When Competition Backfires: Broker Learning and Commission Fees*

Xindan Li†

Baolian Wang‡

Wei Xiong§

Xuewei Yang¶

Yinhong Zhao‖

October, 2025

*We thank James Choi, Chris James, Paul Goldsmith-Pinkham, Hongwei Xing and seminar participants at Hunan University, Princeton University, Shanxi University and Yale University for their valuable comments. Xindan Li acknowledges financial support from the National Natural Science Foundation of China (No. 72495151). Xuewei Yang acknowledges financial support from the National Natural Science Foundation of China (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 in a plain-vanilla Chinese setting. We document three facts: substantial price dispersion across investors, higher fees for naïve investors, and widening dispersion over time. To explain these patterns, we develop a dynamic model in which brokers compete for naïve and sophisticated clients while learning types through repeated interactions. The model shows that an incumbent's information advantage can sustain persistent market power—even over sophisticated investors—and that additional broker entry can backfire by raising markups on naïve investors and widening dispersion. Exploiting branch-level variation in rival presence, we provide causal evidence consistent with these predictions. Quantitatively, although the average number of brokers per city more than doubled over the sample, intensified competition slightly offset—rather than accelerated—the secular decline in fees.

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

1 Introduction

Fees paid by consumers for financial services are 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; Grubb et al., 2024). 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 competition and repeated interactions on price dispersion—remains an open 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 and can be renegotiated later.

China eliminated its fixed brokerage commission regime in 2002, allowing brokers to negotiate flexible rates within a regulatory range. This deregulation triggered a sustained decline in commission fees, mirroring a similar trend observed in the U.S. 27 years earlier. Despite the downward pressure on fees and the growing diversification of brokers into asset management and other business lines, brokerage commissions still account for over one-third of industry revenue, underscoring their continued importance to the sector.

Given the deregulation and the increasing number of brokers competing for clients, one might expect competition to compress the dispersion in commission rates. To investigate this central issue, we analyze 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 several important facts about commission-rate dispersion.

First, there is substantial variation in the commission rates charged across investors, even among those with similar characteristics.

Second, commission rates are systematically higher for less sophisticated investors. The initial rate set at account opening decreases with investor age, portfolio size and the per capital GDP of the investor's region. Interestingly, this initial rate also predicts subsequent behavioral patterns—including the disposition effect, lottery preference, underdiversification, local bias, extrapolation, and the rank effect—indicating that initial rates disproportionately burden less sophisticated investors.

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. Sophis-

tication 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 extend the canonical static framework of Stahl (1989) into a dynamic oligopolistic pricing model with broker learning and investor renegotiation. 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 solicits a new quote from the incumbent and—if dissatisfied—may seek outside quotes. Each broker sets its commission strategically based on a noisy posterior belief about the investor's type, in contrast to the static mixed-strategy equilibrium of Stahl (1989), where firms face anonymous demand and know only the population shares of shopper/searcher types. The broker's equilibrium pricing policy trades off offering low fees to attract sophisticated shoppers against extracting rents from naïve clients, takes rivals' policies as given, and internalizes the option value of future renegotiation.

We characterize the equilibrium and derive several implications. First, a broker's optimal quote is decreasing in its 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 sophis-

ticated 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 implication parallels heterogeneous-consumer search models in other markets (Rosenthal, 1980; Stahl, 1989; Carlin, 2009), though the empirical evidence is mixed (Chandra and Tappata, 2011; Lach and Moraga-González, 2017). Because brokerage accounts are typically tied to local branches—even when opened online—competition for clients is naturally geographically bounded. Leveraging detailed account-level data across cities, we conduct a sharp test using a branch–quarter panel of commission rates for newly opened accounts, measuring local competition by the number of brokers operating in each 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 treatment effects. These results confirm the model's prediction that competition can backfire by amplifying price discrimination.

To quantify the effects of learning and competition, we complement the reduced-form evidence with a structural model estimated via simulated method of moments (SMM) and use it for counterfactuals. The specification allows effective marginal costs to decline over time and incorporates branch-level heterogeneity in investor sophistication and search costs. Targeted moments include the frequency of commission adjustments, tail quantiles of commissions, branch-level means and their covariances with trading volume, and the

dispersion–client-tenure gradient. The estimated parameters are economically plausible and reproduce the key cross-sectional and time-series patterns.

The estimated model lets us decompose the secular decline in commissions. Most of the drop reflects shifts in investor composition and falling broker marginal costs. Although the average number of brokers per city more than doubled, heightened competition slightly offset—rather than accelerated—the decline in fees, reinforcing that competition can backfire: it not only widens dispersion but also dampens reductions in average fees.

The model also supports counterfactual analysis. In one exercise, we eliminate the incumbent's information advantage by revealing its current belief about an investor to all rivals. With informational parity, the incumbent sets substantially lower markups—especially for clients it perceives as savvy—to remain competitive. This underscores cross-type externalities: while it is common in behavioral IO for naïve consumers to be exploited (Grubb, 2015), our setting shows that savvy consumers are also penalized through pooling with naïve consumers and the incumbent's persistent information-based market power.

We also assess the fee-minimizing level of competition. The optimal number of brokers is typically below observed levels, indicating over-entry rather than insufficient competition. By contrast, investor education—raising the share of sophisticated investors—consistently compresses markups market-wide and yields larger fee reductions than plausible changes in market structure. Thus, policies that enhance financial literacy may be more effective than pro-entry interventions for lowering commissions, despite the latter's conventional appeal for financial inclusion.

2 Institutional Background and Descriptive Analysis

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 brokerages receive client orders and

route them to either the Shanghai or Shenzhen Stock Exchange.[1] Both exchanges operate centralized limit order books that match nearly all trades directly. Brokers charge a flat commission proportional to the RMB transaction amount, subject to a minimum of 5 RMB. Commissions apply symmetrically to both buy and sell orders.

Brokerage firms typically charge different commission rates to different clients. At account opening, they quote a fee rate that is often based on information provided in the account application or during an initial meeting between the investor and an account manager. Attentive investors who are aware of the fee structure may attempt to negotiate this rate.

Brokerage firms do not publicly disclose their fee schedules. However, investors may become aware that others are paying lower fees—either at the same firm or at competing firms. At any time, an investor can attempt to renegotiate their commission rate by contacting the brokerage, typically via phone or an in-person visit. A common tactic is to request a lower rate by threatening to transfer the account to another firm. The brokerage then decides whether to accommodate the request and, if so, by how much. If the investor remains unsatisfied, they can initiate an account transfer, a process that generally takes one to two business days and is usually free of charge.

Until 2015, Chinese securities law prohibited investors from holding more than one brokerage account. Even after the restriction was lifted, few investors opted to open a second account. Over our eleven-year sample period, account disappearances are rare in our brokerage data–fewer than 30,000 accounts, representing approximately 4% of the full sample we describe next.[2]

Since May 1, 2002, Chinese regulations have capped commission rates at 30 basis points (bps) and established a minimum equal to the combined fees levied by the China

[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, these trades differ from those involving domestically listed stocks, particularly in fee structures, and represent only a small share of total trading activity. We therefore exclude them from our analysis. The third stock exchange in mainland China, the Beijing Stock Exchange, began operations on November 15, 2021, which falls outside our sample period.

[2]Accounts may disappear from our data for two reasons: either the investor ceases trading and cancels the account, or the investor transfers the account to another broker. Our data do not allow us to distinguish between these two scenarios.

Securities Regulatory Commission (CSRC) and the relevant stock exchange. At the beginning of our sample period, these regulatory fees totaled approximately 1.6 bps, declining to 0.7 bps by the end.[3] Brokerage firms collect these regulatory fees as part of the overall commission and remit them to the CSRC and the exchanges.

2.2 Data and Descriptive Analysis

We analyze an account-level dataset from a brokerage firm in China. This dataset is similar in structure and scope to those used in prior research on Chinese investors (Feng and Seasholes, 2004, 2005; Frydman and Wang, 2020; An et al., 2024), as well as to the U.S. dataset analyzed by Odean (1998). The data originate from a medium-sized brokerage with multiple branches across China. During our sample period, brokerage commissions accounted for approximately 40% of the firm's total revenue, which is typical for the industry in China.[4]

The dataset comprises three files: a trade file, a position file, and an investor demographic file. Our sample period spans from January 2007 to December 2017. The trade file records investor identifiers, trade timestamps, asset symbols, buy/sell indicators, trade prices, share quantities, and–critically for our analysis–commission amounts (denominated in RMB). The position file reports daily holdings for each account. The demographic file includes basic investor characteristics, specifically age and gender. We restrict our analysis to trades involving common stocks and exclude both short sales and margin trading.

Table 1 reports summary statistics for the sample of over 700,000 investors. The average investor age is 45.17 years old, and 53.5% are male. As of December 2017–the end of our sample period–the average account age is 9.72 years. The 75th percentile of account age is 14.08 years, indicating that one-quarter of investors have remained with the broker

[3]The CSRC fee changed once during our sample period: it was 0.4 bps prior to January 1, 2012, and reduced to 0.2 bps thereafter. Exchange fees also evolved. Before June 1, 2012, the Shanghai Stock Exchange charged 1.10 bps and the Shenzhen Stock Exchange charged 1.22 bps. Starting June 1, 2012, both exchanges adopted a unified rate, which was subsequently adjusted to 0.87 bps on June 1, 2012, 0.696 bps on September 1, 2012, and 0.487 bps on August 1, 2015.

[4]To preserve anonymity, we do not disclose the precise number of investors served by the broker. The reported empirical analysis is based on a randomly drawn sample. We verify that the results obtained from the full dataset are quantitatively similar.

for over 14 years. The median log portfolio size is 10.07, corresponding to approximately 23,600 RMB, highlighting the predominance of small investors. Nonetheless, more than 10% of investors hold portfolios exceeding 200,000 RMB. The average annual portfolio turnover is 722.5%, consistent with the 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 dispersion: the standard deviation is 10.69 bps, the 75th percentile is 28 bps, and the 25th percentile is 3 bps—representing nearly a tenfold difference.

Relative to the overall market, our broker serves a disproportionately higher share of small investors. According to the 2017 yearbook published by the Shanghai Stock Exchange, the distribution of investors by portfolio size is as follows: 55.28% hold less than 0.1 million RMB; 30.11% hold between 0.1 and 0.5 million; 7.23% hold between 0.5 and 1 million; 5.08% hold between 1 and 3 million; 1.58% hold between 3 and 10 million; 0.50% hold more than 10 million. In contrast, the corresponding shares in our brokerage sample are 72.31%, 21.82%, 3.42%, 1.78%, 0.43%, and 0.11%, respectively.

We attribute this discrepancy to our broker's branch network being more concentrated in relatively underdeveloped regions. Importantly, as we will show later, our main results remain qualitatively similar when we condition on portfolio size. This suggests that differences in investor composition are unlikely to materially affect our conclusions.

Fig. 1 plots the average commission rates of investors in our sample (dotted line) from 2007 to 2017. For comparison, the figure also includes two benchmark series representing average commission rates in the broader market (solid and dashed lines). For the brokerage firm in our dataset, we consider only commissions from stock trading.

To construct market-wide benchmarks, we collect total commission revenue for all brokerage firms from the annual reports of the Securities Association of China. However, this aggregate figure includes commissions from trading not only stocks but also other securities such as funds, bonds, and options/warrants. As a result, we compute two ratios to bound the market's average stock commission rate: (1) total commissions divided by total trading volume across all securities (solid line), and (2) total commissions divided

by total stock trading volume (dashed line). These two ratios serve as upper and lower bounds, respectively, for the market-wide average commission rate. As shown in Fig. 1, the solid and dashed lines closely track each other–a pattern that is expected given that stock trading accounted for approximately 93% of total trading volume on the exchanges during our sample period.

The figure reveals a pronounced downward trend in commission fees over the sample period. The market average commission rate declined from approximately 15 basis points in 2007 to 3 basis points in 2017. The average commission rate in our sample followed a similar trajectory, falling from around 24 basis points to just above 5 basis points. This decline likely reflects both falling order-handling costs and intensifying competition among brokers. Notably, the brokerage in our sample consistently charges slightly higher commission rates than the market average, likely due to its disproportionately large share of small retail investors.

2.3 The Trend in China and the U.S.

Prior to 2002, the China Securities Regulatory Commission (CSRC) mandated a uniform commission rate of 30 bps on stock trades. This fixed-rate regime ended on May 1, 2002, when the CSRC deregulated commissions. Although the 30 bps ceiling remained in place, brokers were thereafter allowed to negotiate rates with clients. Initially, however, only a small fraction of clients received discounted commissions, as shown later in Fig. 2.

The first major wave of broad-based commission reductions occurred after the 2007–2008 stock market boom and bust. As trading volumes declined and market activity slowed, several major brokerage firms initiated a "commission war" to gain market share, placing significant downward pressure on fees. A second wave followed the CSRC's 2013 reform, which allowed investors to open brokerage accounts online. This regulatory shift enabled brokerages to adopt aggressive digital marketing strategies, triggering another sharp drop in commissions during 2014–2015. Both episodes coincide with sharp declines in average commission rates shown in Fig. 1.

