What Drives Stock Prices in a Bubble?*

Weihua Chen
Shanghai Stock Exchange

Shushu Liang
Harvard University

Donghui Shi
Shanghai Stock Exchange

December 21, 2022

(click here for latest version)

Abstract

To shed light on the formation, expansion, and deflation of bubbles, we 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, we estimate a structural model of heterogeneous investor demand. The model allows us 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. We highlight the ways in which our structural model quantifies the forces in Kindleberger (1978)'s classic narrative.

*Corresponding author: Shushu Liang (shushuliang@g.harvard.edu, Department of Economics, Harvard University). We are extremely grateful to John Campbell, Xavier Gabaix, Mark Egan, and Adi Sunderam for their invaluable support and guidance. We thank Malcolm Baker, Adrien Bilal, Gabriel Chodorow-Reich, Robin Greenwood, Sam Hanson, Sebastian Hillenbrand, Elisabeth Kempf, Josh Lerner, N. Gregory Mankiw, Kenneth Rogoff, Marco Sammon, Andrei Shleifer, Emil Siriwardane, Jeremy Stein, Ludwig Straub, Boris Vallee, Luis Viceira, Jonathan Wallen, Neng Wang, and seminar participants at the Harvard Finance Seminars, Harvard Macro Seminars, and BlackRock Applied Research Award Meeting for their thoughtful comments and discussions. We thank the Shanghai Stock Exchange for data and technical support.

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 2022). 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 section of stocks for better identification and empirical implementation.[1] 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.[2]

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

1. 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.

2. 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.

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.

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 2022).[3] 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), 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. 1990; Shleifer and Summers 1990). We demonstrate the price impacts from retail investors and quantify the magnitudes of retail trading behavior contributing to stock booms

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

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.

Organization of the paper

The remainder of the paper is organized as follows. In Section 1, we describe our data and provide more background information about the 2015 Chinese stock market bubble. In Section 2, we define boom-bust stocks and examine their differing characteristics compared to other stocks. In Section 3, we lay out a framework to estimate investors' heterogeneous asset demand functions. In Section 4, we quantify the different channels contributing to stock boom-busts through variance decomposition. In Section 5, we conduct counterfactual analyses. We conclude and discuss future research directions in Section 6.

1 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.

1.1 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.[4] In addition to their 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 10,000 accounts (Odean 1998; Odean 1999). 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 dataset closest to ours is the Indian stock holdings used by Campbell, Ramadorai, and Ranish (2014) and Campbell, Ramadorai, and Ranish (2019). 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 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 B for details of data construction.

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

1.3 The 2015 Chinese Stock Market Bubble

Let us take a look at the 2015 Chinese stock market bubble. Panel (a) in Figure 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 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.[5] 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.

1.4 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 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 2 are cumulative abnormal returns. 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 reason we control for the CAPM beta here is that we want to examine other stock characteristics

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

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.

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

2 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.

2.1 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 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.[6] Stocks have an average abnormal return boom of 43%, and 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%.

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.

2.2 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

6. 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.

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:

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

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 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 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.

2.3 Suggestive Evidence on Who May Have Influenced Prices

Figure 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 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.

3 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.

3.1 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:

$$\ln\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = \alpha_{i,t} + \beta_{0,i,t}\left(\mathrm{me}_t(n) - \mathrm{be}_t(n)\right) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n) + \epsilon_{i,t}(n), \tag{2}$$

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 3.2.2

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 3.2.4, 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:

$$P_t(n)S_t(n) = \mathrm{ME}_t(n) = \sum_{i=1}^{I} w_{i,t}(n)A_{i,t}. \tag{3}$$

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 (2), we can take log on both sides of the Equation (3) and rewrite the market clearing condition as:

$$\mathbf{p}_t = \mathbf{f}(\mathbf{p}_t) = \log\left(\sum_{i=1}^{I} A_{i,t}\mathbf{w}_{i,t}(\mathbf{p}_t)\right) - \mathbf{s}_t, \tag{4}$$

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.

3.2 Demand Function Estimation

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

3.2.1 Regression Specification

The demand function we need to estimate is Equation (2):

$$\ln\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = \alpha_{i,t} + \beta_{0,i,t}\left(\mathrm{me}_t(n) - \mathrm{be}_t(n)\right) + \sum_{k=1}^{K} \beta_{k,i,t}x_{k,t}(n) + \epsilon_{i,t}(n). \tag{2}$$

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 (2) is estimated using an instrument for $\text{me}_t(n)$, which we will describe in Section 3.2.4.

3.2.2 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 2.2, 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 2.

3.2.3 Parameterization

The demand function (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 B.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 B.7 and B.8 illustrate the procedures used to construct investor types in detail.

3.2.4 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 (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 $\beta_{0,i,t}$.

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:

$$\widehat{\mathrm{me}}_t(n) = \log\left(\sum_{j \in Stable} A_{j,t} \frac{\mathbb{I}_{j,t}(n)\,\mathrm{BE}_t(n)}{\sum_{m=1}^{N} \mathbb{I}_{j,t}(m)\,\mathrm{BE}_t(m)}\right), \tag{5}$$

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:

$$A_{j,t}, \mathbb{I}_{j,t}(n) \perp \epsilon_{i,t}(n) \mid \boldsymbol{x}_t(n), \forall j \in \{\text{Stable institutions}\}. \tag{6}$$

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."[7] 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 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.

7. 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).

Instead, if we used all institutions to construct the mandate IV, some institutions such 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 $\epsilon_{i,t}(n)$.

3.2.5 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.

3.3 Estimation Results

In this subsection, we report the demand estimation results. Summary statistics for the estimated demand function parameters are reported in Table 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.

3.3.1 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 $1 - \hat{\beta}_{0,i,t}$.[8] 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:

$$1 - \hat{\beta}_{0,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1} \left[i \in I, t \in T \right], \tag{7}$$

8. See detailed derivation in Appendix Section A.

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 $\mathbb{1}[i \in I, t \in T]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. 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 7 plots the estimated $\hat{\gamma}_{I,T}$ (coefficient estimates are reported in Column (1) of Table 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 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 ($0.1/0.25 = 0.4$).

3.3.2 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 2.2.

Consider stock profitability and CAPM beta. Preferences/beliefs about other stock characteristics follow similar patterns and are reported in Appendix C. We first examine the preferences/beliefs differences between retail investors and institutions. As in Section 3.3.1, we regress the previously estimated characteristic preferences/beliefs on the interaction between the investor's type (retail vs. institution) and time period indicators:

$$\hat{\beta}_{k,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}\left[i \in I, t \in T\right], \tag{8}$$

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 8 (coefficient estimates are reported in Columns (2) and (3) of Table 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 2.2. 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 9 plots the preferences/beliefs of retail investors with different levels of wealth (coefficient estimates are reported in Table 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 10 plots the average preferences/beliefs of retail investors entering the market at different time periods (coefficient estimates are reported in Table 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.

4 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 and then report decomposition results.

4.1 Return Variance Decomposition

According to the asset demand system in 3.1, 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 $\boldsymbol{\epsilon}_t$:

$$\mathbf{p}_t = \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_t\right). \tag{9}$$

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

$$\mathbf{r}_t = \mathbf{p}_t - \mathbf{p}_{t-1} = \Delta\mathbf{p}_t(\mathbf{s}) + \Delta\mathbf{p}_t(\mathbf{x}) + \sum_{i=1}^{I}\left(\Delta\mathbf{p}_t(\mathbf{A}_i) + \Delta\mathbf{p}_t(\boldsymbol{\beta}_i) + \Delta\mathbf{p}_t(\boldsymbol{\epsilon}_i)\right), \tag{10}$$

where

$$\begin{aligned}
\Delta\mathbf{p}_t(\mathbf{s}) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_{t-1}, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_{t-1}, \mathbf{x}_{t-1}, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\Delta\mathbf{p}_t(\mathbf{x}) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_{t-1}, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\sum_{i=1}^{I}\Delta\mathbf{p}_t(\mathbf{A}_i) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_{t-1}, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\sum_{i=1}^{I}\Delta\mathbf{p}_t(\boldsymbol{\beta}_i) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_{t-1}\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_{t-1}, \boldsymbol{\epsilon}_{t-1}\right), \\
\sum_{i=1}^{I}\Delta\mathbf{p}_t(\boldsymbol{\epsilon}_i) &= \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_t\right) - \mathbf{g}\left(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\beta}_t, \boldsymbol{\epsilon}_{t-1}\right).
\end{aligned} \tag{11}$$

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 (4). The Jacobian matrix used in the iteration has the following analytical form:

$$\frac{\partial f(\mathbf{p};n)}{\partial p(m)} = \begin{cases} \frac{\sum_{i=1}^{I}\beta_{0,i}A_i w_i(\mathbf{p};n)(1-w_i(\mathbf{p};n))}{\sum_{i=1}^{I}A_i w_i(\mathbf{p};n)} & \text{if } m = n \\ \frac{\sum_{i=1}^{I}-\beta_{0,i}A_i w_i(\mathbf{p};n)w_i(\mathbf{p};m)}{\sum_{i=1}^{I}A_i w_i(\mathbf{p};n)} & \text{if } m \neq n \end{cases} \tag{12}$$

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.

4.2 Decomposition Results

Decomposition results are reported in Table 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 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.

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 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 2.3 (Figure 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), 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 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 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 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).[9]

Wealth reallocation among existing retail investors does not seem important.

Table 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 2022). We show that this is not the case for cross-sectional stock returns.

5 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.

9. 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.

5.1 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 11. 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 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.

5.2 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 12 reports the counterfactual price paths under such a scenario.

In Figure 12, 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 (05/2015) is reduced by 32%.

The green line in Figure 12 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%.

6 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. Understanding the price impacts of retail trading behavior is important for related policy discussions and market regulations.

References

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

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

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

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.*

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

Barber, Brad M, and Terrance Odean. 2013. "The behavior of individual investors." In *Handbook of the Economics of Finance,* 2:1533–1570. Elsevier.

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.

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.

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, Tarun Ramadorai, and Benjamin Ranish. 2019. "Do the rich get richer in the stock market? Evidence from India." *American Economic Review: Insights* 1 (2): 225–40.

———. 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).*

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.

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

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

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.

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."

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, 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.

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.

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

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

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

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.

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.

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.

———. 1999. "Do investors trade too much?" *American economic review* 89 (5): 1279–1298.

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

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.

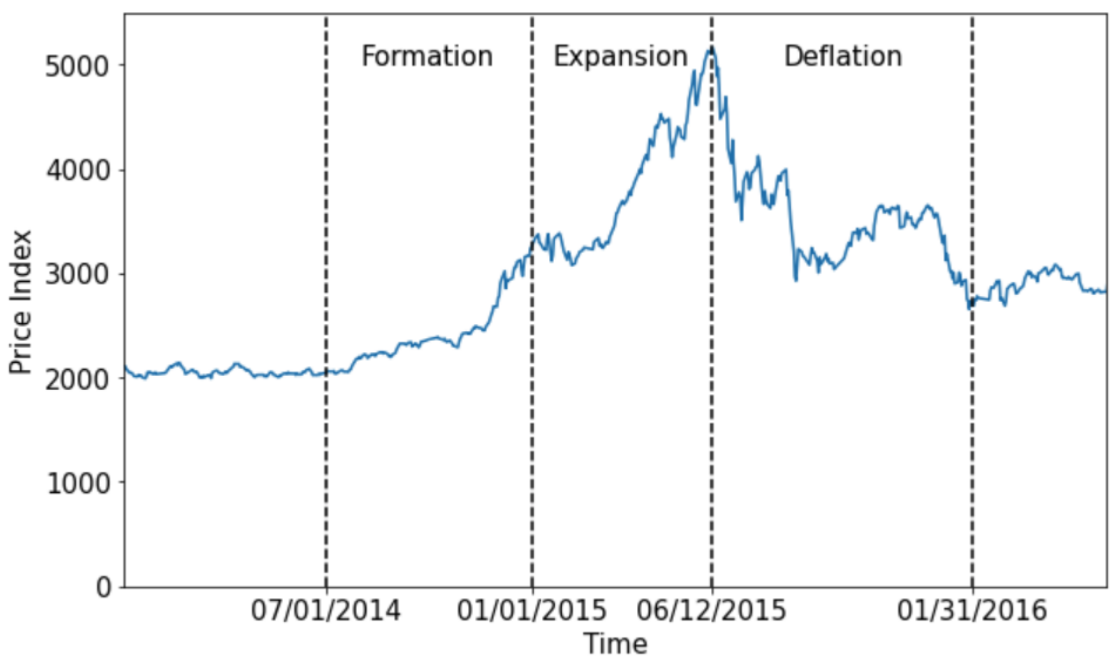
Xiong, Wei, and Jialin Yu. 2011. "The Chinese warrants bubble." *American Economic Review* 101 (6): 2723–53.

