75.25	2.84	5.76	8.93	7.22
$RCP^{K=2}_{Weighted}$	71.31	4.00	7.74	10.63	6.32
$RCP^{K=3}_{Weighted}$	70.78	3.50	8.28	12.25	5.19
$RCP^{K=6}_{Weighted}$	62.25	4.41	14.58	15.34	3.41
$RCP^{K=12}_{Weighted}$	72.75	-3.51	12.82	14.83	3.11

Notes: This table reports the decomposition results for different RCP measures. The first 5 rows are results for RCP calculated using simple arithmetic average. The bottom 5 rows are results for RCP calculated using weighted average with the held stock value as weights.

Table B.4: *Summary Statistics of the Decomposition Results*

	Within(%)	Between(%)	Cross(%)	Entry(%)	Exit(%)
Obs.	10	10	10	10	10
Mean	72.19	1.34	8.78	13.02	4.68
Std.Dev.	4.73	3.20	3.46	2.92	1.76
Min	62.25	-5.24	4.65	8.93	2.45
25%	70.91	1.32	5.84	10.85	3.18
50%	73.56	2.25	8.01	12.78	4.78
75%	75.12	3.33	11.64	15.22	6.13
Max	77.96	4.41	14.58	17.19	7.22

Notes: This table reports the summary statistics of the decomposition results from the 10 different RCP measures.

Table B.5: *6-Month RCP Changes: Decomposition Results For Different RCP Measures*

RCP measure	within(%)	between(%)	cross(%)	entry(%)	exit(%)
RCP calculated using simple arithmetic average					
$RCP^{K=1}_{Arithmetic}$	62.03	2.88	5.14	18.22	11.74
$RCP^{K=2}_{Arithmetic}$	63.06	3.10	4.20	17.72	11.94
$RCP^{K=3}_{Arithmetic}$	64.33	1.96	3.61	18.79	11.32
$RCP^{K=6}_{Arithmetic}$	62.66	2.71	4.69	19.58	10.35
$RCP^{K=12}_{Arithmetic}$	66.32	1.26	1.11	22.68	8.63
RCP calculated using weighted average					
$RCP^{K=1}_{Weighted}$	58.31	4.32	6.64	17.96	12.77
$RCP^{K=2}_{Weighted}$	60.20	4.60	5.20	17.36	12.64
$RCP^{K=3}_{Weighted}$	61.78	3.25	4.75	18.20	12.03
$RCP^{K=6}_{Weighted}$	60.57	3.72	5.84	19.07	10.79
$RCP^{K=12}_{Weighted}$	65.05	2.48	2.08	21.67	8.73

Notes: This table reports the decomposition results for different RCP measures, use the 6-Month RCP changes as the decomposition target. Formally, the decomposition we carry out is:

$$\Delta_6\rho_t^M = \underbrace{\sum_{i\in Both}\omega_{i,t-6}\Delta_6\rho_{i,t}}_{\Delta_6\text{within}} + \underbrace{\sum_{i\in Both}\rho_{i,t-6}\Delta_6\omega_{i,t}}_{\Delta_6\text{between: wealth change}} + \underbrace{\sum_{i\in Both}\Delta_6\rho_{i,t}\Delta_6\omega_{i,t}}_{\Delta_6\text{cross}}$$
$$+ \sum_{i\in Entry}\rho_{i,t}\omega_{i,t} - \sum_{i\in Exit}\rho_{i,t-6}\omega_{i,t-6},$$

where $\Delta_6 x_t = x_t - x_{t-6}$.

The first 5 rows are results for RCP calculated using simple arithmetic average. The bottom 5 rows are results for RCP calculated using weighted average with the held stock value as weights. For each RCP measure, we compute each component's contribution to the aggregate retail RCP changes through variance decomposition.