The evolution of brokerage commissions in China closely parallels developments in the U.S., though on an earlier timeline. Prior to 1975, commissions in the U.S. were ef-

fectively fixed. On May 1, 1975—commonly referred to as "Mayday"—the U.S. Securities and Exchange Commission (SEC) abolished fixed commissions and required brokers to negotiate fees with clients. This reform gave rise to discount brokers such as Charles Schwab, which competed with full-service firms by focusing on low-cost execution rather than research and advisory services. The proliferation of internet-based trading in the 1990s further compressed commissions by automating order flow and reducing overhead. See Jones (2002) for a comprehensive historical overview of U.S. commission trends through 1998. The most recent shift came with the emergence of zero-commission platforms such as Robinhood, which forced established incumbents to adopt similar pricing models (Schwarz et al., 2025).

The parallels between the two markets are notable. In both China and the U.S., commission rates declined sharply in response to regulatory liberalization and technological innovation. The CSRC's 2002 deregulation mirrors the SEC's 1975 reforms, while the CSRC's 2013 liberalization—by enabling online account openings and intensifying competition—resembles the disruptive role played by Robinhood in the U.S. market. Nevertheless, unlike in the U.S., commission rates in China have not yet reached zero.

Given the pronounced downward trend in commission, one might expect competition to have simply driven fees down. Yet U.S. evidence shows that deregulation reduced average fees and increased dispersion. Following Mayday, studies by Schveineir and Smith (1980), Tinic and West (1980), and Blum and Lewellen (1983) document that small trades often faced higher commissions than before. Because small orders are more likely to be placed by less sophisticated investors, these findings suggest that deregulation and competition can lower overall costs while widening pricing heterogeneity. However, this issue has not been examined systematically in the U.S. context. Our analysis fills this gap by focusing on how competition affects commission dispersion in China.

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 discrim-

ination 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

We now examine whether brokers charge higher commissions to less sophisticated investors. We first explore how the initial rate set for new accounts varies with investor characteristics. Table 2 reports regressions of the initial commission rate on age, gender, portfolio size, and the investor's home-city GDP per capita,[5] with year–month fixed effects. Initial rates are lower for older investors, those with larger portfolios, and those from higher-GDP regions, and are higher for males.[6] Interpreting age, portfolio size and

[5]Our data do not record investors' residences, so we proxy location with the city of account opening, as most investors open accounts at local branches.

[6]The commission rate is subject to a regulatory cap of 30 bps, and a substantial share of investors—especially early in the sample—pay this maximum (see Figure 2). Any rate below 30 bps therefore constitutes a

home-city GDP per capita as proxies for sophistication, the patterns are consistent with brokers charging higher rates to less sophisticated clients.

Furthermore, we specifically 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 3, 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 3 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.[7] Consistent with prior literature, our sample investors exhibit strong behavioral biases.

discount. One may view pricing as a two-step decision: (i) whether to grant a discount (extensive margin) and (ii) conditional on granting one, the size of the discount (intensive margin). Table IA.1 shows that investor demographics are associated with both margins in patterns similar to those in Table 2.

[7]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.2 and IA.3 in the Internet Appendix.

Table 4 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 4 support the hypothesis that brokers charge higher commission rates to less sophisticated investors.[8]

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 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

[8] In Table IA.4, we re-specify the analysis using two variables tied to the 30 bps cap: (i) a discount indicator (1 if the initial rate < 30 bps) and (ii) the discount size (30 bps − initial rate). Both the extensive margin (granting a discount) and the intensive margin (discount size) predict subsequent behavioral biases.

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 3.

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 group. Additionally, the line representing the differences indicates a steady widening between the two groups over time.[9] In Table IA.5 of the Internet Appendix, the widening gap between the two groups over time is explained by two forces: (i) sophisticated investors renegotiate more frequently (the extensive margin), and (ii) conditional on renegotiation, they obtain larger reductions in commission rates (the intensive margin). Figure IA.2 of the Internet Appendix further shows that the intensive margin effect is present across nearly all one–basis-point bins of the pre-change commission rates.

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 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

[9]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 4. Over time, the ratio decreases from 0.95 to 0.87.

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 4); (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 extends the classic oligopolistic pricing framework of Stahl (1989) by in-

corporating investor heterogeneity, dynamic broker learning, and repeated renegotiation between brokers and investors. Investors differ in their levels of sophistication, captured by their search costs for switching to an alternative broker. While brokers cannot directly observe an investor's true type, they gradually infer it through repeated interactions over time–unlike the static, one-shot environment in Stahl (1989).

In equilibrium, brokers strategically offer lower commission rates to investors perceived as sophisticated and higher rates to those perceived as naive. As a result, our solution concept is a Bayesian Nash equilibrium rather than Stahl's mixed-strategy equilibrium. In our framework, price dispersion arises endogenously from heterogeneity in brokers' beliefs about investor sophistication, rather than from strategic randomization.

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). 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.

New investors enter the market according to a Poisson process with constant rate λ^{in}, while each existing investor exits the market (i.e., ceases trading altogether) at a constant Poisson rate λ^{out}. It is publicly known that a portion $\bar{\pi}$ of new entrants are smart, while the remaining $(1-\bar{\pi})$ are naive. These entry and exit dynamics give rise to a steady-state distribution of market participants, which is the focus of our model.

We study the tenure of an investor with a particular broker and denote it by the time

variable $t \in [0, \infty)$. At $t = 0$, when an investor first approaches the broker, the broker offers a personalized commission quote. The investor accepts and commits to the broker if the quoted rate does not exceed her reservation price; otherwise, she continues to search for quotes from alternative brokers.

We emphasize that the two types of investors differ in their search costs for brokers.[10] 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.[11]

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.[12]

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

[10] In line with the tradition in the price-dispersion literature, we model search costs as the key behavioral friction. The insights extend to settings with heterogeneous investors: an equivalent formulation features one type that is highly informed about commission rates and searches across brokers, and another whose awareness is more limited and treats fees at or below r_d as roughly equivalent.

[11] 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.

[12] 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.

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.

After opening an account with a broker, an investor may periodically pay attention to the commission rate at a Poisson rate α, which is identical for both types of investors.[13] 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. We assume that an incumbent broker observes a stream of information about θ:

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

[13] This assumption of equal attention rates simplifies the model derivation, even though smart investors typically renegotiate more frequently and secure larger fee reductions each time.

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 $t = 0$, we assume that each broker has a brief initial interaction with the investor to learn their type, which contributes an additional t_0 time of signals. Equivalently, one may assume each investor arrives at $t = -t_0$, but broker offers the initial quote at $t = 0$. y_{jt} over the period from $-t_0$ to 0 before responding to a quote request. A shorter observation period t_0 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 initial interaction 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).

Before the initial interaction, brokers hold a common perceived prior, $\tilde{\pi}$, on the probability that an searching investor is smart. This prior differs from the smart entrant fraction $\bar{\pi}$, since investors who switch brokers are necessarily smart, creating a selection effect. Formally, brokers' perceived prior is

$$\tilde{\pi} = \frac{\lambda^{in}}{\lambda^{in} + \lambda^{sw}}\bar{\pi} + \frac{\lambda^{sw}}{\lambda^{in} + \lambda^{sw}},$$

where

$$\lambda^{sw} = \frac{n}{n+1}\frac{\lambda^{in}}{\lambda^{out}}\bar{\pi}\alpha =: C(n,\alpha,\lambda_{out})\lambda_{in}\bar{\pi}.^{14}$$

Here, λ^{sw} denotes the flow rate of switching investors, obtained as the product of the total stock of investors at other firms $\frac{n}{n+1}\frac{\lambda^{in}}{\lambda^{out}}$, their probability of being smart $\bar{\pi}$, and the rate of renegotiation α. Brokers always begin the learning process from this common perceived prior. For simplicity, we adopt brokers' perceived prior $\tilde{\pi}$ as a key primitive parameter in future quantification exercises.[15]

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 $x_{jt} = \ln\left(\frac{\pi_{jt}}{1-\pi_{jt}}\right)$ and $\tilde{x} = \ln\left(\frac{\tilde{\pi}}{1-\tilde{\pi}}\right)$. By Ito's Formula,

$$dx_{jt} = -\frac{1}{2\sigma^2}\left(\theta - \frac{1}{2}\right)dt + \frac{1}{\sigma}dW_{jt}$$

This means that

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

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.[16] In contrast, the broker expects a smart investor to seek competing offers from the other n

[14] It is straight forward to see that $\bar{\pi} = \frac{\tilde{\pi}}{1+C(n,\alpha,\lambda_{out})(1-\tilde{\pi})}$, where $C(n,\alpha,\lambda_{out}) = \frac{n}{n+1}\frac{\alpha}{\lambda^{out}}$.

[15] While $\tilde{\pi}$ is in general larger than $\bar{\pi}$, the two parameters $\tilde{\pi}$ and $\bar{\pi}$ are highly correlated and can correspond to one another through a one-on-one mapping. Our results will be robust to either specification.

[16] 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).

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

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

where p is the commission rate, ρ is the broker's discount rate, and λ^{out} is the investors' exit rate.[17] For simplicity, we let $\tilde{\rho} = \rho + \lambda^{out}$ in future presentation.

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

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

In this expected payoff, $V(p;\pi_A(\pi))$ 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.[18] Specifically,

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

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

[17] Let τ be the survival time of the investor in the stock market. Note that

$$\mathbb{E}\left[e^{-\rho t}\mathbb{1}_{\{\tau\ge t\}}\right] = e^{-\rho t}\Pr(\tau \ge t) = e^{-(\rho+\lambda^{out})t},$$

with τ following a Poisson distribution.

[18] Our setting exhibits a 'winner's blessing' rather than the familiar 'winner's curse.' Because naive investors are present, acceptance is negatively correlated with sophistication: the posterior share of smart types among acceptors satisfies $\pi_A(\pi) \le \pi$. Adverse selection is therefore not our central concern.

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.[19]

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) at the steady state of the dynamic system, 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 competing broker's belief is higher than the current broker's belief, π. This equation further implies:[20]

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

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

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

[19] 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.

[20] This is derived by applying chain rule:

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

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:[21]

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

where we define the characteristics function:

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

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:

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

The broker's subjective discount rate, ρ, and the exit rate, λ^{out} determines that the value function $V(p;\pi)$ should appreciate by $(\rho+\lambda^{out})V$ over any time interval dt, as given by the

[21] 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

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

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:

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

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 $\mu(\pi)$:

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

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

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

We also have the following boundary conditions:

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

While we can not analytically solve v, μ, a numerical solution of these functions is illustrated in Fig. IA.4.

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.

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 among brokers. Even when one 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.[22]

[22]While Stahl (1989) also delivers a positive gap between the price floor and marginal cost, the underlying

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

mechanism differs. In Stahl (1989), firms have no information about consumer type and must always price for the possibility that the buyer is naive. In our setting, by contrast, even when one broker has learned that a given investor is smart, it still sets a positive markup because competing brokers do not share that information.

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. 5a 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. 5b, 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. 5c examines the impact of different values of $\tilde{\pi}$, 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 $\tilde{\pi}$ 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. 5d 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. 6, 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., $\pi \to 1$) but higher markups for naive investors (i.e., $\pi \to 0$). A natural implication is that price dispersion among investors may increase in response to competition, which we can empirically measure and test.

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. 6 shows that when the fraction of sophisticated investors in the population is higher (i.e., $\tilde{\pi} = 0.8$ versus $\tilde{\pi} = 0.2$), brokers abandon using lower markups to compete for sophisticated investors at lower values of π. This occurs because a higher $\tilde{\pi}$ 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 evi-

dence 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 demonstrated in Fig. 6, 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 across different cities.

Before 2013, regulations prohibited online account opening, requiring investors to visit local branches—effectively confining competition to local markets. After 2013, online opening was allowed, but accounts were typically assigned to the branch nearest the investor's residence, and commission renegotiations were handled by that branch's managers. Consequently, despite rising online adoption, competition remained largely regional through 2017, the end of our sample.

We measure competition by the number of brokers with a branch presence in a prefecture-level city.[23] Although our data come from a single brokerage, its branch network spans many cities with varying rival presence. This cross-city variation allows us to examine at the branch–quarter level, the distribution of commissions for newly opened accounts. We adopt a quarterly frequency to ensure enough new accounts per branch to measure fee distributions reliably.

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

[23] 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.

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. 7 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. 7 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 5, we estimate the same regression while incorporating year-quarter fixed effects (absorbing the secular fee decline) and city fixed effects. The results in Table 5 corroborate the descriptive patterns observed in Fig. 7.

Together, Fig. 7 and Table 5 demonstrate that the number of brokers in a city has effects 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. 7 and Table 5 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 6 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. 7 and Table 5.

In columns (2) and (4), we present the dynamic coefficients, which are also plotted in Fig. 8. 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.

Overall, using cross-city variation in rival presence, we find that stronger local competition is associated with greater commission dispersion, consistent with our model's prediction that competition can backfire in narrowing dispersion. This does not imply competition is irrelevant: it likely contributed to the secular decline in average fees even as it widened cross-investor spreads. To assess competition's role systematically, we structurally estimate our model and conduct counterfactual analyses in the next section.

