When Competition Backfires: Broker Learning and Commission Fees*

Xindan Li†

Baolian Wang‡

Wei Xiong§

Xuewei Yang¶

Yinhong Zhao‖

August, 2025

*We thank Chris James, and seminar participants at Princeton University and Hunan University for their valuable comments. Xindan Li acknowledges financial support from the National Natural Science Foundation of Chian (No. 72495151). Xuewei Yang acknowledges financial support from the National Natural Science Foundation of Chian (No. W2411065).

†School of Management and Engineering, Nanjing University; Jiangsu Key Laboratory of Digital Finance; Institute of New Finance, Nanjing University. Email: xdli@nju.edu.cn.

‡Warrington College of Business, the University of Florida. Email: baolian.wang@warrington.ufl.edu.

§Princeton University. Email: wxiong@princeton.edu.

¶School of Management and Engineering, Nanjing University; Jiangsu Key Laboratory of Digital Finance; Institute of New Finance, Nanjing University. Email: xwyang@nju.edu.cn.

‖Princeton University. Email: williamzyh@princeton.edu.

When Competition Backfires: Broker Learning and Commission Fees

Abstract

This paper examines how learning and competition shape brokerage commission fees, using a plain-vanilla setting from China's brokerage industry. We document three key facts: substantial price dispersion across investors, naïve investors paying higher fees, and widening dispersion over time. To explain these patterns, we develop a dynamic model where brokers compete for naïve and sophisticated investors while learning their types through repeated interactions. The model implies that more broker competition can backfire, raising markups. Using branch-level data, we provide causal evidence supporting this prediction. Counterfactuals show that, in markets with heterogeneous investors, improving financial literacy compresses markups more effectively than broker entry, highlighting the importance of investor education.

Keywords: Competition; Learning; Price Discrimination.
JEL Classification: L10; D40; G24; D12; G41

1 Introduction

Fees paid by consumers for financial services is a central issue in finance. There is extensive evidence showing substantially heterogeneous prices paid by consumers for (nearly) identical financial products or services. For example, the index funds tracking the S&P 500 index offer virtually identical portfolio returns before fees, but the fees vary by an order of magnitude across providers (Hortacsu and Syverson, 2004). In retail deposits, banks pay markedly different rates on the same basic products within and across local markets, and deposit rates adjust sluggishly, leading to persistent cross-bank dispersion (Hannan and Berger, 1991; Drechsler et al., 2017; Egan et al., 2017). For identical mortgage loans (the same borrower type, same loan, same day), the 10th–90th percentile rate gap is about 54 bps (Bhutta et al., 2020).

To explain such substantial price dispersion, the literature has highlighted the roles of consumer naïveté, limited attention, and bounded rationality. With these frictions, producers can shroud attributes and use add-on pricing (e.g., back-end loads and 12b-1 fees in mutual funds; overdraft fees in checking) to target inattentive or naïve consumers (Ellison, 2005; Gabaix and Laibson, 2006; Ellison and Ellison, 2009; Stango and Zinman, 2014). They may also introduce complex structured products whose embedded features and risks are hard to value, enabling large markups and wide price dispersion (Carlin, 2009; Henderson and Pearson, 2011; Célérier and Vallée, 2017).

Many financial services involve repeated consumer–provider interactions, but how repetition affects price dispersion remains unclear. Consumers who initially overlook or misunderstand terms may learn over time and recognize excessive fees, prompting renegotiation or switching that should compress dispersion as relationships mature. Yet search frictions and switching costs make comparison and exit costly, allowing dispersion to persist (Klemperer, 1995; Kim et al., 2003). Even when consumers occasionally pay the fixed costs to search, any price reductions are likely episodic rather than continuous. How these forces net out—and the overall effect of repeated interactions on price dispersion—remains an open empirical question.

In this paper, we study dispersion in brokerage commission fees in China. Since the

early 1990s, China's stock market has expanded, with roughly 100 brokerages serving a large retail base. The core function is straightforward: transmit client orders to the exchange for execution. Because Chinese exchanges use centralized limit-order books, brokers route to a single venue—unlike the complex, multi-venue best-execution architecture in the U.S. Basic execution is therefore relatively homogeneous and transparent, leaving limited scope for obfuscation. Customers pay a flat, ad valorem commission on each trade. The rate is set at account opening—within an allowed range at the broker's discretion—and can be renegotiated later. This setting is well suited to examine how repeated interactions and broker competition shape fee dispersion. Although many brokers have diversified into asset management and other businesses—and commissions have trended downward due to regulation—commissions still account for over one-third of industry revenue, making brokerage fees central to understanding the industry.

We use account-level data from a medium-sized brokerage with branches across China, covering all retail accounts from January 2007 to December 2017. The dataset includes fees and transactions, allowing us to document basic facts about commission-rate dispersion. First, there is substantial variation in the commission rates charged across investors. Second, commission rates are systematically higher for less sophisticated investors: the rate set at account opening predicts subsequent behavioral patterns—disposition effect, lottery preference, under-diversification, local bias, extrapolation, and the rank effect—indicating that initially higher rates disproportionately burden less sophisticated traders later on.

Third, although average commissions fell sharply from about 15 bps in 2007 to about 3 bps in 2017, dispersion increased: the coefficient of variation (standard deviation over mean) rose, and the fee gap between sophisticated and naïve investors widened. Sophistication is measured using a composite index based on the behavioral patterns described above. This widening dispersion challenges the common view that learning and competition should reduce price dispersion over time.

Motivated by these facts, we build a dynamic oligopolistic pricing model, in the spirit of Stahl (1989), to study how investor learning and renegotiation shape commissions when investors repeatedly bargain with brokers. There are $n + 1$ competing brokers and two

types of investors, naive and sophisticated. Naïve investors have a high reservation commission rate and accept any quote at or below that threshold, searching only if a quote exceeds it. Sophisticated investors always solicit quotes from all brokers and choose the lowest. Brokers do not observe investor type ex ante; instead, they infer it from the initial quoting interaction and from subsequent trading activity within the firm, which gives the incumbent broker an informational advantage over rivals.

After opening an account with some broker, each investor receives Poisson opportunities to renegotiate. At such times, the investor requests a new quote from the incumbent and, if dissatisfied, may seek outside quotes. Brokers set commission rates strategically to balance attracting sophisticated shoppers with low prices against extracting rents from naïve clients with higher prices, taking rivals' pricing policies as given.

We characterize the equilibrium and derive several implications. First, a broker's optimal quote is decreasing in its posterior belief that the investor is sophisticated. Second, even when an incumbent learns through repeated interactions that an investor is almost surely sophisticated, it still charges a positive markup because rivals lack the same information; this informational advantage prevents a full race to the bottom. These two implications align with the data: naïve investors face higher fees, and the lower tail of the fee distribution exhibits persistently positive markups.

The third implication concerns the asymmetric effects of broker learning. As brokers gain more information about investor types, they compete more aggressively for sophisticated investors, lowering fees for them but imposing a negative externality on naïve investors, who face higher markups.

The fourth implication is more striking: competition itself can backfire. As the number of brokers increases, rivalry for sophisticated clients reduces their markups, but the declining chance of winning them shifts brokers toward extracting rents from naïve investors. In the limit, with many brokers, each posts the full markup to all investors—so greater competition ultimately raises markups and lowers consumer welfare.

This counterintuitive prediction is a well-known result in models with consumer heterogeneity (Rosenthal, 1980; Stahl, 1989; Carlin, 2009), though its empirical validity has been debated (Chandra and Tappata, 2011; Lach and Moraga-González, 2017). We test

this prediction using branch-quarter data on commission rates for new accounts, measuring competition by the number of brokers operating in a prefecture-level city.

Descriptively, cities with more brokers exhibit lower lower-tail percentiles and higher upper-tail percentiles of commissions, implying wider inter-percentile ranges and greater normalized dispersion. To establish causality, we use a stacked difference-in-differences design exploiting staggered broker entry: treated branches are those experiencing a new entrant after at least a year without entry, matched to controls with similar pre-entry competition and no entry within a two-year window. Entry significantly raises dispersion, with dynamic estimates showing no pre-trends and persistent, gradually strengthening effects. These results confirm the model's prediction that competition can backfire by amplifying price discrimination.

While competition clearly affects price dispersion, repeated interactions make its impact on commission rates non-monotone and heterogeneous. To capture these dynamics, we complement the reduced-form evidence with a structural analysis. We estimate model parameters via simulated method of moments (SMM) and then conduct counterfactuals. The specification allows marginal costs to decline over time (capturing the secular fall in fees) and incorporates branch-level heterogeneity in investor sophistication and search costs. Targeted moments include the frequency of commission adjustments, extreme-quantile commissions, branch-level means and their covariances with trading volume, and the dispersion–tenure relationship. The estimated parameters are economically plausible and reproduce key spatial and temporal patterns.

Using the estimated model, we conduct counterfactuals to assess how competition and investor composition shape fees and markups. Broker entry does not uniformly benefit consumers: when many investors are naïve, additional entry raises their markups even as sophisticated investors gain, producing heterogeneous and non-monotonic effects. The fee-minimizing number of brokers is finite for realistic sophistication shares and almost always below observed levels, indicating excessive rather than insufficient competition. By contrast, investor education—raising the share of sophisticated investors—consistently compresses markups across the market and delivers larger fee reductions than plausible changes in market structure. These findings suggest that policies enhancing financial

literacy may be more effective than pro-entry interventions in lowering commissions, despite the latter's conventional appeal for financial inclusion.

2 Background and Data

2.1 The Chinese Stock Brokerage Market

We study the retail brokerage market in China, which consists of about 100 brokerage firms providing essentially identical services. These firms receive client orders and route them to either the Shanghai or Shenzhen Stock Exchanges.[1] Chinese stock exchanges operate using limit order books, matching nearly all trades directly through the exchanges. Brokers charge a flat commission rate proportional to the RMB trading volume, with a minimum of 5 RMB. Both buyers and sellers are subject to these commissions.

During our sample period (2007-2017), the regulation of brokerage commissions remained unchanged. Since May 1, 2002, regulations have capped commission rates at 30 bps and set a minimum equal to the combined fees charged by the China Securities Regulatory Commission (CSRC) and the respective exchange. At the start of our sample period, these fees totaled approximately 1.6 bps, decreasing to 0.7 bps by its end.[2] Brokerage firms collect these fees on behalf of the CSRC and the exchanges as part of the overall brokerage commissions.

Brokerage firms typically charge different fees to different clients. At the time of account opening, they provide a quoted a fee rate, often based on information submitted through account applications or an initial meeting between the account manager and the investor. Investors may negotiate this rate if they are attentive to it and willing to bargain.

[1] Since 2014, retail investors have been able to trade certain stocks listed on the Hong Kong Stock Exchange through the Shanghai-Hong Kong connect Program and the Shenzhen-Hong Kong Connect Program. However, trading these stocks differs from trading domestically listed stocks, as brokerages charge different fees. Additionally, these trades constitute only a small fraction of all activity and are excluded from our analysis. The third stock exchange in mainland China, the Beijing Stock Exchange, commenced trading on November 15, 2021, which falls outside our sample period.

[2] The CSRC fees changed once during our sample period. Before January 1, 2012, the CSRC charged 0.004%; it was reduced to 0.002% thereafter. Exchange fee rates also evolved. Before June 1, 2012, the Shanghai Stock Exchange charged 0.0110% and the Shenzhen Stock Exchange charged 0.0122%. Starting June 1, 2012, both exchanges adopted a uniform rate, which was adjusted multiple times: to 0.0087% on June 1, 2012, 0.00696% on September 1, 2012, and 0.00487% on August 1, 2015.

Brokerage firms do not voluntarily disclose fee rates to the public. However, investors may learn that others pay lower fees, either at the same firm or at competing firms. At any time, an investor can renegotiate their fee rate by contacting the brokerage manager, either via phone or an office visit. The investor may request a lower rate by threatening to move their account elsewhere. In response, the brokerage firm decides whether to accommodate the request and, if so, by how much. If the investor remains dissatisfied, they can transfer their account to another broker. Such transfers usually take one to two business days and are typically free of charge.

Securities law in China prohibited investors from holding more than one account until 2015. Even after this restriction was lifted, few investors chose to open a second account. Over our eleven-year sample period, we observe a very small fraction of account disappearances in our broker data, fewer than 30,000 accounts, or about 4% of the full sample, which we next describe.[3]

2.2 Data and Descriptive Analysis

We use an account-level data set from a security brokerage company in China. Our data set is very similar to the Chinese data set used by other researchers (Feng and Seasholes, 2004, 2005; Frydman and Wang, 2020; An et al., 2024), as well as to the U.S. data set used by Odean (1998). The data come from a medium-sized brokerage company that has multiple branches throughout China. Brokerage revenue accounts for approximately 40% of its total revenue in the sample period, which is typical in China.[4]

The data set consists of three files: a trade file, a position file, and an investor demographic file. Our sample period is from January 2007 to December 2017. The trade file includes details such as investor identification, trade date and time, asset symbol, buy/sell indicator, trade price, number of shares traded, and, crucially for our analysis, the commissions (in RMB). The position file contains daily position information for each account.

[3]Accounts can disappear from our data for two reasons: either the investor stops trading and cancels their account, or the investor transfers their account to a different broker. Our data do not distinguish between these two scenarios.

[4]To protect the broker's anonymity, we do not disclose the exact number of investors served by the broker. The reported empirical analysis is based on a randomly selected sample. We confirm that the results obtained from the full dataset are very similar.

The demographic file contains information on investor age and gender. We restrict our analysis to the trading of common stocks. We exclude short sales and margin trading.

Table 1 reports the summary statistics for the sample, which consists of over 700,000 investors. The average investor age is 45.17 years, and 53.5% are male. As of the end of the sample period (December 2017), the average account age is 9.72 years. The 75th percentile of account age is 14.08 years, indicating that 25% of investors have been with the broker for more than 14 years. While the median log portfolio size is 10.07—corresponding to roughly 23,600 RMB—reflecting a prevalence of small investors, more than 10% of investors hold portfolios exceeding 200,000 RMB. The average annual portfolio turnover is 722.5%, consistent with figures reported in prior literature (e.g., Frydman and Wang 2020; Liu et al. 2022). The last row of the table summarizes commission rates as of December 2017. The average rate is 14.43 basis points, with substantial cross-sectional variation: the standard deviation is 10.69 bps; the 75th percentile is 28 bps, and the 25th percentile is 3 bps—a nearly tenfold difference.

Relative to the market, our broker serves disproportionately more small investors. Based on the yearbook published by the Shanghai Stock Exchange, at the of 2017, the shares of investors with a portfolio smaller than 0.1 million RMB, between 0.1 and 0.5 million, between 0.5 and 1 million, between 1 and 3 million, between 3 and 10 million, and bigger than 10 million are 55.28%, 30.11%, 7.23%, 5.08%, 1.58%, and 0.50%, respectively. The respective shares for our broker are 72.31%, 21.82%, 3.42%, 1.78%, 0.43%, and 0.11%, respectively. We confirm that the difference is due to our broker having more branches in relatively underdeveloped areas than the market. As we will show later, our main results hold qualitatively similar if within investors with similar portfolio size, suggesting that the difference in investor composition is unlikely to have an important effect on our conclusions.

Fig. 1 illustrates the average commission rates of investors in our sample (dotted line) from 2007 to 2017. The figure also includes the average commission rates for the broader market (solid and dashed lines) as benchmarks. For the brokerage firm covered by our data, only stock trading is considered. To provide context, we collect total commission revenue for all brokerage firms from the annual reports of the Securities Association of

China. However, this data does not separate stock trading commissions from those for other securities, such as funds, bonds, and options/warrants. As a result, we calculate two ratios: (1) dividing total broker commissions by the aggregate trading volume across all securities (solid line), and (2) dividing total commissions by total stock trading volume (dashed line). These two ratios serve as upper and lower bounds for the market's average commission rate. As shown in Fig. 1, the solid and dashed lines are very close, which is expected, as stock trading accounted for approximately 93% of total trading volume on these exchanges during our sample period.

The figure reveals a clear downward trend in commission fees. The market average commission rate declined from about 15 bps in 2007 to 3 bps in 2017. Similarly, the average commission rate in our sample followed this trend, dropping from approximately 24 bps to slightly above 5 bps, likely reflecting decreasing costs of order handling and market competition. We will account for this trend in our empirical analysis. Notably, the brokerage in our data charges a slightly higher commission rate, likely because it serves disproportionately larger number of small investors.

3 Stylized Facts

In this section, we present several stylized facts about price discrimination in commission fees. Specifically, we address the following questions: How significant is price discrimination in fees? Who is most affected? And how does the extent of price discrimination evolve over the course of the investor–broker relationship? These facts provide key insights that inform the construction of our model.

3.1 Dispersion of Commission Rates

The dispersion in commission rates reflects price discrimination among different investors. In Fig. 2, we present the distribution of commission rates across investors. To calculate the commission rate for each investor in a given month, we use their trades during that month, excluding trades with a commission of 5 RMB, the regulatory floor for commission fees.

Panel A includes all investors, while Panel B focuses on new investors—defined as those who opened an account within that month. For each month, we report several percentiles of commission rates across investors, including the 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, and 99% percentiles. Additionally, we show the regulatory upper and lower bounds. The upper bound is 30 bps, while the lower bound equals the sum of the CSRC fee and the exchange fee.

Two major observations emerge from Fig. 2. First, it is evident that different investors pay significantly different commission rates. At the beginning of our sample period, the majority of investors paid 30 bps, the highest allowable rate. By the end of our sample period, about a quarter of investors were still paying this maximum rate, while the first percentile of commission rates had dropped to approximately 2.5 bps–less than 10% of the highest allowable rate.

Restricting the analysis to new accounts reduces the dispersion, but does not eliminate it. For example, at the end of the sample period, some new investors were still paying the highest rate, and the median rate (5 bps) remained double the first percentile.

Second, commission rates are consistently higher than the lowest allowable commission rate, which equals the fees charged by the CRSC and the exchanges. Since all orders are submitted to exchanges electronically, it is reasonable to assume that the marginal cost of submitting an order is equal to these fees. The fact that commission rates exceed this lower bound suggests that competition does not drive brokers to a race to the bottom. Instead, Chinese brokers maintain substantial markups.

Anecdotally, most brokers consider an investor's portfolio size and trading volume when setting commission rates. This is because larger accounts typically make larger trades, generating more fees. In Fig. 3, we present the distribution of commission rates based on an investor's portfolio size or trading volume. To stratify investors, we use the average month-end portfolio size and the average monthly trading volume over the previous six months. For portfolio size (or trading volume), we divide the sample into three groups using cutoff points of 0.2 million and 0.5 million.

As expected, investors with bigger portfolios or higher trading volumes tend to pay lower commission rates, as reflected in smaller percentile values. However, even among

investors with similar portfolio sizes or trading volumes, there remains significant dispersion in commission rates. This suggests that the observed variation in rates cannot be attributed to differences in investor size or trading intensity.

These results are difficult to reconcile with simple competition models. Chinese brokers route orders directly to exchanges, offering a highly homogeneous service. The marginal cost of routing orders is minimal, aside from regulatory fees. Under such conditions, one might expect competition to eliminate both dispersion and markups, with brokers charging rates close to the regulatory minimum. However, the data clearly does not support this expectation.

3.2 Naivete-based Discrimination

In this subsection, we examine the role of investor sophistication in explaining the dispersion of commission rates, specifically investigating whether brokers charge higher rates to less sophisticated investors. We measure investor sophistication using a standard set of prominent behavioral patterns in trading that have been well-documented in the finance literature (Barberis and Thaler, 2003; Barber and Odean, 2013; Liu et al., 2022). These biases include the disposition effect, lottery preference, under-diversification, local bias, extrapolation, and the rank effect. Since these patterns are generally considered suboptimal and most evidence suggests they negatively impact investor performance, we refer to them as biases, following the finance literature. An investor exhibiting a higher degree of these biases is deemed less sophisticated. Details on the computation of these measures are provided in Appendix IA.1.

We present the summary statistics of the aforementioned biases and the initial commission rates in Table 2, where each observation corresponds to an individual investor. The variable *IniRate* represents the initial commission rate assigned to an investor, measured in basis points. Our analysis explores the relationship between this initial rate and the behavioral biases exhibited by investors. Specifically, we analyze whether the initial commission rate set by brokers can predict the biases displayed by investors in their subsequent trading behavior.

Biases are measured for each investor individually. The last row of Table 2 reports the

summary statistics for a composite bias measure. This composite measure is constructed by averaging the discretized percentile ranks (ranging from 1 for the lowest to 100 for the highest) of the six behavioral bias variables. If a bias is missing for an investor, the composite measure is calculated by averaging across the remaining non-missing bias variables. To ensure accurate measurement, biases are calculated using data from the second month after an investor opens their account through the end of the sample period.[5] Consistent with prior literature, our sample investors exhibit strong behavioral biases.

Table 3 reports the regressions, where the unit of analysis is an investor. The dependent variables are the investor biases discussed earlier, and the key independent variable is *IniRate*. All regressions include year-month fixed effects and branch fixed effects, with standard errors clustered by year-month. Panels A and B differ in their control variables. Specifically, we control for portfolio size, investor age, and gender. Portfolio size is measured using data from the month immediately following account opening (i.e., month $t + 1$).

The results show that, for all biases except local bias and extrapolation, the coefficients on *IniRate* are highly significantly positive in both panels. For local bias or extrapolation, however, the coefficients on *IniRate* are statistically insignificant. Overall, the results in Table 3 support the hypothesis that brokers charge higher commission rates to less sophisticated investors.

3.3 Evolution of Commission Fees

As noted earlier, there is a clear downward trend in commission fees during the sample period. It is reasonable to hypothesize that the fee rate for each investor decreases over time. Specifically, given an initial fee rate, the fee rate is expected to decline as the broker-investor relationship lengthens. Moreover, since sophisticated investors have lower search costs and greater bargaining power with brokers, we anticipate a steeper decline in fee rates over time for sophisticated investors compared to their less sophisticated

[5]Our results are robust to alternative measurement windows: biases calculated using data from the first year (or the first two years) after account opening yield similar findings. These robustness results are reported in Tables IA.1 and IA.2 in the Internet Appendix.

counterparts.

We first test the prediction that commission rates decline more for sophisticated investors. To do so, we compare changes in commission rates within cohorts. Specifically, we group investors who open their accounts in the same month as a cohort and track each cohort from the month they open their accounts until the end of our sample period. As discussed earlier, investors rarely disappear from our data, resulting in minimal sample attrition when tracking investors over time. For each investor, we use their first year of trading to estimate their biases, as defined in Appendix IA.1. Within each cohort, we sort investors into two groups based on the median value of their composite bias measure, as defined in Table 2.

Fig. 4a illustrates the dynamics of commission rates for both groups of investors. For each cohort, we calculate the percentage change in average commission rates relative to the initial rates within each group. The reported values represent the averages across all cohorts. Alongside the percentage changes for both groups, we also display the difference between the two groups, with the scale for this difference shown on the right y-axis.

Fig. 4a shows that commission rates have declined for both groups, consistent with the trends observed in Fig. 1. Ten years after account opening, commission rates decrease by 40% for the less sophisticated group and 46% for the more sophisticated groups. Additionally, the line representing the differences indicates a steady widening between the two groups over time.[6]

The results indicate that, compared to less sophisticated investors, more sophisticated investors not only start with lower commission rates but also receive greater downward adjustments over time. Consequently, we expect the dispersion of commission rates to increase over time.