Table B.6: *Use 6-Month RCP Changes: Summary Statistics of the Decomposition Results*

	Within(%)	Between(%)	Cross(%)	Entry(%)	Exit(%)
Obs.	10	10	10	10	10
Mean	62.43	3.03	4.33	19.12	11.09
Std.Dev.	2.41	1.02	1.68	1.75	1.47
Min	58.31	1.26	1.11	17.36	8.63
25%	60.88	2.54	3.76	18.02	10.46
50%	62.35	2.99	4.72	18.50	11.53
75%	64.01	3.60	5.19	19.45	12.00
Max	66.32	4.60	6.64	22.68	12.77

Notes: This table reports the summary statistics of the decomposition results from the 10 different RCP measures, using 6-month RCP changes.

Table B.7: *Summary Statistics for Different Wealth Group Retail Investors, 2011-2019*

Wealth Group (CNY) :	< 100K	100K-500K	500K-5M	5M-10M	> 10M	Total
	(1)	(2)	(3)	(4)	(5)	(6)
No. Accounts (Million)	58.41	18.54	9.60	0.48	0.37	87.39
Trading Volume (CNY Billion)	73.22	57.01	101.19	19.48	45.22	296.12
Trading Volume (%)	25%	19%	34%	7%	15%	100%
Size (CNY Tillion)	0.85	0.76	1.52	0.32	1.09	4.54
Size (%)	19%	17%	33%	7%	24%	100%

Notes: We separate the retail investor sample into 5 wealth groups according to investors' wealth in the stock market. We then calculate the average for all the variables within each group during 2011-2019 to get the table presented numbers. No. Accounts is the total number of accounts for each group from 2011 to 2019. Trading Volume is the daily average traded CNY value of all the stocks. Size is the investors' average CNY value of all holdings of stocks in SSE. The exchange rate is roughly 6.5 CNY = 1 USD.

Table B.8: *Summary Statistics of Stock RCO Decomposition Results*

	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
	(1)	(2)	(3)	(4)	(5)
Within(%)	27.9	16.3	16.1	27.8	38.4
Between(%)	1.4	13.0	-4.4	-0.0	5.8
Cross(%)	-0.2	13.3	-2.5	-0.4	1.5
Entry(%)	46.6	56.0	33.4	56.0	71.4
Exit(%)	24.3	54.8	-0.7	11.9	33.2

Notes: This table reports the summary statistics for 1,481 stock RCO decomposition results. Note that we carry out the variance decomposition for each stock's RCO. Since we have 1,481 stocks in our sample, we have 1,481 decomposition results.

Table B.9: *Mean of Stock RCO Decomposition Results on Different RCP Measures*

RCP measure	Within(%)	Between(%)	Cross(%)	Entry(%)	Exit(%)
	(1)	(2)	(3)	(4)	(5)
RCP calculated using simple arithmetic average					
$RCP^{K=1}_{Arithmetic}$	33.3	2.4	-0.4	24.9	39.7
$RCP^{K=2}_{Arithmetic}$	32.0	2.3	-0.4	28.2	38.0
$RCP^{K=3}_{Arithmetic}$	30.0	2.5	-0.6	33.4	34.8
$RCP^{K=6}_{Arithmetic}$	27.2	3.1	-0.6	36.6	33.8
$RCP^{K=12}_{Arithmetic}$	27.9	1.4	-0.2	46.6	24.3
RCP calculated using weighted average					
$RCP^{K=1}_{Weighted}$	33.6	3.3	-1.0	24.4	39.7
$RCP^{K=2}_{Weighted}$	32.8	3.0	-0.7	28.2	36.7
$RCP^{K=3}_{Weighted}$	31.3	3.2	-1.0	33.0	33.5
$RCP^{K=6}_{Weighted}$	28.7	3.8	-1.0	36.3	32.2
$RCP^{K=12}_{Weighted}$	29.5	2.3	-0.9	45.9	23.3

Notes: This table reports the mean of stock RCO decomposition results for different RCP measures. The first 5 rows are results for RCP calculated using simple arithmetic average. The bottom 5 rows are results for RCP calculated using weighted average with the held stock value as weights.