6 Quantitative Analysis

To systematically assess how competition affects the level and dispersion of commissions, we estimate the model structurally and use the estimated model to study the global properties of competition and its welfare implications.

6.1 Estimation

We estimate the model's parameters using the simulated method of moments (SMM) (Strebulaev and Whited, 2011). To capture regional and time-series heterogeneity in investor composition, we allow brokers' priors $\tilde{\pi}_{kt}$ to vary flexibly at the branch–quarter level (with k indexing branches and t quarters). We also allow brokers' marginal costs MC_t to vary over time and naive investors' search costs κ_k to vary 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-quarter to identify $\tilde{\pi}_{kt}$; 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). Using data from Shenzhen Stock Exchange (2017), we calibrate $\lambda^{out} = 0.0046$ in our sample period. Detailed steps of the estimation procedure are provided in Appendix IA.4.

The estimation results are reported in Table 7. The calibrated discount rate $\rho = 0.0995$ implies that brokers value next year's revenue at roughly 10% less than current-year revenue. The exit hazard $\lambda^{out} = 0.0046$ means that about 0.46% of investors leave the market each year. The attention frequency $\alpha = 0.121$ indicates that roughly 12.1% of investors renegotiate and lower their commissions each year.[24] The elasticity of search costs with respect to trading volume, $m = 0.766$, implies that a 1% increase in volume is associated with a 0.766% increase in search costs.

The initial interaction intensity $t_0 = 0.160$ corresponds to about 58 days (0.160 years) of accumulated information at initial contact. Finally, the signal volatility $\sigma = 0.288$ implies that, starting from a prior $\tilde{\pi} = 0.5$, it takes on average about one year for a broker to be 95% confident that an investor is sophisticated ($\pi > 0.95$).[25]

Our estimates align well with economic intuition in the cross section. First, branches in more urban and economically developed areas have a higher share of sophisticated investors: brokers' prior $\tilde{\pi}_{kt}$ is positively correlated with local population density and GDP per capita (Fig. 9a).[26] Second, controlling for trading volume, branches in more developed regions exhibit lower search costs: naive investors' κ_k is negatively correlated

[24] Although infrequent, each update is sizable—for example, a reduction from 30 bps to 10 bps or from 10 bps to 3 bps—large enough to drive the secular decline in average fees.

[25] For $\theta = 1$ and $\tilde{\pi} = 0.5$, 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 $t = \sigma^2 \left[\ln \frac{\pi}{1-\pi} + \frac{1}{1-\pi}\right]_{0.5}^{0.95} = (0.288)^2 \times 20.95 \approx 1.74$. The noise term scales this by $\frac{1}{1+\sigma^2/2} \approx 0.57$ (via Itô-to-Stratonovich correction), producing $t \approx 0.99$.

[26] Fig. IA.6 shows a similar correlation for the smart-entrant fraction.

with local population density and the number of brokers (Fig. 9b).[27]

6.2 Decomposing the Trend

As shown in Fig. 1, commission rates decline sharply over 2007–2017. What drives this trend? Our quantitative decomposition points to three forces: (i) a shift in investor composition, (ii) falling broker marginal costs, and (iii) rising broker competition.

Fig. IA.5a plots the estimated marginal cost and the smart-entrant share. Estimated marginal costs mimic the drop in per-transaction regulatory fees (Fig. 2a), falling from approximately 7 bps in 2007 to about 1.8 bps in 2017. The smart-entrant share rises steeply: the city-average $\bar{\pi}$ increases from around 5% to nearly 60%, and the perceived prior for inquiring investors $\tilde{\pi}$ from approximately 50% to nearly 90%. Competition also intensifies: Fig. 10b shows the average number of active brokers per city roughly doubles, from 24 to over 52.

Our simulation in Fig. 10a shows that these forces jointly account for the downward path of commissions in Fig. 1. Decomposing the 2007–2017 change (Fig. 10b), shifts in investor composition explain a 19.63 bps decline in commissions charged to new entrants, while falling marginal costs contribute another 5.59 bps.

Notably, despite the substantial rise in broker competition, its contribution is negligible and, if anything, slightly offsets the decline. This accords with our earlier result that competition can backfire—not only widening dispersion but also dampening reductions in average fees.

6.3 Counterfactual Analysis

Our estimated model allows us to run counterfactuals that isolate specific mechanisms.

6.3.1 The Incumbent's Information Advantage

Unlike Stahl (1989), our model features broker learning and an informational wedge between the incumbent and its rivals. To isolate this wedge, we compare the baseline equi-

[27] Search is easier where populations are denser and branch networks thicker; hence we emphasize these dimensions rather than GDP per capita. We also find a mild negative correlation with GDP per capita.

librium with a one-shot information-symmetry shock ("MIT shock"): in a single period, the incumbent's posterior about a given investor, π, is disclosed to all rivals. Rivals then treat π as their prior and quote according to their baseline policies; the incumbent, having learned ex post that the disclosure occurred, best-responds in that period to defend the client.

Fig. 11 plots the incumbent's baseline pricing policy alongside the counterfactual.[28] The incumbent offers a deep discount under the shock, precisely because it has lost its information advantage for that period.

In the baseline, the quoted markup declines monotonically with π: from 100% of the maximum markup at $\pi = 0$ to about 13.6% at $\pi = 1$. Thus, even for a client known to be sophisticated, the incumbent maintains a positive markup (13.6%).

Under the information-symmetry counterfactual, the incumbent charges substantially less, with the discount (relative to baseline) widening from near zero at $\pi = 0$ to 28.3% of the maximum markup by $\pi = 0.70$. For $\pi \geq 0.71$, the quoted markup jumps down to the 13.6% floor, matching the lowest markup rivals would offer and thereby locking in the investor. This sharp shift quantifies the value of the incumbent's informational advantage: when it is removed, the incumbent must cut price materially—particularly for savvy clients—to remain competitive.

This counterfactual isolates the incumbent's information advantage and highlights cross-type externalities. It is unsurprising that naïve investors (low π) face high markups. More striking is that smart investors (high π) are also penalized: because rivals cannot distinguish them from naïve types, they cannot selectively undercut, allowing the incumbent to sustain a positive markup even on known-savvy clients. This contrasts with classic "add-on" models in the behavioral industrial organization literature (e.g., Gabaix and Laibson (2006)), where firms' exploitation of naïve consumers can subsidize the baseline

[28] We compute the counterfactual in three steps. First, we solve for the baseline policy μ^b and value function v^b. Second, for each $\pi^{\text{inc}} \in (0,1)$, we apply μ^b to a counterfactual symmetric belief distribution F_s^{sym} with $\tilde{\pi} = \pi^{\text{inc}}$ to obtain the implied distribution of competing quotes faced by this investor, G_s^{sym}. Third, we choose $\mu^{\text{sym}}(\pi^{\text{inc}})$ to maximize the incumbent's payoff

$$v^b(\pi^{\text{inc}})\Big[1 - \pi^{\text{inc}} + \pi^{\text{inc}}\big(1 - \big(G_s^{\text{sym}}(\pi^{\text{inc}})\big)^n\big)\Big],$$

where n is the number of rival brokers. Figure 11 plots μ^{sym} as a function of π^{inc}.

price and benefit sophisticated shoppers. In our single-price commission setting, no such subsidy margin exists, so pooling instead amplifies the incumbent's market power over sophisticated investors.

6.3.2 Optimal Competition

With the estimated model parameters, we simulate markets under alternative competition environments (varying the number of competing brokers) and investor compositions (varying the proportion of smart investors). This exercise allows us to characterize the optimal level of competition if regulators aims to minimize various outcomes of interest: per capita commission fees, the average markup at entry (new investors), and the average markup after five years.[29] While these fees are transfers to brokers and not direct welfare measures, regulators typically favor lower-fee environments to promote financial inclusion (World Bank, 2013).

The left panels of Fig. 12 plot the average initial markup over a wide grid of parameter values.[30] The right panels report the marginal effect of adding one broker—the global treatment effect—computed as the derivative of each outcome variable with respect to the number of competing brokers.

As a benchmark, when all investors are sophisticated ($\tilde{\pi} \to 1$), more competition unambiguously reduces both per-capita profit and average markup (left panel of Fig. 12a; note the x-axis is reversed, so values farther right correspond to fewer rivals). This accords with classic theory. With naïve investors, however, the marginal effect of an additional broker becomes heterogeneous and non-monotonic.

Naive investors do not search for other brokers, so adding brokers yields little benefit—and may even harm them if brokers shift toward extracting higher markups from naive investors rather than competing for sophisticated investors. For example, when $\tilde{\pi} < 0.94$,

[29] 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 sophisticated. We let $n \in [1, 100]$. 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 $\frac{1}{I}\sum_I \phi$, $\frac{1}{I_s}\sum_{I_s} \phi$, $\frac{1}{I_d}\sum_{I_d} \phi$.

[30] Robustness checks in the Internet Appendix: profits (Fig. IA.8); five-year markups (Fig. IA.9); and cumulative 0–5 year markups by type—smart and naïve (Fig. IA.10, Fig. IA.11).

naive investors gain nothing from having more than two brokers ($n \geq 1$; Fig. 12c). By contrast, sophisticated investors benefit: when $\tilde{\pi} > 0.7$, moving from two to three brokers reduces their average markup, though the marginal gain diminishes quickly as brokers scale back aggressive pricing (Fig. 12b). Thus, while "competition backfiring" is not universal, it holds over a broad and relevant parameter range.

For policy, Fig. 13a plots the broker count that minimizes the average initial markup.[31] When a moderate share of investors is naive ($\tilde{\pi} < 0.82$), two brokers minimize consumer costs: with few shoppers, additional entrants have little incentive to undercut. As $\tilde{\pi}$ rises, the optimal number of competitors increases, diverging only as $\tilde{\pi} \to 1$. As expected, the fee-minimizing n for sophisticated investors always exceeds that for naïve investors.

Finally, Fig. 13b compares these optima with observed broker counts by branch. In most markets, actual broker numbers exceed the fee-minimizing level. This points to a potential regulatory role: rather than encouraging additional entry, policymakers might consider curbing excessive entry. We note, however, that our model abstracts from brokers' entry and exit decisions.

6.3.3 Investor Education

While competition can backfire in reducing consumer markups, our quantitative analysis shows that investor education—by raising the share of sophisticated investors—has far greater impact. In the left plots of Fig. 12, the average markup declines monotonically with the population's mean ability for both smart and naive investors, at every competition level. This matches the theory in Fig. 5c, where higher broker priors generate spillover benefits for the entire client base.

In magnitude, sophistication dominates market structure. Whether a market is concentrated or fragmented, a one–percentage point increase in the smart-investor share reduces average markups more than realistic changes in the number of competitors, such as doubling or halving the number of competitors.

[31] Fig. IA.12 in the Internet Appendix shows analogous results using average lifetime profit.

Policy implication. Prioritize investor education. By enabling more investors to recognize and act on competitive options, literacy initiatives reduce fees more effectively than reforms focused solely on promoting additional entry.

7 Conclusion

This paper studies price dispersion in brokerage commission fees—a seemingly homogeneous service. Using account-level data from China, we document three patterns: (i) substantial, persistent fee dispersion across investors; (ii) systematically higher fees for less sophisticated investors; and (iii) widening dispersion over the broker–investor relationship, even as average fees decline.

We explain these facts with a dynamic oligopoly model in which brokers learn investor types through repeated interactions. Three insights follow. First, brokers use informational advantages to price discriminate, charging higher markups to investors perceived as naive. Second, learning creates a negative externality: as brokers compete more effectively for sophisticated clients, markups on naive clients rise. Third, increased broker entry can backfire: with more firms chasing a limited pool of sophisticated shoppers, the incentive shifts toward extracting rents from captive naive investors, raising markups and dispersion.

Our model departs from the classic framework (e.g., Stahl (1989)) by placing behavior-based pricing at the core. In the classic setting, dispersion arises from mixed strategies over an aggregate demand curve with limited scope to condition on individual histories. By contrast, our pricing is type-contingent, explicitly tied to learned customer posteriors. This delivers distinct comparative statics: persistent incumbent market power—including over sophisticated investors—and the possibility that additional entry amplifies fee dispersion. Exploiting cross-city variation in rival presence, we provide strong empirical evidence for these predictions, filling a gap in the literature.

Our structural estimation and counterfactuals quantify these effects. Using the estimated parameters, we find that the pronounced decline in commissions is driven mainly by (i) shifts in investor composition and (ii) falling broker marginal costs. Although the

average number of brokers per city more than doubled over the sample, intensified competition slightly offset—rather than accelerated—the decline in fees.

Counterfactuals further show that the fee-minimizing number of brokers is typically below observed levels, implying excessive (not insufficient) entry when a sizable share of investors is naïve. By contrast, policies that raise financial literacy—and thereby increase the share of sophisticated investors—consistently compress markups for all clients, with effects larger than plausible changes in market structure via entry or exit.

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*.

Blum, Gerald A and Wilbur G Lewellen (1983) "Negotiated brokerage commissions and the individual investor," *Journal of Financial and Quantitative Analysis*, 18 (3), 331–343.

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.

Grubb, Michael D. (2015) "Failing to Choose the Best Price: Theory, Evidence, and Policy," *Review of Industrial Organization*, 47, 303–340, 10.1007/s11151-015-9476-x.