Fig. 4b illustrates the dynamics of this dispersion. For each cohort and each month, we calculate the dispersion as the cross-sectional standard deviation of commission rates, scaled by the mean commission rate. The scaling ensures comparability across time, given

[6]As an alternative approach, we compute the ratio between the two groups. The results, presented in Fig. IA.1 of the Internet Appendix, show that the ratio is less than one in the first month, indicating that less sophisticated investors face higher commission rates at account opening. This finding aligns with the results in Table 3. Over time, the ratio decreases from 0.95 to 0.87.

the downward trend in average commission rates (see Fig. 1). For each cohort, we then compute the percentage change in dispersion since account opening. Finally, we group all cohorts by event month, setting the measure to zero for the first month. Fig. 4b reports the averages across cohorts, along with the 95% confidence intervals. The figure reveals a clear upward trend, indicating that the dispersion in commission rates increases as investors remain with the broker for longer periods.

Overall, our findings reveal that commission rates for more sophisticated investors decline more significantly over time compared to those for less sophisticated investors. This greater decline likely reflects a stronger awareness of commission fees, more active search efforts, and greater bargaining power among sophisticated investors. However, despite the larger decline in fees for sophisticated investors, the magnitude of the difference is modest—-46% versus 40% after more than 10 years—-across the more and less sophisticated groups. This suggests that even for sophisticated investors, an initially high fee may not be easily bargained down.

Fig. 2a highlights this observation even more clearly. While the lower percentiles of all accounts—such as the 1st and 5th percentiles—experienced significant declines from 2007 to 2017, they remained well above the regulatory lower bound. This demonstrates that even the most sophisticated accounts in our sample were unable to bargain their fee rates down to the minimum.

4 Model

In this section, we introduce a model that captures the stylized facts outlined in Section 3. Specifically, the model seeks to reconcile the following observations: (1) Brokers offer different commission rates to different investors, with rates consistently exceeding marginal costs (Fig. 2a); (2) Investors who trade more frequently tend to benefit from lower commission rates (Fig. 3); (3) Brokers charge higher commission rates to less sophisticated investors (Table 3); (4) Over time, commission rates decline more rapidly for sophisticated investors compared to their less sophisticated counterparts (see Fig. 4a). However, even sophisticated investors are unable to bargain their commission rates down to the

minimum, even after an extended period (Fig. 2a).

Our model builds on the classic oligoplistic pricing framework of Stahl (1989), incorporating investor heterogeneity and dynamic negotiations between brokers and investors. Investors differ in their levels of sophistication, which are captured by their search costs for switching to an alternative broker. While brokers cannot directly observe an investor's true type, they can infer it over time through repeated interactions. In equilibrium, brokers strategically offer lower rates to investors perceived as more sophisticated and higher rates to those perceived as naive.

The model also highlights a novel form of information asymmetry that emerges endogenously from long-term relationships. Incumbent brokers gain a deeper understanding of their existing investors' types through repeated interactions. As a result, even after fully learning an investor's type, an incumbent broker need not offer competitive rates to sophisticated investors, knowing that competitors cannot accurately identify them as such.

4.1 Model Setup

There are two types of investors: smart investors (denoted by s) and naive investors (denoted by d). It is publicly known that a portion $\bar{\pi}$ of the population is smart, while the remaining $(1 - \bar{\pi})$ are naive. The market consists of $n + 1$ brokers, all of whom serve both types of investors and are ex ante homogeneous. These brokers provide identical services and share the same marginal cost MC for executing orders for any investor. Our analysis focuses on a representative broker, referred to as "incumbent broker", while treating the other n brokers as competitors. By symmetry, all brokers adopt the same equilibrium strategies.

Time is continuous with $t = [-t_0, \infty)$. We divide the timeline into two stages: the entry stage ($-t_0 \leq t \leq 0$) and the trading stage ($t > 0$). When a new investor i arrives at $t = -t_0$, she randomly approaches one of the brokers to request a quote on commission rates. The broker responds with a quote at $t = 0$. The investor commits to the broker if the quoted rate is less than or equal to her reservation price; otherwise, the investor continues to seek quotes from other brokers.

We emphasize that the two types of investors differ in their search costs for brokers. Specifically, each type of investors has a reservation price for commissions:

$$r^s < MC < r^d,$$

where r^s applies to smart investors and r^d to naive investors. The difference in their reservation prices ultimately reflects the variation in their search costs.[7]

To make the analysis more compelling, we assume $r^s < MC$, which implies that smart investors will always continue searching for an alternative broker as long as the incumbent does not offer the minimum rate of MC. Consequently, a smart investor will eventually seek quotes from all brokers before committing to the one offering the lowest rate. In contrast, our assumption $MC < r^d$ suggests that naive investors may stop searching for an alternative broker if the incumbent offers a rate lower than r^d.[8]

Intuitively, both reservation prices decrease with the per-unit-of-time trading flow q, as investors who trade more frequently are more motivated to secure competitive rates. Since q is an investor characteristic that brokers can easily observe (e.g., through account size), we treat r^d and r^s as exogenous parameters in our model, publicly known to the investor and brokers. In our quantitative analysis, we parameterize the reservation price of naive investors as a decreasing function of q:

$$r^d = MC + \frac{A\kappa}{q^{1-m}}, \tag{1}$$

where $\kappa > 0$ is a constant representing base level search cost, $A > 0$ is a constant representing market environment, and $m \in [0, 1]$ is the load on trading volume. We provide a micro-foundation of this parameterization in Appendix IA.2.

[7] As demonstrated in the data, there is a downward trend in MC over time. In the model, we treat MC as exogenous and normalize it to remain constant. However, in the quantitative analysis, we model MC as a deterministic function that decreases over time.

[8] Note that when $r^d < MC$, brokers cannot sustain an equilibrium where naive investors do not search, leading the market to revert to a competitive benchmark characterized by Bertrand competition. Conversely, when $r^s > MC$, brokers can set a price at which both smart and naive investors choose not to search, resulting in the paradoxical scenario analyzed by Diamond (1971), where brokers impose monopoly pricing on all investors. In either case, both types of investors are treated identically. Therefore, we focus on the more interesting case of $r^s < MC < r^d$, where smart and naive investors are treated differently, providing an explanation for the observed price discrimination in the data.

After opening an account with a broker during the trading stage ($t > 0$), an investor may periodically pay attention to the commission rate at a Poisson rate α, which is identical for both types of investors. Upon paying attention, the investor initiates a process to renegotiate the commission rate with the broker by requesting a new quote. If dissatisfied with the quote, the investor may seek competing quotes from other brokers and ultimately move her account to the broker offering the lowest quote.

4.1.1 Information Structure

Our model features a key information asymmetry between the incumbent and competing brokers regarding an investor's type. While all brokers symmetrically receive a noisy signal about a new investor's type when providing a quote, the situation differs for existing investors (those already in the trading stage). The incumbent broker gains an informational advantage by continuously observing the investor's trading record up to the current time t. In contrast, a competing broker can only obtain a limited and opaque signal during their first interaction with the investor. Additionally, we assume that competing brokers cannot distinguish between old and new investors when responding to a request for a quote.

We define the type of an investor as $\theta = 1$ if the investor is smart, and $\theta = 0$ if the investor is naive. The type θ is unobservable to any broker. Following Scheinkman and Xiong (2003), we assume that an incumbent broker observes a stream of information about θ:

$$dy_{jt} = \theta dt + \sigma dW_{jt},$$

where W_{jt} is a Brownian motion representing signal noise, and $\sigma > 0$ indicates the magnitude of the noise. With this continuous flow of signals, the broker's inference of the investor's type improves over time. We define the broker's conditional belief at time t about θ as

$$\pi_{jt} = \Pr\big(\theta = 1 | \{y_{js}, s \leq t)\}\big) \in [0, 1],$$

where the subscript j refers to broker j. In this setting, the incumbent broker observes the information flow up to time t.

When a new investor arrives at the entry stage, all brokers observe y_{jt} over the period from $-t_0$ to 0 before responding to a quote request. The shorter observation period t_0 during the entry stage implies that brokers' information about a new investor's type is noisier. When an existing investor seeks a quote from a competing broker, the contacted broker cannot distinguish the old investor from a new one. As a result, the competing broker obtains a noisy signal equivalent to that for a new investor, with the quality of the information determined by the length of the entry stage t_0.

For this binary state with continuous information flow and Gaussian noise, the optimal filter is the Wonham filter (Wonham, 1964; Hansen, 2007), which updates over time as follows:

$$d\pi_{jt} = \frac{\pi_{jt}(1-\pi_{jt})}{\sigma^2}\big(dy_{jt} - \pi_{jt}dt\big). \tag{2}$$

Here, the term $(dy_{jt} - \pi_{jt}dt)$ represents the innovation derived from the most recent signal. The factor $\pi_{jt}(1-\pi_{jt})$ implies that learning is most sensitive to new information when the broker's belief about the investor type is more uncertain (i.e. π_{jt} near 0.5).

The initial condition $\pi_{j,-t_0}$ is the prior of each broker meeting a smart investor. Two sources of selection are in play: a) smart investors search all $n+1$ brokers, while naive investor only searches one broker, so each broker is more likely to see smart investors than the population average; b) among the existing investors, only the smart ones search, so all searching existing investors must be smart. Therefore, we have

$$\pi_{j,-t_0} = (1-\lambda)\frac{(n+1)\bar{\pi}}{1+n\bar{\pi}} + \lambda,$$

where $\bar{\pi}$ represents the proportion of smart investors in the population, $n+1$ represents the total number of brokers, and λ represents the proportion of existing investors who are searching for brokers among all searching investors. This implies that brokers always begin the learning process from this common prior.

The stochastic differential equation (SDE) for π_{jt} can be linearized via a change of variables, implying that π_{jt} follows a logistic-normal distribution. Let $x_{jt} = \ln\left(\frac{\pi_{jt}}{1-\pi_{jt}}\right)$. By

Ito's Formula,

$$dx_{jt} = \begin{cases} -\frac{1}{2\sigma^2}dt + \frac{1}{\sigma}dW_{jt} & \text{if } \theta = 0, \\ \frac{1}{2\sigma^2}dt + \frac{1}{\sigma}dW_{jt} & \text{if } \theta = 1. \end{cases}$$

This means that

$$x_{jt} \sim \begin{cases} N\left(x_{-t_0} - \frac{t+t_0}{2\sigma^2}, \frac{t+t_0}{\sigma^2}\right) & \text{if } \theta = 0, \\ N\left(x_{-t_0} + \frac{t+t_0}{2\sigma^2}, \frac{t+t_0}{\sigma^2}\right) & \text{if } \theta = 1. \end{cases}$$

Thus, $\pi_t = \frac{e^{x_{jt}}}{1+e^{x_{jt}}}$ follows a logistic normal distribution conditional on the investor's type. We denote this distribution as F_s and F_d for smart and naive investors, respectively.

4.1.2 Broker's Problem

When faced with a quote request from an investor—whether a new investor seeking an initial quote or an existing account holder—the broker bases its quote on its belief π about the investor's type. The broker anticipates that a naive investor will not seek competing offers as long as its quote remains below the naive investor's reservation price, r^d.[9] In contrast, the broker expects a smart investor to seek competing offers from the other n brokers.

Each time a smart investor approaches a broker, the broker receives a new signal about her type, influencing the quote she is offered. Consequently, each broker understands that the probability of retaining a smart investor depends on the quoted price p. While a higher price reduces the probability of retaining the investor, it increases the payoff if the investor is retained, creating a trade-off the broker must consider.

Given a broker's belief π for an investor to be smart, the expected value of the investor to the broker is

$$V(p, MC; \pi) = \mathbb{E}_\pi \left[\int_0^\infty e^{-\rho t}(p(t) - MC)q \, dt \,\middle|\, p(0) = p \right], \tag{3}$$

where p is the commission rate, and ρ is the broker's discount rate.

[9]Note that brokers never offer a price higher than $r^d(q)$, in which case even naive investors would seek competing quotes. Such a quote would be strictly suboptimal and never attract any investors in equilibrium, as discussed in Stahl (1989).

The broker's problem is to offer a price p that maximizes its expected payoff:

$$\max_{p} V(p, MC; \pi_A(\pi))\Big[(1-\pi) + \pi\big(1 - G_s(p)\big)^n\Big]. \tag{4}$$

In this expected payoff, $V(p, MC; \pi_A(\pi))$ represents the broker's expected payoff conditional on the investor being retained. Note that $\pi_A(\pi)$ is the Bayesian updated belief after the investor accepts the offer, which signals that the investor is more likely to be naive. Specifically,

$$\pi_A(\pi) = \frac{\pi\big(1 - G_s(p)\big)^n}{1 - \pi + \pi\big(1 - G_s(p)\big)^n} \leq \pi.$$

The terms in the bracket capture the probability of the investor being retained: the term $1 - \pi$ represents the probability the investor is naive and thus being retained, while the term $\pi\big(1-G_s(p)\big)^n$ represents the probability that the investor is smart and the broker's quote p is lower than the investor's competing offers from the other n brokers. Specifically, $G_s(p)$ is the distribution of competing quotes from other brokers, conditional on the investor is smart.[10]

Given the distribution of competing quotes from other brokers G_s, this problem determines each broker's optimal strategy based on its belief about the investor's type, π. We seek a symmetric Bayes-Nash Equilibrium (BNE), where all brokers adopt the same strategy, $p = \phi(\pi)$, which is monotonically decreasing. This strategy, in turn, determines the quote distribution G_s, satisfying a fixed-point condition.

From the perspective of a smart investor, each time she approaches a new broker, the broker's belief is a new random draw from the distribution F_s. Combining this distribution with the brokers' equilibrium quote strategy, $p = \phi(\pi)$, yields

$$G_s(p) = G_s\big(\phi(\pi)\big) = 1 - F_s(\pi).$$

This equation implies that, in equilibrium, the probability of a competing quote being lower than the current broker's quote is equivalent to the probability that the compet-

[10] If the buyer is naive, then the buyer would not search for the competitor pricing at all, so the competitor's price distribution when the buyer is naive G_d, does not matter for the outcome.

ing broker's belief is higher than the current broker's belief, π. This equation further implies:[11]

$$G_s'(p) = -\frac{f_s(\pi)}{\phi'(\pi)}.$$

Differentiating the expected payoff in Eq. (4) with respect to p and evaluating it at $p = \phi(\pi)$, the first-order condition is

$$\left(V_p(p, MC; \pi_A(\pi)) + V_{\pi_A}(p, MC; \pi_A(\pi))\frac{\partial \pi_A}{\partial p}\right)\left[1 - \pi + \pi\big(1 - G_s(p)\big)^n\right] = \pi\, n\big(1 - G_s(p)\big)^{n-1} G_s'(p) V(p, MC; \pi_A(\pi)). \quad (5)$$

Importantly, in equilibrium, the monotonicity of the function ϕ ensures a one-to-one mapping between the quote strategy and the broker's belief.

We can substitute these into the first order condition and, after rearranging, obtain a differential equation for the quote strategy. This equation resembles an initial value problem:[12]

$$\phi'(\pi) + H(\pi)\left[\frac{V + (1 - \pi_A)V_{\pi_A}}{V_p}\right]\Bigg|_{(\phi(\pi), MC; \pi_A(\pi))} = 0, \quad \phi(0) = r^d, \quad (6)$$

where we define the characteristics function:

$$H(\pi) = \frac{n F_s^{n-1}(\pi)\, \pi f_s(\pi)}{1 - \pi + \pi F_s^n(\pi)} > 0.$$

[11] This is derived by applying chain rule:

$$G_s'(p) = \frac{dG_s(p)}{dp} = \frac{dG_s(\phi(\pi))}{d\pi}\frac{d\pi}{d\phi(\pi)} = \frac{d(1 - F_s(\pi))}{d\pi}\frac{d\pi}{d\phi(\pi)} = -\frac{f_s(\pi)}{\phi'(\pi)}.$$

[12] Note that

$$\frac{\partial \pi_A}{\partial p} = \frac{-\pi(1-\pi)n(1 - G_s(p))^{n-1} G_s'(p)}{[1 - \pi + \pi(1 - G_s(p))^n]^2} = \frac{\pi(1-\pi)nF_s(\pi)^{n-1} f_s(\pi)}{[1 - \pi + \pi F_s(\pi)^n]^2 \phi'(\pi)} = \frac{H(\pi)}{\phi'(\pi)}\frac{1 - \pi}{1 - \pi + \pi F_s(\pi)^n} = \frac{H(\pi)(1 - \pi_A(\pi))}{\phi'(\pi)}.$$

Plug this into Eq. (5) and rearrange the terms, we get

$$V_p + V_{\pi_A}\frac{H(\pi)(1 - \pi_A)}{\phi'(\pi)} = -H(\pi)\frac{V}{\phi'(\pi)} \Rightarrow \phi'(\pi)V_p + H(\pi)\Big[V + (1 - \pi_A)V_{\pi_A}\Big] = 0.$$

The boundary condition $\phi(0) = r^d$ ensures that when the investor is surely naive, the broker quotes a fee rate r to make the investor indifferent from seeking other quotes.

Given the broker's pricing strategy, it is straightforward to derive the expected value of an investor to the broker by converting Eq. (3) to a Hamilton-Jacobian-Bellman (HJB) Equation:

$$\begin{aligned}\rho V(p, MC; \pi) =& \underbrace{(p - MC)q}_{\text{flow payoff}} + \underbrace{\frac{\pi^2(1-\pi)^2}{2\sigma^2}\frac{\partial^2 V(p, MC; \pi)}{\partial \pi^2}}_{\text{belief diffusion}} \\ &+ \underbrace{\alpha\left[V(\phi(\pi), MC; \pi_A(\pi))(1 - \pi + \pi F_s^n(\pi)) - V(p, MC; \pi)\right]}_{\text{jump upon renegotiation}}.\end{aligned} \tag{7}$$

The broker's subjective discount rate, ρ, determines that the value function $V(p, MC; \pi)$ should appreciate by ρV over any time interval dt, as given by the left-hand side of the HJB equation.

The terms on the right-hand side represent the expected fluctuations. After retaining an investor, for any time interval dt, the incumbent broker receives the payoff flow $(p - MC)qdt$, as captured by the first term.

The broker also observes the information flow $y_t dt$, leading to an update of the broker's belief about the investor's type. Under the optimal filter rule Eq. (2), the belief π follows a diffusion process. Since it is a martingale, only the diffusion term enters the value function after applying Ito's Lemma, as captured by the second term.

During the time interval dt, the investor has a αdt chance of paying attention to commissions. Upon this, she initiates a negotiation process by requesting a new quote, and the incumbent broker responds with a new price. There is a probability of $(1 - \pi + \pi F_s^n(\pi))$ of retaining the investor. Probability $1 - \pi$ corresponds to the investor being naive and thus always retained. Probability π corresponds to the investor being smart, seeking competing offers from other brokers, and having a $F_s^n(\pi)$ probability of being retained. Together, these considerations form the jump term when the investor pays attention.

4.1.3 Equilibrium

The equilibrium in this model satisfies the following conditions:

- **Investor Search Decisions**: Investors seek quotes from brokers based on their reserve prices and accept the lowest quote they receive.
- **Brokers' Learning**: Brokers optimally infer investor types from the signals they observe.
- **Brokers' Pricing Strategy**: Brokers optimally respond to investors' requests for quotes based on their dynamic profit maximization and their belief about investor types.

Equations (6) and (7) jointly determine a representative broker's pricing strategy and value function. Standard dynamic programming arguments ensure the existence and uniqueness of the value and policy functions.

To facilitate our analysis, we normalize the broker's pricing strategy by the potential markup $r^d - MC$:

$$\mu(\pi) \equiv \frac{\phi(\pi; r^d) - MC}{r^d - MC}.$$

Here, μ represents the percentage realization of potential markup by the current price.

We conjecture that the value of the investor to the broker consists of two components:

$$V(p, MC; \pi) = \underbrace{\frac{1}{\rho + \alpha}(p - MC)q}_{\text{value from current price}} + \underbrace{v(\pi)(r^d - MC)q}_{\text{value from fees after renegotiation}} . \tag{8}$$

The first component represents the expected value from the current price p, which will be reset once the investor becomes attentive to commission fees. This value is discounted at a rate equal to the broker's discount rate ρ plus the rate at which the investor becomes attentive α. The second term captures the expected value after the investor becomes attentive and the price is renegotiated. We anticipate this term to be independent of the current price p and to depend solely on the broker's belief about the investor's type. Thus, we specifies v as a function of π that does not depend on the current price p.

Intuitively, $v(\pi)$ is positive because the broker expects to retain the investor with a positive probability. However, $v(\pi)$ decreases with π, as the future price is expected to be lower when the investor is more likely to be sophisticated.

Plugging $\mu(\pi)$ and $v(\pi)$ into the original Eq. (6) and Eq. (7), we can eliminate the variables p, r^d, q on both sides and obtain the following equations for (π) and $\mu(\pi)$:

$$(\rho+\alpha)v(\pi) = \frac{\pi^2(1-\pi)^2}{2\sigma^2}v''(\pi) + \alpha\left(1-\pi+\pi F_s^n(\pi)\right)\left[v(\pi_A) + \frac{\mu(\pi)}{\rho+\alpha}\right], \tag{9}$$

$$\mu'(\pi) = -(\rho+\alpha)H(\pi)\left[\frac{\mu(\pi)}{\rho+\alpha} + v(\pi_A) + (1-\pi_A)v'(\pi_A)\right]. \tag{10}$$

Notably, p does not appear in Eq. (9), validating the conjectured form in Eq. (8).

We also have the following boundary conditions:[13]

$$\mu(0) = 1, v(0) = \frac{\alpha}{\rho(\rho+\alpha)}, v'(0) = -\frac{\alpha}{\rho(\rho+\alpha)}, v(1) = \frac{\alpha}{\rho(\rho+\alpha)}\mu(1) \geq 0.$$

While we can not analytically solve v, μ, a numerical solution of these functions is illustrated in Fig. 5.

4.2 Model Implications

This subsection summarizes the model's implications across three key dimensions: the broker's pricing strategy, the impact of the broker's learning about investor types, and the effects of competition.

[13] Plug in $\pi = 0$ and $\pi = 1$, we can trivially obtain all the boundary conditions on $\mu(0)$, $v(0)$, and $v(1)$, we take derivative of Eq. (9) on both sides and obtain

$$\begin{aligned}(\rho+\alpha)v'(\pi) = &\frac{\pi^2(1-\pi)^2}{2\sigma^2}v'''(\pi) + \frac{\pi(1-\pi)(1-2\pi)}{\sigma^2}v''(\pi) \\ &+ \alpha(-1+n\pi F_s^{n-1}(\pi) + F_s^n(\pi))\left[v(\pi_A) + \frac{\mu(\pi)}{\rho+\alpha}\right] + \alpha(1-\pi+\pi F_s^n(\pi))\left[v(\pi_A) + \frac{\mu(\pi)}{\rho+\alpha}\right]'.\end{aligned}$$

Plug in $\pi = 0$ and with $F_s(0) = 0$, we get

$$(\rho+\alpha)v'(0) = \alpha * (-1) * (\frac{\alpha}{\rho(\rho+\alpha)} + \frac{1}{\rho+\alpha}) + \alpha(v'(0) * \pi_A'(0) + \mu'(0)/(\rho+\alpha) = -\frac{\alpha}{\rho},$$

which means that $v'(0) = -\frac{\alpha}{\rho(\rho+\alpha)} = -v(0)$.

4.2.1 Brokers' Pricing Strategies

The model offers several insights into brokers' pricing strategies that align with our data observations, and these implications can be analytically validated.

First, brokers charge lower fees to investors they perceive as smart (i.e., those with higher π). This model implication can be directly proved by demonstrating that $\phi(\pi)$ is monotonically decreasing, as established in Proposition 4.1. Since naive investors are more likely to be classified as such, their fees are, on average, higher than those of sophisticated investors (in terms of first-order stochastic dominance). This result aligns with the empirical facts presented in Section 3.2, which shows that brokers tend to charge higher fees to less sophisticated investors.