Table B.10: *Relationship Between SRCP/IRCP and Other Investor Characteristics*

Outcome Var:	SRCP	IRCP	SRCP-IRCP
	(1)	(2)	(3)
Disposition Effect	0.219***	-0.111***	0.330***
	(0.014)	(0.024)	(0.032)
Trade Frequency	0.003***	-0.002**	0.005***
	(8.60e-04)	(7.17e-04)	(0.001)
VolatilityStockHeld	0.361***	0.112***	0.249***
	(0.026)	(0.006)	(0.028)
Log(AvgInvest)	1.845***	-2.481***	4.326***
	(0.099)	(0.076)	(0.139)
Age	-0.387***	-0.051**	-0.336***
	(0.031)	(0.021)	(0.030)
Age_t^2	0.003***	6.94e-04***	0.002***
	(2.51e-04)	(1.63e-04)	(2.85e-04)
Experience	-2.571***	1.076***	-3.646***
	(0.184)	(0.369)	(0.403)
$Experience_t^2$	0.122***	-0.031	0.153***
	(0.011)	(0.024)	(0.026)
Female	-1.059***	2.794***	-3.853***
	(0.153)	(0.101)	(0.169)
Time Fixed Effects	Y	Y	Y
R-Squared	0.054	0.031	0.028
Observations.	662,470,542	662,470,542	662,470,542

Notes: Unit of observation is individual-by-month. Sample contains 18,162,319 investor accounts from the Shanghai Stock Exchange from Jan 2011 to Dec 2019. Standard errors (in parentheses) are double-clustered at the investor and date levels. All regressions control for Number Stocks and HHI up to the 3rd polynomials. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively. The R-Squared values reported do not include the impact of fixed effects.

Table B.11: *The Impacts of SRCP/IRCP on Future Investor Returns*

Outcome Var:	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+4}	Ret_{t+5}	Ret_{t+6}
	(1)	(2)	(3)	(4)	(5)	(6)
Ret_t	-0.002	-0.003	-0.004	0.002	0.001	0.007**
	(0.009)	(0.009)	(0.006)	(0.005)	(0.005)	(0.003)
$SRCP_t$	-0.424***	-0.323***	-0.273***	-0.244***	-0.233***	-0.223***
	(0.108)	(0.071)	(0.051)	(0.047)	(0.045)	(0.041)
$IRCP_t$	-0.256**	-0.194***	-0.153***	-0.121**	-0.105**	-0.083*
	(0.109)	(0.074)	(0.052)	(0.049)	(0.049)	(0.050)
Disposition $Effect_t$	-0.094**	-0.100**	-0.106**	-0.105**	-0.102**	-0.099**
	(0.046)	(0.043)	(0.043)	(0.042)	(0.043)	(0.042)
$VolatilityStockHeld_t$	-0.028	-0.021	-0.030	-0.025	-0.020	-0.015
	(0.039)	(0.033)	(0.027)	(0.027)	(0.028)	(0.027)
$Log(AvgInvest)_t$	0.001	0.015	0.027	0.032	0.036	0.036
	(0.051)	(0.052)	(0.050)	(0.048)	(0.046)	(0.046)
HHI_t	-0.157*	-0.050	-0.047	-0.040	-0.038	-0.038
	(0.082)	(0.035)	(0.031)	(0.030)	(0.029)	(0.029)
Trade $Frequency_t$	-0.056**	-0.041*	-0.027	-0.022	-0.032	-0.031*
	(0.024)	(0.024)	(0.025)	(0.023)	(0.021)	(0.018)
Age_t	0.048**	0.048**	0.047**	0.046*	0.045*	0.043*
	(0.020)	(0.023)	(0.023)	(0.024)	(0.024)	(0.023)
$Experience_t$	0.239***	0.198***	0.185***	0.177***	0.168***	0.158***
	(0.042)	(0.032)	(0.028)	(0.027)	(0.025)	(0.023)
$Female_t$	0.078***	0.064***	0.060***	0.058***	0.055***	0.053***
	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Time Fixed Effects	Y	Y	Y	Y	Y	Y
No. Observations	554,687,312	526,636,040	509,836,753	496,064,711	483,908,705	472,889,370
R-squared	0.001	8.15e-04	6.79e-04	5.71e-04	5.25e-04	5.33e-04

Notes: Unit of observation is investor-by-month. Sample contains 18,147,831 investor accounts from the Shanghai Stock Exchange from Jan 2011 through Dec 2019. Standard errors (in parentheses) are double-clustered at the individual and time levels. All the variables except returns (Ret_{t+k}) are standardized to the number of standard deviations from the mean. ***, **, * denote significance at the 1%, 5%, 10% levels, respectively. The R-Squared values reported do not include the impact of fixed effects.