Grubb, Michael D., Darragh Kelly, Jeroen Nieboer, Matthew Osborne, and Jonathan Shaw (2024) "Sending Out an SMS: Automatic Enrollment Experiments for Overdraft Alerts," *The Journal of Finance*, 80, 10.1111/jofi.13404.

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.

Jones, Charles M (2002) "A century of stock market liquidity and trading costs," *Available at SSRN 313681*.

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.

Schveineir, John C and Keith V Smith (1980) "The impact of Mayday on diversification costs," *The Journal of Portfolio Management*, 6 (4), 28–36.

Schwarz, Christopher, Brad Barber, Xing Huang, Philippe Jorion, and Terrance Odean (2025) "The "actual retail price" of equity trades," *The Journal of Finance*.

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.

Shenzhen Stock Exchange (2017) "Shenzhen Stock Exchange Fact Book 2017."

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.

Tinic, Seha M and Richard R West (1980) "The securities industry under negotiated brokerage commissions: Changes in the structure and performance of New York Stock Exchange member firms," *The Bell Journal of Economics*, 29–41.

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.

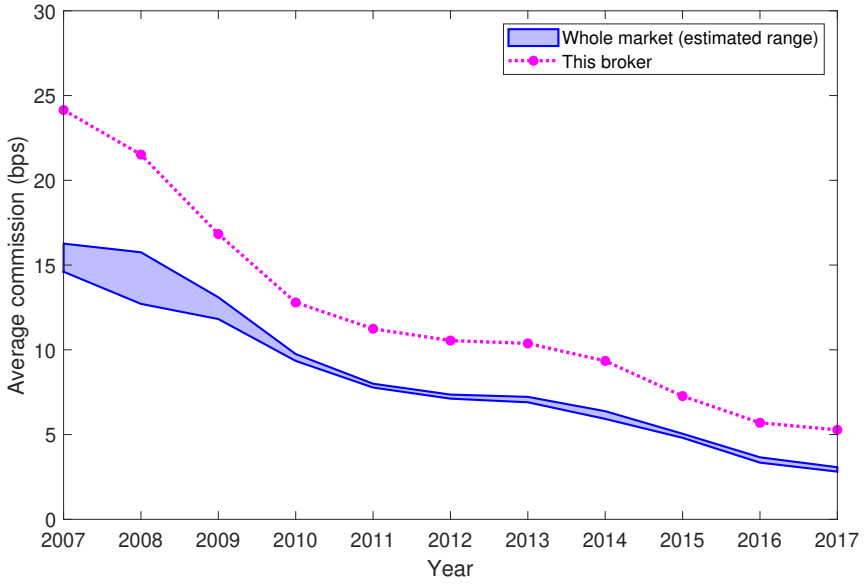
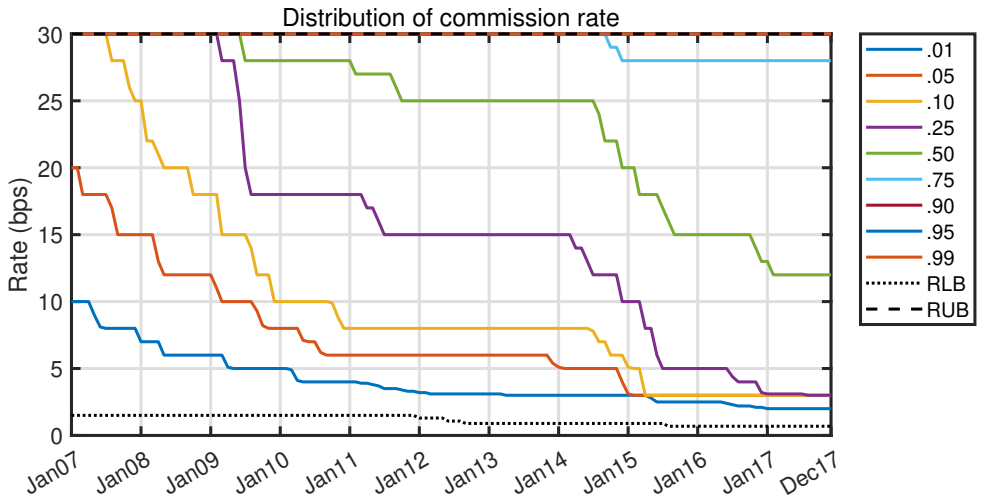
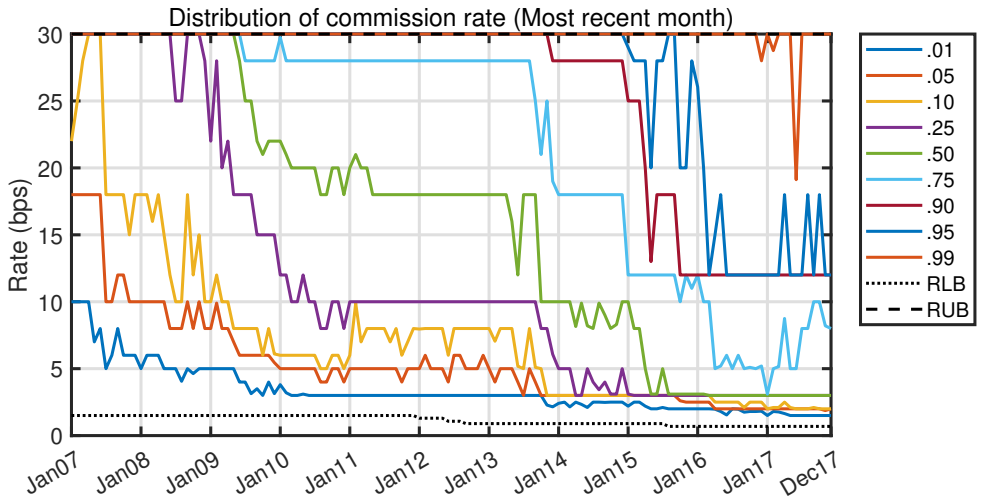


Figure 1: The average commission rates over time
This figure displays the average commission rates of the investors in our sample (dotted line) and the average commission rates of the market (shaded area within the solid line), from 2007 to 2017. For our broker, we include only stock trading. The exact national average of retail stock trading commissions is not directly observed, so we bound it with two methods in the shaded area. The lower bound is calculated as the total commission revenue across all brokers divided by the aggregate trading volume (including stocks, funds, bonds, and options/warrants). The upper bound is 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.

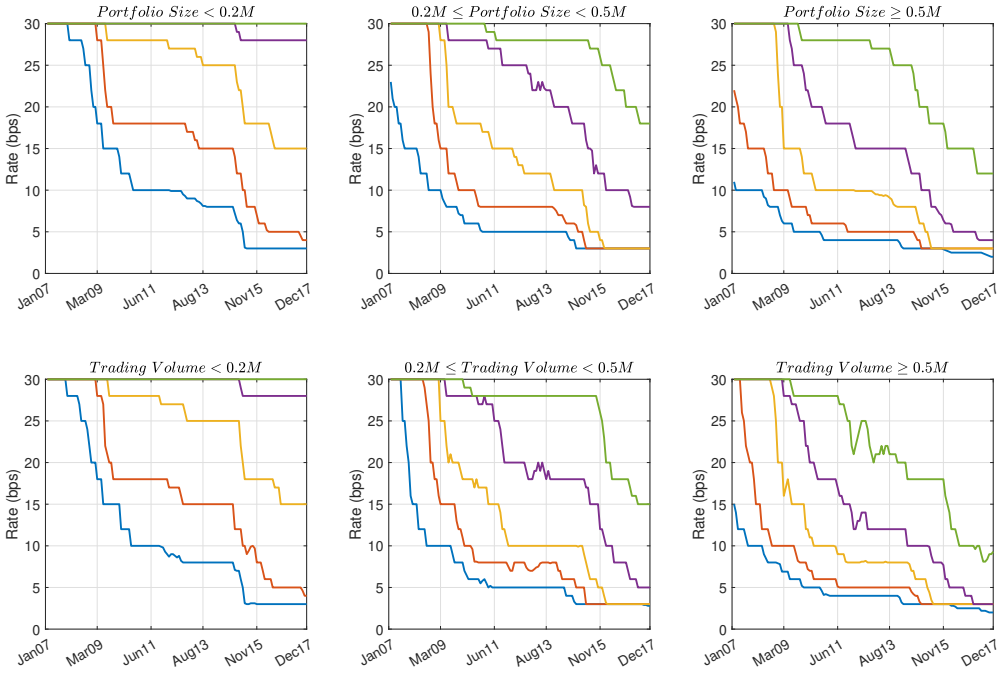
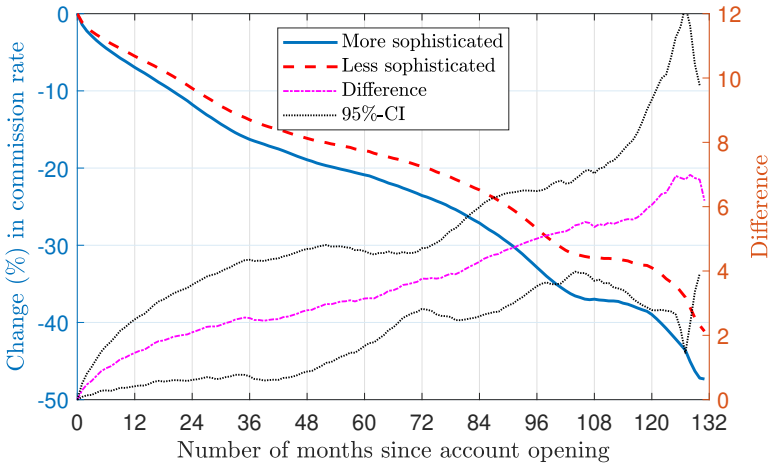
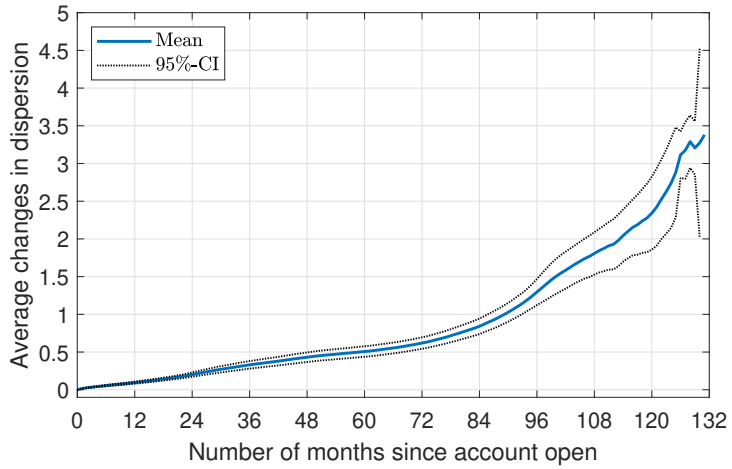


Figure 3: Dispersion of commission rates, by portfolio size and trading volume
This figure displays the distribution of commission rates across investors by subsamples. For each subsample in each month, we calculate the commission rate for each investor and report several percentiles, i.e., 10th, 25th,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



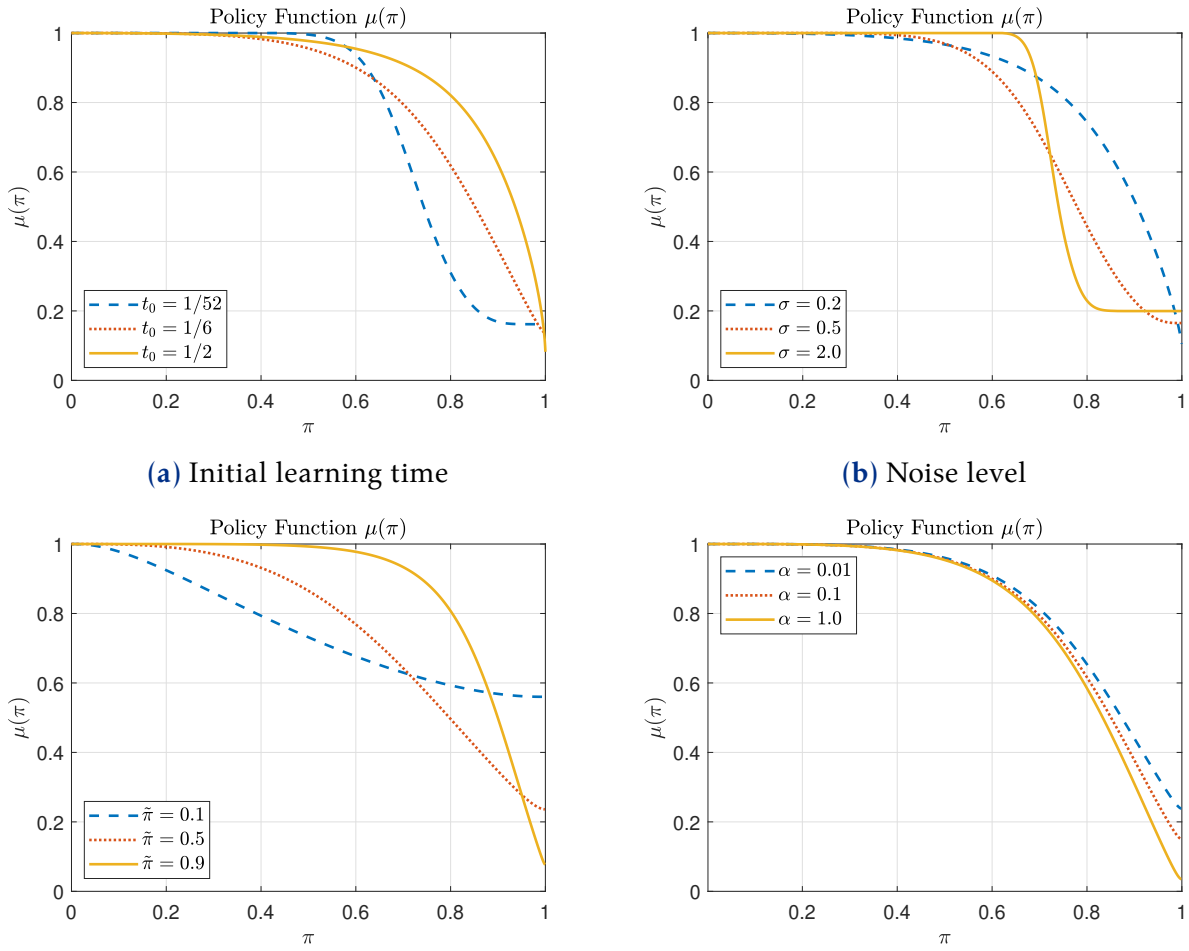
(a) Percentage changes in commission rates



(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 3. 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.