Proposition 4.1. *$\phi(\pi)$ decreases monotonically with π. Consequently, G_d has first-order stochastic dominance over G_s.*

Proof: See Section IA.3 in the Internet Appendix.

Second, we show that the fees charged by brokers are always above marginal costs, even for investors they perceive as perfectly sophisticated, consistent with our empirical findings in Section 3.1. This indicates that brokers enjoy a consistently positive markup for all investors. This outcome arises from the information asymmetry between the incumbent and competing brokers. Even when the incumbent broker knows an investor is sophisticated, it also knows that competing brokers lack this information, allowing it to avoid offering fees at marginal cost to the sophisticated investor.

Proposition 4.2. *$\mu(\pi) > 0$, which is equivalent to $\phi(\pi) > MC$, for any $\pi \in [0, 1)$.*

Proof: See Section IA.3 in the Internet Appendix.

Third, our model highlights a spillover effect of naive investors on sophisticated investors. If the reserve price of naive investors, r^d, decreases—indicating that naive investors become savvier—brokers respond by charging lower fees to all investors. Due to information frictions, brokers cannot perfectly distinguish between naive and sophisticated investors. As a result, they must lower fees across the board to retain naive investors.

Proposition 4.3. *$\phi(\pi; r^d)$ is strictly increasing with r^d for any $\pi \in [0, 1)$.*

Proof: Note that $\phi(\pi; r^d) = MC + (r^d - MC)\mu(\pi)$, $\mu(\pi) > 0$, *and* $\mu(\pi)$ *does not depend on* r^d. *Thus,* $\phi(\pi; r^d)$ *strictly increases with* r^d.

Lastly, we show that price discrimination between smart and naive investors intensifies as investors remain longer with a broker. This occurs because the broker learns investor types more accurately over time. We prove that the gap between the median fees charged to smart and naive investors—serving as a proxy for price discrimination—widens over time. The same intuition applies to the entire price distributions. This implication is consistent with the empirical facts presented in Section 3.3.

Proposition 4.4. *Suppose the median fee rate for smart investors at time* t *is* $p_s(t)$, *and the median fee rate for naive investors at time* t *is* $p_d(t)$. *The inter-median difference,* $p_d(t) - p_s(t)$, *increases over time as* t *grows.*

Proof: See Section IA.3 in the Internet Appendix.

4.2.2 Impact of Broker Learning

From a broker's perspective, the goal is to distinguish naive investors from sophisticated ones to implement differential pricing—charging naive investors high fees close to their reservation price while offering more competitive fees to attract sophisticated investors and compete with other brokers. Learning influences this market system through both incumbent broker and competing brokers: it helps the incumbent broker better separate naive investors from sophisticated ones, but it also aids competing brokers in competing for sophisticated investors.

These two effects may counterbalance each other, making it challenging to derive comparative statics of the learning effects in closed-form analytical expressions. However, we can examine these effects through numerical simulations, using model parameters estimated from our quantitative analysis, which we will describe in detail later.

Specifically, Fig. 6a illustrates the impact of the initial learning time t_0—the amount of information brokers have about a new investor—on $\mu(\pi)$) for three different values of t_0. This information reduces the incumbent broker's information advantage over competing brokers, thereby lowering its markup on sophisticated investors. By diminishing

the incumbent broker's ability to retain sophisticated investors, it consequently shifts its focus toward extracting more profit from naive investors by imposing higher markups on them. These opposite effects on sophisticated and naive investors can be seen from the monotonically decreasing pattern in the markup for $\pi \to 1$ (i.e., investors perceived as perfectly sophisticated) and the monotonically increasing pattern for $\pi \to 0$ (i.e., investors perceived as perfectly naive). For intermediate values of π, the interplay of these two effects may result in non-monotonic patterns.

Similarly, a lower noise level σ facilitates learning and weakens the information advantage of the incumbent broker, similar to the effect of increasing t_0. As illustrated in Fig. 6b, lowering σ has monotonic but opposite effects on the markups for sophisticated investors (i.e., $\pi \to 1$) and naive investors (i.e., $\pi \to 0$).

Fig. 6c examines the impact of different values of π_0, the fraction of sophisticated investors in the population. When sophisticated investors make up a smaller portion of the pool, competing brokers are more likely to classify a new investor as naive, thereby increasing the incumbent broker's information advantage over the sophisticated investor it identifies. As a result, the incumbent broker can charge a higher markup on sophisticated investors. Reflecting this effect, the figure shows that a higher π_0 leads to a monotonically decreasing markup on sophisticated investors (i.e., $\pi \to 1$). This effect, in turn, results in a monotonically increasing markup on naive investors (i.e., $\pi \to 0$). For intermediate values of π, the interplay of these offsetting effects on naive and sophisticated investors leads to non-monotonic patterns.

Finally, Fig. 6d shows that all investors benefit from more attention paid to commission fees—i.e., a larger value of α.

Overall, our findings suggest that sophisticated investors benefit from increased information availability to brokers. Enhanced information enables competing brokers to more effectively differentiate sophisticated investors from naive ones, resulting in more competitive pricing. This competitive pressure also forces the incumbent broker to reduce its fees for sophisticated investors. However, this advantage for sophisticated investors imposes a negative externality on naive investors, who face higher markups as a consequence.

4.2.3 Impact of Competition

Standard arguments suggest that competition among firms typically benefits consumers by driving prices down. Interestingly, our model reveals more nuanced effects of competition. To illustrate these subtleties, we first present an analytical result for the limiting case as n approaches infinity. With an infinite number of brokers in the market, each broker faces intense competition for investors. The following proposition demonstrates that, under such extreme competition, each broker adopts a surprising strategy: charging all investors the full markup.

Proposition 4.5. *When $n \to \infty$, $\mu(\pi)$ pointwise converges to 1 for any $\pi \in (0, 1)$.*

Proof: See Appendix B.

The reason competition may backfire lies in the heterogeneity between naive and sophisticated investors. Brokers can easily retain naive investors by charging fees below their reservation prices, while they must actively compete to attract sophisticated investors. As competition intensifies, retaining sophisticated investors becomes increasingly difficult. In the limiting case of infinitely many competitors, the probability of attracting a sophisticated investor becomes so small that brokers find it more effective to focus exclusively on extracting the full markup from naive investors, effectively giving up on competing for sophisticated investors altogether.

More generally, when each broker competes with a finite number of rivals, the impact of competition reflects two opposing forces. On one hand, lowering the price p attracts sophisticated investors. On the other hand, lowering p reduces the profit earned from each retained investor, including both naive and sophisticated ones. These offsetting effects result in a non-monotonic relationship between the number of competitors n and brokers' pricing strategies.

Fig. 7, based on model parameters estimated from our quantitative analysis, illustrates brokers' pricing strategies across different levels of competition, represented by varying values of n. The figure highlights the opposite effects of stronger competition on sophisticated and naive investors: greater competition, reflected in a larger number of competing brokers n, leads to lower markups for sophisticated investors (i.e., $\phi \to 1$) but higher

markups for naive investors (i.e., $\pi \to 0$).

As before, the effect of competition on investors with intermediate values of π is non-monotonic, as it reflects the aggregation of the two offsetting forces acting on naive and sophisticated investors. Notably, Fig. 7 shows that when the fraction of sophisticated investors in the population is higher (i.e., $\pi_0 = 0.8$ versus $\pi_0 = 0.2$), brokers abandon using lower markups to compete for sophisticated investors at lower values of π. This occurs because a higher π_0 makes it easier for competing brokers to identify sophisticated investors, thereby intensifying the competition among brokers. As a consequence, the incumbent broker places greater emphasis on extracting higher markups from naive investors rather than lowering markups to compete for sophisticated investors.

The model's implication of competition's potential adverse effects contrasts with the classical economic intuition that competition benefits consumers. However, this result is well-documented in a broad class of models with consumer heterogeneity (Rosenthal, 1980; Varian, 1980; Stahl, 1989; Carlin, 2009). In such settings, greater competition can shift firms' differential pricing strategies, potentially harming consumer welfare. Despite these theoretical insights, their empirical relevance remains uncertain due to mixed evidence in the literature (Chandra and Tappata, 2011; Lach and Moraga-González, 2017). This uncertainty may arise from the sensitivity of the theoretical results to parameter choices and the extent of heterogeneity in specific contexts. Empirically testing this counterintuitive prediction is thus crucial, a task we undertake in the next two sections.

5 Empirical Analysis

As discussed in Section 4.2.3, competition may have unintended effects on commission fees: increased competition among brokers can result in lower fees for sophisticated investors but higher fees for naive investors, thereby increasing price dispersion. In this section, we empirically investigate how local brokerage competition influences the dispersion of commission fees.

Consistent with the model, we measure competition by the number of brokers oper-

ating in the market, defined at the prefecture-level city.[14] We examine the distribution of commission fees for new accounts opened at each branch office of the broker covered in our data, on a quarterly basis. We select a quarterly frequency to ensure a sufficient number of new accounts per branch-quarter observation to reliably measure the distribution of fees.

We first present descriptive evidence showing that greater competition is associated with increased dispersion in fees, followed by causal evidence to support this relationship.

5.1 Descriptive Evidence

We begin by presenting descriptive evidence. For each branch-quarter, we measure the distribution of commission rates for investors who open accounts in that branch during the specified quarter. Specifically, we calculate several percentiles (i.e., 10th, 25th, 75th, and 90th) and the differences between the 90th and the 10th percentiles. Additionally, we compute the standard deviation of commission rates. To ensure comparability over time, we normalize these statistics by the mean commission rate of the respective branch-quarter.

Fig. 8 displays scatter plots of these commission rate distribution measures against the number of brokers operating in the city where the focal brokerage branch is located. To include a branch-quarter in our analysis, we require that at least 50 investors opened accounts at that branch during the quarter. Each subplot of Fig. 8 reports the correlation coefficient between the respective distribution statistic and the log number of brokers. The solid red lines represent linear regressions, where the dependent variable is a distribution statistic, and the independent variable is the log number of brokers.

In Table 4, we estimate the same regression while incorporating year-quarter fixed effects and city fixed effects. The results in Table 4 corroborate the descriptive patterns observed in Fig. 8.

Together, Fig. 8 and Table 4 demonstrate that the number of brokers in a city has effects

[14] A typical prefecture-level city in China is comparable to the State of Connecticut in terms of geographic size and population. Most brokerage branches are concentrated in the central urban areas of these cities, where the majority of local investors open their accounts.

on the lower and upper tail percentiles of commission rates. Specifically, the number of brokers is strongly negatively correlated with the lower tail percentiles but is strongly positively correlated with the upper tail percentiles. Consequently, the number of brokers is positively correlated with the dispersion measures, including the inter-percentile ranges (e.g., the difference between the 90th and 10th percentiles) and the standard deviation. Overall, these findings align with the key implication of our model: stronger competition among brokers can exacerbate price discrimination and, paradoxically, backfire.

5.2 Causal Evidence

While Fig. 8 and Table 4 reveal salient patterns consistent with the model implication that competition may exacerbate price discrimination, correlation does not necessarily imply causation. Cities with varying numbers of brokers may differ endogenously along other dimensions that could also influence commission rates. To address potential endogeneity concerns, we implement a stacked difference-in-differences identification strategy in this subsection to isolate the causal effect of new entry on the distribution of commission rates.

We compile a comprehensive list of all brokerage branches in China for each quarter from January 2007 to December 2017, using data provided by Flush iFinD, a major financial information data platform in China. We define treated brokerage branches as those experiencing a new broker entry into their city during the current quarter, provided there was no broker entry in the previous year. To create a control group, we use one-to-three matching, in which each treated branch is matched with three control branches whose pre-event number of competing brokers (i.e., the number of brokers in the same city) is closest to that of the treated branch. Additionally, we require that the control branches experience no new broker entry during the two-year window surrounding the treated entry. Using this approach, we identify 65 treatment events.

To estimate the treatment effect, we use the following regression model:

$$Y_{it} = \beta\, Treat_i * Post_t + \gamma_{\text{cohort}(i)} + \lambda_{\text{branch}(i)} + \delta_{\text{event}(t)} + \epsilon_{it}, \tag{11}$$

where the dependent variable, $Y_{i,t}$, is a measure of the dispersion of commission rates.

We consider two measures: (1) the standard deviation of the commission rates divided by the mean commission rates; (2) the difference between the 90th and 10th percentiles of commission rates, also divided by the mean commission rates. These dispersion measures are calculated at the branch-quarter level based on new investors, defined as those opening accounts in that branch and quarter. $Treat_i$ equals 1 for treated branches and 0 otherwise, while $Post_t$ equals 0 for pre-event quarters and 1 for post-event quarters. The model includes cohort, branch, and event-time fixed effects.

Columns (1) and (3) of Table 5 report the regression results. The coefficient on $Treat * Post$ is significantly positive, indicating that new broker entry increases the dispersion of commission rates. This finding aligns with the descriptive evidence presented in Fig. 8 and Table 4.

In columns (2) and (4), we present the dynamic coefficients, which are also plotted in Fig. 9. Importantly, the coefficients for the pre-period are insignificant, supporting the parallel trend assumption of our difference-in-differences design. A significant increase in the treatment coefficient is observed in the quarter of new broker entries, with the treatment effect estimates remaining significantly positive in most post-event quarters. Additionally, the treatment effect appears to increase during the first four quarters, likely reflecting a gradual onset of the treatment effect.

6 Quantitative Analysis

Our reduced-form analysis confirms the model's prediction that competition can indeed backfire in the data. However, the effect of competition may be non-monotonic, as suggested by our model in Fig. 7 and reflected in the mixed empirical findings of the literature (Chandra and Tappata, 2011; Lach and Moraga-González, 2017). To systematically assess the impact of competition on brokerage fees, it is therefore useful to estimate the structural parameters of the model and use the estimated model to examine the global properties of the competition effect and its welfare implications.

6.1 Estimation

We estimate the model's parameters using the simulated method of moments (SMM) (Strebulaev and Whited, 2011). The marginal cost MC_t is allowed to vary over time (t denotes quarters) to capture the downward trend in commission fees in the Chinese stock market. Population smartness $\bar{\pi}_k$ and naive investors' search costs κ_k varies by branches (k denotes branches) to reflect regional heterogeneity across branches.

We match a set of moments to identify the model parameters: average frequency of commission adjustments to identify α; the commission rates of investors with trading volume approaching infinity in each quarter t to identify MC_t; the branch-level mean and the covariance between commission rates and trading volume at each branch to identify κ_k and m; the average initial markup level at each branch to identify $\bar{\pi}_k$; the variance of initial commission rates to identify t_0; and the covariance between commission rate variance and account age to identify σ. The brokers' discount rates are calibrated to $\rho = 0.0995$ following Gormsen and Huber (2025). Detailed steps of the estimation procedure are provided in Appendix IA.4.

The estimation results are reported in Table 11. The calibrated discount rate, $\rho = 0.0995$, implies that brokers value next year's future revenue at roughly 10% less than current-year revenue. The estimated average attention frequency, $\alpha = 0.121$, indicates that about 12.1% of investors renegotiate and reduce their commission rates each year.[15] The estimated trading-volume elasticity of search costs, $m = 0.766$, suggests that investors with 1% more trading volume is expected to have 0.766% higher search costs. The initial interaction intensity, $t_0 = 0.124$, corresponds to about 45 days (0.124 years) of accumulated information in early broker-investor interactions. Finally, the estimated signal volatility, $\sigma = 0.285$, implies that with a prior $\pi_0 = 0.5$, it takes on average about 0.97 years for a broker to be 95% confident that an investor is sophisticated ($\pi > 0.95$).[16]

Our estimates align well with economic intuition in both spatial and temporal dimen-

[15] While this frequency appears low, each update is substantial—for example, a reduction from 30 bps to 10 bps or from 10 bps to 3 bps—large enough to drive the overall downward trend in average commissions.

[16] For $\theta = 1$ and $\pi_0 = 0.5$, the deterministic approximation of (2) yields $d\pi_t/dt = \pi_t(1-\pi_t)^2/\sigma^2$. Integrating $\int_{0.5}^{0.95} \frac{d\pi}{\pi(1-\pi)^2} = \int_0^t \frac{dt}{\sigma^2}$ gives $t = \sigma^2 \left[\ln \frac{\pi}{1-\pi} + \frac{1}{1-\pi}\right]_{0.5}^{0.95} = (0.285)^2 \times 20.95 \approx 1.70$. The noise term scales this by $\frac{1}{1+\sigma^2/2} \approx 0.57$ (via Itô-to-Stratonovich correction), producing $t \approx 0.97$.

sions. First, we find a higher proportion of smart investors in branches located in more urban and economically developed regions. As shown in Fig. 11, population smartness is significantly positively correlated with local population density and GDP per capita. Second, we find lower search costs in branches controlling on trading volume located in branches in more economically developed regions. As shown in Fig. 12, search costs of naive investors are significantly negatively correlated with local population density and number of brokers.[17] Third, our estimates of marginal costs over time closely track the downward trend in regulatory fees that brokers pay per transaction—a key component of marginal cost—as illustrated in Fig. 2a.

With the estimated model parameters, we simulate markets under alternative competition environments (by varying the number of competing brokers) and investor compositions (by varying the proportion of smart investors). Our variables of interest are per capita commission fees, the average markup on commissions for new investors, and the average markup for investors after five years.[18] While these fees are transfers to brokers and thus not direct measure of welfare, regulators generally favor a low-fee environment to promote financial inclusion (World Bank, 2013). We present the counterfactual results on broker competition and investor composition in the following two subsections.

6.2 The Effect of Competition

Having shown both theoretically and empirically that competition can backfire, we now quantify its impact on commission fees under varying levels of broker competition and investor composition. The left panels of Fig. 13, Fig. 14, and Fig. 15 display, respectively, per capita profit, average initial markup, and average markup after five years earned by a single broker. The right panels present the global treatment effects of adding one more broker on these outcomes across a broad range of parameter values, calculated as the

[17] Intuitively, investor search is easier in regions with denser population and with more brokerage branches. We thus focus on these two dimensions rather than GDP per capita. In fact, we also find a mild negative correlation between search costs and GDP per capita.

[18] Holding all other parameters fixed, for each $(\bar{\pi}, n)$ pair, we simulate $I = 1,000,000$ investors—of whom $I_s = I \cdot \bar{\pi}$ are smart—in a market with $n \in [1, 100]$ competitors, tracking their paths of μ. We assign each investor a q based on type (smart or naive) and the empirical distribution of q for that type, allowing us to back out r^d and ϕ. From the simulation, we compute measures such as per capita commission fees $\frac{1}{I}\sum_I(\phi q)$, $\frac{1}{I_s}\sum_{I_s}(\phi q)$, $\frac{1}{I_d}\sum_{I_d}(\phi q)$, as well as the average initial commission rate $\frac{1}{I}\sum_I \phi$, $\frac{1}{I_s}\sum_{I_s} \phi$, $\frac{1}{I_d}\sum_{I_d} \phi$.

derivative of each outcome variable with respect to the number of competing brokers.

As a benchmark case, when all investors are sophisticated ($\bar{\pi} \to 1$), increased competition unambiguously reduces both per capita profit and average markup, as shown in the top left panels of Fig. 13, Fig. 14, and Fig. 15. (Note that n decreases as the axis moves outward.) This pattern aligns with classical economic theory, which predicts that competition benefits the market when all participants are sophisticated. However, in the presence of naive investors, the effect of an additional broker becomes heterogeneous and non-monotonic.

Naive investors do not shop for lower commissions, so adding more than one rival yields negligible benefits for them—and may even harm them if brokers focus on extracting higher markups from naive investors rather than competing for sophisticated investors. For example, as shown in Panel C of Fig. 15, when $\bar{\pi} < 0.92$, naive investors gain no benefit from having more than two brokers ($n \geq 1$). By contrast, smart investors benefit from additional brokers: as shown in Panel B, when $\bar{\pi} > 0.5$, moving from two to three brokers reduces their average markup, though the marginal benefit diminishes quickly as brokers scale back aggressive price competition. Thus, while the prediction that competition may backfire is not universal, it holds across a broad and economically relevant range of parameters.

To draw policy implications, Fig. 16 plots the number of brokers that minimizes either total profit or average markup. When a sufficiently large share of the market is naive ($\bar{\pi} < 0.60$ for profit or $\bar{\pi} < 0.70$ for markup), two competing brokers suffice to minimize consumer costs, as shown in Panel A. This is because naive investors do not take leverage competition to secure lower commissions, brokers have little incentive to undercut each other when only a small fraction of investors are sophisticated. As $\bar{\pi}$ increases, the optimal number of competitors rises, diverging to infinity only in the limit as $\bar{\pi} \to 1$. Consistent with intuition, the optimal competition level for sophisticated investors is always higher than that for naive investors, as shown in Panel B.

Finally, Fig. 17 compares the theoretical optima with the observed number of brokers faced by each branch in our data. In nearly all cases, real-world markets host more brokers than would minimize fees. Absent entry and exit dynamics in our model, brokers would

find it profitable to serve naive investors as long as profits cover fixed operating costs—costs that appear too low to discipline market structure. This gap points to a potential role for regulatory intervention: rather than promoting further competition, regulators might consider limiting excessive entry in such markets.

6.3 The Effect of Investor Education

While market competition may backfire in reducing consumer markups, our quantitative analysis shows that investor education—by boosting the fraction of sophisticated investors—is far more effective. As shown in Fig. 15, the average markup declines monotonically with the population's mean ability, for both smart and naive investors, and at every level of competition.[19] This quantitative finding aligns with our theoretical result in Fig. 6c, which predicts that higher population smartness yields spillover benefits for the entire investor base.

Investor sophistication has a far greater impact on commission levels than any plausible variation in market structure. Whether a market is concentrated among a few brokerage firms or highly fragmented, a one-percentage-point increase in the share of smart investors reduces the average markup more than could be achieved by, for example, doubling or halving the number of competitors. In practice, improving financial literacy shifts the entire fee schedule downward more substantially than any realistic change in competition alone. Intuitively, only smart investors can take advantage of competitive alternatives to secure lower fees, so competition has a first-order effect on markups only when the investor base is sufficiently sophisticated ($\bar{\pi} \to 1$) and is effectively negligible for naive investors.

Our quantitative analysis highlights the substantial potential of investor education, suggesting that regulators should prioritize policies that enhance financial literacy as a primary tool for improving market outcomes. By enabling more investors to recognize and act on competitive opportunities, such initiatives can reduce fees more effectively

[19] The exception is the top-left panel of Fig. 13, where the pattern breaks mechanically because smart investors trade more frequently than naive ones; in our simulation, greater trading by smart investors raises total broker profits.

than structural reforms focused solely on increasing competition.

7 Conclusion

TBA

References

An, Li, Joseph Engelberg, Matthew Henriksson, Baolian Wang, and Jared Williams (2024) "The portfolio-driven disposition effect," *The Journal of Finance*, 79 (5), 3459–3495.

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

Barberis, Nicholas and Richard Thaler (2003) "A survey of behavioral finance," *Handbook of the Economics of Finance*, 1, 1053–1128.

Ben-David, Itzhak and David Hirshleifer (2012) "Are investors really reluctant to realize their losses? Trading responses to past returns and the disposition effect," *The Review of Financial Studies*, 25 (8), 2485–2532.

Bhutta, Neil, Andreas Fuster, and Aurel Hizmo (2020) "Paying too much? Price dispersion in the US mortgage market," *Federal Reserve Board Working Paper*.

Carlin, Bruce I. (2009) "Strategic price complexity in retail financial markets," *Journal of Financial Economics*, 91, 278–287.