Table B.12: *Predict Future Industry Returns Using IRCO*

Outcome Var:	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}
	(1)	(2)	(3)
Investor-Characteristic Ownership Signals			
$IRCO_t$	-1.401***	-1.139***	-1.240***
	(0.412)	(0.402)	(0.446)
Age_t	1.070***	0.851***	0.570**
	(0.351)	(0.195)	(0.272)
$Experience_t$	0.671**	0.808***	1.108***
	(0.338)	(0.313)	(0.407)
$Female_t$	0.788**	0.588*	0.590*
	(0.399)	(0.333)	(0.334)
Disposition $Effect_t$	-0.521	-0.686	-0.486
	(0.639)	(0.718)	(0.666)
$Wealth_t$	0.181	0.053	0.195
	(0.389)	(0.401)	(0.278)
Trade $Frequency_t$	0.722**	0.341	-0.077
	(0.331)	(0.508)	(0.486)
Number $Stocks_t$	-0.214	-0.302	-0.365
	(0.466)	(0.411)	(0.457)
Investor $Return_t$	-0.169	-0.626	-0.648
	(0.702)	(0.748)	(0.739)
Controls			
Retail $Share_t$	0.154	0.017	-0.252
	(0.360)	(0.333)	(0.295)
$Volatility_t$	-1.267	-0.529	-0.550
	(0.870)	(0.750)	(0.669)
Stock Past $Return_t$	1.009	0.689	0.701
	(0.637)	(0.527)	(0.607)
$Turnover_t$	-0.666	-0.488	-0.268
	(0.791)	(0.634)	(0.528)
$\beta_{Market,t}$	1.046**	0.349	0.313
	(0.462)	(0.470)	(0.518)
$\beta_{size,t}$	1.745**	1.152	1.256
	(0.866)	(0.824)	(0.811)
$\beta_{value,t}$	-0.566	-0.146	-0.122
	(0.487)	(0.522)	(0.516)
Time Fixed Effects	Y	Y	Y
R-squared	0.023	0.016	0.017
No. Observations	2,996	2,968	2,940

Notes: Unit of observation is industry-by-month. All the variables are taken quantile rank transformation.