(a) Initial learning time

(b) Noise level

(c) Proportion of smart investors

(d) Attention on commissions

Figure 5: Numerical solutions under different market conditions
This figure shows numerical solutions of policy function under different market conditions. Baseline parameters are $\rho = 0.0995, \lambda^{out} = 0.0046, \alpha = 0.1213, t_0 = 0.1599, \sigma = 0.2880, \tilde{\pi} = 0.7488$, and $n = 1$.

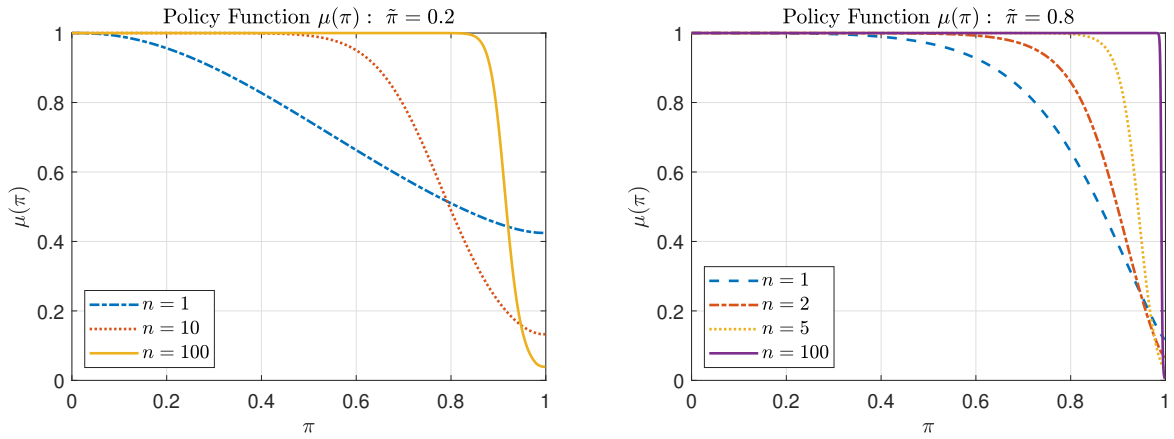


Figure 6: Numerical solutions under different competition environments
This figure shows numerical solutions of policy function under different competition environments. Baseline parameters are $\rho = 0.0995, \lambda^{out} = 0.0046, \alpha = 0.1213, t_0 = 0.1599, \sigma = 0.2880, \tilde{\pi} = 0.2$ (left panel) or 0.8 (right panel).

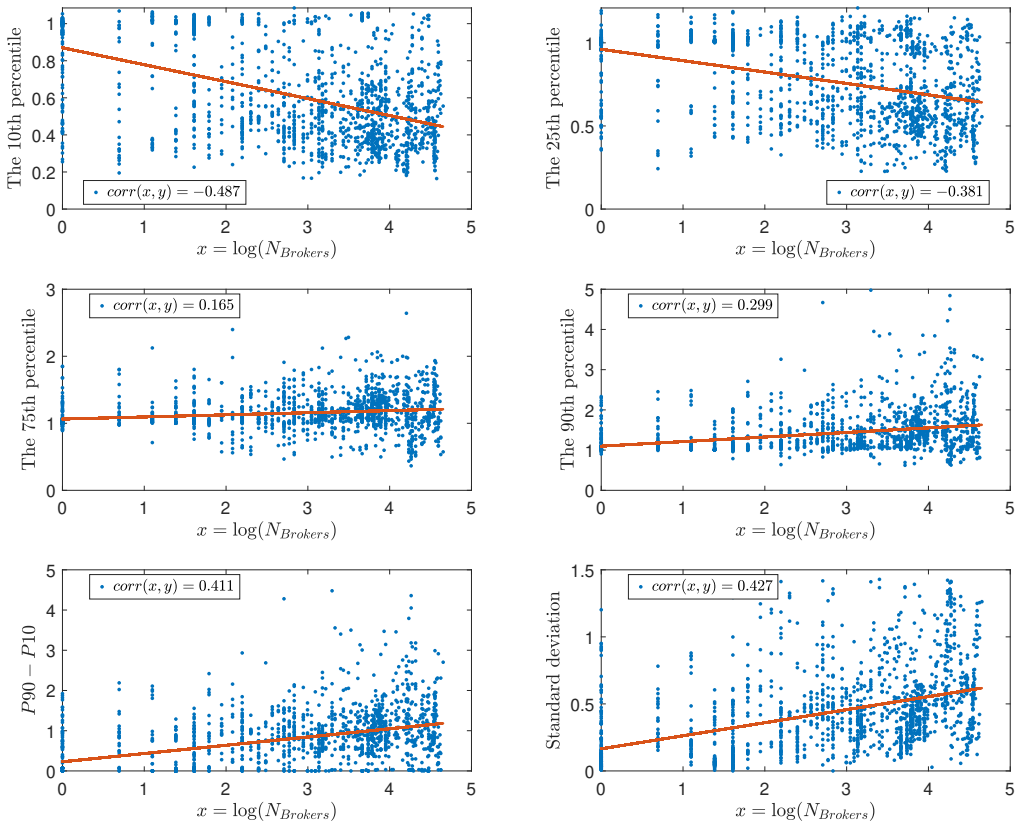


Figure 7: Commission rates dispersion and the number of brokers
This figure presents scatter plots of various commission rate distribution statistics—either percentiles or standard deviations—plotted against the logarithm of the number of brokers in a city. These statistics are computed at the branch-quarter 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.

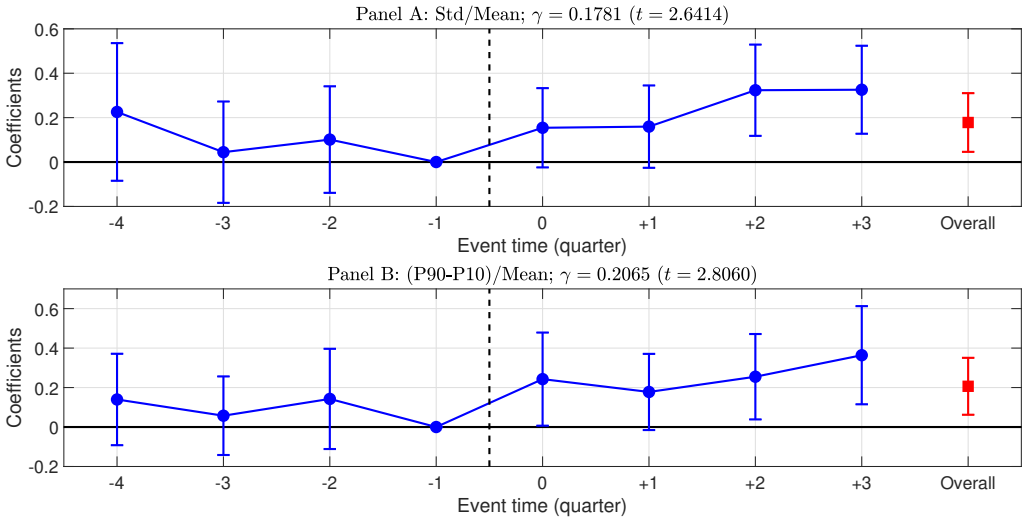
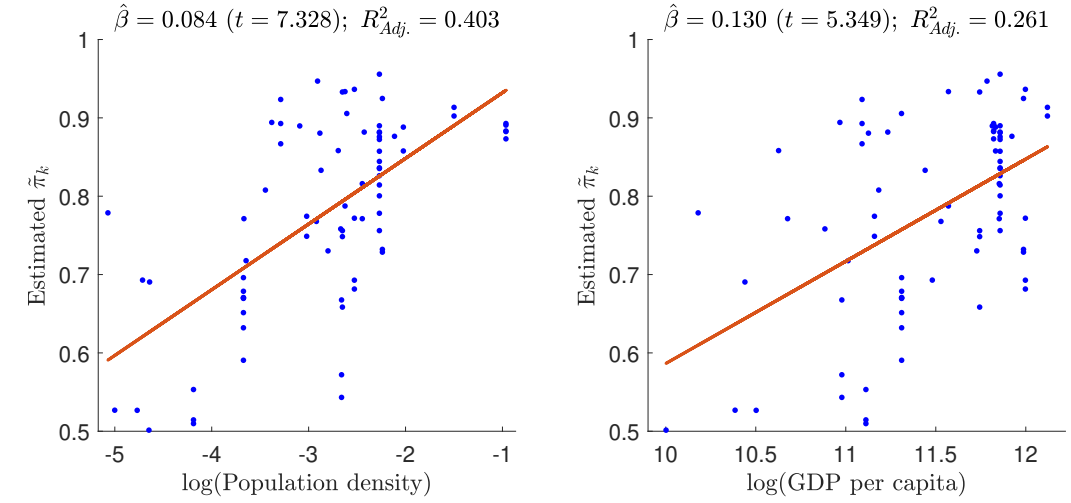
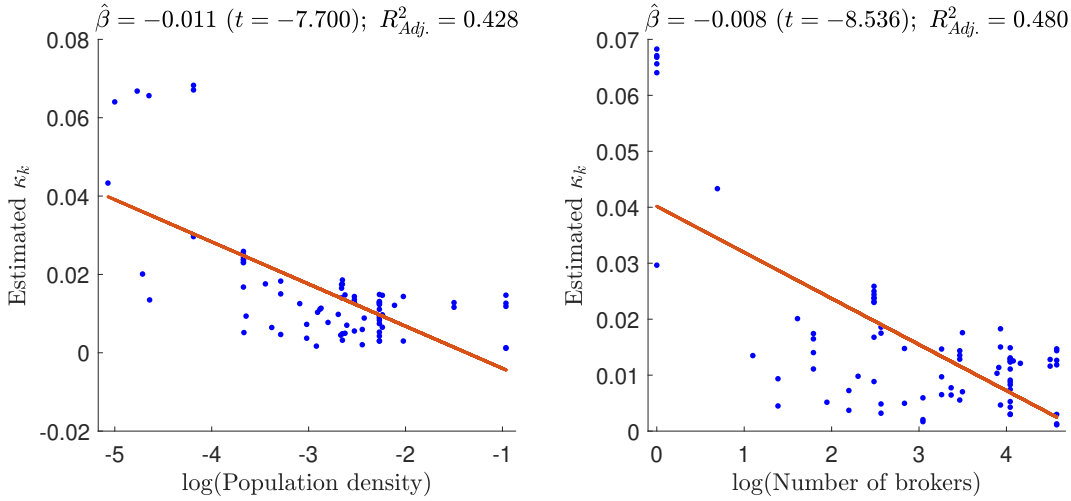


Figure 8: Dynamic treatment effects of broker entry on price dispersion
This figure displays dynamic coefficients and the corresponding 95% confidence intervals from stacked DiD analyses. The isolated red error bar in each plot shows the overall estimate of the treatment effect. We define treated units 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.