Chandra, Ambarish and Mariano Tappata (2011) "Consumer search and dynamic price dispersion: an application to gasoline markets," *The RAND Journal of Economics*, 42, 681–704.

Célérier, Claire and Boris Vallée (2017) "Catering to investors through security design: Headline rate and complexity," *Quarterly Journal of Economics*, 132 (3), 1469–1508.

Diamond, Peter A. (1971) "A model of price adjustment," *Journal of Economic Theory*, 3, 156–168.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl (2017) "The deposits channel of monetary policy," *Quarterly Journal of Economics*, 132 (4), 1819–1876.

Egan, Mark, Ali Hortaçsu, and Gregor Matvos (2017) "Deposit competition and financial fragility: Evidence from the us banking sector," *American Economic Review*, 107 (1), 169–216.

Ellison, Glenn (2005) "A model of add-on pricing," *Quarterly Journal of Economics*, 120 (2), 585–637.

Ellison, Glenn and Sara Fisher Ellison (2009) "Search, obfuscation, and price elasticities on the internet," *Econometrica*, 77 (2), 427–452.

Feng, Lei and Mark S Seasholes (2004) "Correlated trading and location," *The Journal of Finance*, 59 (5), 2117–2144.

——— (2005) "Do investor sophistication and trading experience eliminate behavioral biases in financial markets?" *Review of Finance*, 9 (3), 305–351.

Frydman, Cary and Baolian Wang (2020) "The impact of salience on investor behavior: Evidence from a natural experiment," *The Journal of Finance*, 75 (1), 229–276.

Gabaix, Xavier and David Laibson (2006) "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *Quarterly Journal of Economics*, 121 (2), 505–540.

Gormsen, Niels Joachim and Kilian Huber (2025) "Corporate Discount Rates," *American Economic Review*, 115, 2001–2049.

Hannan, Timothy H. and Allen N. Berger (1991) "The rigidity of prices: Evidence from the banking industry," *American Economic Review*, 81 (4), 938–945.

Hansen, Lars Peter (2007) "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97, 1–30.

Hartzmark, Samuel M (2015) "The worst, the best, ignoring all the rest: The rank effect and trading behavior," *The Review of Financial Studies*, 28 (4), 1024–1059.

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

Henderson, Brian J. and Neil D. Pearson (2011) "The dark side of financial innovation: A case study of the pricing of a retail financial product," *Journal of Financial Economics*, 100 (2), 227–247.

Hortacsu, Ali and Chad Syverson (2004) "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *The Quarterly Journal of Economics*, 119, 403–456.

Kim, Moshe, Doron Kliger, and Bent Vale (2003) "Estimating switching costs: the case of banking," *Journal of Financial Intermediation*, 12 (1), 25–56.

Klemperer, Paul (1995) "Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade," *Review of Economics Studies*, 62 (4), 515–539.

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

Lach, Saul and José L Moraga-González (2017) "Asymmetric Price Effects of Competition," *The Journal of Industrial Economics*, 65, 767–803.

Liu, Hongqi, Cameron Peng, Wei A Xiong, and Wei Xiong (2022) "Taming the bias zoo," *Journal of Financial Economics*, 143 (2), 716–741.

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

Rosenthal, Robert W. (1980) "A Model in which an Increase in the Number of Sellers Leads to a Higher Price," *Econometrica*, 48, 1575.

Scheinkman, José A. and Wei Xiong (2003) "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111, 1183–1220.

Seru, Amit, Tyler Shumway, and Noah Stoffman (2010) "Learning by trading," *The Review of Financial Studies*, 23 (2), 705–739.

Shefrin, Hersh and Meir Statman (1985) "The disposition to sell winners too early and ride losers too long: Theory and evidence," *The Journal of Finance*, 40 (3), 777–790.

Stahl, Dale O (1989) "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review*, 79, 700–712.

Stango, Victor and Jonathan Zinman (2014) "Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees," *Review of Financial Studies*, 27 (4), 990–1030.

Strebulaev, Ilya A. and Toni M. Whited (2011) "Dynamic Models and Structural Estimation in Corporate Finance," *Foundations and Trends® in Finance*, 6, 1–163.

Varian, Hal R. (1980) "A Model of Sales," *The American Economic Review*, 70, 651–659.

Wonham, W. M. (1964) "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 2, 347–369.

World Bank (2013) "Global financial development report 2014 : financial inclusion," 11.

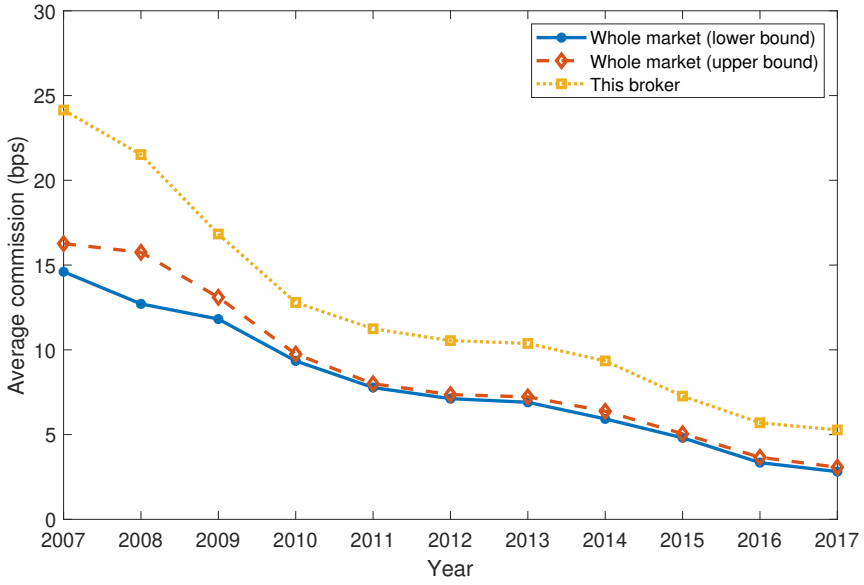
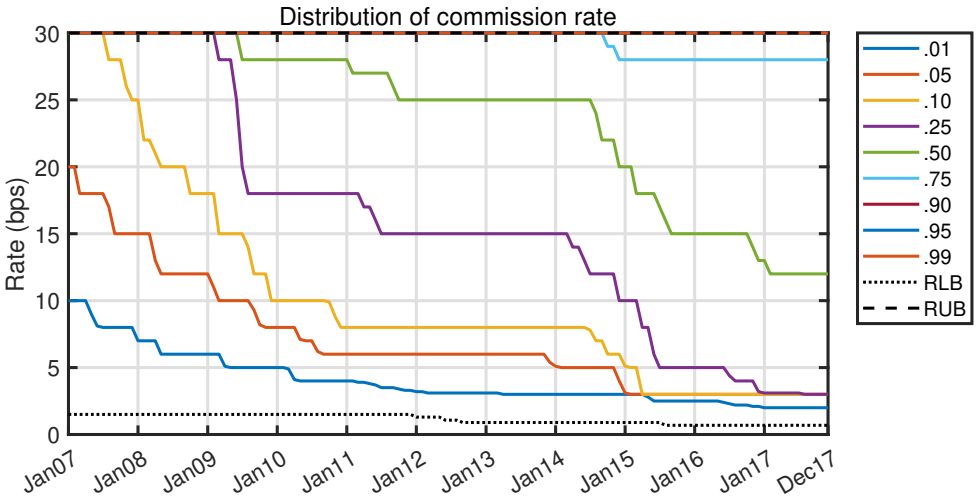
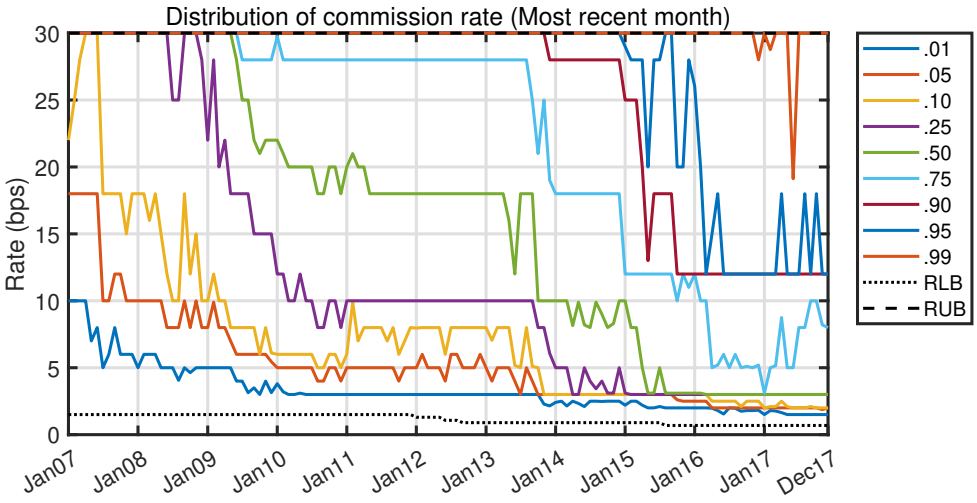


Figure 1: The average commission rates over time
This figure displays the average commission rates of the investors in our sample (dotted line) and the average commission rates of the market (solid and dashed lines), from 2007 to 2017. For our broker, we include only stock trading. The solid line represents the commission rates calculated as the total commission revenue across all brokers divided by the aggregate trading volume (including stocks, funds, bonds, and options/warrants). The dashed line represents the commission rates calculated as the total commission revenue across all brokers divided by the aggregate stock trading volume.



(a) all accounts



(b) new accounts

Figure 2: Dispersion of commission rates
This figure displays the distribution of commission rates across investors. For each month, we calculate the commission rate for each investor and report several percentiles, i.e., 1st, 5th, 10th, 25th, 75th, 90th, 95th, and 99th. RLB (dotted line) represents the regulatory lower bound. The regulatory upper bound (not shown in the figure) is 30 bps. Panel (a) is based on all investors, and Panel (b) is based on new accounts. New accounts for a month are defined as accounts opened in that month.

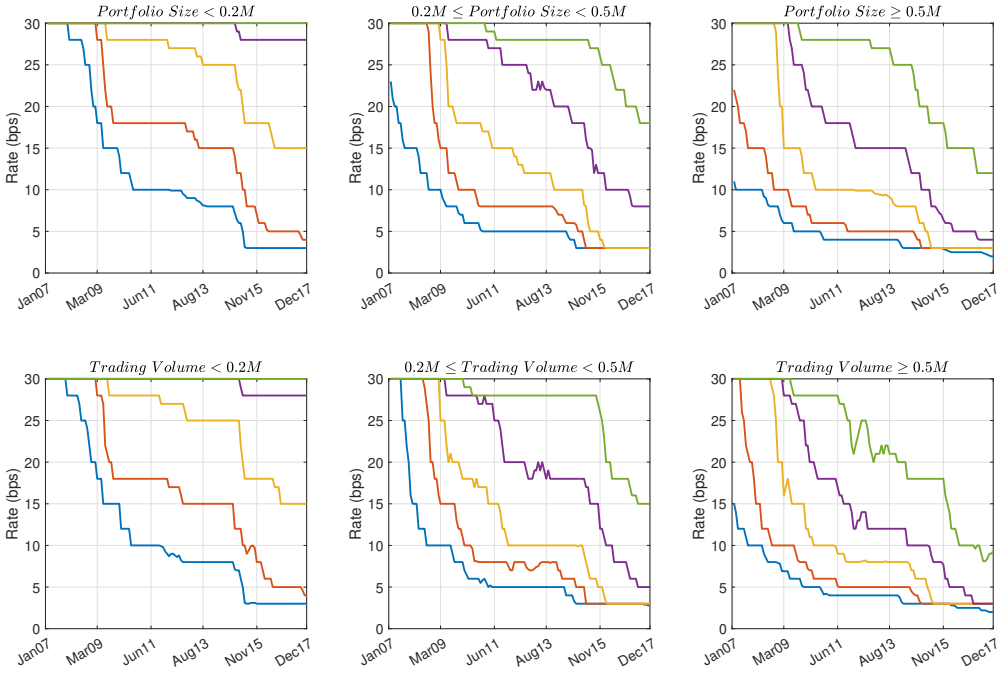
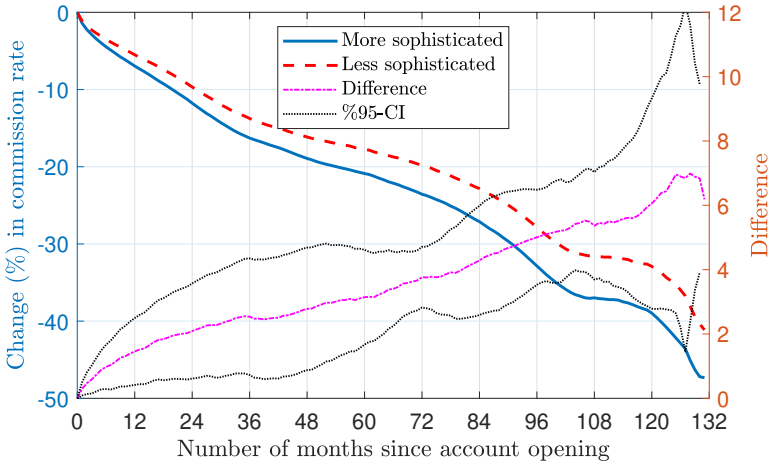
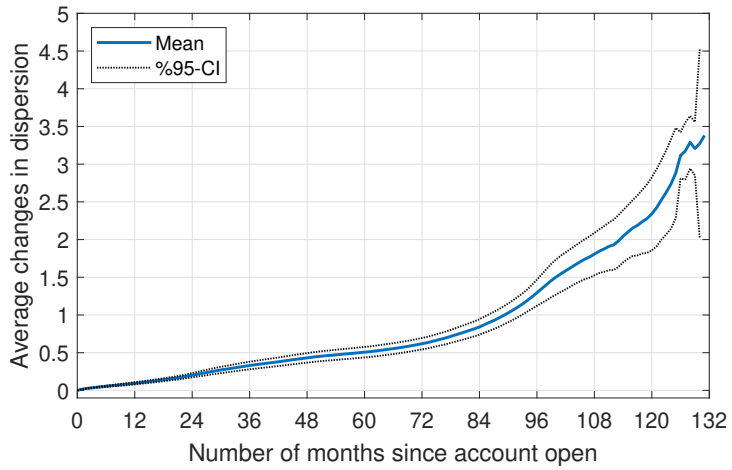


Figure 3: Dispersion of commission rates, by portfolio size and trading volume
This figure displays the distribution of commission rates across investors by subsamples. For each subsample in each month, we calculate the commission rate for each investor and report several percentiles, i.e., 10th, 25th, 50th, 75th, and 90th.



(a) Percentage changes in commission rates



(b) Dispersion of commission rates by account age

Figure 4: Evolution of commission fees

Panel (a) shows the changes in average commission rates by investor groups. For each cohort of investors (defined as investors who open their accounts in the same month), we sort them into two groups by the median value of their composite bias measure, defined as per Table 2. For both groups, we report the percentage changes in commission rates relative to their initial commission rates. In addition, we also present the difference between the two groups. The scale for the difference is shown on the right y-axis. Panel (b) presents the percentage changes in the cross-sectional dispersion of commission rates. For each cohort of investors, we calculate the dispersion of commission rates as the standard deviation divided by the mean rate, month by month. Then we calculate, for each cohort, the percentage change in the dispersion of commission since account opening. In the end, we group all cohorts by event month and report the cross-cohort averages for each event month, together with the 95% confidence intervals.

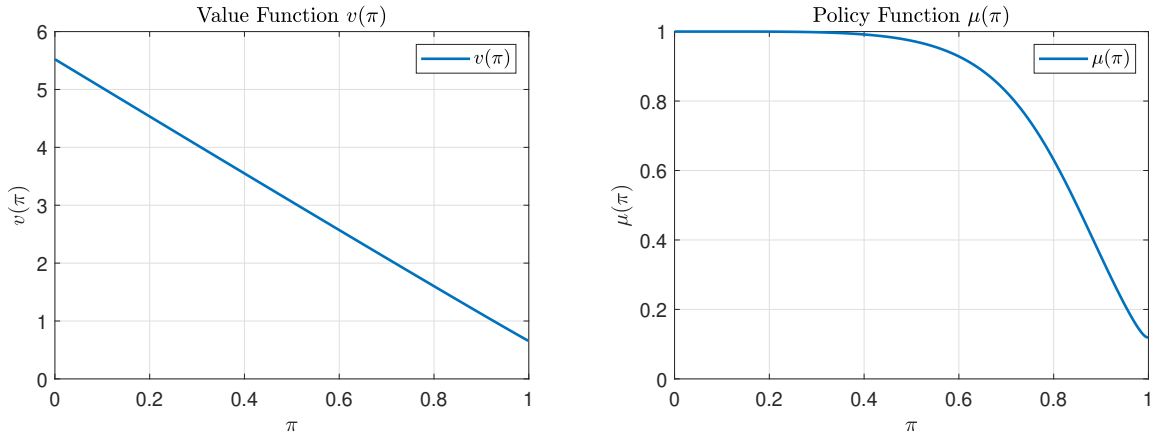
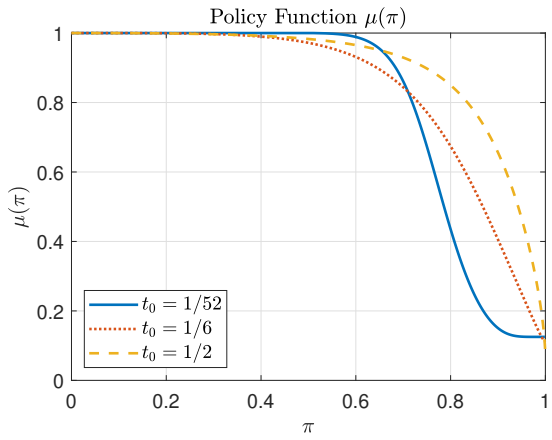
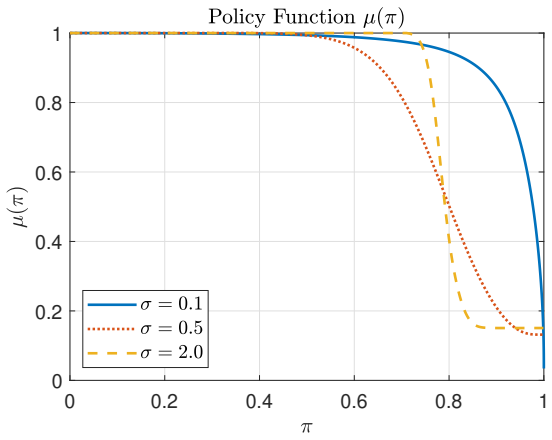


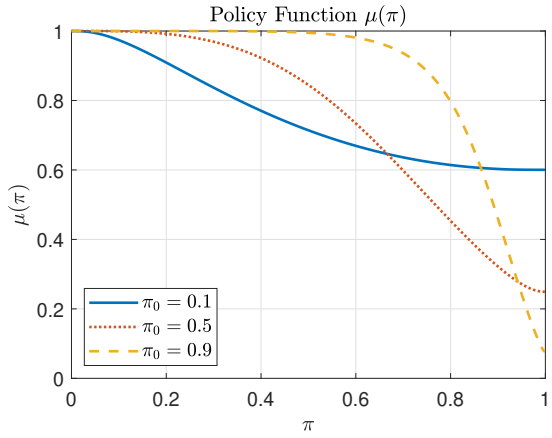
Figure 5: Numerical solutions of value function and policy function
This figure shows numerical solutions of value function and policy function with baseline parameters $\rho = 0.0995$, $\alpha = 0.1213$, $t_0 = 0.1240$, $\sigma = 0.2851$, $\bar{\pi} = 0.8062$, and $n = 1$. v and π are solved simultaneously by a shooting algorithm that makes a initial guess of $v'(0)$.



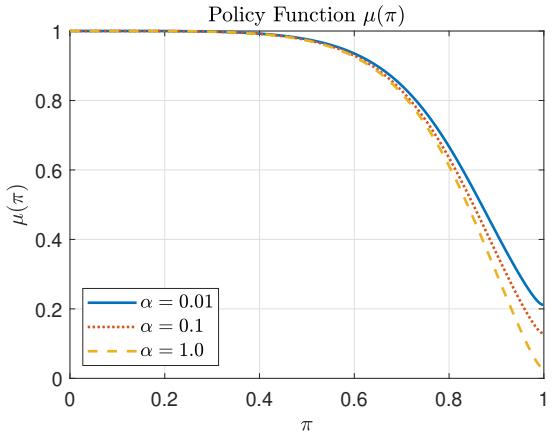
(a) Initial learning time



(b) Noise level



(c) Proportion of smart investors



(d) Attention on commissions

Figure 6: Numerical solutions under different market conditions
This figure shows numerical solutions of policy function under different market conditions. Baseline parameters are $\rho = 0.0995, \alpha = 0.1213, t_0 = 0.1240, \sigma = 0.2851, \bar{\pi} = 0.8062$, and $n = 1$.

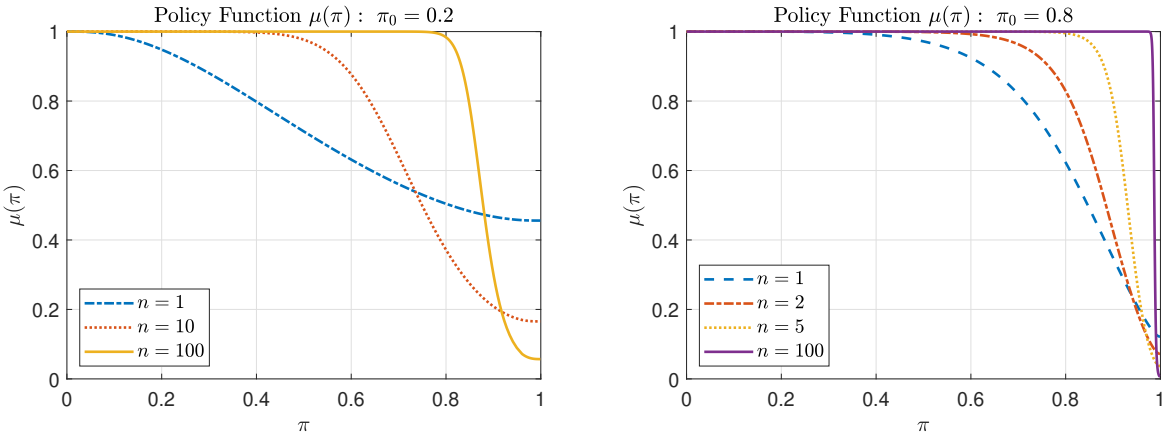


Figure 7: Numerical solutions under different competition environments
This figure shows numerical solutions of policy function under different competition environments. Baseline parameters are $\rho = 0.0995, \alpha = 0.1213, t_0 = 0.1240, \sigma = 0.2851, \bar{\pi} = 0.2$ (left panel) or 0.8 (right panel).