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?.

Figure 1: Time Series of Shanghai Stock Exchange Composite Index

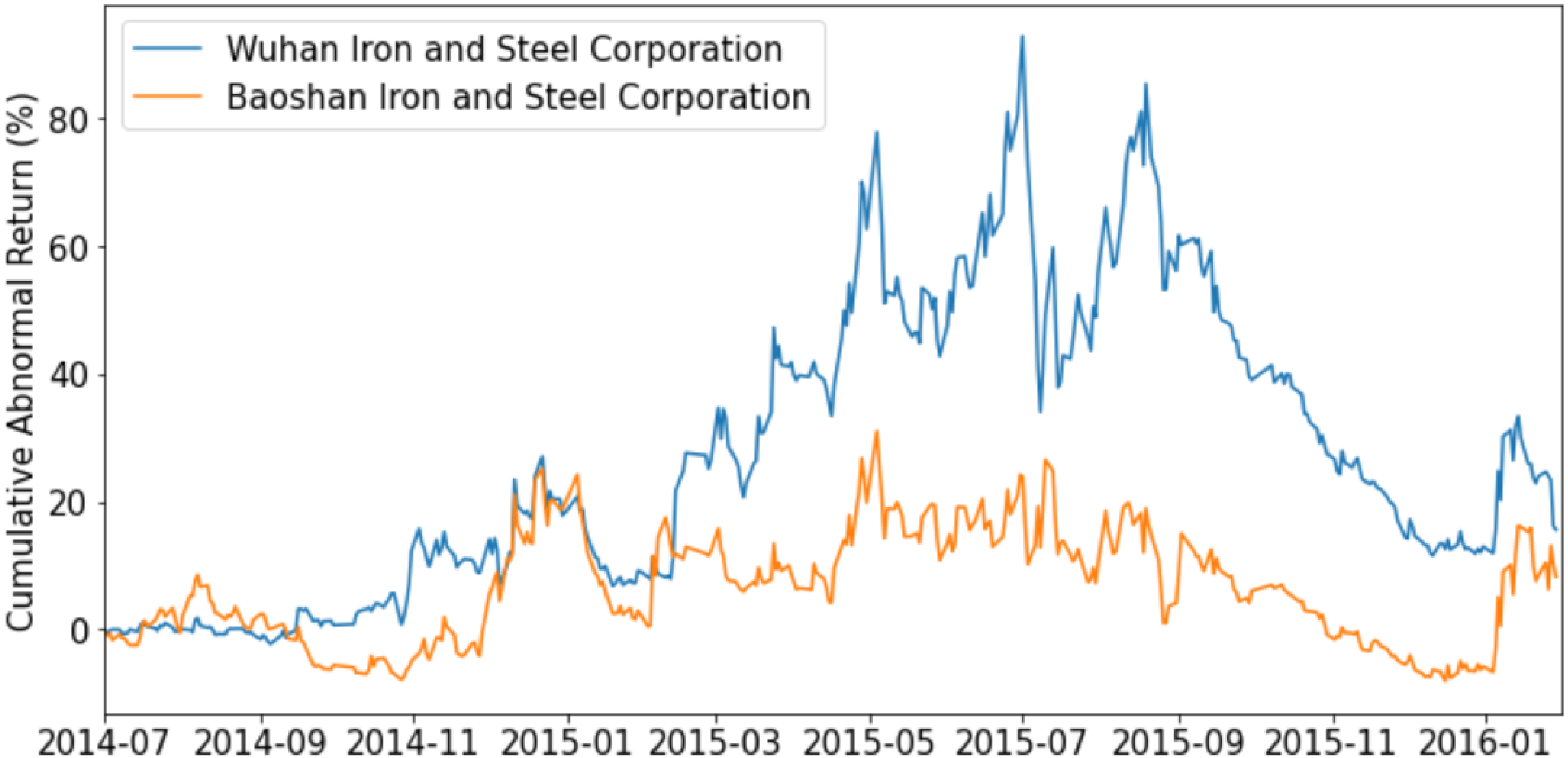
(a) Price Index from 2011 through 2019

(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).

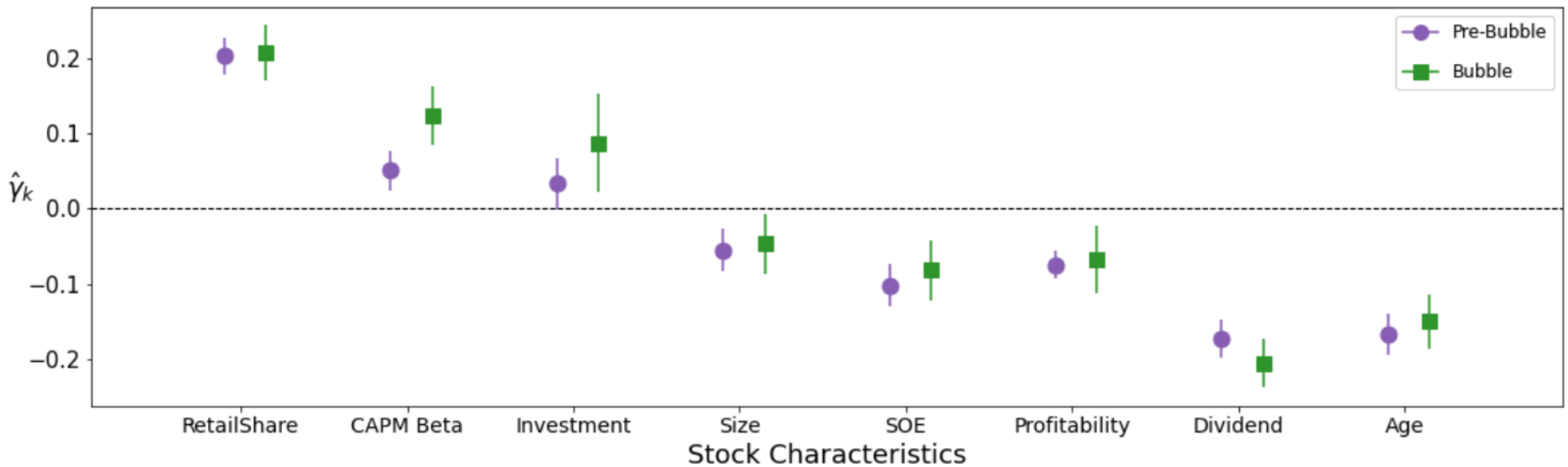
Figure 2: Cumulative Abnormal Returns for Two Illustrative Stocks during the 2015 Chinese Stock Market Bubble



	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.

Figure 3: Differences in Characteristics between Boom-Bust Stocks and Other Stocks

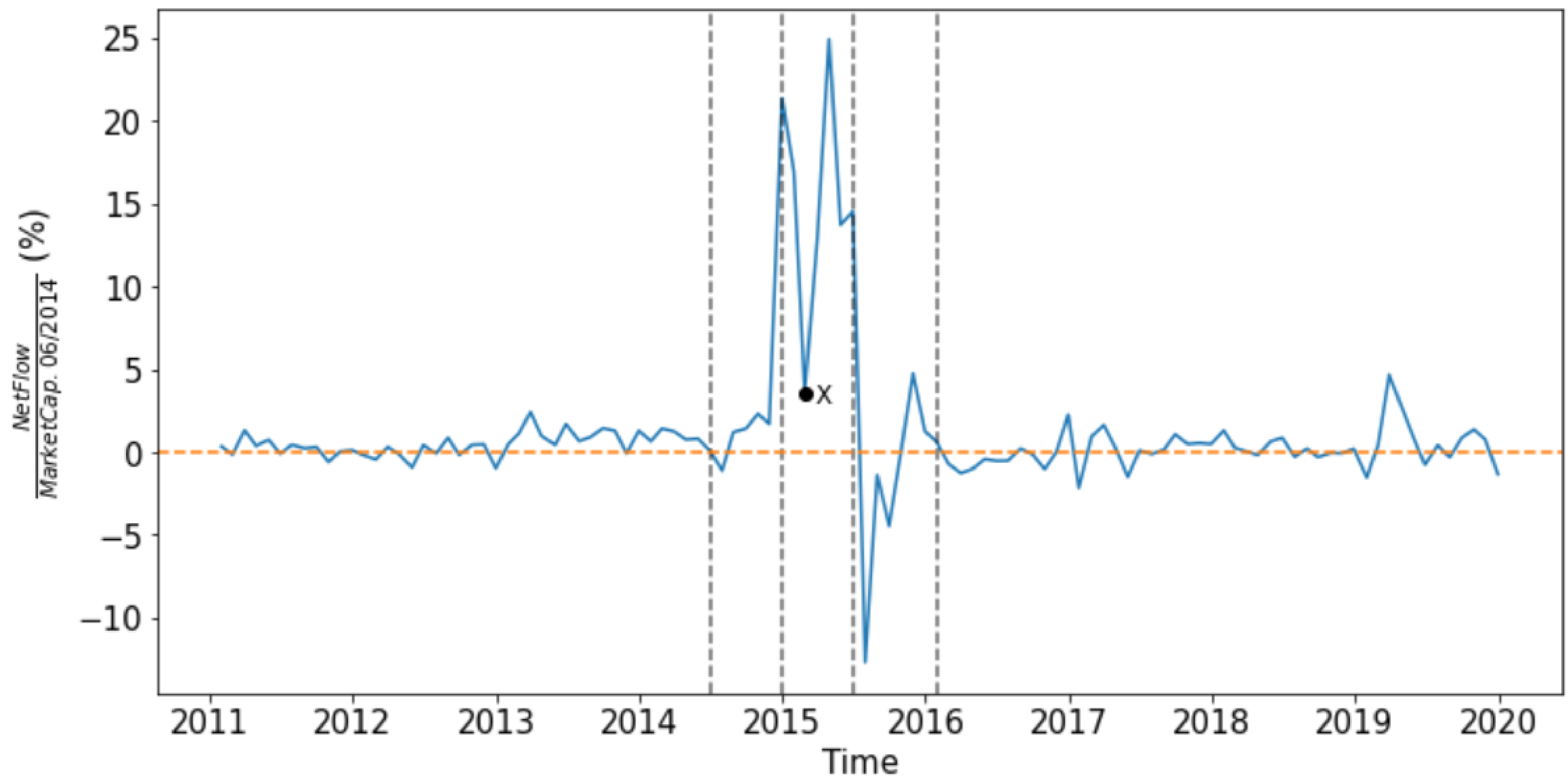


Notes: This figure plots the differences in characteristics between boom-bust stocks and other stocks (the estimated γ_k), both during the bubble period (07/2014-01/2016) and the pre-bubble period (01/2011-06/2014). γ_k is estimated from the following regression:

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

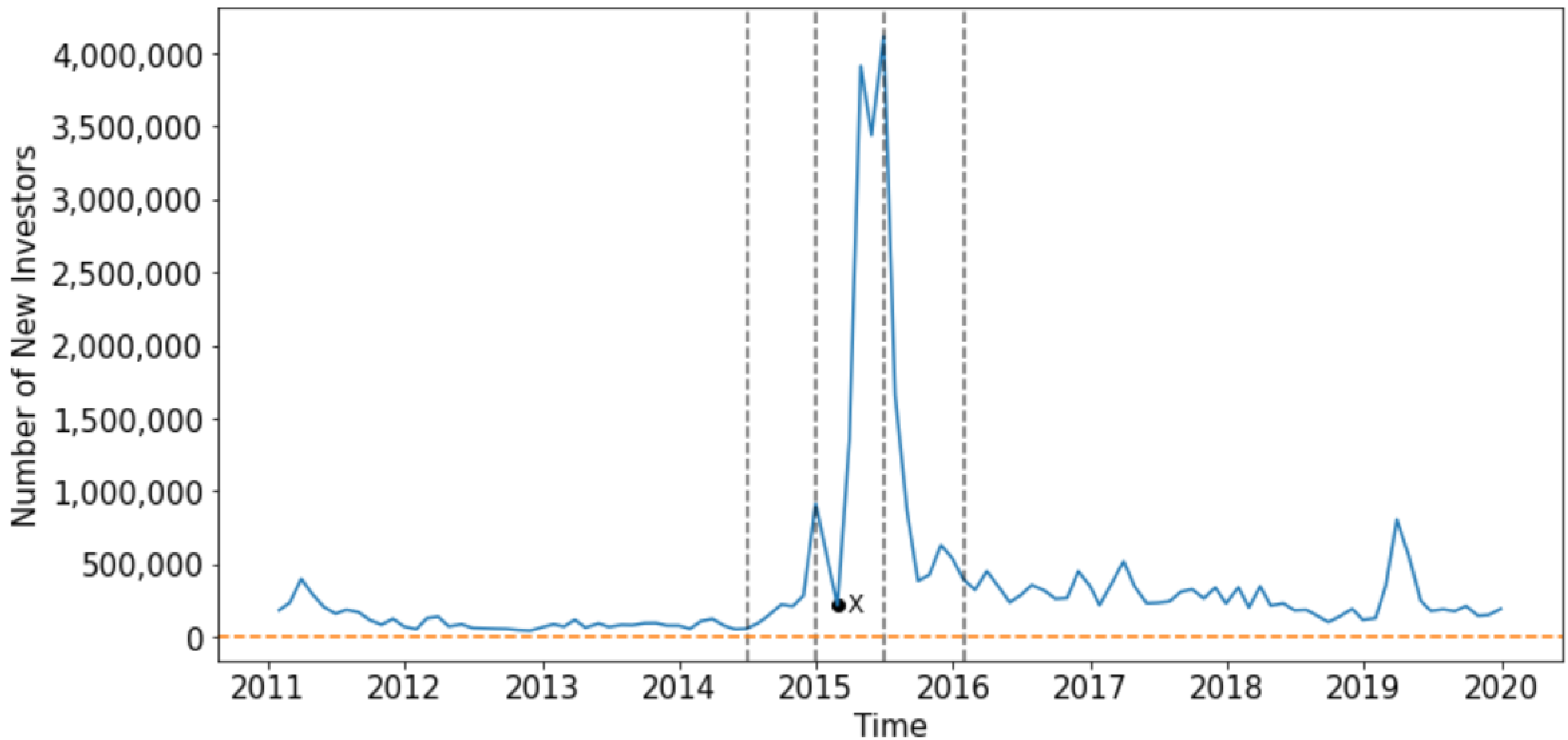
where $x_{k,t}(n)$ is stock n's characteristic x_k at time t. $\mathbb{1}_{\text{Boom-Bust Stock}}(n)$ indicates whether stock n is classified as a boom-bust stock. $\delta_{k,t}$ are time fixed effects. Observations are at the stock-by-month level. Heteroskedasticity-robust standard errors are reported, 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.

Figure 4: Retail Investors' Net Capital Flows into Boom-Bust Stocks



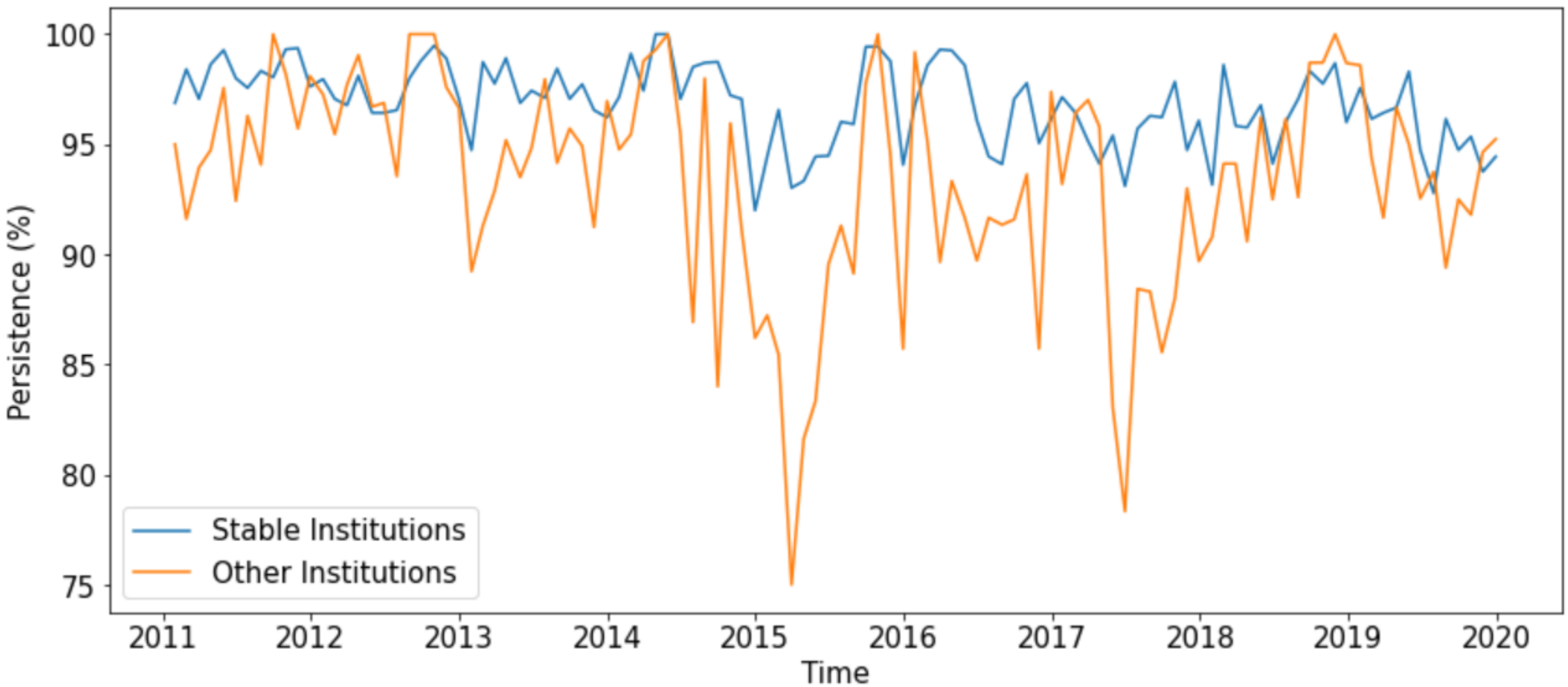
Notes: This figure plots retail investors' net capital flows into boom-bust stocks. Institutions serve as the counterpart. Net capital flows are calculated as the total value purchased minus the total value sold. We divide the net capital flows by the stocks' market capitalization prior to the bubble (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 5: Number of New Retail Investors Entering the Shanghai Stock Exchange



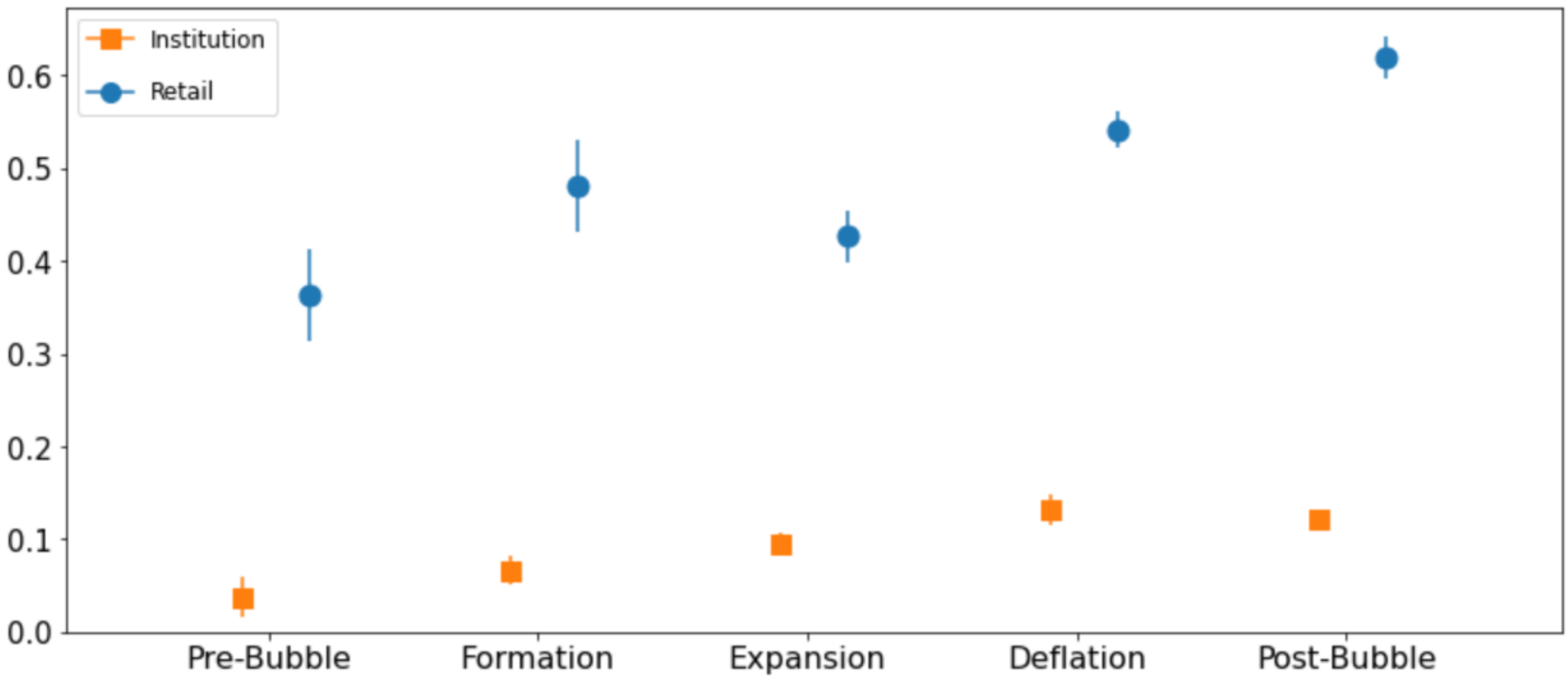
Notes: This figure shows a time series of the number of new retail investors entering the Shanghai Stock Exchange. The grey dashed lines indicate different bubble phases: the formation, the expansion, and the deflation of the bubble. During the beginning of 2015, there were shortly imposed policy restrictions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

Figure 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.