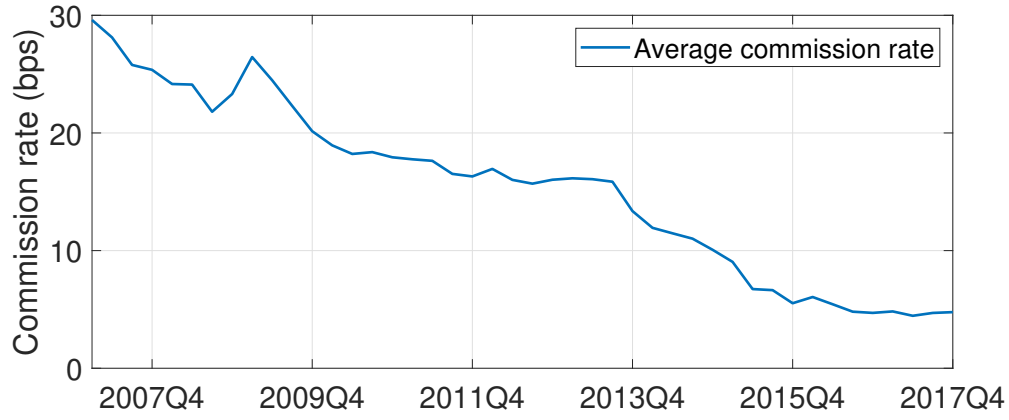


(a) Perceived prior $\tilde{\pi}_k$

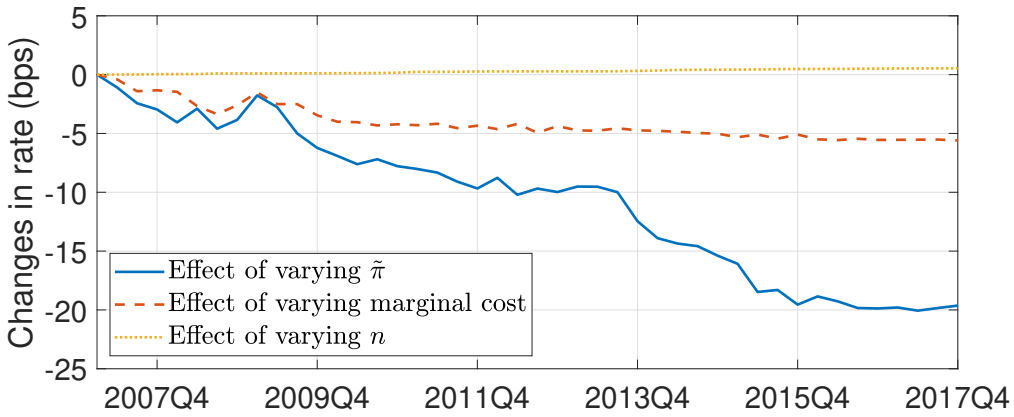


(b) Naive investors' search costs κ_k

Figure 9: Geographic variation of estimated parameters $\tilde{\pi}_{kt}$ and κ_k
Panel (a) of the figure presents scatter plots of $\tilde{\pi}_k$, time-series means of the estimated perceived prior $\tilde{\pi}_{kt}$ 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. When computing $\tilde{\pi}_{kt}$ from $\tilde{\pi}_{kt}$, we set $\lambda_{out} = 0.0046$. Panel (b) of the figure presents scatter plots of the estimated κ_k also 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.



(a) Average commission rate from simulation



(b) Effects of varying investor composition, marginal cost, and number of brokers

Figure 10: Decomposition of the downward trend in commission fees
Panel (a) of this figure simulates the average commission rate for new entrants over the sample period, using the estimated parameters reported in Table 7. The average commission rate declines by 24.83 basis points, from 29.60 bps in 2007 Q1 to 4.77 bps 2017 Q4, replicating the empirical pattern in Fig. 1. Panel (b) decomposes this decline into contributions from changes in investor composition, marginal cost, and the number of brokers. Each component in Panel (b) is simulated by varying one factor at a time while holding the others fixed at their values estimated for the first quarter of 2007. The change attributable to shifts in investor composition $\tilde{\pi}$ is −19.63 bps, while the change due to variation in marginal cost is −5.59 bps. Changes in the level of competition have only a minimal effect in explaining the downward trend.

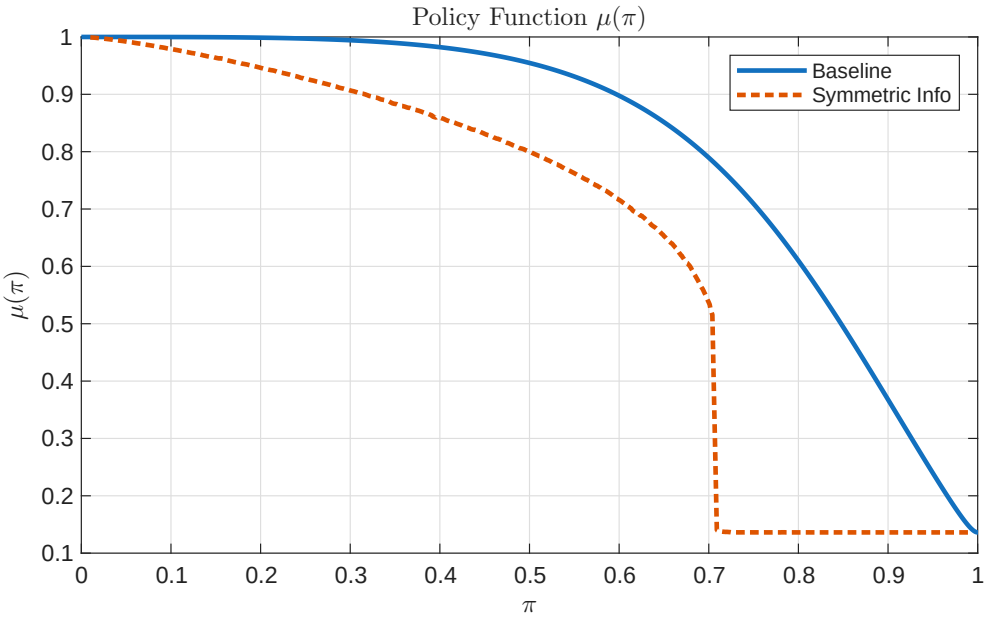
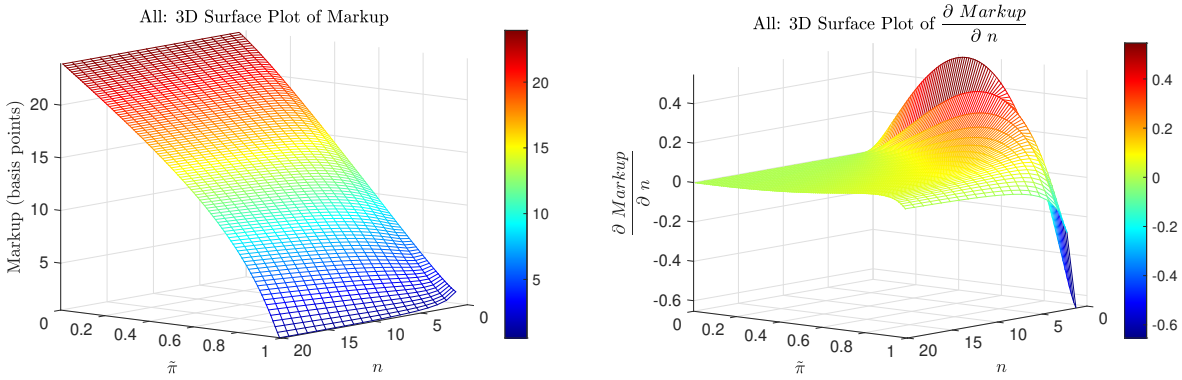
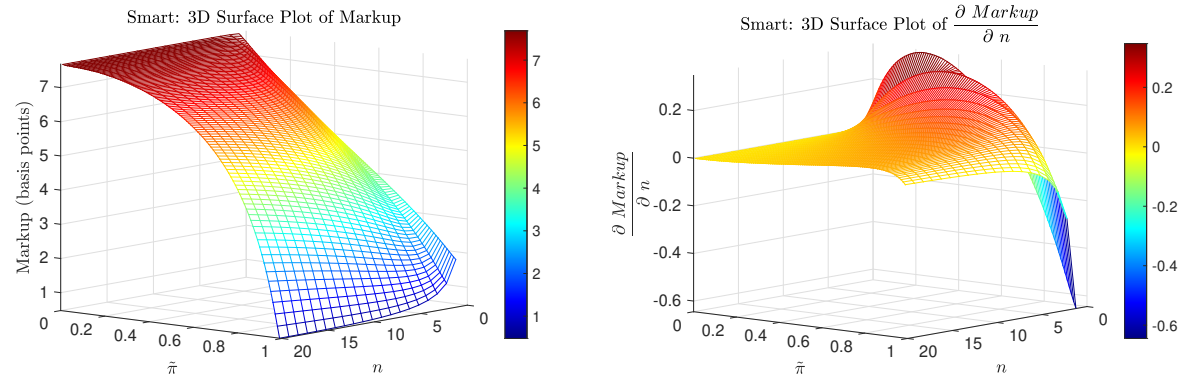


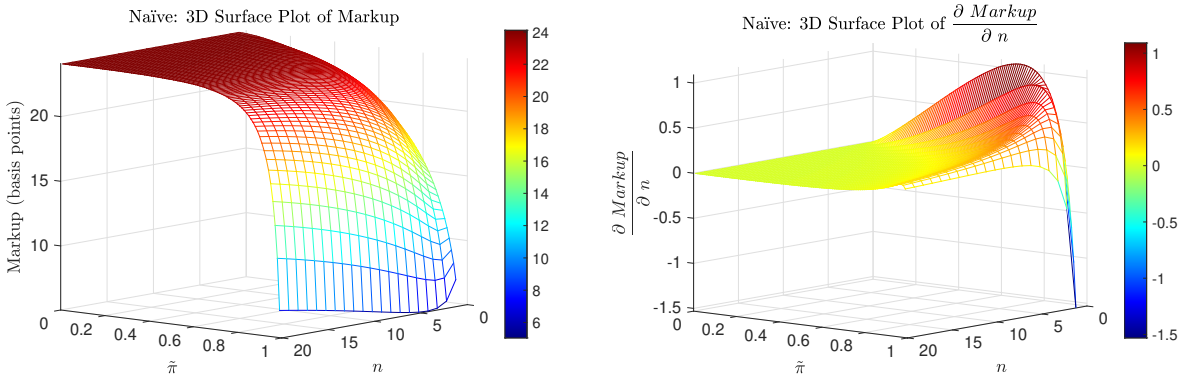
Figure 11: Incumbent's policy function under symmetric information
This figure shows numerical solutions of the incumbent's policy function under an one-shot information symmetry shock. Suppose the incumbent's posterior (π) is disclosed to all its rivals, who then use this as the prior and quote according to their baseline strategies to price this investor after some initial interaction. The "Symmetric Info" line plots the incumbent brokers' best response in this first-order deviation case. Baseline parameters: $\rho = 0.0995$, $\lambda^{out} = 0.0046$, $\alpha = 0.1213$, $t_0 = 0.1599$, $\sigma = 0.2880$, $\tilde{\pi} = 0.7488$, and $n = 1$.



(a) All investors

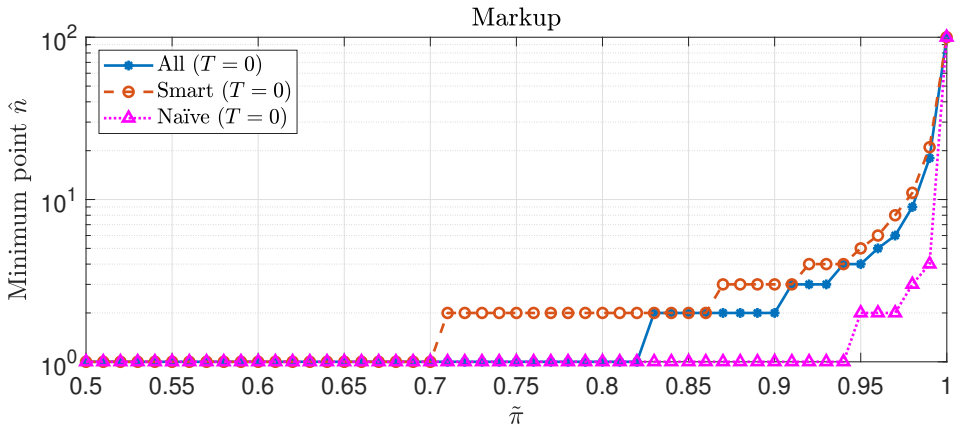


(b) Smart investors

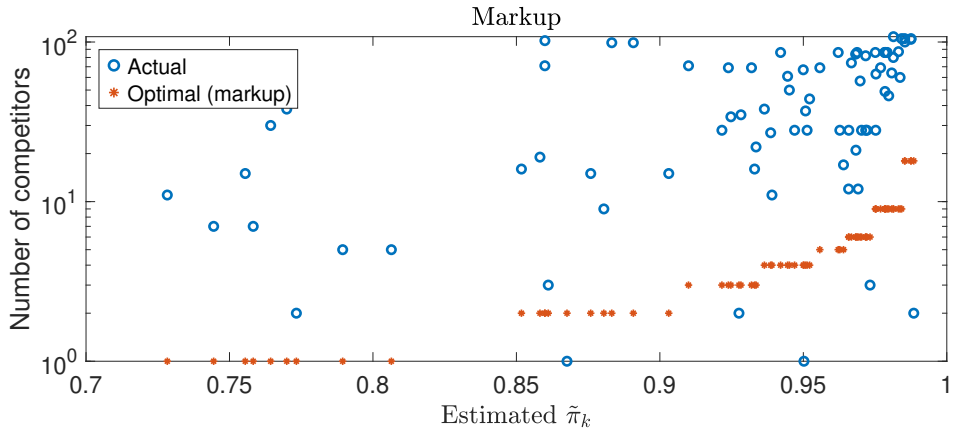


(c) Naïve investors

Figure 12: Competition effect on average initial markup
The left plot in each panel displays average initial markup levels as initial belief $\tilde{\pi}$ and number of competitors n changes. The right plot in each panel shows the corresponding partial derivative with respect to n across the parameter space, which is equivalent to the treatment effect of another brokerage firm entry on markup. Simulation with alternative outcome variables is shown in the Internet Appendix Fig. IA.7, IA.8, IA.9, IA.10, IA.11.