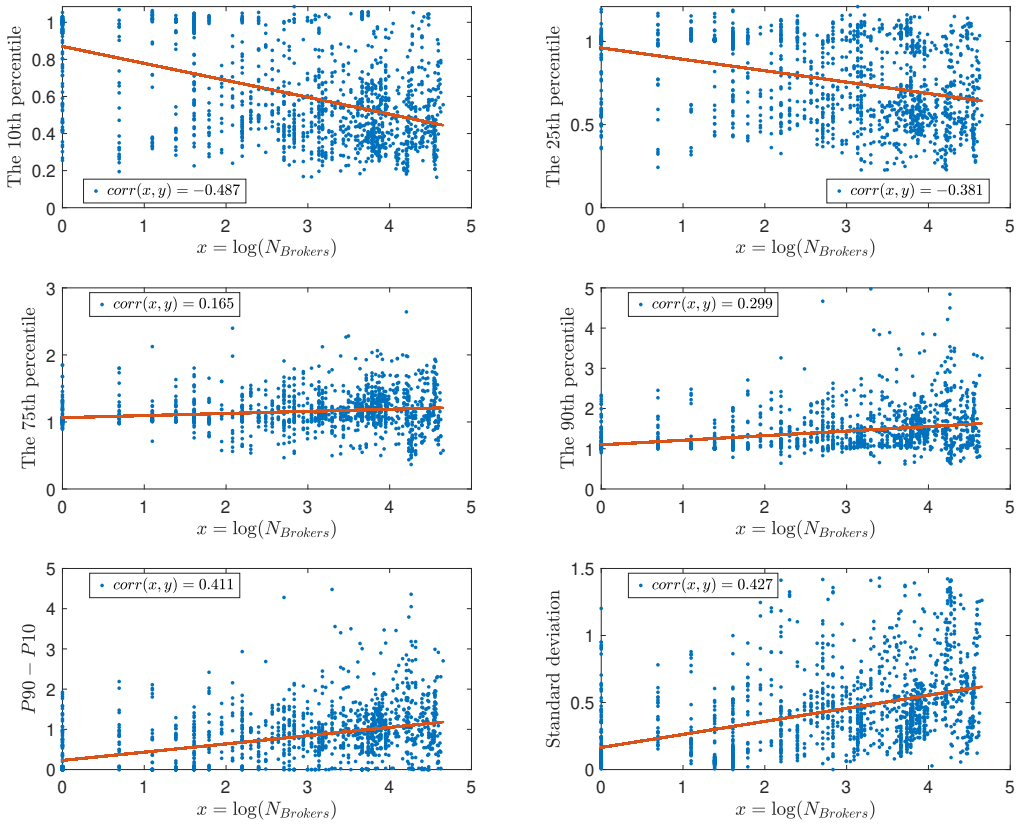


Figure 8: Commission rates dispersion and the number of brokers

This figure presents scatter plots of various commission rate distribution statistics—either percentiles or standard deviations—plotted against the logarithm of the number of brokers in a city. These statistics are computed at the branch-quarter level, based on new investors who opened accounts in that branch-quarter. The analysis is restricted to branch-quarters with at least 50 new investors. Each panel reports the correlation coefficient between the distribution statistic and the log broker count. The solid red lines depict linear regressions of the distribution statistics on the log broker count. "$P90 - P10$" denotes the difference between the 90th and 10th percentiles. All distribution statistics have been normalized by their respective sample means.

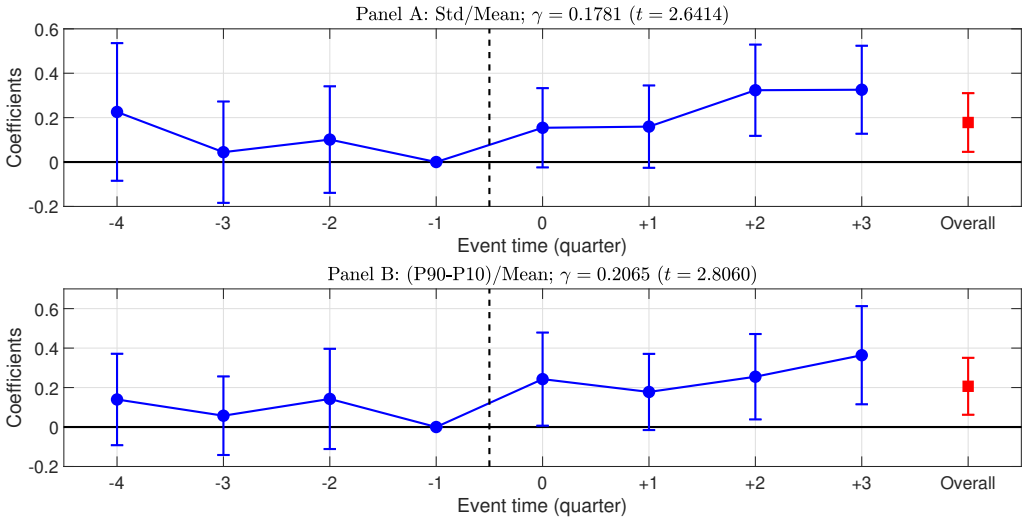


Figure 9: Dynamic treatment effects of broker entry on price dispersion
This figure displays dynamic coefficients and the corresponding 95% confidence intervals from stacked DiD analyses. The isolated red error bar in each plot shows the overall estimate of the treatment effect. We define treated units as those with new broker entries in the current quarter and no broker entry in the previous year. The control group must have no broker entries during the 2-year event window. We use one-to-three matching, where each treated unit is matched with the three control units whose pre-event number of brokers (within the same city) is closest to that of the treated unit. In Panel A, we calculate the dispersion as the standard deviation of commission rates divided by the mean commission rate. In Panel B, we calculate the dispersion as the difference between the 90th and the 10th percentiles of commission rates, divided by the mean commission rate. For the dispersion calculation, we include only new investors, i.e., those who open their accounts in a quarter. We include cohort, branch, and event-time fixed effects in all analyses.

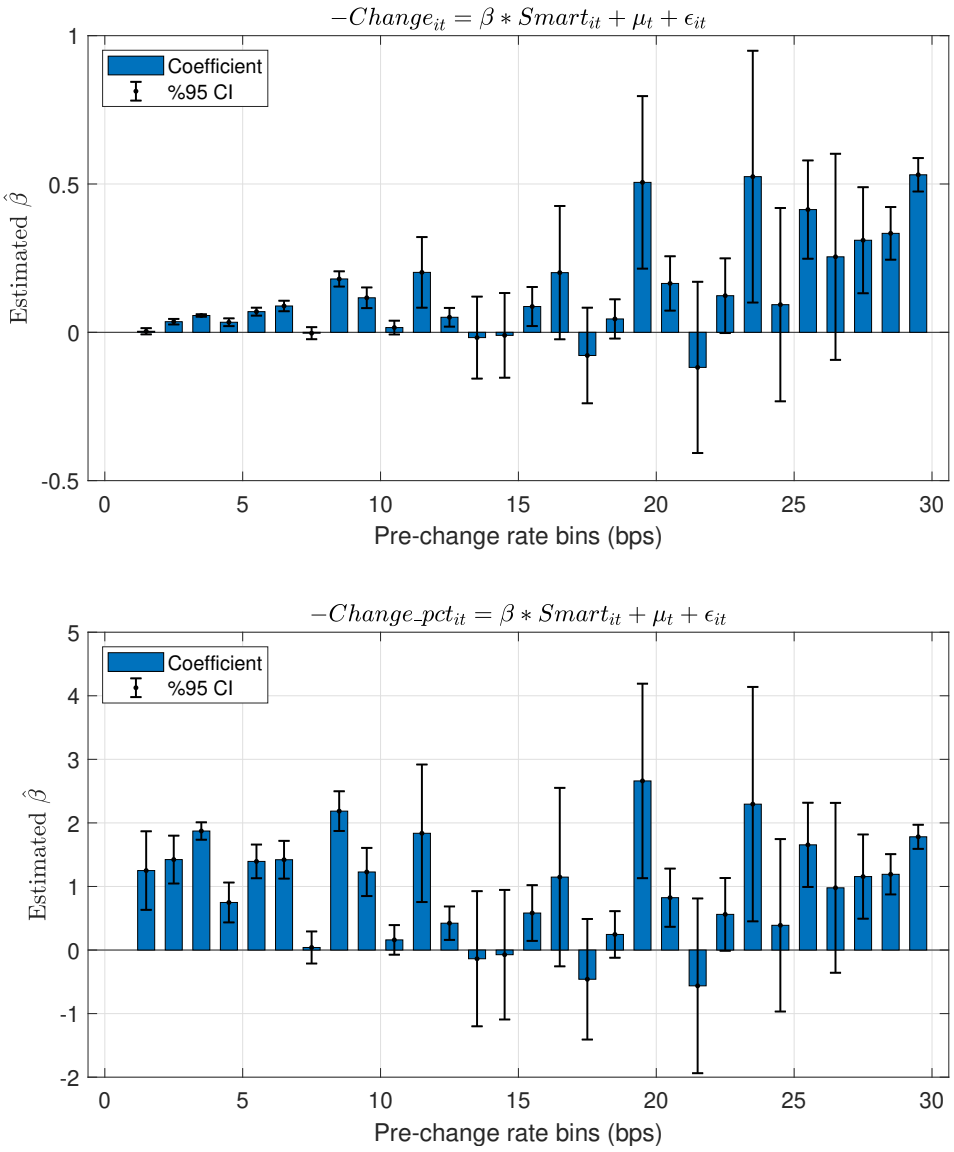


Figure 10: Percentage change on commission rates: Smart versus Naive
For each case of rate renegotiation, we record its pre-change rate. We then group all observations into 30 bins (each with length of 1 bp) according to the level of their pre-change rates. For each cohort of investors (defined as investors who renegotiate their rates in the same month), we sort them into two groups (more sophisticated and less sophisticated) by the median value of their composite bias measure, defined as per Table 2. In each bin of pre-change rate, we run a regression of changes in commission rates on the smart dummy (1 for more sophisticated, 0 for others) with year-month fixed effect. This figure plots the estimated values $\hat{\beta}$ together with their associated %95 confidence intervals for each bin of pre-change commission rates. The confidence intervals are computed using standard errors clustered by year-month. The dependent variable in the top panel is the negative value of changes in commission rate. The dependent variable in the bottom panel is the negative value of relative percentage change.

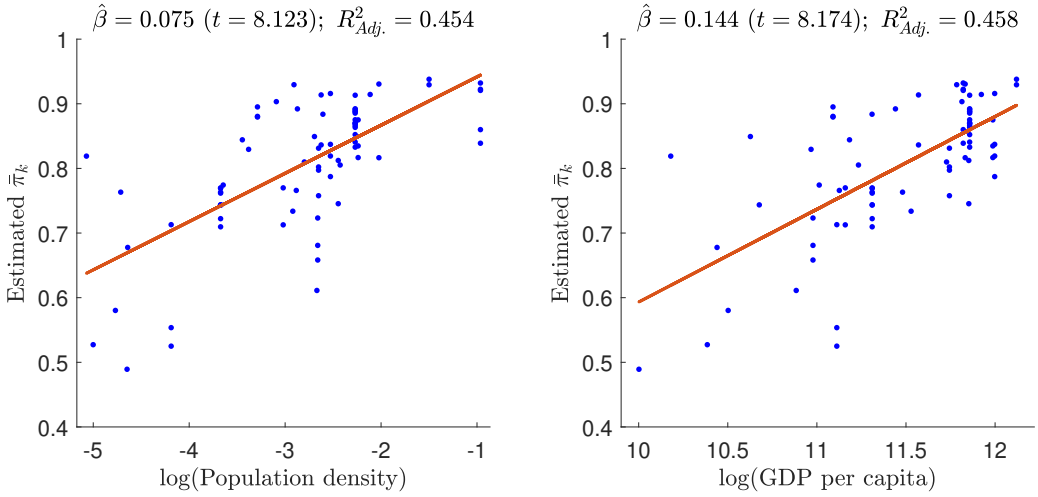


Figure 11: Proportion of smart investors versus population density and GDP per caiptal

This figure presents scatter plots of the estimated values $\bar{\pi}_k$ against two explanatory variables: (1) the logarithm of population density of branch k's host city and, (2) the logarithm of GDP per capita of the host city. The solid line in each plot represents the OLS fit of the sample. We report the estimated slope $\hat{\beta}$, the associated t-value, and the adjusted R^2 in the title of each plot.

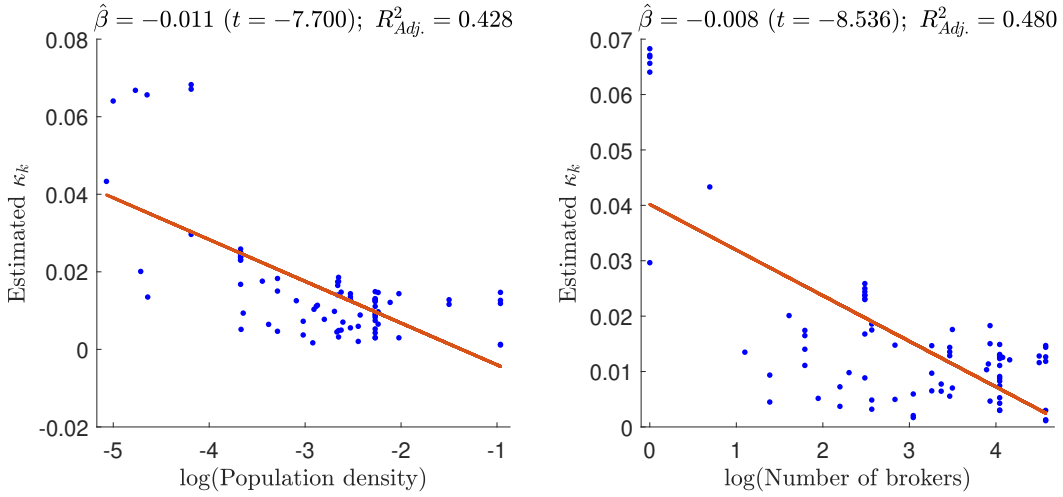


Figure 12: Search cost of naïve investors versus population density and number of brokers

This figure presents scatter plots of the estimated values κ_k against two explanatory variables: (1) the logarithm of population density of branch k's host city and, (2) the logarithm of number of brokers of the host city. The solid line in each plot represents the OLS fit of the sample. We report the estimated slope $\hat{\beta}$, the associated t-value, and the adjusted R^2 in the title of each plot.

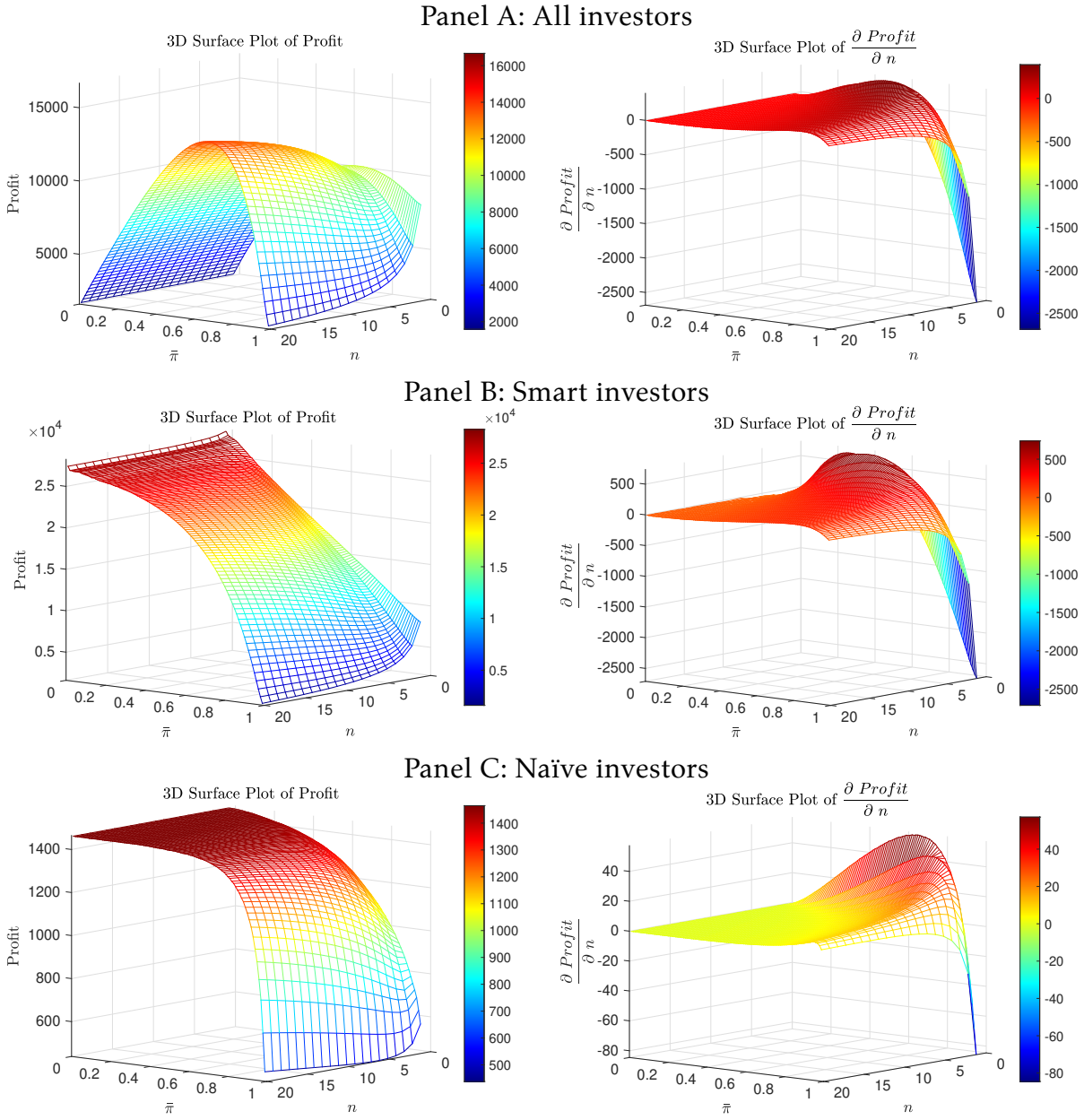


Figure 13: Competition effect on average profit per account
The left plot in each panel displays lifetime net profit per account as initial belief $\bar{\pi}$ and number of competitors n changes. The right plot in each panel shows the corresponding partial derivative with respect to n. The solid line with marker in the left plot of Panel A shows the minimum point of n for each value of $\bar{\pi}$.

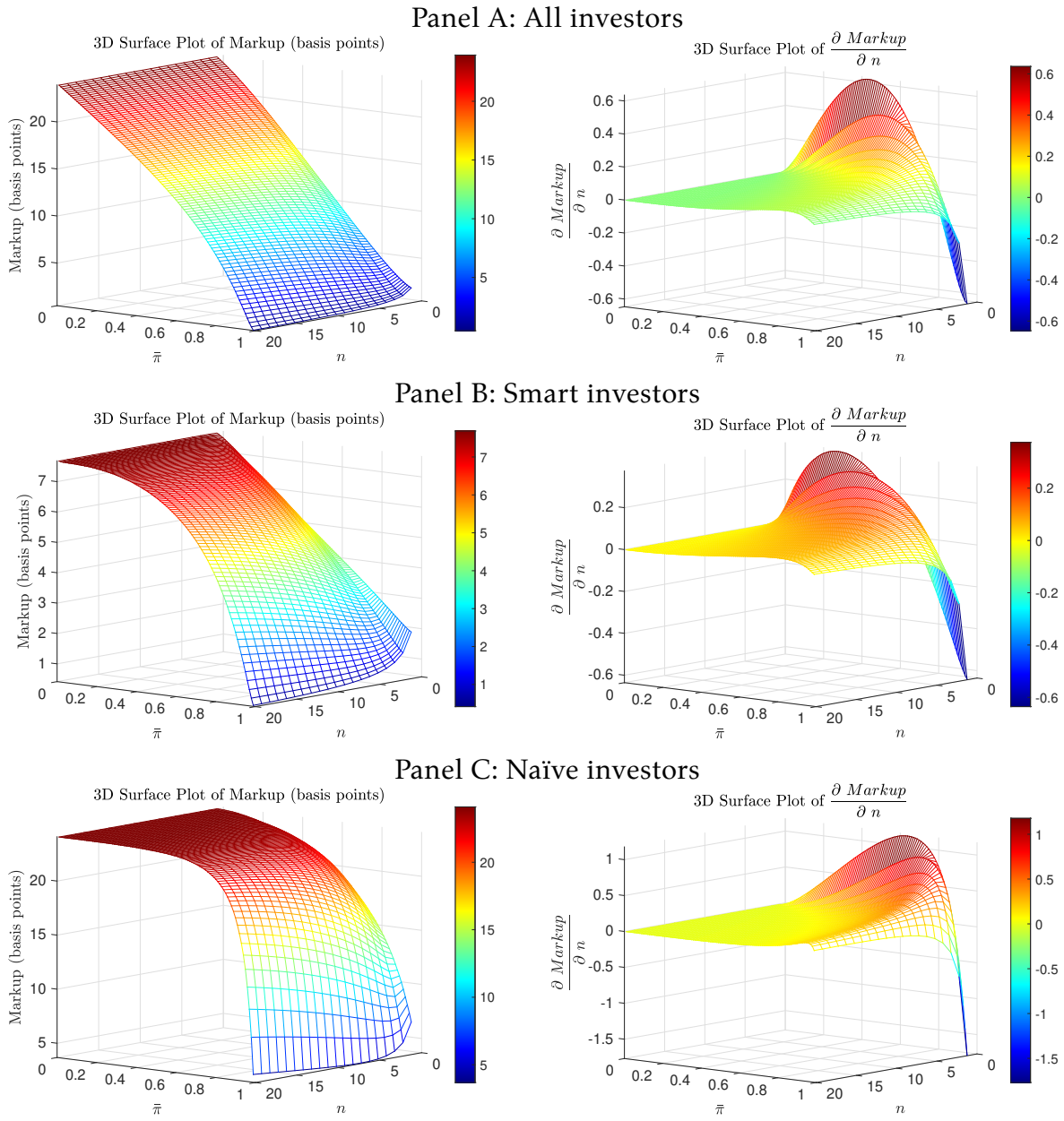


Figure 14: Competition effect on average initial markup
The left plot in each panel displays average initial markup levels as initial belief $\bar{\pi}$ and number of competitors n changes. The right plot in each panel shows the corresponding partial derivative with respect to n. The solid line with marker in the left plot of Panel A shows the minimum point of n for each value of $\bar{\pi}$.