Table B.13: *Predict Future Stock Returns Using SRCO*

Outcome Var:	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+1}	Ret_{t+1}
	(1)	(2)	(3)	(4)	(5)
Investor-Characteristic Ownership Signals					
SRCO_t	-0.508*	-0.243	-0.238	0.620	-0.116
	(0.303)	(0.285)	(0.325)	(0.391)	(0.743)
$\text{SRCO}_t \times \text{Retail Share}_t$				-2.230***	0.499
				(0.607)	(1.237)
$\text{SRCO}_t \times \text{Retail Share}_t \times \text{Stock Past Return}_t$					-4.208*
					(2.232)
Age_t	0.667***	0.212	0.189	0.644***	0.649***
	(0.181)	(0.151)	(0.149)	(0.179)	(0.179)
Experience_t	0.724***	0.514**	0.421*	0.719***	0.684***
	(0.263)	(0.211)	(0.240)	(0.263)	(0.259)
Female_t	0.361**	-0.077	-0.087	0.341*	0.318*
	(0.180)	(0.145)	(0.152)	(0.180)	(0.177)
$\text{Disposition Effect}_t$	-0.462	-0.409	-0.333	-0.418	-0.339
	(0.354)	(0.360)	(0.391)	(0.356)	(0.359)
Wealth_t	0.361	0.058	0.202	0.366	0.354
	(0.227)	(0.183)	(0.173)	(0.228)	(0.228)
Trade Frequency_t	-0.188	-0.144	-0.236	-0.175	-0.121
	(0.227)	(0.208)	(0.224)	(0.228)	(0.223)
Number Stocks_t	-0.320	-0.288	-0.007	-0.324	-0.363
	(0.242)	(0.255)	(0.260)	(0.241)	(0.236)
Investor Return_t	-0.247	0.044	0.037	-0.324	-0.442
	(0.536)	(0.520)	(0.531)	(0.532)	(0.537)
Controls					
Retail Share_t	-0.332	0.034	-0.057	0.836	0.834
	(0.406)	(0.386)	(0.369)	(0.518)	(0.748)
Volatility_t	-0.379	-0.906**	-0.841**	-0.387	-0.434
	(0.376)	(0.374)	(0.354)	(0.377)	(0.380)
$\text{Stock Past Return}_t$	-0.678	-0.314	-0.232	-0.626	0.468
	(0.430)	(0.369)	(0.352)	(0.428)	(1.106)
Turnover_t	-0.929**	-0.604	-0.511	-0.910**	-0.892**
	(0.450)	(0.418)	(0.393)	(0.448)	(0.446)
$\beta_{Market,t}$	0.002	-0.233	-0.332	-0.013	-0.013
	(0.307)	(0.295)	(0.287)	(0.308)	(0.309)
$\beta_{size,t}$	1.064*	0.747	0.855	1.009	1.004
	(0.622)	(0.604)	(0.592)	(0.629)	(0.629)
$\beta_{value,t}$	-0.133	0.127	0.126	-0.074	-0.092
	(0.383)	(0.371)	(0.358)	(0.386)	(0.387)
Time Fixed Effects	Y	Y	Y	Y	Y
Industry Fixed Effects	Y	Y	Y	Y	Y
R-squared	0.006	0.003	0.003	0.006	0.007
No. Observations	110,640	108,357	106,347	110,640	110,640

Notes: Unit of observation is stock-by-month. Sample contains 1,481 stocks at the Shanghai Stock Exchange from Jan 2011 to Dec 2019. Standard errors (in parentheses) are double clustered at the stock and time levels. All the variables are taken the quantile rank transformation.

Table B.14: *The Explanatory Power of RCP for Future Return Chasing Purchases*

Variable	Coef.	Std. Error	R-squared
RCP calculated using simple arithmetic average			
$\text{RCP}^{K=1}_{Arithmetic}$	0.157	0.015	0.014
$\text{RCP}^{K=2}_{Arithmetic}$	0.171	0.014	0.018
$\text{RCP}^{K=3}_{Arithmetic}$	0.186	0.012	0.022
$\text{RCP}^{K=6}_{Arithmetic}$	0.201	0.010	0.028
$\text{RCP}^{K=12}_{Arithmetic}$	0.221	0.020	0.037
RCP calculated using weighted average			
$\text{RCP}^{K=1}_{Weighted}$	0.148	0.014	0.014
$\text{RCP}^{K=2}_{Weighted}$	0.162	0.013	0.017
$\text{RCP}^{K=3}_{Weighted}$	0.176	0.012	0.020
$\text{RCP}^{K=6}_{Weighted}$	0.189	0.010	0.027
$\text{RCP}^{K=12}_{Weighted}$	0.207	0.019	0.035

This table reports the regression coefficient ρ, the standard error for ρ, and regression R-squared for each RCP measure. The first 5 rows are results for RCP calculated using simple arithmetic average. The bottom 5 rows are results for RCP calculated using weighted average with the held stock value as weights. For each RCP measure, we run the following regression:

$$\text{RCP}^{\text{New Purchased}}_{i,t(p)} = \alpha + \rho\text{RCP}_{i,t(p)-1} + \epsilon_{i,t(p)},$$

where $t(p)$ are the timestamps when investor i purchases stocks. $\text{RCP}^{\text{New Purchased}}_{i,t(p)}$ is investor i's newly purchased stocks' RCP at time $t(p)$, using only newly purchased stocks to calculate RCP. $\text{RCP}_{i,t(p)-1}$ is investor i's RCP before the purchase date $t(p)$.