(a) Optimal competition level for smart and naive investors



(b) Optimal competition for all investors v.s. actual competition

Figure 13: Optimal and actual competition levels across branches
Panel (a) of this figure shows the optimal number of competitors that minimize the average markup for different values of $\tilde{\pi}$ for all investors, smart investors, and naive investors. For each branch k, the optimal number of competitors are the minimum point $\hat{n}$ when the associated $\tilde{\pi}$ is closest to the $\tilde{\pi}_k$ that minimizes the average initial markup. Panel (b) of this figure shows the actual and optimal number of competitors that minimize the average markup for all investors across all branches. We use a logarithmic scale for the y-axis to improve clarity. The x-axis represents the perceived prior $\tilde{\pi}_k$ for branch k calculated as the time-series mean of the estimated $\tilde{\pi}_{kt}$. 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
Log 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: Regressions of the initial commission rate on observed characteristics

This table examines the relationship between the initial commission rate and investor characteristics, including age, gender, portfolio size, and the per capita GDP (in thousand Yuan) of their city of residence, using linear regression. *Initial rate* is represented in basis points. *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. *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. In all regressions, we include year-month fixed effects. Standard errors are clustered by year-month. Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)
	Initial rate	Initial rate	Initial rate	Initial rate	Initial rate
Age	-0.046***	—	—	—	-0.019***
	(-12.725)				(-8.025)
Male	—	0.213***	—	—	0.089
		(3.380)			(1.604)
Portfolio size	—	—	-0.792***	—	-0.625***
			(-15.534)		(-13.526)
GDP per capita	—	—	—	-0.051***	-0.048***
				(-11.975)	(-11.161)
Obs	426,786	426,786	419,486	442,236	391,355
Adj-R^2	0.559	0.557	0.626	0.617	0.625
Year-month FE	Yes	Yes	Yes	Yes	Yes

Table 3: 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 4: 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 5: 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. 7. 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 6: 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 7: 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.5a
$\tilde{\pi}_{kt}$	Perceived prior at each branch-quarter pair (k,t)	See Fig. 9a and IA.5a
κ_k	Naive investor's base search cost	See Fig. 9b
m	Load of trading volume in naive investor's search cost	0.7655
t_0	Intensity of initial interaction	0.1599
σ	Rate of learning on investor type	0.2880

Internet Appendix to

"When Competition Backfires: Broker Learning and Commission Fees"

Xindan Li

Baolian Wang

Wei Xiong

Xuewei Yang

Yinhong Zhao

October, 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] We measure $\lambda^{out} = 0.0046$ using public data from Shenzhen Stock Exchange. Thus, $\tilde{\rho} = \rho + \lambda^{out} = 0.1041$ in our analysis.

2. We calibrate λ^{out} using data from Shenzhen Stock Exchange (2017). Page 473 of the report documents each year's total number of investors and new investors. We estimate the investors' exit rate by taking the average for $t = 2007, 2008, ...2017$

$$\text{Exiting Investors}_t = \text{New Investors}_t + \text{All Investors}_{t-1} - \text{All Investors}_t,$$

$$\lambda^{out} = \frac{1}{11}\sum_{t=2007}^{2017}\frac{\text{Exiting Investors}_t}{\text{All Investors}_t}.$$

3. We measure n_{kt} directly from our data as the average number of competing brokers in each prefecture over time.

4. 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

[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.

commission rate is updated twice per year for the average investor, then $\alpha = 2$.

5. 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

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

   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% of their commission rates as our estimate of MC_t.

6. 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, $r^d_{kt}(q)$.

7. 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

[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.

the constant m. Finally, we estimate

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

This is equivalent to matching the following moments:

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

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

8. With the estimated $\bar{\kappa}_k$ and MC_t, we compute the empirical markup for each investor i in quarter t as

$$\widehat{\mu_{kt}(\pi_{it})} = \frac{p_{it} - MC_t}{r^d_{kt}(q_{it}) - MC_t} = \frac{p_{it} - MC_t}{\bar{\kappa}_k / q^{1-m}}.$$

9. 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.

10. We iterate over each branch-quarter pair (k, t) , guessing a value of $\tilde{\pi}_{kt}$ as the inner loop. Given $\tilde{\pi}_{kt}$ and the empirical n_{kt}, we simulate the policy function μ_{kt} for each branch. Using this function, we further simulate $|I|$ investors in branch-quarter (k, t) and compute the theoretical average markup of initial commissions, $\hat{\mu}_{kt}$. We then search for the $\tilde{\pi}_{kt}$ that equates the theoretical and empirical means:

$$\frac{1}{|I|}\sum_{i \in I_{k,t}} \widehat{\mu_{kt}(\pi_{it})} = \bar{\mu}_{kt}.$$

11. Given the estimated $\tilde{\pi}_{kt}$ from the inner loop, we recalculate the policy function μ_{kt}. We then simulate $|I|$ 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.

12. Meanwhile, we measure $\hat{\sigma}_{th}$, the standard deviation of $\widehat{\mu_{kt}(\pi_{it})}$ 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

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

13. 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 $\bar{\kappa_k}$

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

We also recover the smart entrant fraction parameter $\bar{\pi}_{kt}$ from brokers' perceived prior $\tilde{\pi}_{kt}$ from

$$\bar{\pi}_{kt} = \frac{\tilde{\pi}_{kt}}{1 + \frac{n_{kt}}{n_{kt}+1}\frac{\alpha}{\lambda^{out}}(1-\tilde{\pi}_{kt})}.$$

IA.5 Additional Tables and Figures

Table IA.1: Extensive and intensive margin analyses of initial commission rates
This table examines the relationship between the initial commission rate and investor characteristics, including age, gender, portfolio size, and the per capita GDP (in thousand Yuan) of their city of residence, using linear regression. The dependent variable *Discount dummy* is a dummy that equals 1 if and only if the initial commission rate is strictly lower than the regulatory cap (i.e., 30 bps). The dependent variable *Discount size* is defined as the difference between the regulatory cap (i.e., 30 bps) and the actual initial commission rate. In the lower panel, we only include observations with *Discount size* > 0. *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. *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. In all regressions, we include year-month fixed effects. Standard errors are clustered by year-month. Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)
Panel A: Extensive margin					
	Discount dummy	Discount dummy	Discount dummy	Discount dummy	Discount dummy
Age	0.063***	—	—	—	0.019
	(3.257)				(0.919)
Male	—	-0.644***	—	—	-0.441***
		(-3.368)			(-2.695)
Portfolio size	—	—	1.695***	—	1.309***
			(11.321)		(11.366)
GDP per capita	—	—	—	0.126***	0.126***
				(9.505)	(8.867)
Obs	426,793	426,793	419,492	442,243	391,361
Adj-R^2	0.553	0.553	0.596	0.572	0.591
Year-month FE	Yes	Yes	Yes	Yes	Yes
	(1)	(2)	(3)	(4)	(5)
Panel B: Intensive margin					
	Discount size	Discount size	Discount size	Discount size	Discount size
Age	0.067***	—	—	—	0.032***
	(12.351)				(8.939)
Male	—	-0.255***	—	—	-0.107
		(-3.047)			(-1.424)
Portfolio size	—	—	0.819***	—	0.653***
			(13.291)		(10.562)
GDP per capita	—	—	—	0.063***	0.061***
				(12.892)	(11.510)
Obs	229,211	229,211	232,812	243,776	205,755
Adj-R^2	0.313	0.305	0.390	0.424	0.442
Year-month FE	Yes	Yes	Yes	Yes	Yes

Table IA.2: Regressions of Behavioral Biases on the Initial Commission Rate
This table is the counterpart of Table 4. All settings are the same as those in Table 4 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.3: Regressions of Behavioral Biases on the Initial Commission Rate
This table is the counterpart of Table 4. All settings are the same as those in Table 4 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

Table IA.4: Behavioral biases and the regulatory cap

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. *Discount dummy* is a dummy that equals 1 if and only if the initial commission rate is strictly lower than the regulatory cap (i.e., 30 bps). *Discount size* (in percentage points) is defined as the gap between the regulatory cap and the initial commission rate. *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: Regressions on *Discount dummy*						
Discount dummy	-0.013***	0.987***	-0.758***	-0.040***	-0.002	0.302**	-0.383***
	(-9.000)	(6.825)	(-5.711)	(-5.301)	(-0.703)	(2.483)	(-2.658)
Portfolio size	-0.001***	-0.127***	-0.264***	-0.019***	-0.001***	0.061**	-0.008
	(-10.984)	(-7.477)	(-31.454)	(-18.787)	(-6.280)	(2.548)	(-0.405)
Age	0.001***	-0.042***	-0.020***	-0.005***	0.000***	-0.029***	0.040***
	(13.440)	(-4.458)	(-5.682)	(-17.416)	(3.170)	(-4.655)	(7.440)
Male	-0.036***	2.160***	0.062	0.046***	-0.004***	-0.502***	-0.627***
	(-28.216)	(11.375)	(1.221)	(15.957)	(-3.280)	(-9.001)	(-7.734)
Obs	406,198	419,355	406,222	418,705	406,222	406,222	222,703
Adj-R^2	0.044	0.050	0.028	0.048	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
	Panel B: Regressions on *Discount size* (conditional on *Discount size* > 0)						
Discount size	-0.129***	8.745***	-12.271***	-0.564***	-0.051***	-0.465	-2.508**
	(-8.247)	(7.666)	(-17.211)	(-15.481)	(-2.744)	(-0.715)	(-2.473)
Portfolio size	-0.001***	-0.155***	-0.254***	-0.018***	-0.002***	0.071**	0.001
	(-9.218)	(-6.658)	(-21.412)	(-12.029)	(-5.307)	(2.130)	(0.041)
Age	0.001***	-0.054***	-0.031***	-0.005***	0.000***	-0.044***	0.042***
	(7.283)	(-4.806)	(-8.731)	(-17.477)	(2.960)	(-5.844)	(5.077)
Male	-0.037***	2.729***	0.115*	0.050***	-0.005***	-0.305***	-0.822***
	(-19.915)	(12.057)	(1.842)	(12.435)	(-3.153)	(-5.147)	(-8.176)
Obs	216,286	224,615	217,346	224,245	217,346	217,346	116,469
Adj-R^2	0.038	0.035	0.026	0.061	0.052	0.027	0.002
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table IA.5: Sophistication and commission renegotiation

This table reports results of the relationship between investor sophistication and commission renegotiation. Columns (1) and (2) examine the renegotiation probability (i.e., extensive margin), and columns (3) and (4) examine the magnitudes of commission rate reduction conditional on renegotiation (i.e., intensive margin). In columns (1) and (2), the dependent variable is I_{Change}, which equals one if investor i received a rate cut in month t, and zero otherwise. In columns (3) and (4), the dependent variable is the negative value of relative percentage changes in commission rates. Specifically, we report the results based on the following regression: $Y_{it} = \beta\ Smart_{it} + \gamma\ Controls + \epsilon_{it}$. 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)
	I_{Change}	I_{Change}	$\frac{-Change}{\text{Pre-rate}}$	$\frac{-Change}{\text{Pre-rate}}$
Smart	0.128***	0.023***	1.355***	0.762***
	(12.574)	(3.011)	(13.179)	(8.268)
Asset_{t-1}	—	0.077***	—	1.209***
		(6.803)		(18.297)
Turnover_{t-1}	—	0.086***	—	0.249***
		(8.767)		(4.485)
Obs	25,010,847	25,010,847	352,672	352,672
Adj-R^2	0.264	0.268	0.369	0.382
Bin FE	Yes	Yes	Yes	No
Time FE	Yes	Yes	Yes	Yes

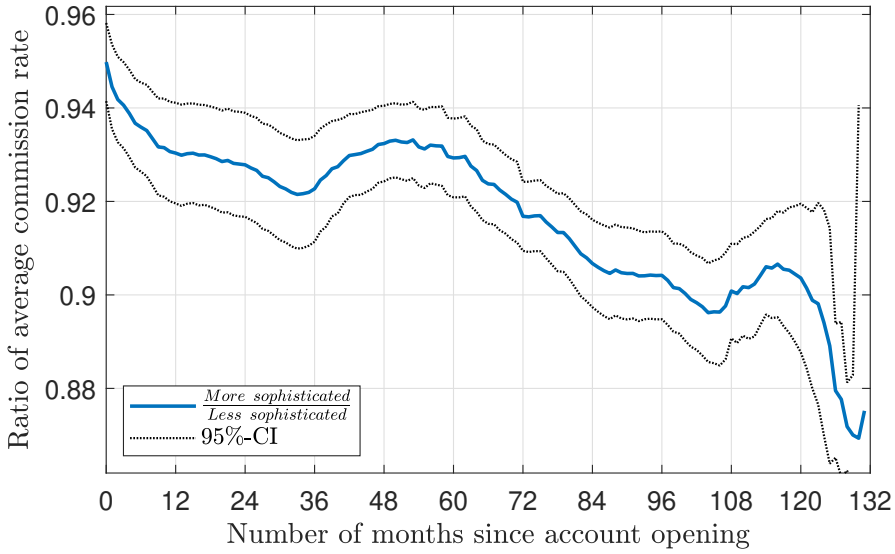


Figure IA.1: Ratio between average commission rates by investor group
This figure shows the evolution of commission rates by investor groups. For each cohort of investors (defined as investors who open their accounts in the same month), we sort them into two groups by the median value of their composite bias measure, defined as per Table 3. We report the ratio between the two groups' average commission rates.