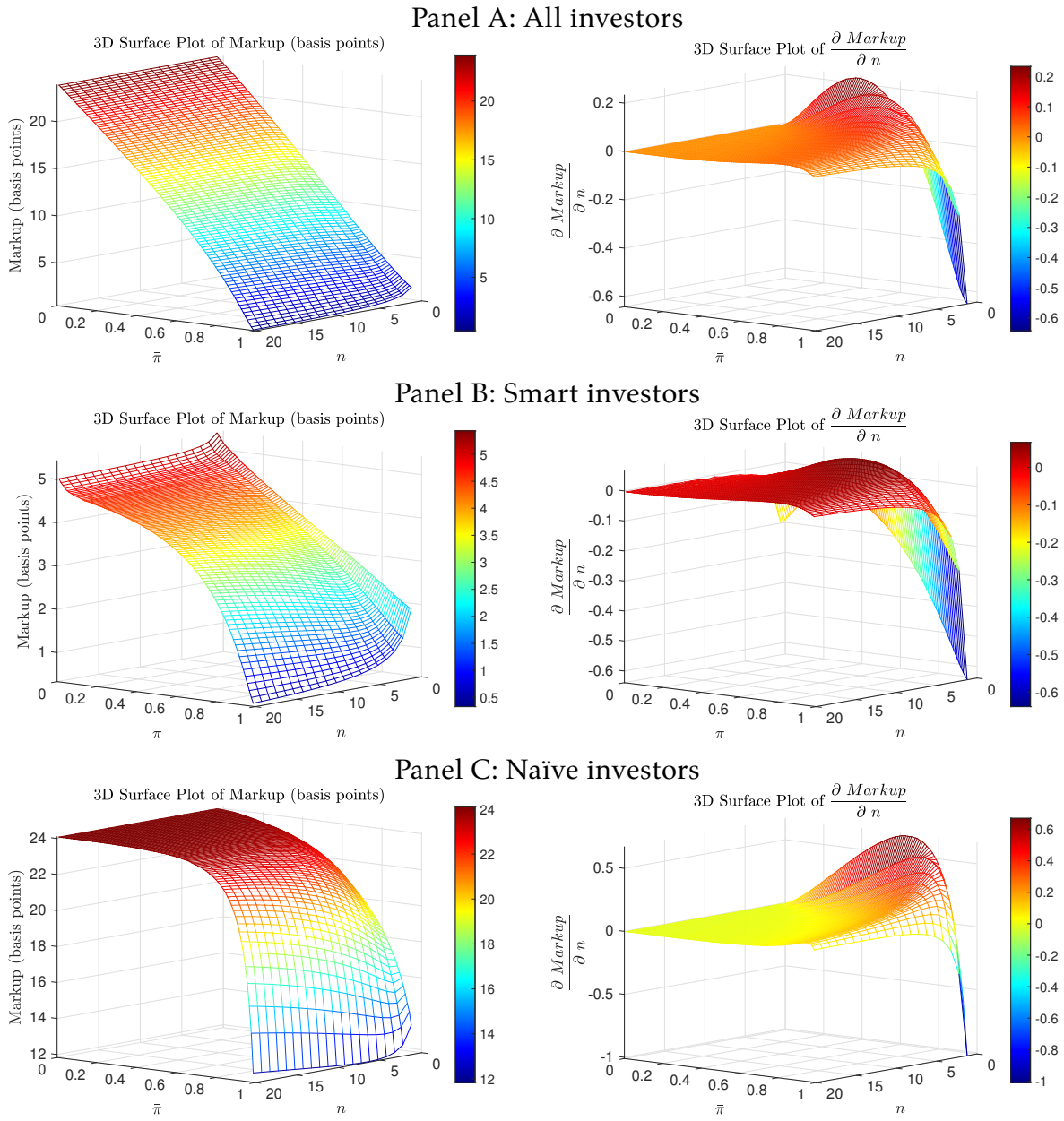


Figure 15: Competition effect on average markup at $T = 5$

The left plot in each panel displays average markup levels at $T = 5$ as initial belief $\bar{\pi}$ and number of competitors n changes. The right plot in each panel shows the corresponding partial derivative with respect to n. The solid line with marker in the left plot of Panel A shows the minimum point of n for each value of $\bar{\pi}$.

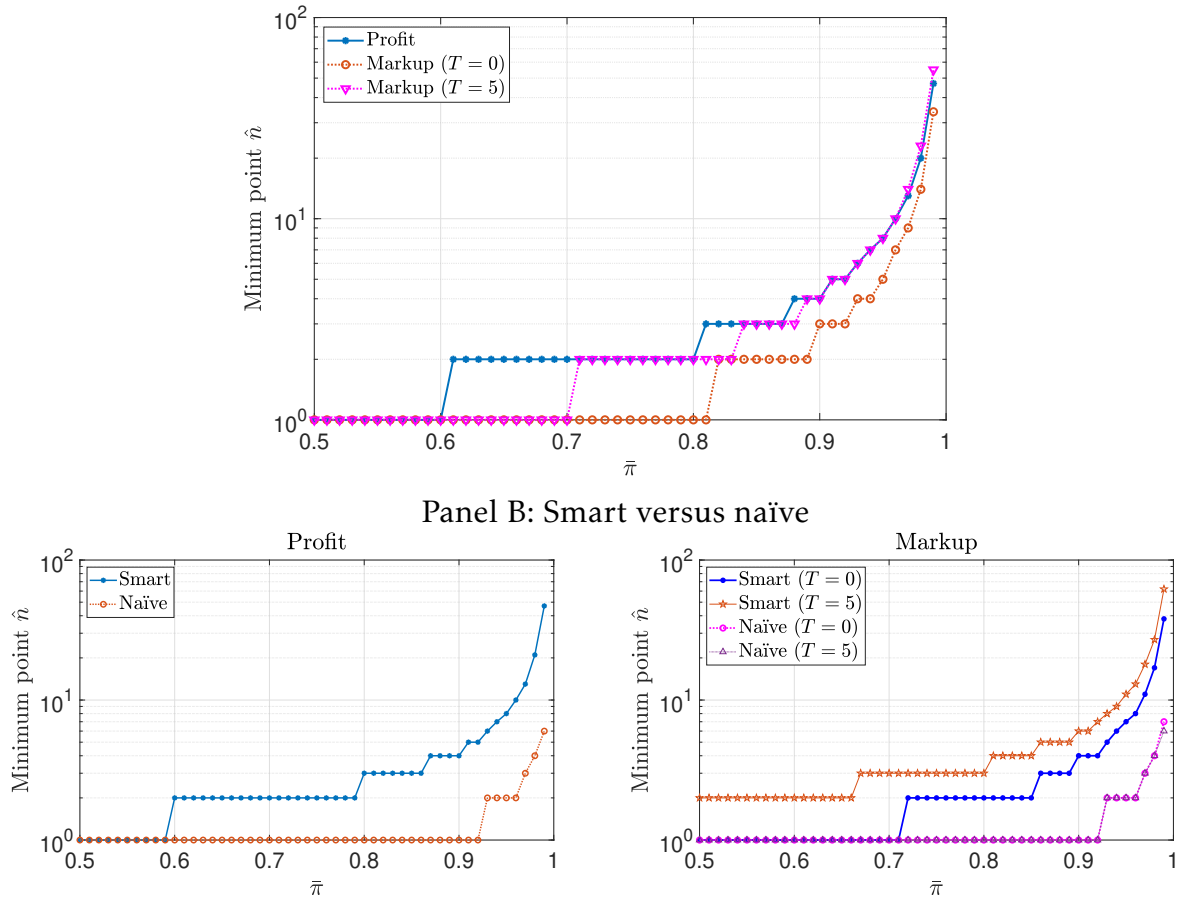


Figure 16: Optimal competition level for different values of $\bar{\pi}$
Panel A of this figure shows, for all investors, the optimal number of competitors for different values of $\bar{\pi}$. The solid line represents the optimal numbers of competitors that minimize average lifetime profit, while the dotted line indicates those that minimize the average markup. We use a logarithmic scale for the y-axis to improve clarity. In Panel B, we report results for smart (solid lines) and naïve (dotted lines) investors separately. Left (right) plot presents the optimal numbers of competitors that minimize average profit (markup).

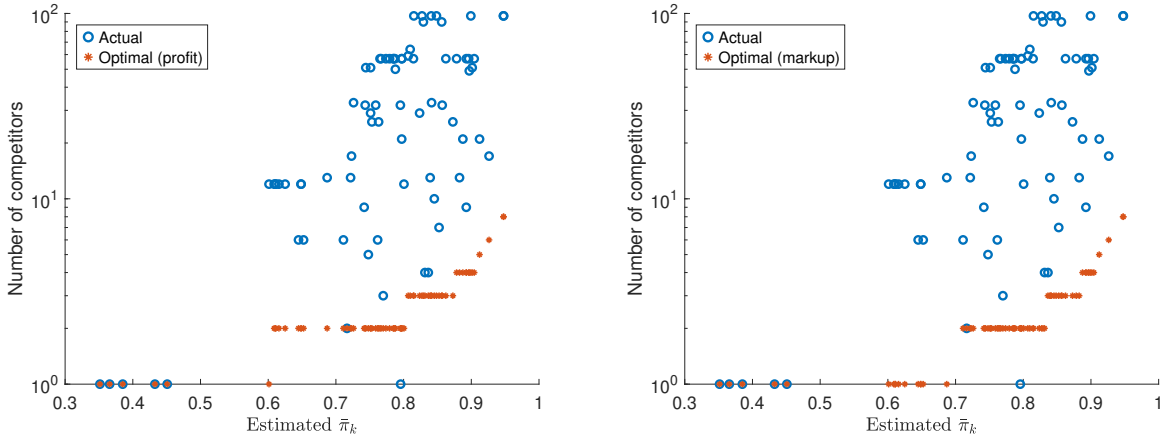


Figure 17: Actual and optimal competition levels across branches
This figure shows the actual and optimal number of competitors for all branches. The x-axis represents the estimated proportion of smart investors $\bar{\pi}_k$ for branch k. For each branch k, the optimal number of competitors are the minimum point $\hat{n}$ from Fig. 16 when the associated $\bar{\pi}$ is closest to $\bar{\pi}_k$. The actual number of competitors is calculated as the time-series average of end-of-quarter figures.

Table 1: Summary statistics

This table reports the summary statistics of the sample. All the statistics are calculated based on the cross-section of investors. We report the number of observations, sample means, standard deviations, and several percentiles. Age, Account Age, and Commission are computed at the end of 2017. Male equals one for male and zero otherwise. For portfolio size and turnover, we first calculate the time series average (across months) of end-of-month total assets and monthly turnover ratio, respectively. Age and Account Age are measured in years. Portfolio size is reported in logarithms. We report the annualized number for the turnover ratio. Commission is reported in basis points.

Variables	NoObs	Mean	Std.	10th	25th	Median	75th	90th
Age (Dec 2017)	682,596	45.173	16.679	28.691	36.280	45.871	54.802	64.874
Male	642,337	0.535	0.499	0.000	0.000	1.000	1.000	1.000
Account age (Dec 2017)	682,596	9.717	5.932	2.583	3.250	9.917	14.083	18.583
Portfolio size	712,058	9.891	2.000	7.314	8.756	10.067	11.216	12.216
Turnover	711,109	7.225	12.132	0.214	0.978	3.243	8.415	17.805
Commission (Dec 2017)	629,616	14.430	10.689	3.000	3.000	12.000	28.000	30.000

Table 2: Summary statistics of behavioral biases

This table reports the summary statistics of investor-level behavioral biases. We report the number of observations, sample means, standard deviations, and several percentiles for initial commission rate (in basis points) and the behavioral bias variables. The behavioral bias measures are computed using the full sample. Lottery, Extrapolation, and Rank effect are reported in percentage points. A composite bias measure is constructed by averaging the discretized percentile ranks (ranging from 1 for the lowest to 100 for the highest) of the six behavioral bias variables. The last row presents summary statistics for the composite bias measure. See Appendix IA.1 for the detailed definitions of these biases.

Variables	NoObs	Mean	Std.	10th	25th	Median	75th	90th
Initial rate	453,757	19.833	11.022	3.000	9.900	28.000	30.000	30.000
Disposition effect	458,312	0.739	0.212	0.500	0.631	0.781	0.897	0.964
Lottery	459,592	16.878	16.339	0.000	5.451	13.773	23.424	35.559
Underdiversification	481,545	-0.716	0.564	-1.480	-1.075	-0.660	-0.248	-0.000
Local bias	450,205	0.099	0.441	-0.207	-0.083	0.033	0.183	0.409
Extrapolation	459,592	10.071	14.098	-5.486	1.817	9.477	17.587	26.370
Rank effect	237,148	5.098	23.698	-19.231	-3.030	5.533	14.074	28.938
Composite bias	457,549	53.336	15.920	36.200	43.500	51.500	60.400	71.167

Table 3: Regressions of behavioral biases on the initial commission rate

This table reports results from linear regressions of investor biases on their initial commission rate. The biases are computed using data from month $t + 2$ to the end of our sample, with t being the account opening month). See Appendix IA.1 for the detailed definitions of these biases. *Initial rate* is represented in percentage points. *Portfolio size* is the natural logarithm of the total asset value at the end of the month immediately after the account opening, i.e., month $t + 1$ for investors who opened their accounts in month t. *Age* denotes the investor's age (in years) as of the account opening date. *Male* is a dummy that equals 1 for men and 0 for women. In all regressions, we include year-month fixed effects and branch fixed effects. Standard errors are clustered by year-month. Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1) Disposition effect	(2) Turnover	(3) Lottery	(4) Under diversifi.	(5) Local bias	(6) Extrapo lation	(7) Rank effect
	Panel A: No control						
Initial rate	0.083***	-5.698***	9.140***	0.477***	-0.011	0.203	1.852***
	(8.762)	(-10.886)	(12.131)	(20.161)	(-0.968)	(0.359)	(2.649)
Obs	425,026	441,473	426,209	440,725	419,002	426,209	230,522
Adj-R^2	0.040	0.061	0.026	0.022	0.040	0.039	0.001
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Panel B: With controls						
Initial rate	0.092***	-6.651***	8.229***	0.393***	-0.009	0.060	2.318***
	(9.182)	(-14.622)	(12.830)	(13.899)	(-0.816)	(0.100)	(3.283)
Portfolio size	-0.001***	-0.249***	-0.299***	-0.025***	-0.002***	0.085**	0.005
	(-5.026)	(-8.718)	(-27.621)	(-16.219)	(-8.192)	(2.176)	(0.224)
Age	0.001***	-0.038***	-0.016***	-0.005***	0.000***	-0.030***	0.040***
	(13.344)	(-4.114)	(-4.565)	(-16.560)	(3.495)	(-4.400)	(7.061)
Male	-0.036***	2.157***	0.050	0.045***	-0.004***	-0.502***	-0.629***
	(-28.614)	(11.511)	(0.980)	(15.577)	(-3.286)	(-8.937)	(-7.762)
Obs	406,197	419,353	406,221	418,703	406,221	406,221	222,702
Adj-R^2	0.044	0.054	0.030	0.057	0.041	0.025	0.002
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4: Panel regressions of commission rates dispersion on the number of brokers in a city

This table reports the results from panel regressions of various commission rate distribution variables on the logarithm of the number of brokers in a city. The dispersion variables are defined as per Fig. 8. In all regressions, we include year-quarter fixed effects and city fixed effects. Standard errors are clustered by city. Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1) $\frac{P10}{Mean}$	(2) $\frac{P25}{Mean}$	(3) $\frac{P75}{Mean}$	(4) $\frac{P90}{Mean}$	(5) $\frac{P90-P10}{Mean}$	(6) $\frac{Std}{Mean}$
$log(N_{Brokers})$	-0.376***	-0.206***	0.252***	0.247***	0.624***	0.150***
	(-3.997)	(-7.233)	(3.932)	(4.519)	(6.261)	(3.238)
Obs	2,642	2,642	2,642	2,642	2,642	2,642
Adj-R^2	0.361	0.327	0.119	0.219	0.306	0.443
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: Difference-in-differences analyses of dispersion

This table reports results from the following difference-in-differences regressions of the dispersion of commission rates: $Y_{it} = \beta\, Treat_i * Post_t + \gamma_{\text{cohort}(i)} + \lambda_{\text{branch}(i)} + \delta_{\text{event}(t)} + \epsilon_{it}$, where the dependent variable is a measure of the dispersion of commission rates. We consider two measures: (1) the standard deviation of the commission rates divided by the mean commission rates; (2) the difference between the 90th and the 10th percentiles of the commission rates, also divided by the mean commission rates. The dispersion measures are calculated at the branch-quarter level based on new investors, defined as those opening accounts in that branch and quarter. $Treat_i$ equals one for treated branches and zero otherwise. $Post_t$ equals zero for the pre-event quarters and one otherwise. We include cohort, branch, and event-time fixed effects in the regression. Standard errors are clustered by branch. Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)
	Std/Mean	Std/Mean	$\frac{P90-P10}{Mean}$	$\frac{P90-P10}{Mean}$
Treat×Post	0.178***	—	0.206***	—
	(2.641)		(2.806)	
Treat×T_{-4}	—	0.226	—	0.140
		(1.427)		(1.181)
Treat×T_{-3}	—	0.044	—	0.057
		(0.381)		(0.564)
Treat×T_{-2}	—	0.101	—	0.142
		(0.826)		(1.099)
Treat×T_0	—	0.154*	—	0.243**
		(1.694)		(2.017)
Treat×T_1	—	0.160*	—	0.178*
		(1.686)		(1.805)
Treat×T_2	—	0.324***	—	0.255**
		(3.085)		(2.309)
Treat×T_3	—	0.326***	—	0.364***
		(3.221)		(2.868)
Obs	1,282	1,282	1,282	1,282
Adj-R^2	0.392	0.396	0.419	0.420
Cohort FE	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes
Event-time FE	Yes	Yes	Yes	Yes

Table 6: Rate renegotiation: Smart versus Naive

This table reports results from the following regression of the relative changes in commission rates: $-Change_pct_{it} = \beta_1\ Smart_{it} + \beta_2\ Smart_{it} * AccAge_{it} + \gamma\ Controls + \epsilon_{it}$, where the dependent variable is the negative value of relative percentage changes in commission rates. For each cohort of investors (defined as investors who renegotiate their rates in the same month), we sort them into two groups (more sophisticated and less sophisticated) by the median value of their composite bias measure, defined as per Table 2. *Smart* equals one for more sophisticated investors and zero otherwise. We sort observations into 30 one-bp bins based on their pre-change rates and include bin fixed effects in all regressions. Control variables include the logarithm of lagged portfolio size and the logarithm of lagged turnover value. Standard errors are clustered by time (year-month). Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)
	$\frac{-Change}{\text{Pre-rate}}$	$\frac{-Change}{\text{Pre-rate}}$	$\frac{-Change}{\text{Pre-rate}}$	$\frac{-Change}{\text{Pre-rate}}$	$\frac{-Change}{\text{Pre-rate}}$
Smart	1.355***	0.762***	0.276*	0.572***	0.406***
	(13.179)	(8.268)	(1.820)	(4.074)	(2.718)
Asset_{t-1}	—	1.209***	1.188***	1.153***	1.021***
		(18.297)	(17.919)	(18.138)	(13.146)
Turnover_{t-1}	—	0.249***	0.262***	0.230***	0.240***
		(4.485)	(4.764)	(4.497)	(3.937)
Smart×AccAge	—	—	0.132***	0.070**	0.108***
			(4.267)	(2.474)	(3.572)
Obs	352,672	352,672	352,672	352,494	315,512
Adj-R^2	0.369	0.382	0.384	0.400	0.421
Bin FE	Yes	Yes	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	No
AccAge FE	No	No	Yes	No	No
Bin*AccAge	No	No	No	Yes	No
Bin*Time*AccAge	No	No	No	No	Yes

Table 7: Rate renegotiation: Smart versus Naive

This table reports results from the following regression of changes in commission rates: $-Change_{it} = \beta_1\ Smart_{it} + \beta_2\ Smart_{it} * AccAge_{it} + \gamma\ Controls + \epsilon_{it}$, where the dependent variable is the negative value of changes in commission rates. For each cohort of investors (defined as investors who renegotiate their rates in the same month), we sort them into two groups (more sophisticated and less sophisticated) by the median value of their composite bias measure, defined as per Table 2. *Smart* equals one for more sophisticated investors and zero otherwise. We sort observations into 30 one-bp bins based on their pre-change rates and include bin fixed effects in all regressions. Control variables include the logarithm of lagged portfolio size and the logarithm of lagged turnover value. Standard errors are clustered by time (year-month). Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)
	$-Change$	$-Change$	$-Change$	$-Change$	$-Change$
Smart	0.331***	0.189***	0.202***	0.150***	0.093*
	(15.629)	(8.029)	(4.498)	(3.127)	(1.793)
$Asset_{t-1}$	—	0.248***	0.245***	0.240***	0.216***
		(20.201)	(19.229)	(19.575)	(14.292)
$Turnover_{t-1}$	—	0.082***	0.082***	0.075***	0.071***
		(5.555)	(5.646)	(5.570)	(4.387)
Smart×AccAge	—	—	-0.004	0.013	0.027***
			(-0.458)	(1.517)	(3.093)
Obs	352,672	352,672	352,672	352,494	315,512
Adj-R^2	0.417	0.429	0.430	0.447	0.441
Bin FE	Yes	Yes	Yes	No	No
Time FE	Yes	Yes	Yes	Yes	No
AccAge FE	No	No	Yes	No	No
Bin*AccAge	No	No	No	Yes	No
Bin*Time*AccAge	No	No	No	No	Yes

Table 8: Post-change gap: Smart versus Naive

This table reports results from the following regression of post-change gaps in commission rates: $GapPos_{it} = \beta_1\ Smart_{it} + \beta_2\ Smart_{it} * AccAge_{it} + \gamma\ Controls + \epsilon_{it}$, where the dependent variable equals one if the post-change gap is positive and zero otherwise. Here gap is defined as the difference between investors' commission rates and the prevailing rate of new investors within the same branch who opened their accounts during the most recent 6-month period. For each cohort of investors (defined as investors who renegotiate their rates in the same month), we sort them into two groups (more sophisticated and less sophisticated) by the median value of their composite bias measure, defined as per Table 2. *Smart* equals one for more sophisticated investors and zero otherwise. Control variables include the logarithm of lagged portfolio size and the logarithm of lagged turnover value. Standard errors are clustered by time (year-month). Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)
	GapPos	*GapPos*	*GapPos*	*GapPos*
Smart	-2.009***	-0.461***	1.095***	1.177***
	(-6.303)	(-2.788)	(5.048)	(5.544)
$Asset_{t-1}$	—	-1.428***	-1.527***	-1.524***
		(-8.899)	(-9.030)	(-8.666)
$Turnover_{t-1}$	—	-0.872***	-0.774***	-0.781***
		(-6.754)	(-6.414)	(-6.271)
Smart×AccAge	—	—	-0.448***	-0.470***
			(-9.282)	(-10.011)
Obs	352,672	352,672	352,672	352,415
Adj-R^2	0.052	0.086	0.092	0.098
Time FE	Yes	Yes	Yes	No
AccAge FE	No	No	Yes	No
Time*AccAge	No	No	No	Yes

Table 9: Post-change gap: Smart versus Naive

This table reports results from the following regression of post-change gaps in commission rates: $GapPct_{it} = \beta_1\ Smart_{it} + \beta_2\ Smart_{it} * AccAge_{it} + \gamma\ Controls + \epsilon_{it}$, where the dependent variable is the relative post-change gap (gap divided by prevailing rate, in percent). Here gap is defined as the difference between investors' commission rates and the prevailing rate of new investors within the same branch who opened their accounts during the most recent 6-month period. For each cohort of investors (defined as investors who renegotiate their rates in the same month), we sort them into two groups (more sophisticated and less sophisticated) by the median value of their composite bias measure, defined as per Table 2. *Smart* equals one for more sophisticated investors and zero otherwise. Control variables include the logarithm of lagged portfolio size and the logarithm of lagged turnover value. Standard errors are clustered by time (year-month). Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)
	GapPct	*GapPct*	*GapPct*	*GapPct*
Smart	-6.161***	-1.575***	1.408***	1.629***
	(-9.148)	(-4.844)	(3.800)	(4.157)
$Asset_{t-1}$	—	-4.723***	-4.952***	-4.912***
		(-13.446)	(-12.464)	(-11.909)
$Turnover_{t-1}$	—	-2.345***	-2.096***	-2.113***
		(-6.955)	(-7.070)	(-6.769)
Smart×AccAge	—	—	-0.880***	-0.959***
			(-7.177)	(-7.906)
Obs	352,672	352,672	352,672	352,415
Adj-R^2	0.043	0.093	0.102	0.105
Time FE	Yes	Yes	Yes	No
AccAge FE	No	No	Yes	No
Time*AccAge	No	No	No	Yes

Table 10: Post-change gap: Smart versus Naive

This table reports results from the following regression of post-change gaps in commission rates: $Gap_{it} = \beta_1\ Smart_{it} + \beta_2\ Smart_{it} * AccAge_{it} + \gamma\ Controls + \epsilon_{it}$, where the dependent variable is the post-change gap (in basis points). Here gap is defined as the difference between investors' commission rates and the prevailing rate of new investors within the same branch who opened their accounts during the most recent 6-month period. For each cohort of investors (defined as investors who renegotiate their rates in the same month), we sort them into two groups (more sophisticated and less sophisticated) by the median value of their composite bias measure, defined as per Table 2. *Smart* equals one for more sophisticated investors and zero otherwise. Control variables include the logarithm of lagged portfolio size and the logarithm of lagged turnover value. Standard errors are clustered by time (year-month). Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)
	Gap	*Gap*	*Gap*	*Gap*
Smart	-0.695***	-0.214***	-0.262***	-0.268***
	(-15.727)	(-7.399)	(-6.713)	(-6.659)
$Asset_{t-1}$	—	-0.523***	-0.535***	-0.527***
		(-24.165)	(-25.257)	(-25.402)
$Turnover_{t-1}$	—	-0.232***	-0.219***	-0.216***
		(-9.923)	(-10.117)	(-9.731)
Smart×AccAge	—	—	0.010	0.007
			(1.187)	(0.809)
Obs	352,672	352,672	352,672	352,415
Adj-R^2	0.155	0.220	0.223	0.231
Time FE	Yes	Yes	Yes	No
AccAge FE	No	No	Yes	No
Time*AccAge	No	No	No	Yes

Table 11: Parameter estimates from MSM
This table reports parameter estimates from method of simulated moments. The estimation procedure is detailed in Section IA.4.