Table B.15: *Variation Explained by Investor Fixed Effects for Different RCP Measures*

Measure	R-squared
RCP calculated using simple arithmetic average	
$\text{RCP}^{K=1}_{Arithmetic}$	0.451
$\text{RCP}^{K=2}_{Arithmetic}$	0.431
$\text{RCP}^{K=3}_{Arithmetic}$	0.432
$\text{RCP}^{K=6}_{Arithmetic}$	0.468
$\text{RCP}^{K=12}_{Arithmetic}$	0.523
RCP calculated using weighted average	
$\text{RCP}^{K=1}_{Weighted}$	0.449
$\text{RCP}^{K=2}_{Weighted}$	0.429
$\text{RCP}^{K=3}_{Weighted}$	0.429
$\text{RCP}^{K=6}_{Weighted}$	0.464
$\text{RCP}^{K=12}_{Weighted}$	0.518

This table reports the R-squared for each RCP measure fixed effects regression. The first 5 rows are results for RCP calculated using simple arithmetic average. The bottom 5 rows are results for RCP calculated using weighted average with the held stock value as weights. For each RCP measure, we run the following regression:

$$RCP_{i,t} = \delta_i + \epsilon_{i,t},$$

where δ_i are the investor fixed effects.

Table B.16: *Persistence of Stock RCO*

Outcome Var:	RCO_t	RCO_t	RCO_t	RCO_t	RCO_t	RCO_t
	(1)	(2)	(3)	(4)	(5)	(6)
RCO_{t-1}	0.881***	0.892***	0.889***	0.887***	0.890***	0.890***
	(0.028)	(0.065)	(0.066)	(0.067)	(0.071)	(0.071)
RCO_{t-2}		-0.012	0.010	0.009	0.012	0.015
		(0.047)	(0.038)	(0.036)	(0.036)	(0.035)
RCO_{t-3}			-0.022	-0.070**	-0.072**	-0.068**
			(0.049)	(0.029)	(0.028)	(0.027)
RCO_{t-4}				0.054*	0.016	0.015
				(0.031)	(0.021)	(0.022)
RCO_{t-5}					0.039	0.008
					(0.024)	(0.018)
RCO_{t-6}						0.032
						(0.025)
Adj. R-squared	0.800	0.799	0.796	0.794	0.792	0.790
Observations	114,979	113,493	112,016	110,542	109,071	107,610

Notes: Column (1) to (6) are correspondingly AR(1) to AR(6) regressions. Unit of observation is stock-by-month. Sample contains 1,481 stocks at the Shanghai Stock Exchange from Jan 2011 to Dec 2019. Standard errors (in parentheses) are double clustered at the stock and time levels. The ***, **, * denote significance at the 1%, 5%, 10% level correspondingly.

Table B.17: *The Returns on Triple Sort of RCO, Past Returns and Retail Share(%)*

	RetailShare1 (Low)			RetailShare2			RetailShare3 (High)		
	PastRet1	PastRet2	PastRet3	PastRet1	PastRet2	PastRet3	PastRet1	PastRet2	PastRet3
RCO1 (Low)	8.36	12.62	8.46	3.09	7.40	-0.86	13.83	10.38	5.06
RCO2	-1.23	5.23	11.93	1.93	13.12	7.60	16.47	9.35	-0.85
RCO3	-2.08	-3.21	1.76	-6.25	5.81	6.20	4.42	5.21	1.37
RCO4	0.32	-5.41	1.09	-5.04	-0.11	-3.55	5.03	-1.58	-8.51
RCO5 (High)	-13.66	4.63	-1.99	-2.79	-7.34	-10.62	9.44	-6.90	-25.64
RCO1-RCO5	22.03	7.99	10.46	5.89	14.74	9.76	4.39	17.27	30.70