Figure 7: Estimated Demand Elasticities of Retail Investors and Institutions over Different Periods



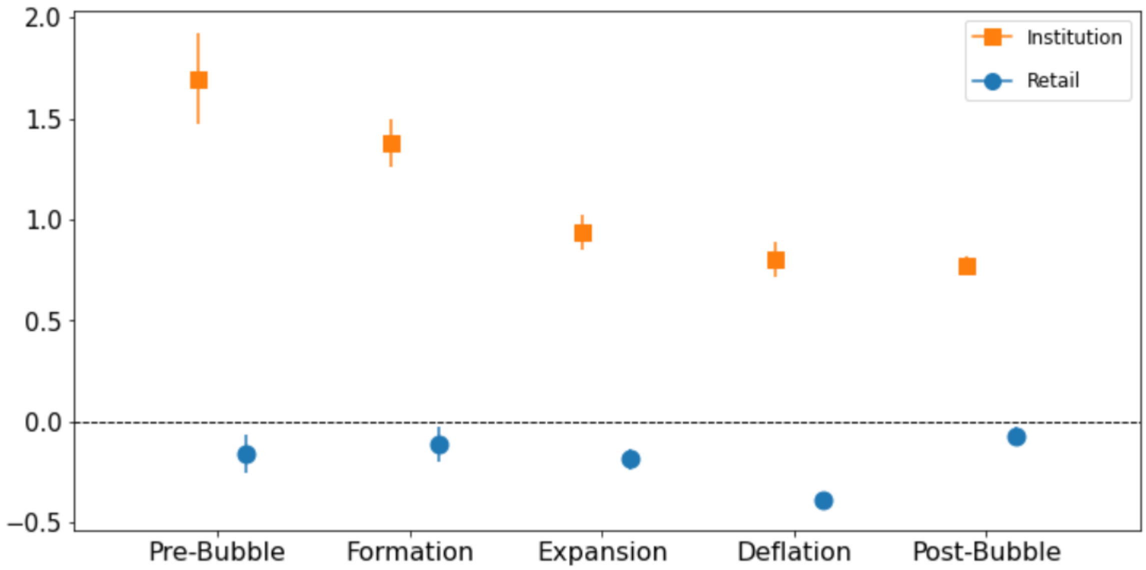
Notes: This figure reports the average demand elasticities ($\hat{\gamma}_{I,T}$) of retail investors and institutions, over different time periods. $\hat{\gamma}_{I,T}$ is estimated by regressing the demand elasticity $(1 - \hat{\beta}_{0,i,t})$ on the interaction between the investor's type (retail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

$$1 - \hat{\beta}_{0,i,t} = \sum_{I,T} \gamma_{I,T} \mathbb{1}\left[i \in I, t \in T\right], \tag{13}$$

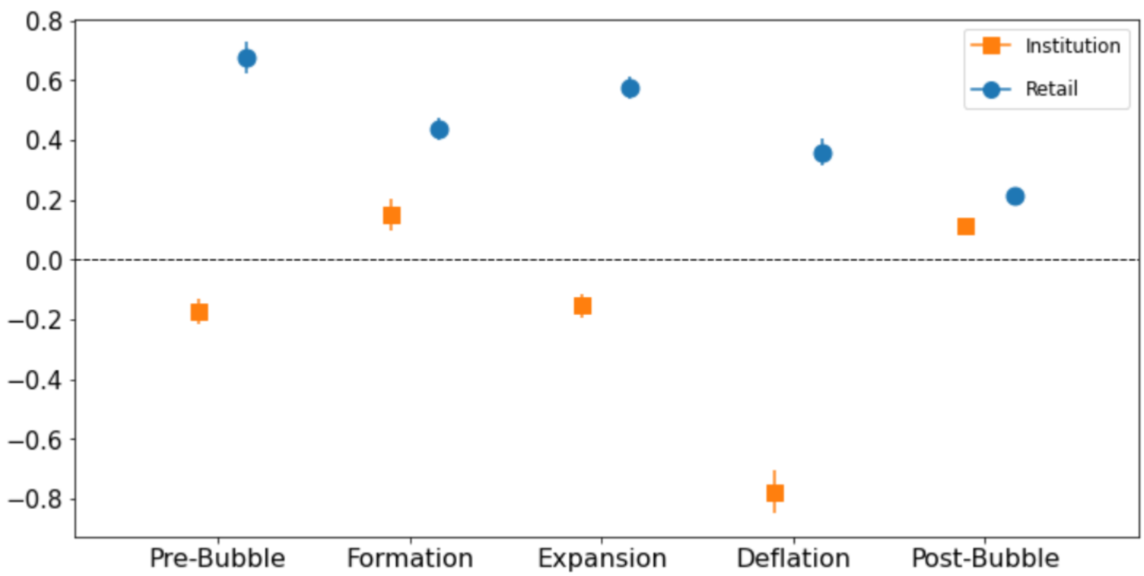
where $1 - \hat{\beta}_{0,i,t}$ is the previously estimated demand elasticity for investor i at time t, $I \in$ {Retail, Institution}, $T \in$ {Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable indicating whether investor i belongs 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.

Figure 8: Preferences/Beliefs of Retail Investors and Institutions

(a) Preferences/Beliefs about Profitability



(b) Preferences/Beliefs about CAPM Beta



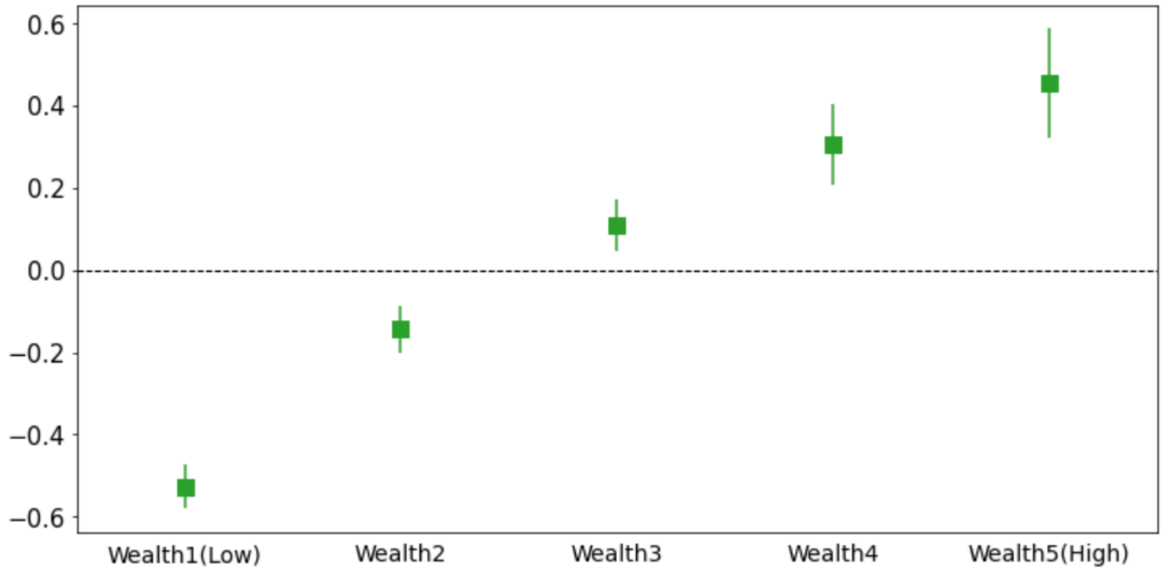
Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_{I,T}$) of retail investors and institutions over different time periods. $\hat{\gamma}_{I,T}$ is estimated by the following regression:

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

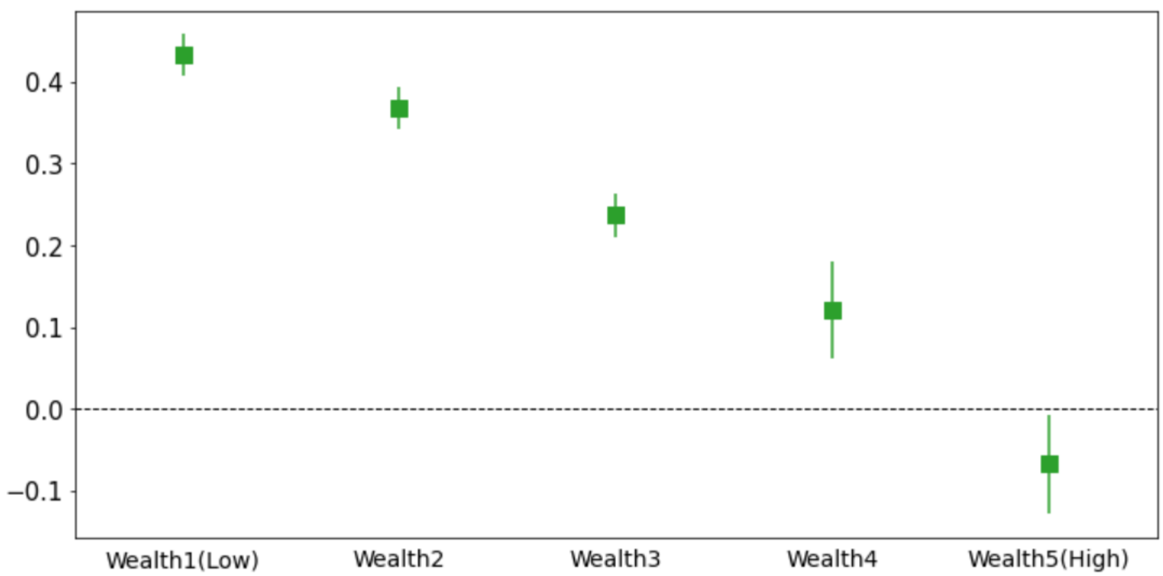
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $I \in \{\text{Retail, Institution}\}, T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable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

Figure 9: Preferences/Beliefs of Retail Investors with Different Levels of Wealth

(a) Preferences/Beliefs about Profitability



(b) Preferences/Beliefs about CAPM Beta



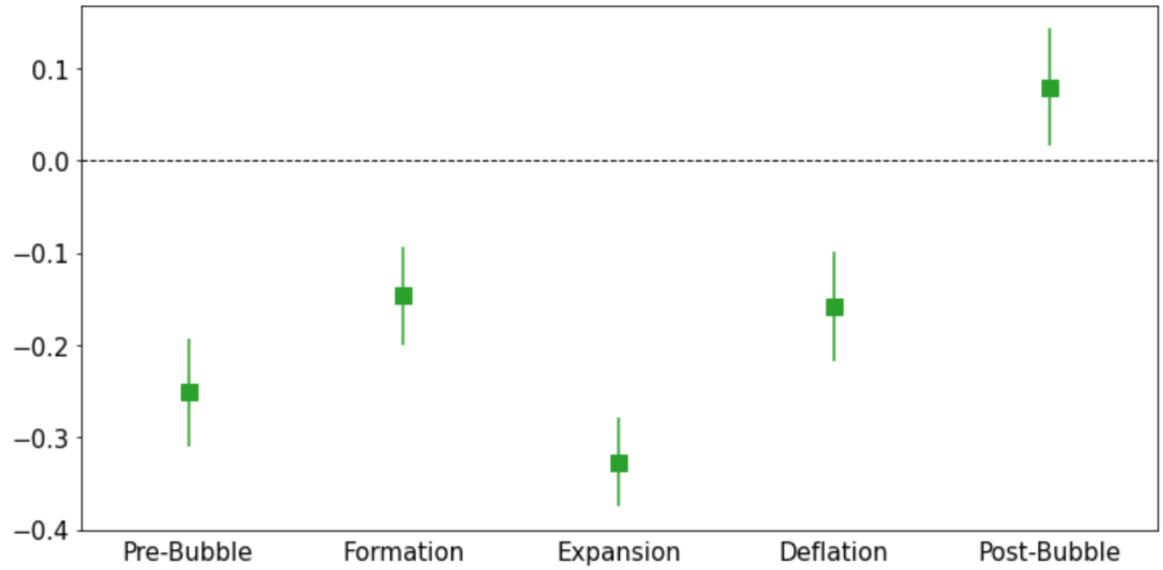
Notes: This figure reports the preferences/beliefs ($\hat{\gamma}_W$) of retail investors with different levels of wealth. We classify retail investors into five groups according to their average stock market wealth in the previous year. $\hat{\gamma}_W$ is estimated by the following regression:

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

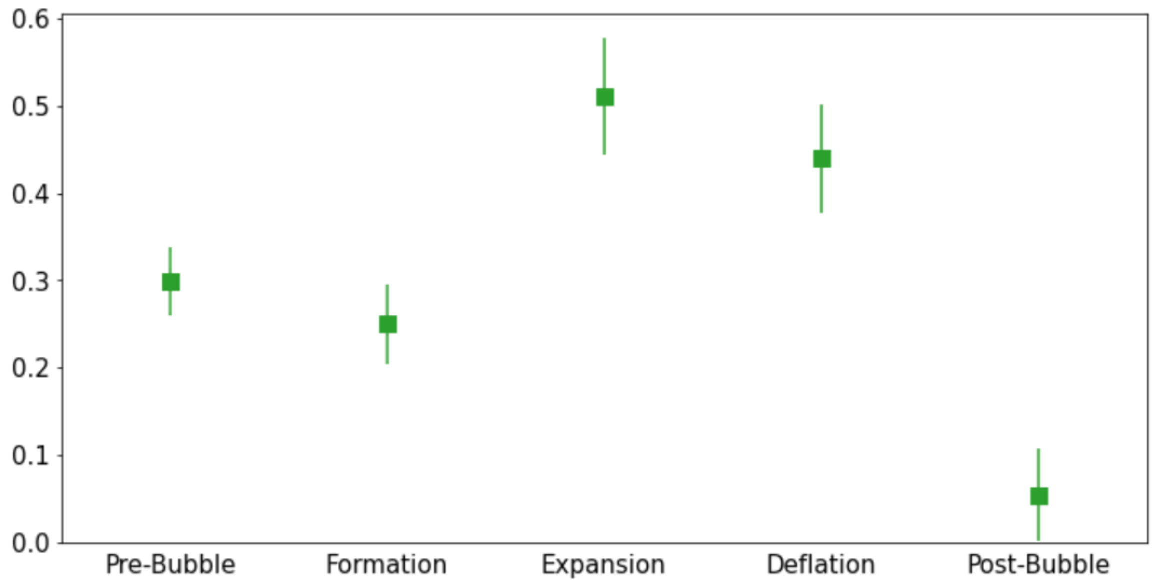
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $W \in \{$Wealth1 (Low), Wealth2, Wealth3, Wealth4, Wealth5 (High)$\}$, and $\mathbb{1}\left[i \in W\right]$ is a dummy variable indicating whether investor i belongs to wealth 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.

Figure 10: Preference/Beliefs of Retail Investors in Different Cohorts

(a) Preferences/Beliefs about Profitability



(b) Preferences/Beliefs about CAPM Beta

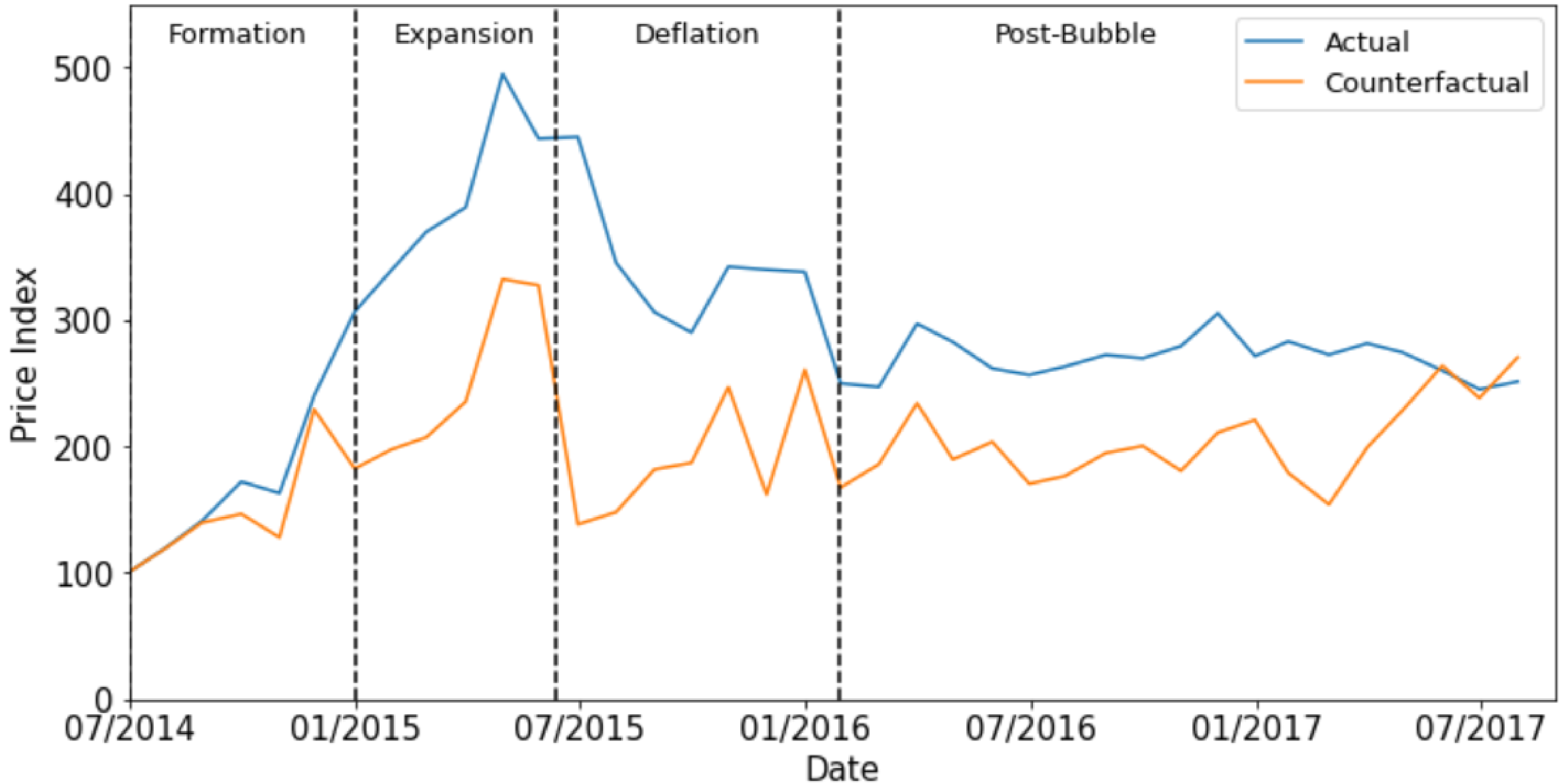


Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_C$) of retail investors entering the market at different time periods. We classify retail investors by their market entry cohort: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble 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:

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

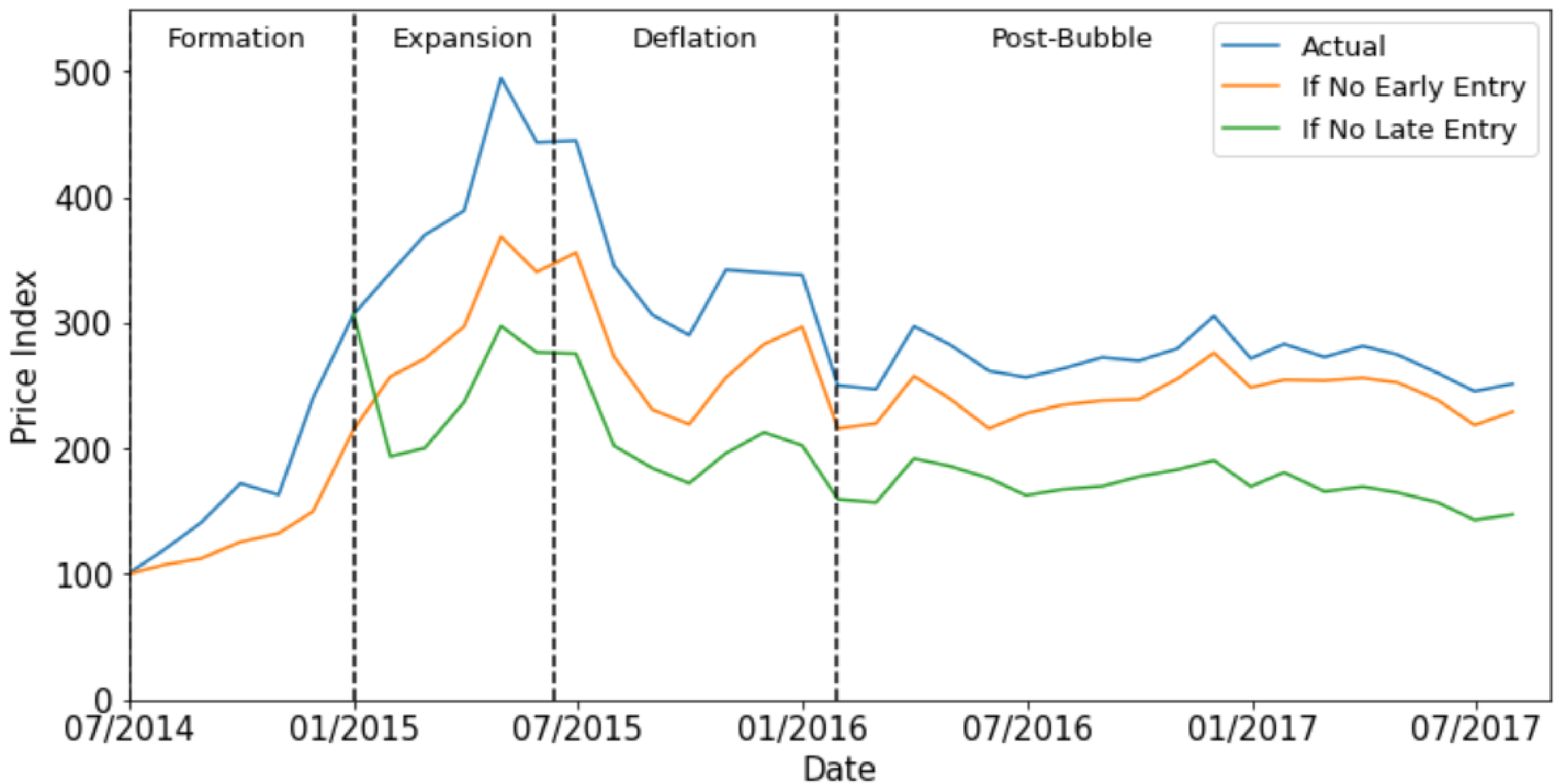
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $C \in$ {Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}, $\mathbb{1}\left[i \in C\right]$ is a dummy variable indicating whether investor i belongs to cohort C, and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

Figure 11: 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 ($\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \boldsymbol{\epsilon}_t$) are unchanged. We compute the counterfactual price vector under this scenario.

Figure 12: 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.

Table 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.

Table 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.

Table 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 3.2.4.

Table 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, $I \in \{\text{Retail, Institution}\}, T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, 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. 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.

Table 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, $W \in$ {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.

Table 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.

Table 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 ($\mathbf{A}$)	-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 ($\boldsymbol{\epsilon}$)	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 ($\mathbf{A}$)	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 ($\boldsymbol{\epsilon}$)	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 ($\mathbf{x}$)	21.2***	-7.1	1.5	3.0
	(5.9)	(7.2)	(3.4)	(3.2)
Shares outstanding ($\mathbf{s}$)	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.

Internet Appendix For

"What Drives Stock Prices in a Bubble?"