Parameter	Meaning	Estimates
α	Rate of investor attention	0.1213
MC_t	Marginal cost at time t	See Fig. IA.2
$\bar{\pi}_k$	Population smartness at each branch k	See Fig. 11
κ_k	Naive investor's base search cost	See Fig. 12
m	Load of trading volume in naive investor's search cost	0.7655
t_0	Intensity of initial interaction	0.1240
σ	Rate of learning on investor type	0.2851

Internet Appendix to

"When Competition Backfires: Broker Learning and Commission Fees"

Xindan Li

Baolian Wang

Wei Xiong

Xuewei Yang

Yinhong Zhao

August, 2025

In this appendix, we provide some supplementary illustrations, technical proofs, and ancillary empirical results.

IA.1 Definition of Investor Biases

In this appendix, we provide details about how to compute the measures of investor biases that we consider in this paper.

A. The Disposition Effect

The first bias we consider is the disposition effect — the tendency to sell stocks that earn capital gains rather than stocks that incur capital losses — which is perhaps the most robust empirical finding on investor trading (Shefrin and Statman, 1985; Odean, 1998). To examine the disposition effect, we first construct a holding sample that contains observations for all investor-stock-day triples. Following the literature (Seru et al., 2010; Ben-David and Hirshleifer, 2012), we count every purchase of a stock as the beginning of a new position, and a position ends on the date the investor liquidates all of her holdings. Investor-days on which an investor does not trade are included in the analysis. Purchases and sales are aggregated on a daily basis for each investor-stock. Those days on which a position was sold (including partial sales) are defined as selling days. We restrict our analyses to common stocks and exclude positions for which we do not have information on the purchase price (value-weighted average price), which primarily arise if investors purchased stocks before the start of our sample period. We calculate holding-period returns after adjusting for stock splits and dividends. Information on stock prices and distribution comes from the China Stock Market and Accounting Research Database (CSMAR). On day t when an investor can sell a stock, it is considered at a gain if the stock's purchase price is lower than the closing price on day $t-1$. We then calculate the proportion of gains realized (PGR) and the proportion of losses realized (PLR) based on the investor-stock-day sample. The disposition effect is defined as $PGR/(PGR+PLR)$.

B. Lottery preference

We calculate each investor's lottery preference using the value-weighted percentage of lottery stocks an investor buys, where lottery stocks are defined following Kumar (2009). Specifically, at the end of each month t, across all domestically listed Chinese stocks, across all domestically listed Chinese stocks, those with a stock price at the end of month $t-1$ below the median, idiosyncratic volatility higher than the median, and idiosyncratic

skewness higher than the median are classified as lottery stocks. We calculate idiosyncratic volatility and idiosyncratic skewness using daily data from the past six months. Idiosyncratic volatility is the variance of the residuals obtained by fitting a one-factor model, where the factor is the market excess return. Idiosyncratic skewness is the scaled measure of the third moment of the residuals obtained by fitting a two-factor model to daily stock returns, where the two factors are excess market returns and squared excess market returns (Harvey and Siddique, 2000).

C. Underdiversification

For an investor, we count the number of stocks in her portfolio every day and use the time-series average as our measure of diversification. For all other bias measures, a higher value is associated with a stronger bias. To ease exposition, we use the negative value of the natural logarithm of the average number of stocks as our measure of underdiversification.

D. Local bias

It is well documented that investors tend to buy stocks geographically close to them. Our measure of local bias is the benchmark-adjusted average distance between an investor's residence city and the headquarters of the stocks they bought. We assume that an investor lives in the city where her brokerage branch is located, as Chinese investors typically open an account near where they reside (Feng and Seasholes, 2004). We first calculate the value-weighted average distance (in kilometers) between an investor and her purchased stocks. The benchmark is the market capitalization-weighted average distance between this investor's city and the headquarters of all the listed stocks. The information about the firm headquarters is from CSMAR. Our local bias measure is defined as $Log(1 + benchmark\ distance) - Log(1 + average\ distance)$. Investors for whom the benchmark distance is longer than the average distance are more likely to exhibit local bias.

E. Extrapolation

Investors tend to buy stocks with abnormally high past returns (Barber and Odean, 2013). Our extrapolation measure is calculated as volume-weighted average stock excess returns, where the excess return of a stock is calculated as the return on the stock over the 120 trading days preceding the purchase minus the market return over the same period.

F. The rank effect

The "rank effect" refers to the fact that investors have a greater tendency to sell their best and worst positions when ranked by holding-period returns (Hartzmark, 2015). We estimate the rank effect at the individual level using the following regression:

$$Sell_{i,k,t} = \varphi_i + \gamma_i * Extreme_{i,k,t-1} + \epsilon_{i,k,t}, \tag{IA.1}$$

where $Sell_{i,k,t}$ is a dummy variable that equals 1 if investor i sells stock k (fully or partially) on day t and 0 otherwise; $Extreme_{i,k,t-1}$ is a dummy variable that takes the value of one if stock k has the highest or lowest return in investor i's portfolio on day $t-1$ and zero otherwise. Following (Hartzmark, 2015), we restrict the sample to investor-days in which the investor sold at least one stock and there are at least five stocks in the investor's portfolio. A higher value of $\hat{\gamma}_i$ indicates a higher rank effect.

IA.2 Endogenous Search

This section attempts to endogenize investors' reserve price from a model of search costs. We model investor's problem every time investors pay attention to commissions either in the entry stage or the trading stage. Investors evaluate the broker's offer with imperfect dynamic foresight. They assume that they would hold this offer in their trading in all future periods. With no implication for trading, the disutility from commissions is

$$u(p,q) = \int_{t=0}^{\infty} e^{-\rho t}(-pq)dt = -\frac{q}{\rho}p.$$

Following Stahl (1989), we suppose smart investors have search costs $\kappa_s = 0$, and naive investors have search costs $\kappa_d > 0$. Moreover, investors with larger trading volume are often richer and have a higher opportunity cost for search, so we further parameterize $\kappa_d = \kappa q^m$, with $m \in [0,1]$ being a constant to be determined later.

Given an observed price z, the investor will weigh the expected surplus against its exogenous search cost to decide if it continues to search, i.e.,

$$\int_{\phi(1)}^{z} \frac{q}{\rho}(z-p)dG_i(p) = \kappa_i \iff z = \mathbb{E}_{p\sim G_i}[p|p \le z] + \frac{\rho\kappa_i}{q}. \tag{IA.2}$$

This generates a unique reserve price $z = r_i, i = s, d$. Consumer i would continue to search if $z > r_i$ and stop searching if $z < r_i$. With $\kappa_s = 0$, smart investors continue to search with probability one. Meanwhile, brokers would never respond with a price higher than naive investors' reserve price. So for any $p \le r^d$, $\mathbb{E}_{p\sim G_d}[p|p \le z] = \mathbb{E}_{p\sim G_d}[p] = \mathbb{E}_{\pi\sim F_d}[\phi(\pi; r^d)]$, and the reserve price of naive investors r^d should satisfy

$$r^d = \mathbb{E}_{\pi\sim F_d}[\phi(\pi; r^d)] + \frac{\rho\kappa_d}{q} \tag{IA.3}$$

$$= MC + (r^d - MC)\mathbb{E}_{\pi\sim F_d}[\mu(\pi)] + \frac{\rho\kappa}{q^{1-m}} \tag{IA.4}$$

This gives

$$(r^d - MC)\big(1 - \mathbb{E}_{\pi\sim F_d}[\mu(\pi)]\big) = \frac{\rho\kappa}{q^{1-m}}.$$

Notice that $\mathbb{E}_{\pi\sim F_d}[\mu(\pi)] = \int_0^1 \mu(\pi) f_d(\pi) d\pi < \int_0^1 f_d(\pi) d\pi = 1$, so we can rewrite the equation as

$$r^d = MC + \frac{\rho}{1 - \mathbb{E}_{\pi\sim F_d}[\mu(\pi)]} \cdot \frac{\kappa}{q^{1-m}} \equiv MC + \frac{A\kappa}{q^{1-m}},$$

where $A = \frac{\rho}{1-\mathbb{E}_{\pi\sim F_d}[\mu(\pi)]}$ is a constant that does not depend on MC or r^d since $\mu(\pi)$ and $f_s(\pi)$ both don't depend on them. This functional form justifies our parametrization of reserve price in Section 4.

IA.3 Proofs

Proof of Proposition 4.1:

Lemma 3.1.1: The value function v is convex on $(0, 1)$, i.e., $v''(\pi) \geq 0$.

Proof of Lemma 3.1.1: Let $Q(\pi) = \frac{2\sigma^2(\rho+\alpha)}{\pi^2(1-\pi)^2} \geq 0$, $R(\pi) = \alpha(1-\pi+\pi F_s^n(\pi))(v(\pi_A)+\frac{\mu(\pi)}{\rho+\alpha}) \geq 0$ since $v(\pi_A) + \frac{\mu(\pi)}{\rho+\alpha} = \frac{V(\phi(\pi),MC,\pi_A)}{(r^d-MC)q} \geq 0$. We can then write Eq. (9) as a Dirichlet problem

$$-v''(\pi) + Q(\pi)v = R(\pi).$$

We first define a boundary filter ψ that satisfies $\psi''(\pi) = Q(\pi)\psi$ and $\psi(0) = v(0), \psi(1) = v(1)$. The one-dimensional minimum principle guarantees that $\psi(\pi) \geq 0$ since $\psi(0) \geq \psi(1) \geq 0$. Since $Q(\pi) \geq 0$, we know $\psi''(\pi) \geq 0$.

Now let $w = v - \psi$. We have $w(0) = w(1) = 0$, and

$$-w''(\pi) + Q(\pi)w = R(\pi).$$

We can then use a Green function representation to characterize w. The local operator $\mathcal{L}[u] := -u'' + Qu$ has a Green function $G(\pi, s) > 0$ which is $\mathcal{C}^2$ off the diagonal and satisfies for each fixed s, $\mathcal{L}_\pi G(\pi, s) = \delta(\pi - s)$. Since w solves $\mathcal{L}[w] = R$, we can write

$$w(\pi) = \int_0^1 G(\pi, s)R(s)ds,$$

whose second derivative is

$$w''(\pi) = \int_0^1 G_{\pi\pi}(\pi, s)R(s)ds.$$

It is also known that the Green function can be written as

$$G(\pi, s) = \begin{cases} g_1(\pi)g_2(s)/\hat{W} & \text{if } \pi < s, \\ g_1(s)g_2(\pi)/\hat{W} & \text{if } \pi > s, \end{cases}$$

where g_1, g_2 both solve the homogeneous equation $-g'' + Q(\pi)g = 0$. Therefore, $G_{\pi\pi} = Q(\pi)G(\pi, s) \geq 0$, so $w''(\pi) \geq 0$.

Since we have shown that $\psi''(\pi) \geq 0$, so $v''(\pi) = \psi''(\pi) + w''(\pi) \geq 0$. v is convex.

Lemma 3.1.2: Let

$$T(\pi) = \frac{\mu(\pi)}{\rho + \alpha} + v(\pi_A) + (1 - \pi_A)v'(\pi_A),$$

then $T(\pi) \geq 0$ for any $\pi \in [0, 1]$.

Proof of Lemma 3.1.2: We first notice that $T(0) \geq 0$. This is because $\mu(\pi) \leq \mu(0) = 1$ for any π, so $\mu'(0) \leq 0$, which means that $T(0) \geq 0$. Also $T(1) = \frac{\mu(1)}{\rho} \geq 0$.

Suppose there exists $\pi \in [0, 1]$ such that $T(\pi) < 0$. We let $\pi^* = \inf\{\pi : T(\pi) \leq 0\}$. This means that $T(\pi) \geq 0$ for any $\pi \leq \pi^*$, $T(\pi^*) = \mu'(\pi^*) = 0$, and $T'(\pi^*) < 0$. Now compute the derivative of T at π^*:

$$\begin{aligned} T'(\pi^*) &= \frac{\mu'(\pi)}{\rho + \alpha} + v'(\pi_A)\pi_A'(\pi) + (1 - \pi_A)v''(\pi_A)\pi_A'(\pi) - \pi_A'(\pi)v'(\pi_A) \\ &= (1 - \pi_A)v''(\pi_A)\pi_A'(\pi) \geq 0 \end{aligned}$$

since $\pi_A < 1$, $v''(\pi_A) \geq 0$ by Lemma 3.1.1, and $\pi_A'(\pi) \geq 0$ by definition of π_A. This leads to a contradiction. So $T(\pi) \geq 0$ for any π.

Proof of Proposition: From Eq. (10), we have

$$\mu'(\pi) = -\mu(\pi)T(\pi).$$

Since $T(\pi) \geq 0$, and $\mu(0) \geq \mu(1) \geq 0$, it follows that $\mu'(\pi) \leq 0$. This proves that the pricing rule $\phi(\pi) = MC + (r^d - MC)\mu(\pi)$ is decreasing in π.

We have shown that $\pi_t = \frac{e^{x_t}}{1+e^{x_t}}$, where x_t follows a normal distribution conditional each investor type. Notice that $\mathbb{E}[x_t|i = s] > \mathbb{E}[x_t|i = d]$ and $Var[x_t|i = s] = Var[x_t|i = d]$ for any $t > -t_0$, we have F_s has first-order stochastic dominance over F_d. Since ϕ is a monotonically decreasing function and $G_i(\phi(\pi)) = 1 - F_i(\pi)$, we obtain G_d has first-order stochastic dominance over G_s.

Proof of Proposition 4.2:

We have already shown $\mu(\pi)$ is also decreasing in π. Since $v(1) = \frac{1}{\rho}\mu(1) \geq 0$, we have $\mu(1) \geq 0$, which means that $\mu(\pi) \geq 0$ for any $\pi \in [0, 1]$ by monotonicity.

Suppose there exists $\pi \in [0, 1)$ such that $\mu(\pi) = 0$, then let $\pi_0 = \inf\{\pi : \mu(\pi) = 0\}$. Since μ is non-increasing and $\mu(\pi) \geq 0$, we have $\mu(\pi) = \mu'(\pi) = 0$ for any $\pi \in [\pi_0, 1]$. Plug this into Eq. (10), we have $T(\pi) = v(\pi_A) + (1 - \pi_A)v'(\pi_A) = 0$ for any $\pi \in [\pi_0, 1]$ or $\pi_A \in [\pi_A(\pi_0), 1]$. This means that $v(\pi) = c(1 - \pi)$ for $\pi \in [\pi_A(\pi_0), \pi_0]$ and a constant c, so $v''(\pi) = 0$. If $c \neq 0$, then $(\rho + \alpha)(1 - \pi) = \alpha(1 - \pi + \pi F_s(\pi)^n)(1 - \pi_A) = \alpha(1 - \pi)$, so $\rho = 0$, which is a contradiction.

If $c = 0$, then $v(\pi) = 0$ for $\pi \in [\pi_A(\pi_0), 1]$. Now consider $\pi_1 \in (\pi_A(\pi_0), \pi_0)$, then $v(\pi_1) = v'(\pi_1) = v''(\pi_1) = 0, \mu(\pi_1) > 0$. Plug this back to Eq. (9), we have $v(\pi_A(\pi_1)) + \frac{\mu(\pi_1)}{\rho+\alpha} = 0$, which means that $v(\pi_A(\pi_1)) < 0$. Since $v(\pi_A(\pi_0)) = 0$, there exists $\pi_2 \in [\pi_A(\pi_0), \pi_1]$ such that $v(\pi_A(\pi_2)) < 0$ and $v'(\pi_A(\pi_2)) < 0$. Similarly, we still have $v(\pi_A(\pi_2)) + \frac{\mu(\pi_2)}{\rho+\alpha} = 0$. Plug this to *Eq.* (10), we get $T(\pi_2) = v'(\pi_A(\pi_2))(1 - \pi_A(\pi_2)) < 0$. However, we have proved that $T(\pi) \geq 0$, so this leads to another contradiction.

Therefore, we conclude that $\mu(\pi) > 0$, and thus $\phi(\pi; r^d) > MC$ $\pi \in [0, 1)$. Lastly, notice that r_s is neither involved in the broker nor in the investor's problem, so r_s does not affect any outcomes. Moreover, if $r^d(q)$ is a decreasing function as we parameterized $r^d(q) = MC + \frac{\kappa}{q^{1-m}}$, then $\phi(\pi)$ is also decreasing with q.

Proof of Proposition 4.4:

Suppose the distribution of belief conditional on investor type at time t is $\pi_{it} \sim F_{it}$. We have shown that $\pi_{it} = \frac{e^{x_{it}}}{1+e^{x_{it}}} = logit(x_{it})$, and x_{it} follows a normal distribution for each $i = s, d$.

As a primer, monotone functions preserve a one-to-one mapping between quantiles. This means that the $\tau - th$ of smart investors pricing

$$p_s = \phi(\bar{\pi}_{st}) = \phi(logit(\bar{x}_{st})) = \phi\left(logit\left(x_{-t_0} + \frac{t+t_0}{2\sigma^2}\right)\right),$$
$$p_d = \phi(\bar{\pi}_{dt}) = \phi(logit(\bar{x}_{dt})) = \phi\left(logit\left(x_{-t_0} - \frac{t+t_0}{2\sigma^2}\right)\right),$$

where $\bar{\pi}_{it}, \bar{x}_{it}$ are the median of π and x at time t for investor type i. With ϕ a decreasing function and $logit$ an increasing function, p_s is decreasing in t, and p_d is increasing in t. Therefore, $p_s(t) - p_d(t)$ is increasing in t.

Proof of Proposition 4.5:

By the fundamental theorem of calculus, we have

$$\mu(\pi) - \mu(0) = \int_0^{\pi} \mu'(x)dx = -\int_0^{\pi} (\rho + \alpha)H(x)T(x)dx.$$

Since both v and μ are bounded in $[0,1]$, T is bounded in $[0,1]$. We can let $M = \sup_{\pi\in[0,1]}\{T(\pi)\}$. So $\mu(\pi) \geq 1 - M\int_0^{\pi} H(x)dx$. Notice that

$$H(x) = \frac{nF_s^{n-1}(x)\, x f_s(x)}{1 - x + xF_s^n(x)} \leq nF_s^{n-1}(x)\, f_s(x).$$

Also notice that $F_s(x) \leq F_s(\pi) < F_s(1) = 1$. Let $\bar{f}_s = \sup_{x\in[0,1]} f_s(x)$ and $\tilde{F}_s = -\ln F_s(\pi) > 0$, so we have

$$\lim_{n\to\infty} nF_s(\pi)^{n-1} = \lim_{n\to\infty} \frac{n}{e^{\tilde{F}_s n}} = \lim_{n\to\infty} \frac{1}{\tilde{F}_s e^{\tilde{F}_s n}} = 0$$

by l'Hôpital's rule. By definition of limit, for any $\epsilon > 0$, we know there exists N_ϵ such that for any $n > N_\epsilon$, $nF_s(\pi)^{n-1} < \frac{\epsilon}{\bar{f}_s M}$, which means that $nF_s(x)^{n-1} < \frac{\epsilon}{\bar{f}_s M}$ for any $x \in [0,\pi]$ since

F_s is increasing. Therefore, we have

$$\begin{aligned}\mu(\pi) &\geq 1 - M\int_0^{\pi} H(x)dx \\ &\geq 1 - M\bar{f}_s\int_0^{\pi} nF_s^{n-1}(x)dx \\ &\geq 1 - M\bar{f}_s\pi\frac{\epsilon}{\bar{f}_s M} \geq 1-\epsilon.\end{aligned}$$

In conclusion, we proved that for any $\epsilon > 0$, there exists N_ϵ such that for any $n > N_\epsilon$, $0 \leq 1 - \mu(\pi) < \epsilon$. So $\mu(\pi)$ converges point-wise to 1 for $\pi \in [0, 1)$.

IA.4 Estimation Details

1. We set the annual discount rate $\rho = 0.0995$ following Gormsen and Huber (2025).[1]

2. We measure n_k directly from our data as the average number of competing brokers in each prefecture over time. Although the number of competitors may vary slightly due to brokers' entry and exit, we treat n_k as constant to reduce computational burden, using the time average as an approximation.

3. We measure $\hat{\alpha}$ as the average attention frequency in our data, defined as the average number of times each investor's commission rate is updated. For example, if the commission rate is updated twice per year for the average investor, then $\alpha = 2$.

4. From this point on, we focus on the subsample where the commission rate was updated. We first estimate MC_t by taking $q \to \infty$ in Eq. (1), which implies that

$$MC_t = \lim_{q\to\infty} r_{kt}^d(q) = \lim_{q\to\infty} \phi_t(\pi|q).$$

   We approximate this limit by selecting, in each quarter t across all branches, the investors with the top 10% trading volume and taking the mean of the lowest 10%

[1]The discount rate is not identified in most dynamic models. We therefore follow the literature and calibrate its value using existing evidence. While comparable data from China are unavailable, we approximate brokers' discount rates using the average discount rate for Japanese firms, measured directly from earnings call.

of their commission rates as our estimate of MC_t.

5. Note that the reserve price of naive investors is defined for those deemed naive, i.e., $r^d(q) = \phi(\pi = 0|q)$. For each quarter t and branch k, we proxy for $\phi(\pi = 0|q)$ by the highest 10% of commission rates within each trading volume bin. This yields estimates of naive investors' reserve prices at the quarter-branch-volume level, $r^d_{kt}(q)$.

6. Define $\bar{\kappa}_k = A_k \kappa_k$.[2] We estimate two parameters $\bar{\kappa}_k$ and m for each branch using Eq. (1):

$$r^d_{kt}(q) - MC_t = \frac{\bar{\kappa}_k}{q^{1-m}}.$$

Take logs, we obtain

$$\ln(r^d_{kt}(q) - MC_t) = \ln(\bar{\kappa}_k) - (1 - m)\ln q.$$

This is a linear model with $\ln(r^d_{kt}(q) - MC_t)$ as the dependent variable and $\ln(q)$ as the independent variable. The slope term $(1 - m)$ is a constant, but the intercept term $\ln(\bar{\kappa}_k)$ may vary across locations.