Notes: This table reports the triple sort returns by stock RCO, retail share and past returns. Stocks are sorted into 5 quintiles by their RCO in the current month. RCO1 is the group of stocks with the lowest RCO, while RCO5 is the group of stocks with the highest RCO. Stocks are also sorted into 3 terciles according to their retail share in the current month. RetailShare1 means the group with the lowest retail share, while RetailShare3 means the group with the highest retail share. Stocks are also sorted into 3 terciles according to their past one year returns. PastRet1 means the group with the lowest past returns, while PastRet3 means the group with the highest past returns. We then carry out an independent triple sort on RCO, retail share, and past returns, which gives us $5 * 3 * 3 = 45$ groups. Each cell in the table reports the average annualized returns for the corresponding triple sort group. The last row shows the RCO factor returns for different past returns and retail share combinations.

Table B.18: *Relationships between IPO Price, Current Price, and Retail Share*

This table shows the relationship between IPO price, current stock price, and retail share. Column (1) regresses the log of the current price on its IPO price. A one standard deviation increase in Log(IPO Price) (0.784) will increase the current price by 6.7%. Column (2) shows that a stock's IPO price negatively predicts its current retail share: a one standard deviation increase in Log(IPO Price) (0.784) will make the retail share decrease by 0.86 percentage points. Column (3) regresses retail share on log current price. A one standard deviation increase in Log(Price) (0.7) will reduce retail share by 7.3 percentage points. Finally, Column (4) adds the log of the current price on top of Column (2). We can see that once we control the stock's current price, the IPO price has no effect on its retail share, i.e., a stock's IPO price influences its current retail share only through the channel that it correlates with the current stock price.

Outcome Var:	Log(Price)	Retail Share	Retail Share	Retail Share
	(1)	(2)	(3)	(4)
Log(IPO Price)	0.083***	-0.011**		-0.003
	(0.022)	(0.005)		(0.004)
Log(Price)			-0.104***	-0.104***
			(0.006)	(0.006)
Controls				
Past1YearRet	0.341***	-0.086***	-0.050***	-0.051***
	(0.039)	(0.012)	(0.009)	(0.009)
β_{Market}	-0.017	0.032**	0.030**	0.030**
	(0.055)	(0.013)	(0.012)	(0.012)
β_{size}	-0.114***	0.172***	0.161***	0.160***
	(0.035)	(0.009)	(0.008)	(0.008)
β_{value}	-0.328***	-0.011	-0.045***	-0.046***
	(0.022)	(0.007)	(0.007)	(0.007)
Volatility	2.928***	-0.093	0.195***	0.211***
	(0.389)	(0.067)	(0.057)	(0.056)
Turnover	0.146***	0.051***	0.067***	0.066***
	(0.020)	(0.005)	(0.005)	(0.005)
Time Fixed Effects	Y	Y	Y	Y
R-Squared	0.225	0.279	0.369	0.370
Observations.	113,137	113,075	113,131	113,075

Observations are stock by month. Standard errors are double clustered at stock and month levels.

Table B.19: *Correlation Matrix Among Investor-Characteristic Ownership Signals*

	RCO	Age	Experience	Female	DispositionEffect	Wealth	TradeFrequency	NumberStocks
RCO								
Age	-0.17							
Experience	-0.24	0.58						
Female	0.08	0.07	0.12					
Disposition Effect	0.02	-0.23	-0.26	-0.01				
Wealth	-0.03	0.02	-0.05	-0.23	-0.37			
Trade Frequency	0.20	-0.17	-0.14	-0.08	0.02	0.30		
Number Stocks	-0.26	0.31	0.46	0.07	-0.06	0.01	0.08	
Investor Return	0.16	0.16	0.20	0.03	-0.49	0.21	0.21	0.07

Notes: Unit of observation is stock-by-month. Sample contains 1,481 stocks from the Shanghai Stock Exchange from Jan 2011 to Dec 2019. Correlation matrix among ownership signals are reported. Since we take the quantile rank transformation on all the variables, the Pearson correlation and Spearman correlation are of the same value.