A Omitted Derivations

A.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:

$$-\frac{\partial \mathbf{q}_{i,t}}{\partial \mathbf{p}_t} = \begin{bmatrix} -\frac{\partial \mathbf{q}_{i,t}}{\partial p_t(1)}, & \cdots, & -\frac{\partial \mathbf{q}_{i,t}}{\partial p_t(n)} \end{bmatrix}$$
$$= \mathbf{I} - \beta_{0,i,t} \begin{bmatrix} 1 - w_{i,t}(1) & -w_{i,t}(2) & \cdots & -w_{i,t}(n) \\ -w_{i,t}(1) & 1 - w_{i,t}(2) & \cdots & -w_{i,t}(n) \\ \vdots & \vdots & \ddots & \vdots \\ -w_{i,t}(1) & -w_{i,t}(2) & \cdots & 1 - w_{i,t}(n) \end{bmatrix}, \tag{14}$$

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:

$$\frac{1}{|\mathcal{N}_{i,t}|} \sum_{m \in \mathcal{N}_{i,t}} [1 - \beta_{0,i,t}(1 - w_{i,t}(m))] = 1 - \beta_{0,i,t} + \beta_{0,i,t} \frac{1 - w_{i,t}(0)}{|\mathcal{N}_{i,t}|}. \tag{15}$$

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.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

$$-\frac{\partial \mathbf{q}_t}{\partial \mathbf{p}_t} = \begin{bmatrix} -\frac{\partial \mathbf{q}_t}{\partial p_t(1)} & \cdots & -\frac{\partial \mathbf{q}_t}{\partial p_t(n)} \end{bmatrix}$$
$$= \begin{bmatrix} 1-\frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(1)(1-w_{i,t}(1))}{\sum_{i=1}^I A_{i,t}w_{i,t}(1)} & \frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(1)w_{i,t}(2)}{\sum_{i=1}^I A_{i,t}w_{i,t}(1)} & \cdots & \frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(1)w_{i,t}(n)}{\sum_{i=1}^I A_{i,t}w_{i,t}(1)} \\ \frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(2)w_{i,t}(1)}{\sum_{i=1}^I A_{i,t}w_{i,t}(2)} & 1-\frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(2)(1-w_{i,t}(2))}{\sum_{i=1}^I A_{i,t}w_{i,t}(2)} & \cdots & \frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(2)w_{i,t}(n)}{\sum_{i=1}^I A_{i,t}w_{i,t}(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(n)w_{i,t}(1)}{\sum_{i=1}^I A_{i,t}w_{i,t}(n)} & \frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(n)w_{i,t}(2)}{\sum_{i=1}^I A_{i,t}w_{i,t}(n)} & \cdots & 1-\frac{\sum_{i=1}^I \beta_{0,i,t}A_{i,t}w_{i,t}(n)(1-w_{i,t}(n))}{\sum_{i=1}^I A_{i,t}w_{i,t}(n)} \end{bmatrix} \tag{16}$$

B Details on Data Construction

B.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.

B.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.

B.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:

$$\mathrm{RCP}_{i,t} = \sum_{s \in \{S_{i,t}\}} \left[\gamma_s \left(R^s_{t(s)-K,t(s)} - R^M_{t(s)-K,t(s)} \right) \right], \tag{17}$$

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.

B.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.

B.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 $s_t^p(n)$, and the number of shares held by our retail sample as $s_t^s(n)$. We can compute the multiplier $x_t(n)$ for each stock n at month t:

$$x_t(n) = \frac{s_t^p(n)}{s_t^s(n)}. \tag{18}$$

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 $s_{i,t}^s(n)$. We can therefore reweight to get the population holdings $s^p_{i,t}(n)$ for account i at month t:

$$s^p_{i,t}(n) = s^s_{i,t}(n) x_t(n), \forall i. \tag{19}$$

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

B.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.

B.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]

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.

B.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.

C 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.

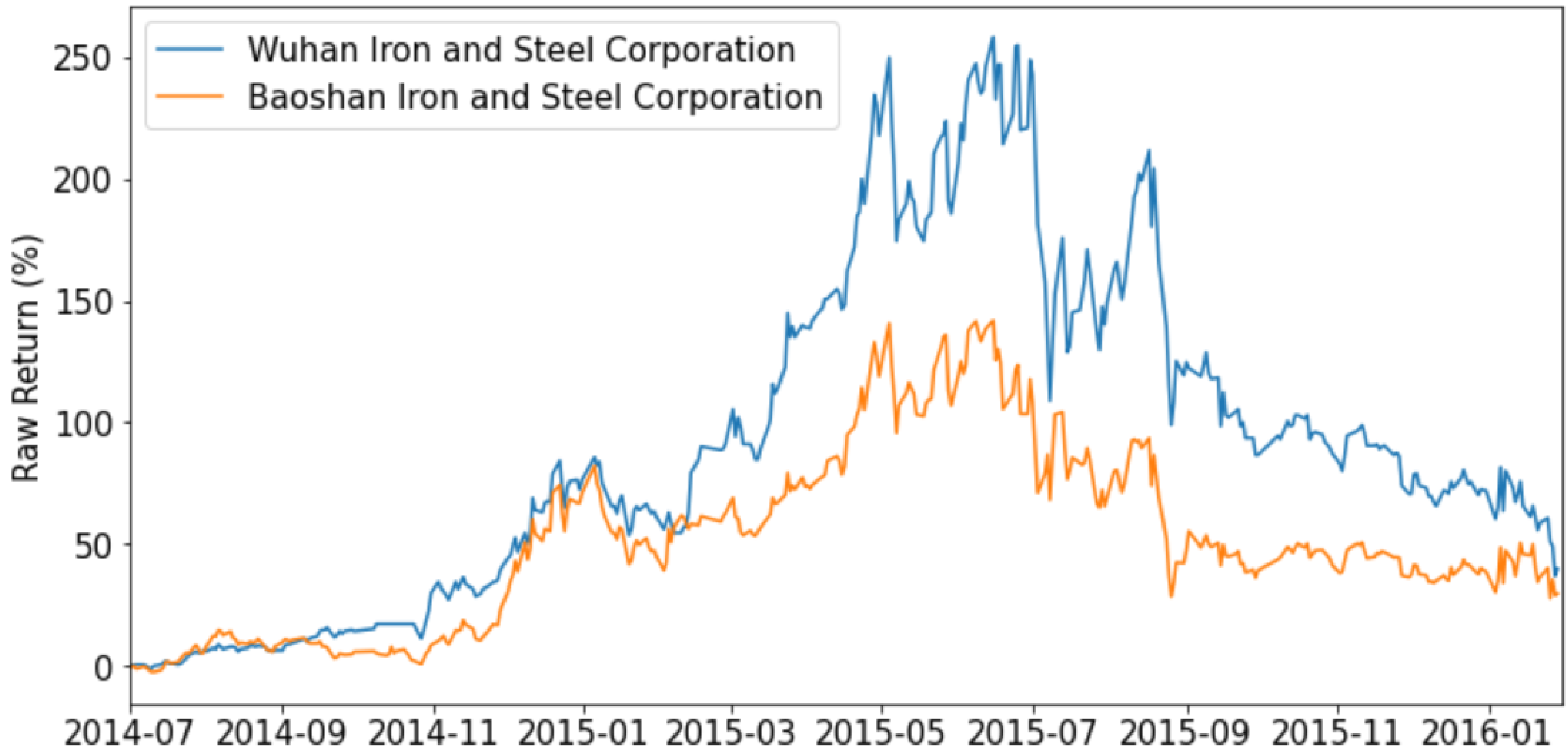
In Appendix Figures A.6 and A.7, we examine the heterogeneity among retail investors by market entry cohort. We see that investors who entered during the formation or expansion phase of the bubble prefer stocks that have lower dividends, are smaller, are younger, and are in the construction industry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

References

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

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

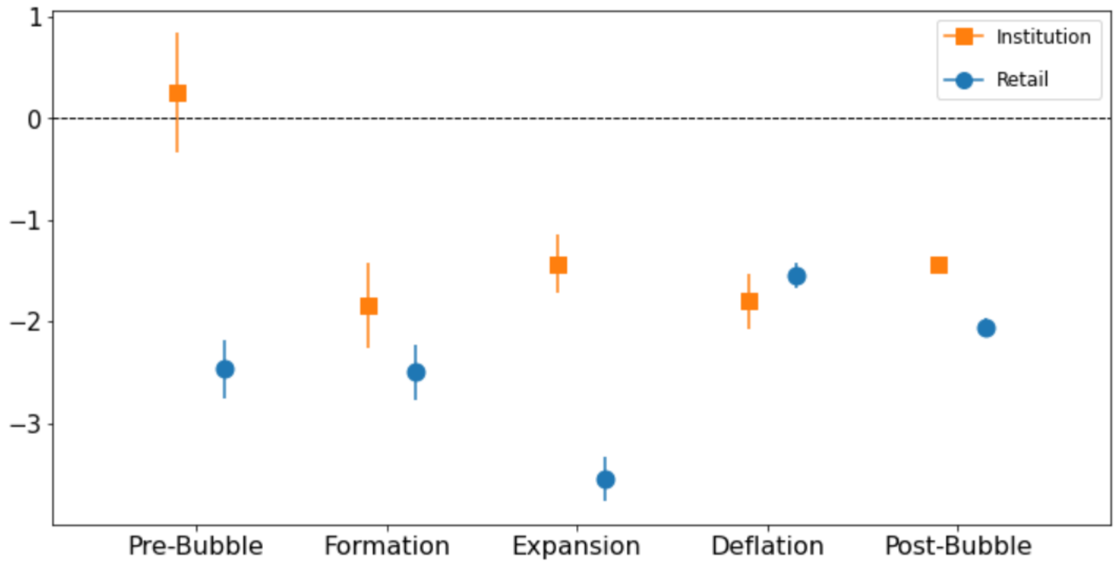
Figure A.1: Cumulative Raw Returns for Two Illustrative Stocks during the 2015 Chinese Stock Market Bubble



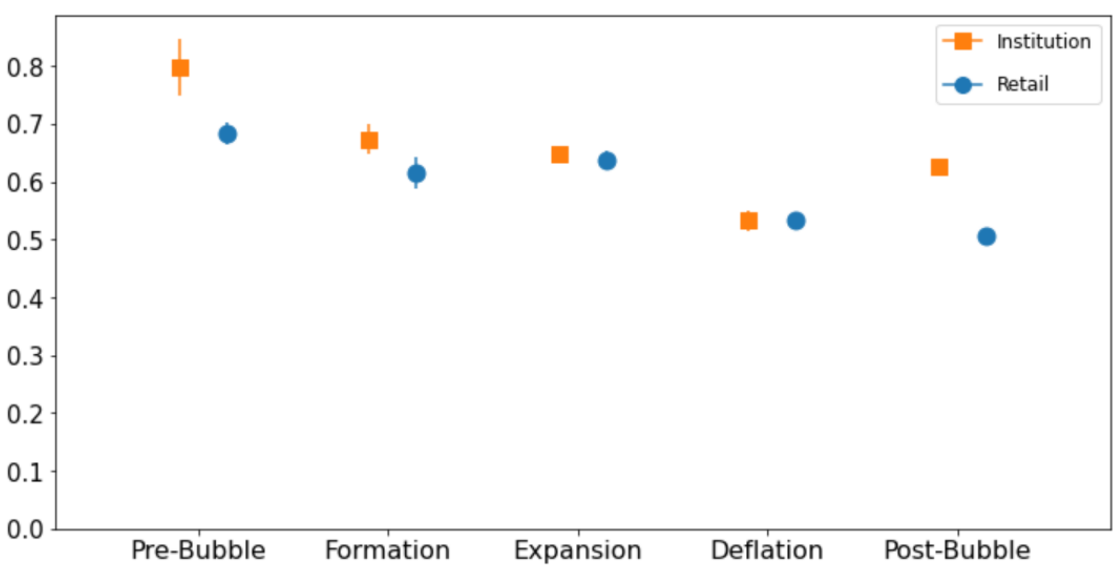
Notes: This figure plots the cumulative raw returns for two illustrative stocks, "Wuhan Iron and Steel Corporation" and "Baoshan Iron and Steel Corporation," during the bubble period.