We employ a within-estimator to estimate $\bar{\kappa}_k$ and m. We first demean both $\ln(r^d_{kt}(q) - MC_t)$ and $\ln(q)$ at each branch k, so that they are mean-zero at the branch level. We then run a linear regression of the demeaned $\ln(r^d_{kt} - MC_t)$ on demeaned $\ln(q)$ without constant, pooling data across all quarters, to recover the $-(1 - m)$ and thus the constant m. Finally, we estimate

$$\ln(\bar{\kappa}_k) = \mathbb{E}_k[\ln(r^d_{kt}(q) - MC_t)] + (1 - \hat{m})\mathbb{E}_k[\ln(q)]. \quad \forall k$$

[2] Note that $A_k = \frac{\rho}{1 - \mathbb{E}_{\pi \sim F_d}[\mu_k(\pi)]}$ depends on the parameter μ_k, which varies across branches.

This is equivalent to matching the following moments:

$$\forall k, \quad \mathbb{E}\left[\ln(r_{kt}^d(q) - MC_t) - \ln(\bar{\kappa}_k) + (1-m)\ln(q)\right] = 0.$$

$$\mathbb{E}\Bigg[\Big(\ln(r_{kt}^d(q) - MC_t) - \mathbb{E}_k[\ln(r_{kt}^d(q) - MC_t)] + (1-m)[\ln(q) - \mathbb{E}_k[\ln(q)]]\Big) \times\Big(\ln(q) - \mathbb{E}_k[\ln(q)]\Big)\Bigg] = 0.$$

7. With the estimated $\bar{\kappa}_k$ and MC_t, we compute the empirical markup for each investor i in quarter t as
$$\widehat{\mu_k(\pi_{it})} = \frac{p_{it} - MC_t}{r_{kt}^d(q_{it}) - MC_t} = \frac{p_{it} - MC_t}{\bar{\kappa}_k / q^{1-m}}.$$

8. We then guess a pair of values (t_0, σ) as the outer loop of the subsequent estimation. These parameters are ultimately calibrated to match the variance of initial commission rates and the covariance of the commission rate variance and account age.

9. We iterate over each branch k, guessing a value of $\bar{\pi}_k$ as the inner loop. Given $\bar{\pi}_k$ and the empirical n_k, we simulate the policy function μ_k for each branch. Using this function, we further simulate I_k investors in branch k for each quarter t and compute the theoretical average markup of initial commissions, $\bar{\mu}_{kt}$. We then search for the $\bar{\pi}_k$ that equates the theoretical and empirical means:
$$\frac{1}{T}\sum_t \frac{1}{I_k}\sum_{i \in I_k} \widehat{\mu_k(\pi_{it})} = \frac{1}{T}\sum_t \bar{\mu}_{kt}.$$

10. Given $\bar{\pi}_k$ from the inner loop, we recalculate the policy function μ_k. We then simulate I_k individuals in each market, each entering at time t, and track their commission paths over the next h quarters. We compute the theoretical average of standard deviation of μ among all investors who entered at time t with age h, denoted σ_{th}. Finally, we regress σ_{th} on h to obtain γ_0^h and γ_1^h.

11. Meanwhile, we measure $\hat{\sigma}_{th}$, the standard deviation of $\widehat{\mu_k(\pi_{it})}$ from data at each time t for each account age h. We then regress $\hat{\sigma}_{th}$ on h to obtain $\hat{\gamma}_0^h$ and $\hat{\gamma}_1^h$. The outer

loop searches over (t_0, σ) to find the values that satisfy

$$\gamma_0^h = \hat{\gamma}_0^h, \quad \gamma_1^h = \hat{\gamma}_1^h.$$

12. Finally, We recover the baseline search cost κ_k by computing the market environment parameter at each branch $A_k = \frac{\rho}{1-\mathbb{E}_{\pi\sim F_d}[\mu_k(\pi)]}$ and divide it from $\bar{\kappa_k}$

$$\kappa_k = \bar{\kappa_k}/A_k.$$

IA.5 Additional Tables and Figures

IA.5.1 TBA

Table IA.1: Regressions of Behavioral Biases on the Initial Commission Rate
This table is the counterpart of Table 3. All settings are the same as those in Table 3 except that we measure the biases using data in the first year after account opening.

	(1) DispEff	(2) Turnover	(3) Lottery	(4) UnderDiver	(5) LocalBias	(6) Extrapola	(7) Rank
	Panel A: No control						
Initial rate	0.075***	-7.091***	11.193***	0.498***	0.033*	-0.574	2.440*
	(6.816)	(-11.886)	(16.416)	(15.887)	(1.908)	(-0.720)	(1.898)
Obs	371,539	412,258	387,260	417,482	380,106	387,260	148,610
Adj-R^2	0.020	0.077	0.038	0.036	0.025	0.070	0.002
	Panel B: With controls						
Initial rate	0.085***	-10.188***	8.880***	0.194***	0.012	0.704	2.614*
	(7.705)	(-17.131)	(12.113)	(5.963)	(0.648)	(0.786)	(1.869)
Portfolio size	0.000	-0.477***	-0.313***	-0.052***	-0.005***	0.327***	-0.026
	(0.231)	(-11.689)	(-14.980)	(-27.079)	(-12.028)	(6.287)	(-0.561)
Age	0.001***	-0.043***	-0.034***	-0.004***	0.001***	-0.026***	0.017**
	(10.238)	(-8.088)	(-6.483)	(-13.559)	(7.290)	(-3.414)	(2.081)
Male	-0.033***	2.002***	0.171***	0.024***	-0.006***	-0.157	-0.687***
	(-23.277)	(8.670)	(2.987)	(5.351)	(-3.987)	(-1.626)	(-4.675)
Obs	353,059	390,490	367,577	395,543	367,577	367,577	141,246
Adj-R^2	0.024	0.082	0.040	0.101	0.027	0.063	0.002

Table IA.2: Regressions of Behavioral Biases on the Initial Commission Rate
This table is the counterpart of Table 3. All settings are the same as those in Table 3 except that we measure the biases using data in the first two years after account opening.

	(1) DispEff	(2) Turnover	(3) Lottery	(4) UnderDiver	(5) LocalBias	(6) Extrapola	(7) Rank
	Panel A: No control						
Initial rate	0.068***	-7.048***	10.738***	0.464***	0.006	0.054	2.908***
	(6.207)	(-11.032)	(17.357)	(15.046)	(0.410)	(0.086)	(3.134)
Obs	395,919	425,270	404,778	426,680	397,571	404,778	175,307
Adj-R^2	0.027	0.062	0.043	0.032	0.028	0.059	0.001
	Panel B: With controls						
Initial rate	0.077***	-9.362***	8.785***	0.227***	-0.005	0.824	3.224***
	(6.816)	(-15.745)	(13.586)	(7.248)	(-0.338)	(1.171)	(3.247)
Portfolio size	-0.000*	-0.366***	-0.300***	-0.042***	-0.004***	0.235***	-0.048
	(-1.679)	(-12.316)	(-23.533)	(-27.677)	(-11.621)	(5.600)	(-1.395)
Age	0.001***	-0.040***	-0.026***	-0.004***	0.000***	-0.022***	0.031***
	(11.971)	(-5.530)	(-6.454)	(-15.805)	(4.725)	(-2.716)	(4.426)
Male	-0.034***	2.023***	0.195***	0.033***	-0.006***	-0.316***	-0.730***
	(-23.130)	(9.576)	(3.863)	(8.564)	(-4.269)	(-5.140)	(-5.770)
Obs	377,219	403,310	384,907	404,680	384,907	384,907	167,589
Adj-R^2	0.031	0.060	0.045	0.088	0.029	0.050	0.002

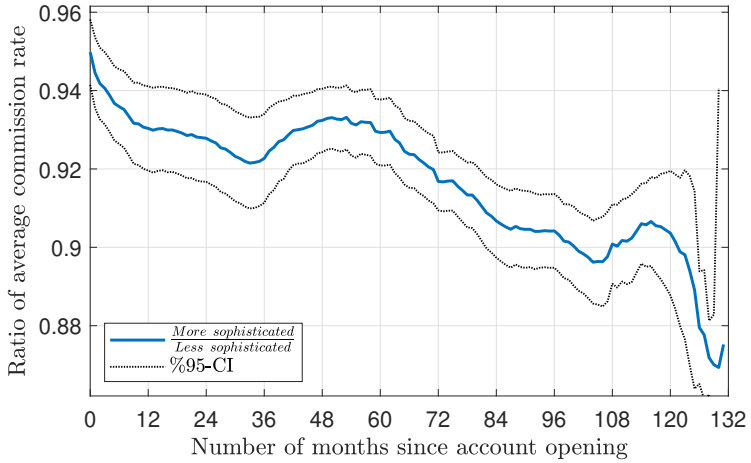


Figure IA.1: Ratio between average commission rates by investor group
This figure shows the evolution of commission rates by investor groups. For each cohort of investors (defined as investors who open their accounts in the same month), we sort them into two groups by the median value of their composite bias measure, defined as per Table 2. We report the ratio between the two groups' average commission rates.

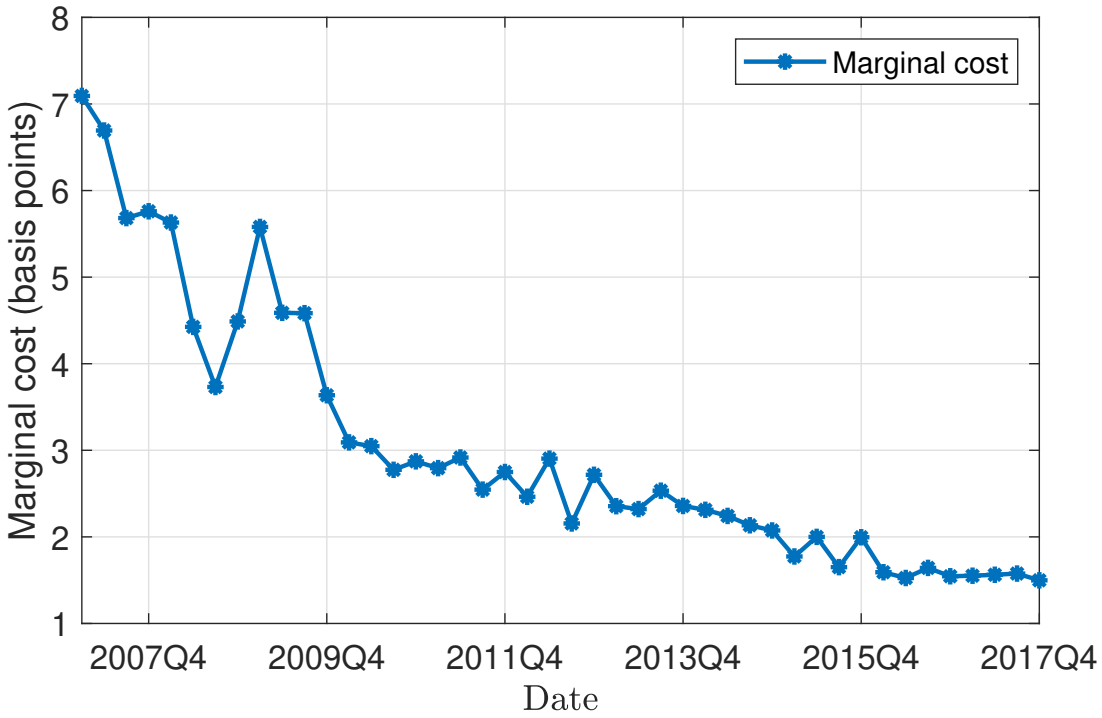


Figure IA.2: Estimated marginal cost
This figure displays the estimated marginal cost (in basis points) for each quarter.

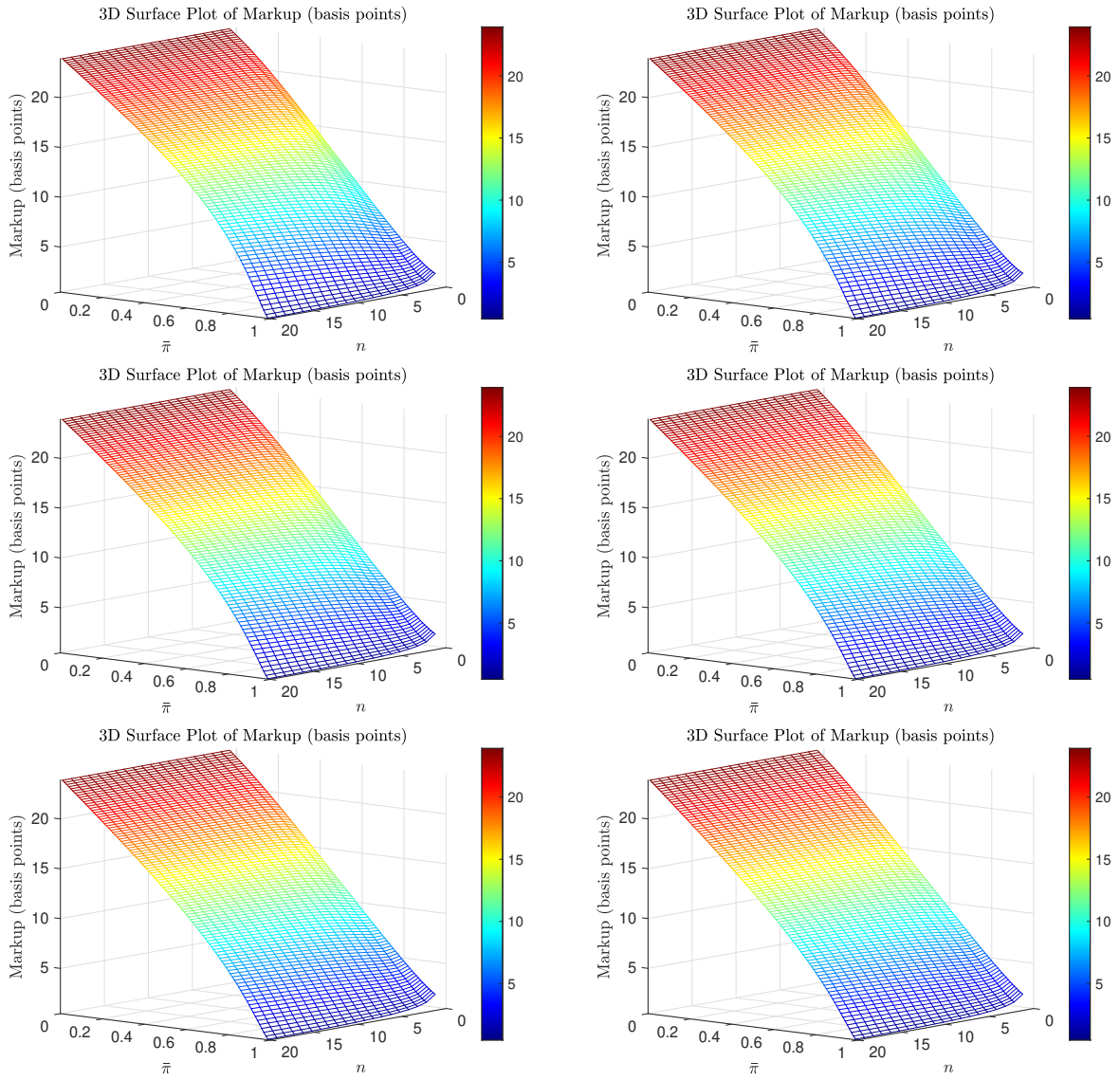


Figure IA.3: Competition effect on average markup
The left plots display average markup levels across all investors at $T = 0, 2, 4$ as initial belief $\bar{\pi}$ and number of competitors n changes. The right plots display average markup levels across all investors at $T = 1, 3, 5$ as initial belief $\bar{\pi}$ and number of competitors n changes.

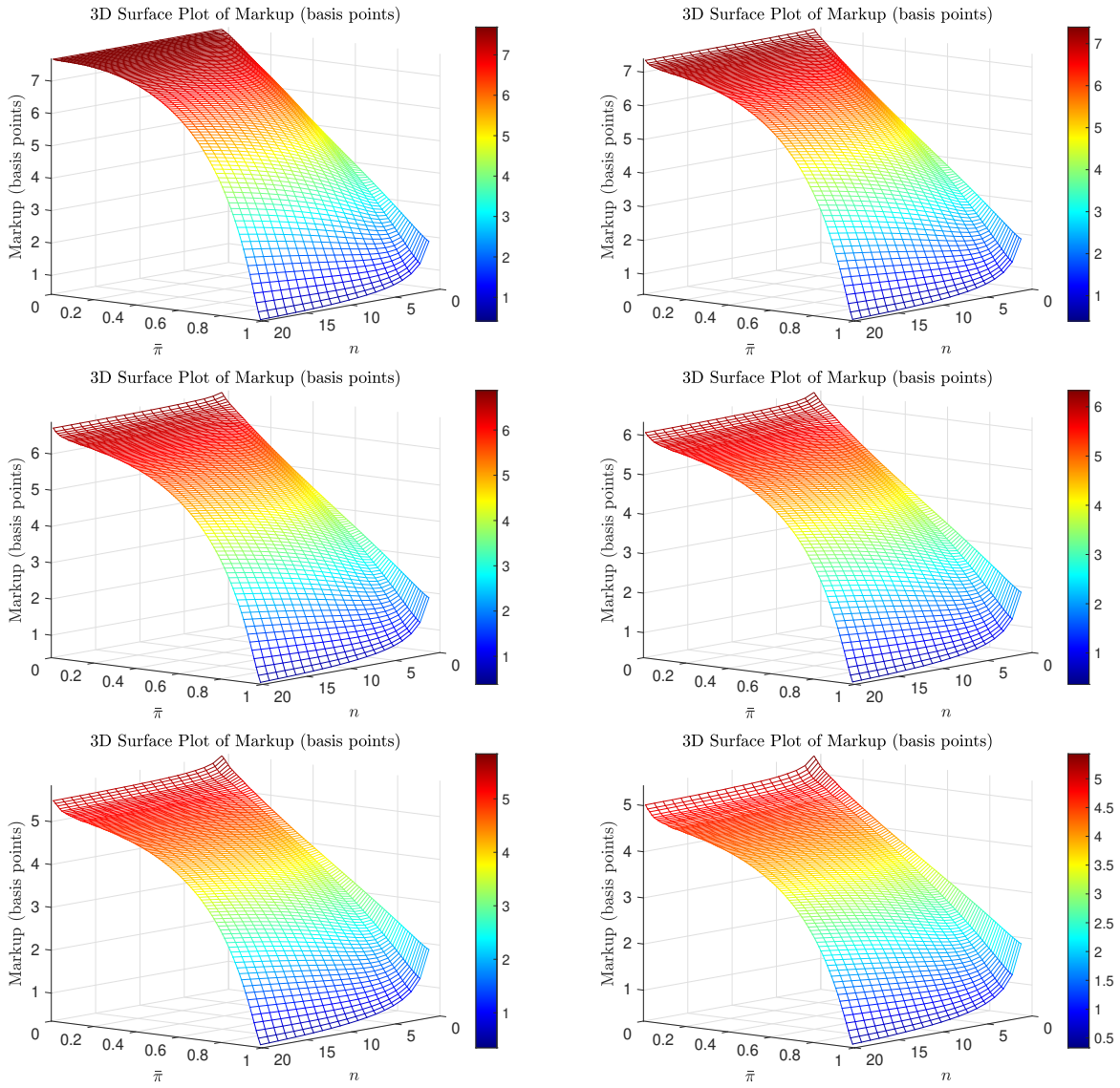


Figure IA.4: Competition effect on smart investors' average markup
The left plots display average markup levels across smart investors at $T = 0, 2, 4$ as initial belief $\bar{\pi}$ and number of competitors n changes. The right plots display average markup levels across smart investors at $T = 1, 3, 5$ as initial belief $\bar{\pi}$ and number of competitors n changes.

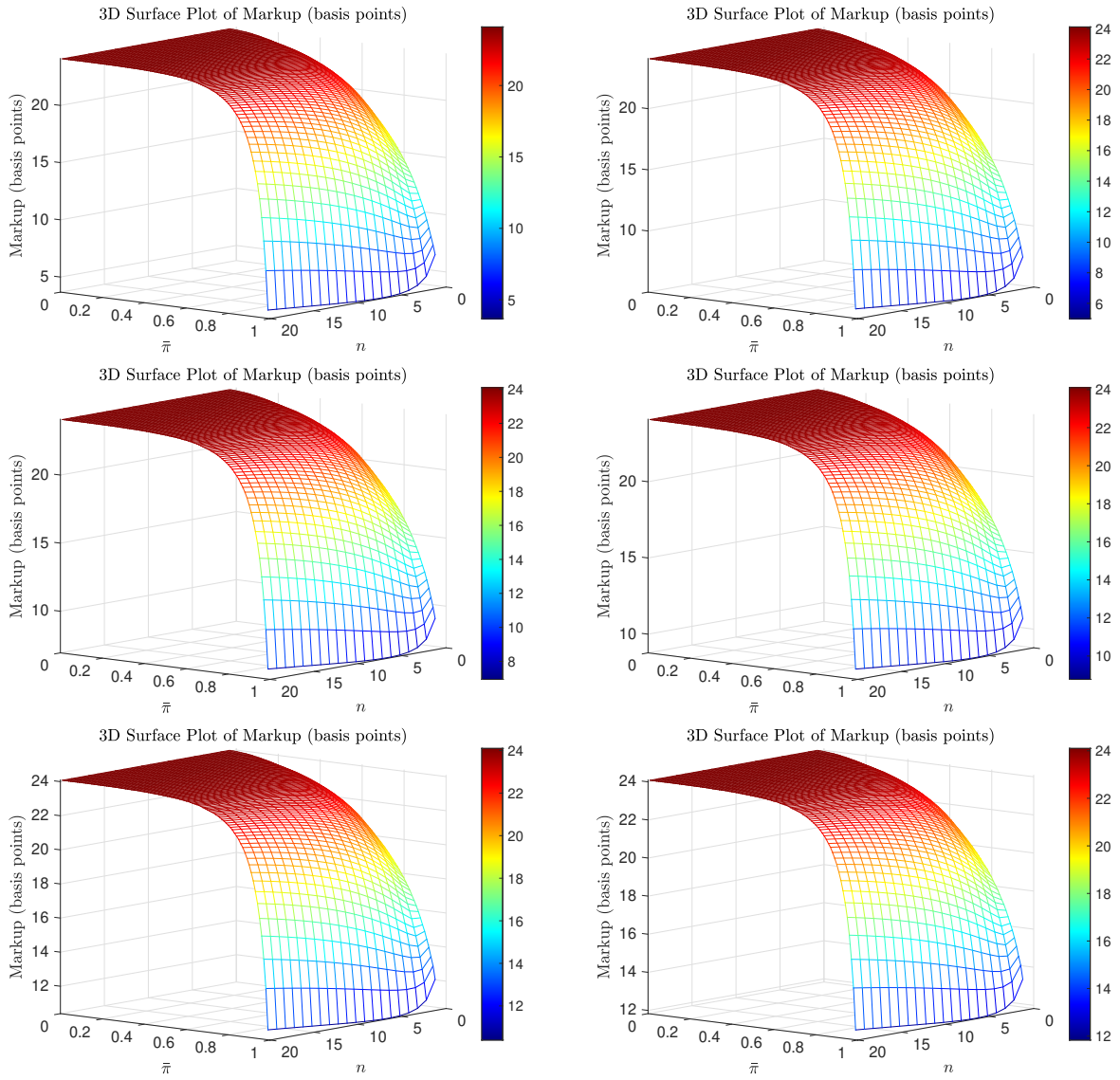


Figure IA.5: Competition effect on naïve investors' average markup
The left plots display average markup levels across naïve investors at $T = 0, 2, 4$ as initial belief $\bar{\pi}$ and number of competitors n changes. The right plots display average markup levels across naïve investors at $T = 1, 3, 5$ as initial belief $\bar{\pi}$ and number of competitors n changes.