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

(a) Preferences/Beliefs about Dividend



(b) Preferences/Beliefs about Size



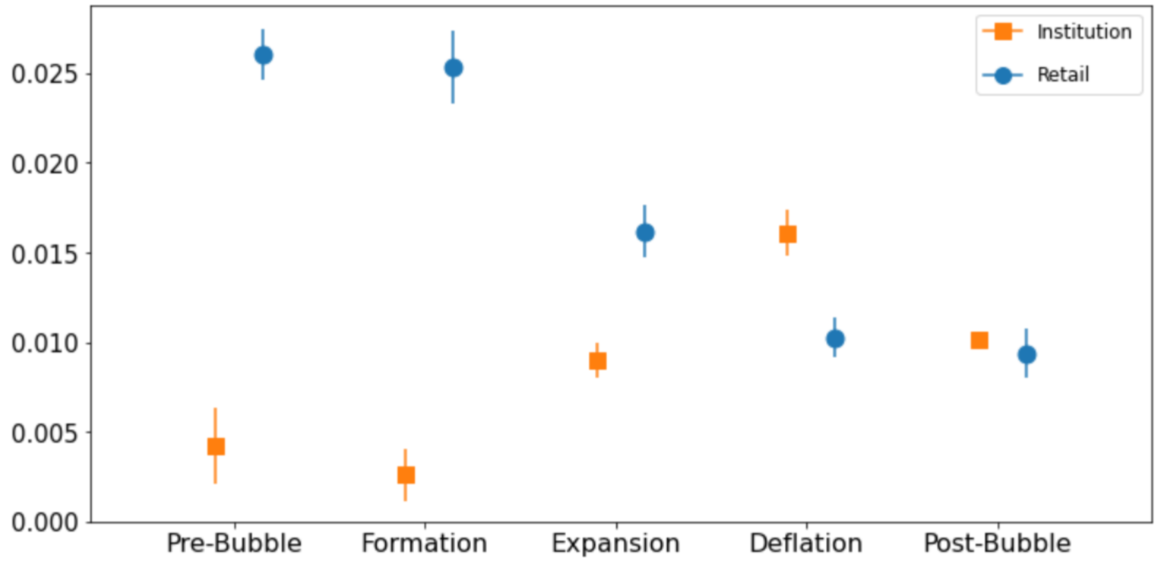
Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_{I,T}$) of retail investors and institutions over different time periods. $\hat{\gamma}_{I,T}$ is estimated by the following regression:

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

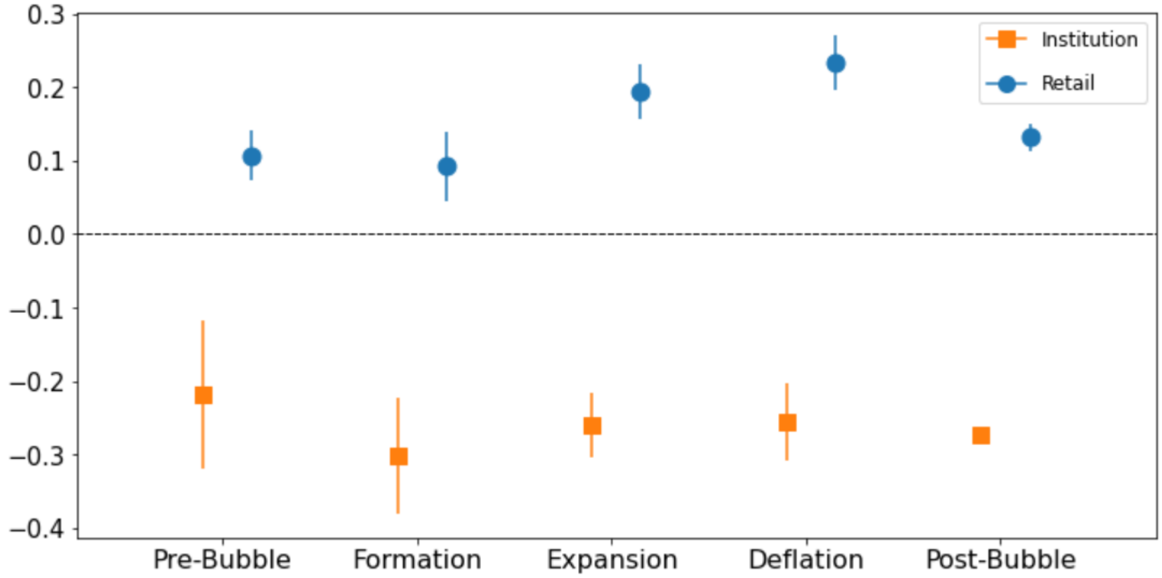
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $I \in \{\text{Retail, Institution}\}, T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable indicating whether investor i belongs to investor type I and time t belongs to time period T. Panel (a) reports investors' preferences/beliefs about 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



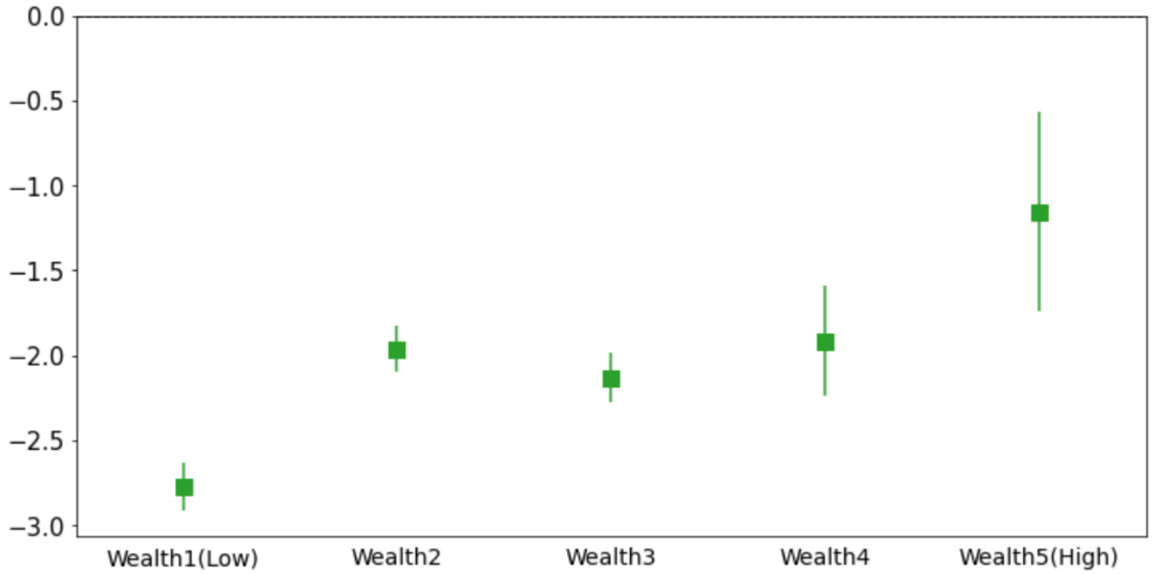
Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_{I,T}$) of retail investors and institutions over different time periods. $\hat{\gamma}_{I,T}$ is estimated by the following regression:

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

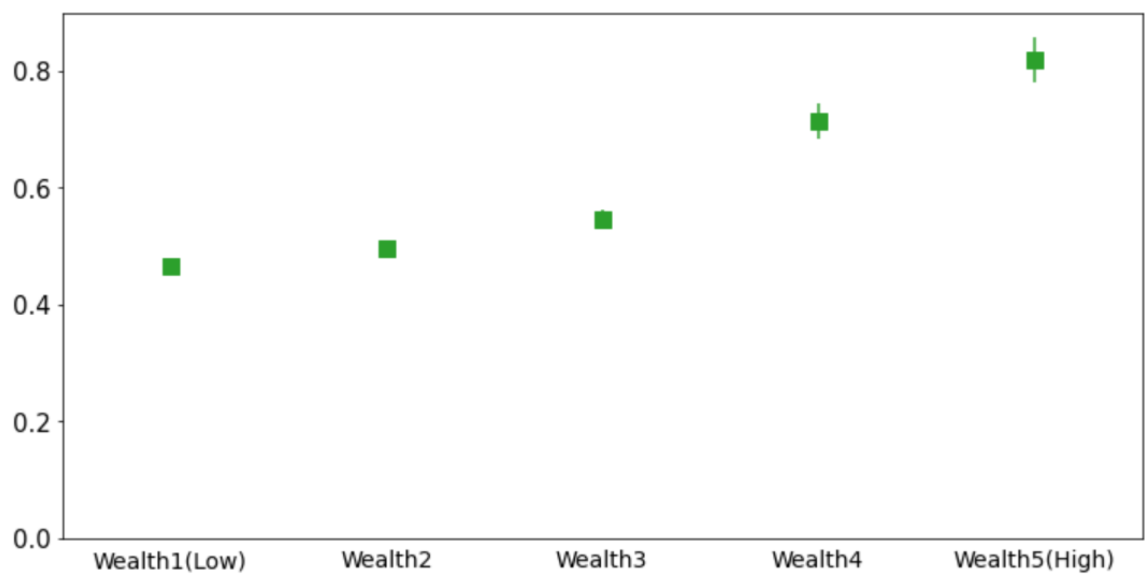
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $I \in \{\text{Retail, Institution}\}, T \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, and $\mathbb{1}\left[i \in I, t \in T\right]$ is a dummy variable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



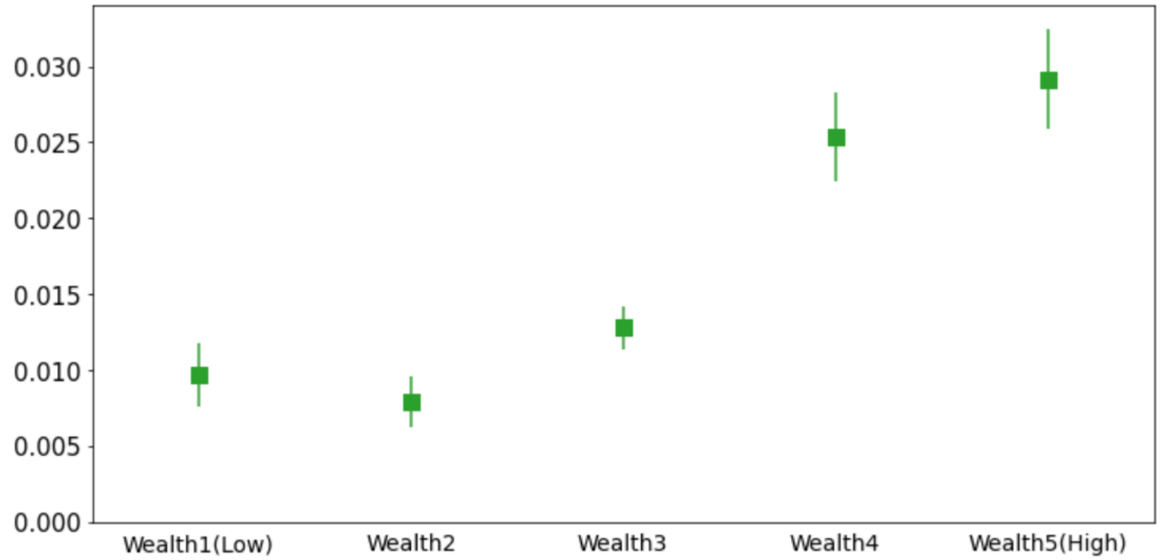
Notes: This figure reports the preferences/beliefs ($\hat{\gamma}_W$) of retail investors with different levels of wealth. We classify retail investors into five groups according to their average stock market wealth in the previous year. $\hat{\gamma}_W$ is estimated by the following regression:

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

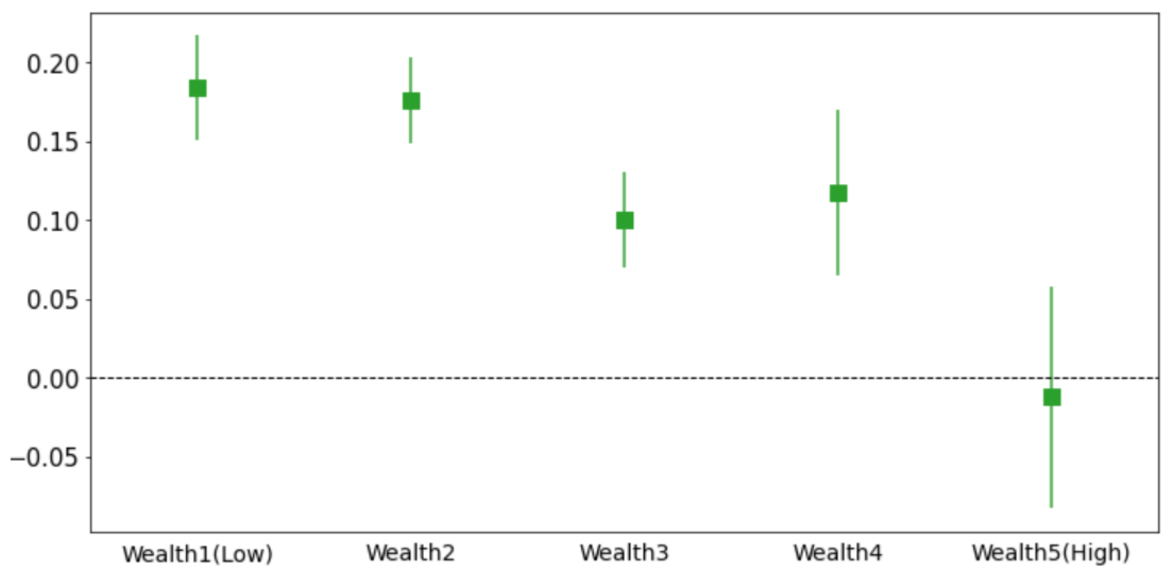
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $W \in \{$Wealth1 (Low), Wealth2, Wealth3, Wealth4, Wealth5 (High)$\}$, and $\mathbb{1}\left[i \in W\right]$ is a dummy variable indicating whether investor i belongs to wealth 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



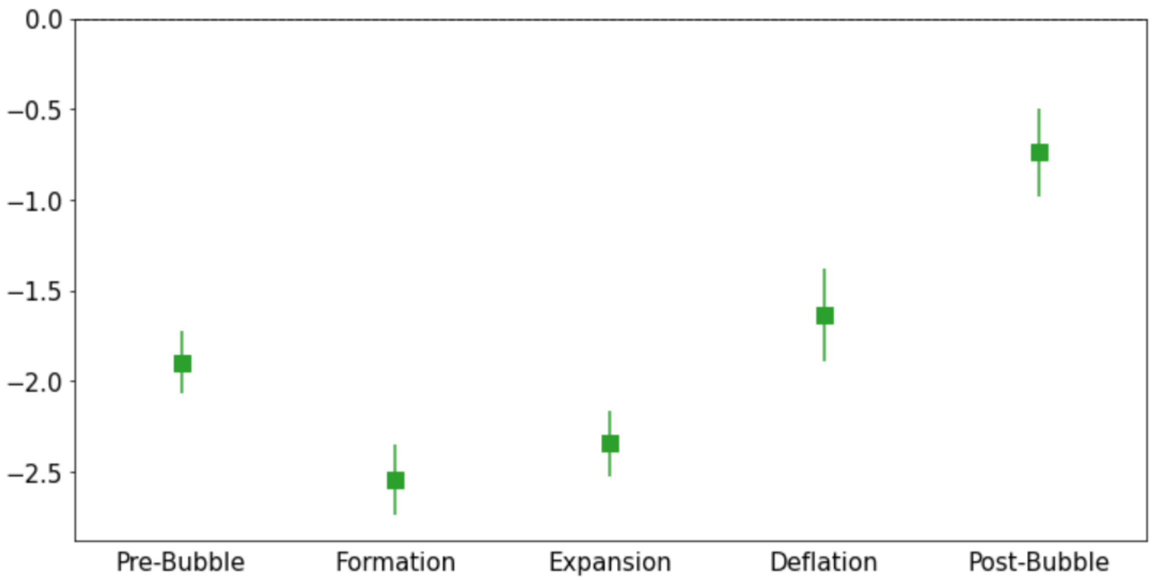
Notes: This figure reports the preferences/beliefs ($\hat{\gamma}_W$) of retail investors with different levels of wealth. We classify retail investors into five groups according to their average stock market wealth in the previous year. $\hat{\gamma}_W$ is estimated by the following regression:

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

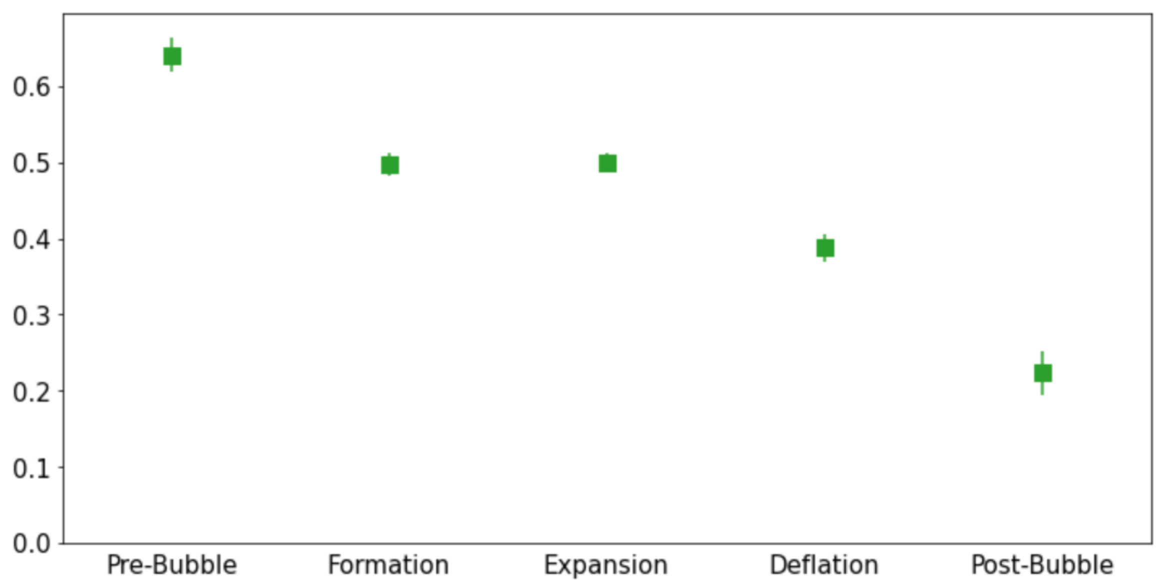
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $W \in \{$Wealth1 (Low), Wealth2, Wealth3, Wealth4, Wealth5 (High)$\}$, and $\mathbb{1}\left[i \in W\right]$ is a dummy variable indicating whether investor i belongs to wealth 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



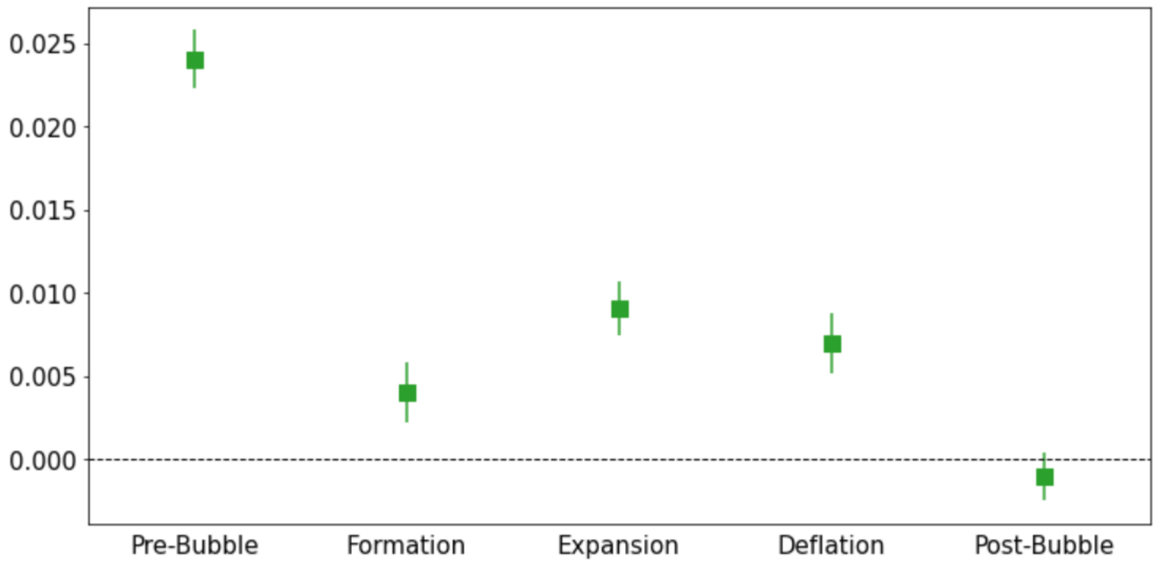
Notes: This figure reports the average preferences/beliefs ($\hat{\gamma}_C$) of retail investors entering the market at different time periods. We classify retail investors by their market entry cohort: the pre-bubble phase, the formation phase, the expansion phase, the deflation phase, and the post-bubble 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:

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

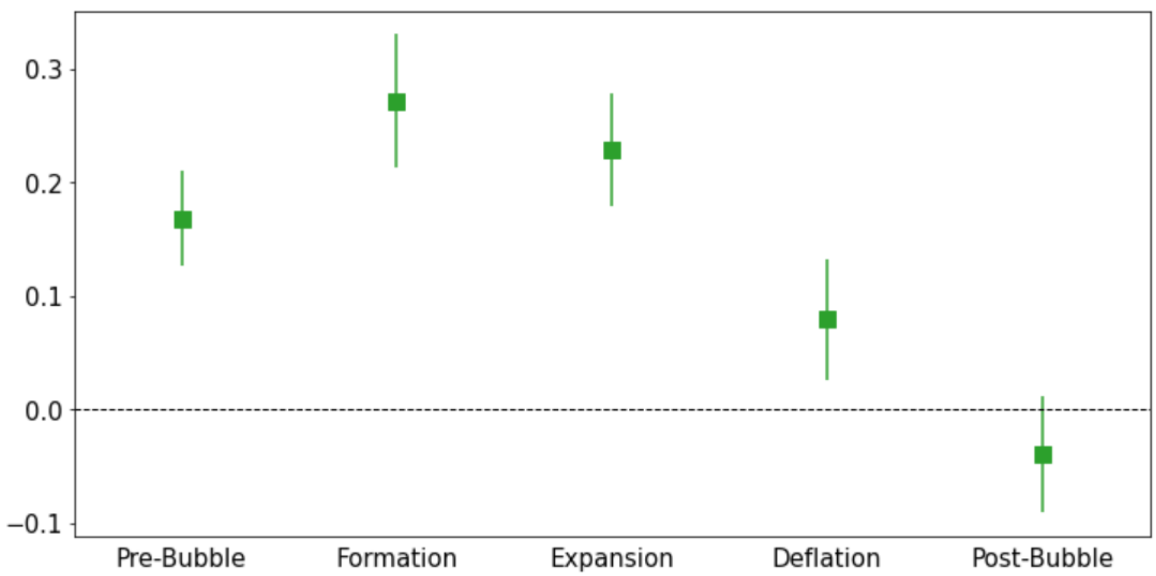
where $\hat{\beta}_{k,i,t}$ is the previously estimated preferences/beliefs about characteristic x_k for investor i at time t, $C \in \{\text{Pre-Bubble, Formation, Expansion, Deflation, Post-Bubble}\}$, $\mathbb{1}\left[i \in C\right]$ is a dummy variable indicating whether investor i belongs to cohort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