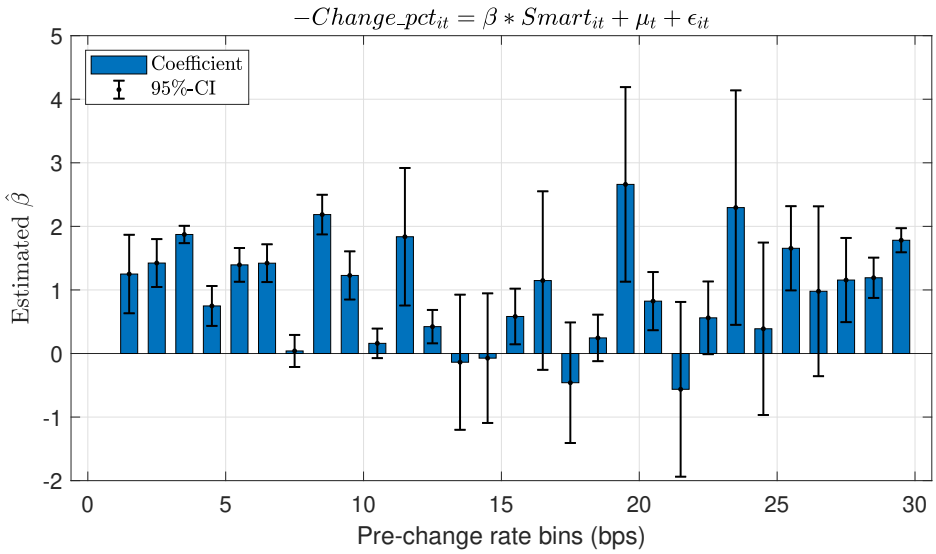


Figure IA.2: Commission rate cuts in renegotiation by sophistication
For each case of rate renegotiation, we record its pre-change rate. We then group all observations into 30 bins (each with length of 1 bp) according to the level of their pre-change rates. For each cohort of investors (defined as investors 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 is the negative value of relative percentage change.

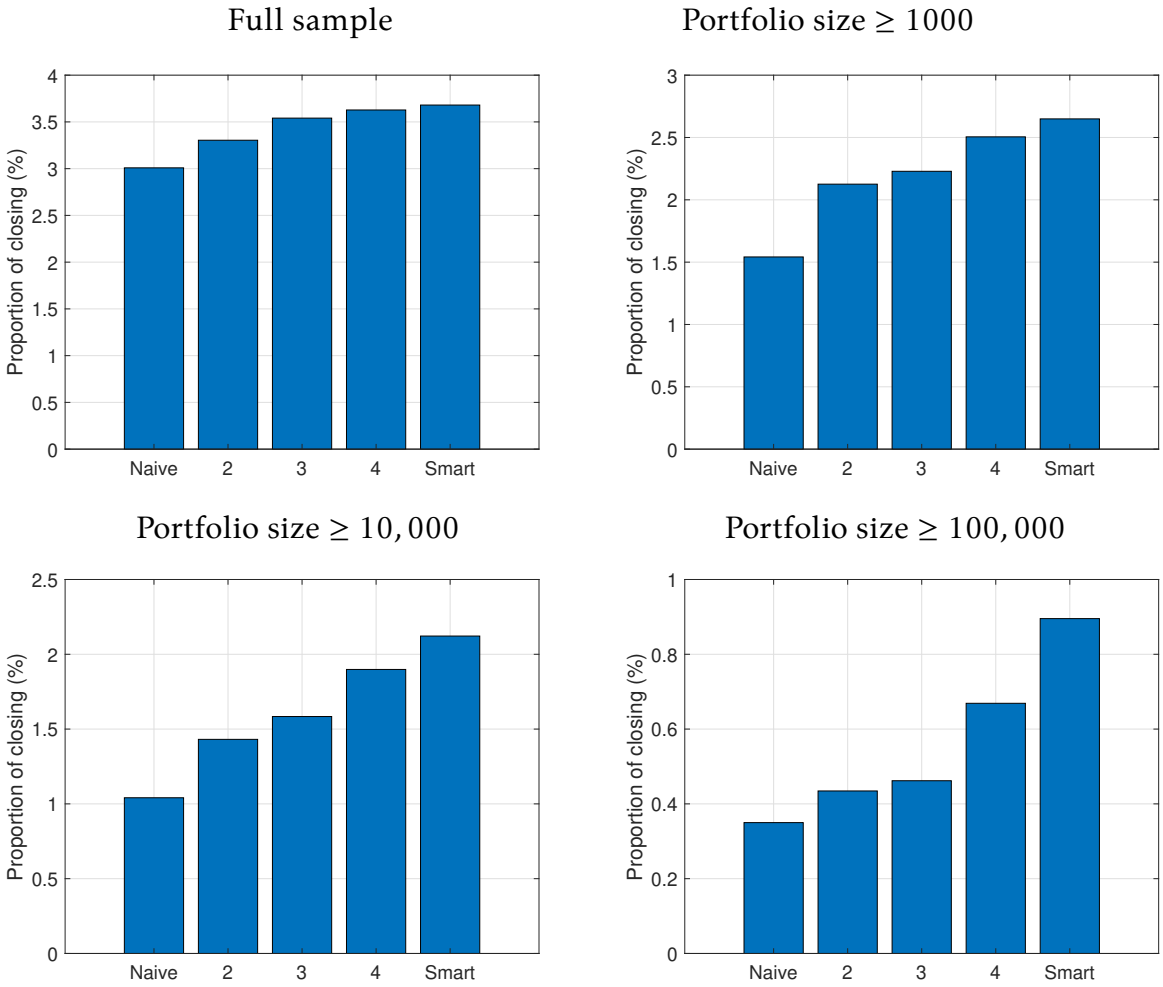


Figure IA.3: Proportion of account closing by sophistication
We report the proportion of account closing for each quintile of sophistication measured by the composite bias measure. Four samples are considered. The first is the full sample. The second incorporates those with average portfolio size no less than 1000 CNY during a three-month period before account closing. Similarly, the third and the forth samples consist of investors whose pre-closing portfolio size is no less than 10,000 CNY and 100,000 CNY, respectively.

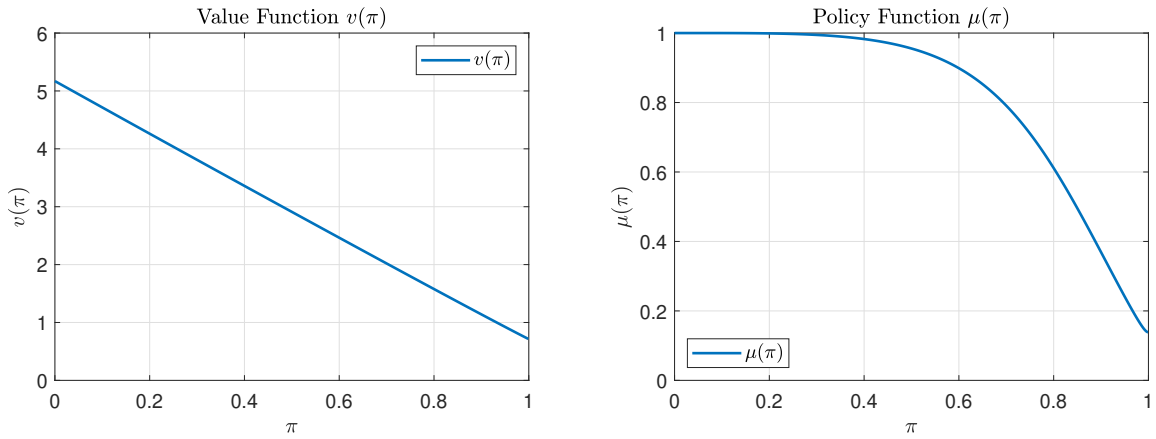
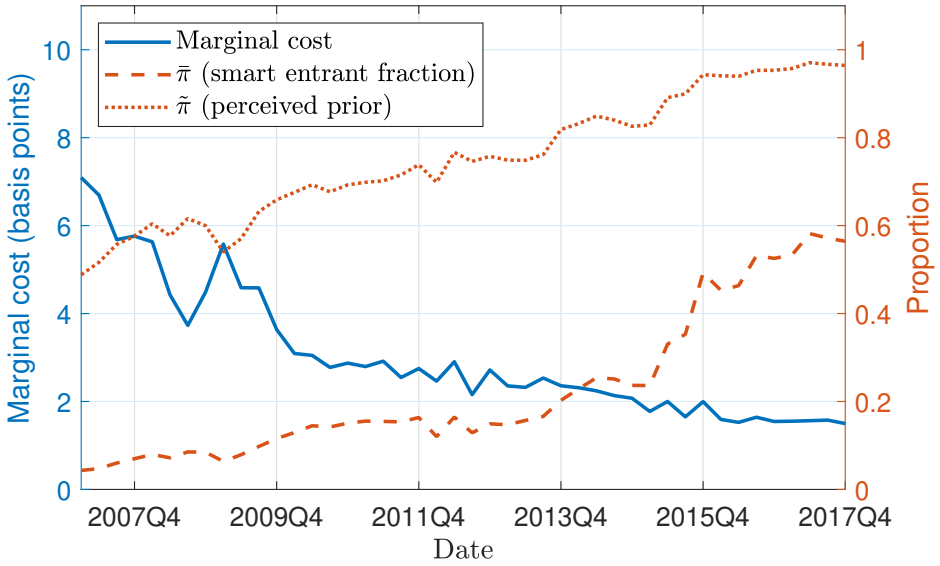
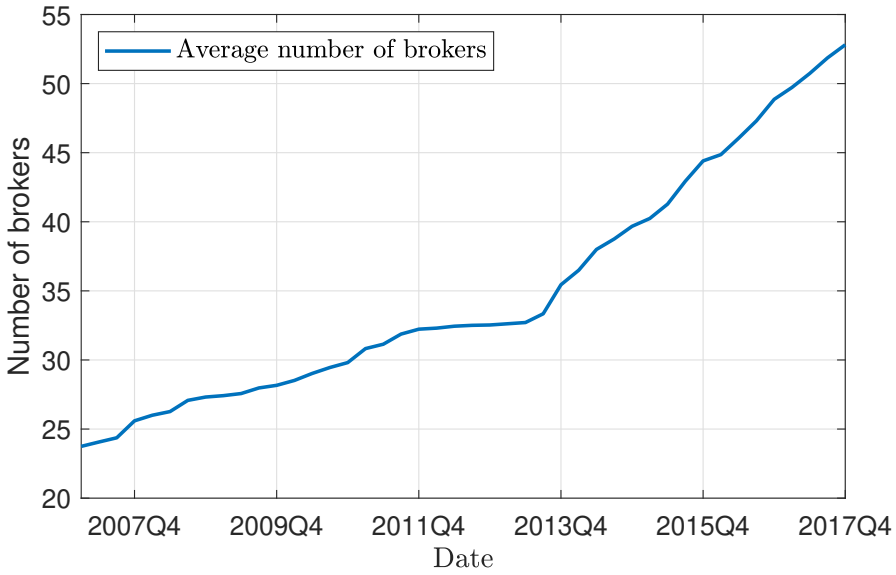


Figure IA.4: Numerical solutions of value function and policy function
This figure shows numerical solutions of value function and policy function with baseline parameters $\rho = 0.0995$, $\lambda^{out} = 0.0046$, $\alpha = 0.1213$, $t_0 = 0.1599$, $\sigma = 0.2880$, $\tilde{\pi} = 0.7488$, and $n = 1$. v and π are solved simultaneously by a shooting algorithm that makes an initial guess of $v'(0)$.



(a) Marginal cost and proportion of smart investors



(b) Average number of competing brokers

Figure IA.5: Time trend of estimated parameters
Panel (a) displays the estimated marginal cost (in basis points, left axis), $\tilde{\pi}_t$ (right axis), cross-sectional means of the estimated perceived prior $\tilde{\pi}_{kt}$, and $\bar{\pi}_t$ (right axis), cross-sectional means of the estimated smart entrant fraction $\bar{\pi}_{kt}$. Panel (b) depicts the time-series measurement of the monthly cross-branch average number of brokers located in the host city.

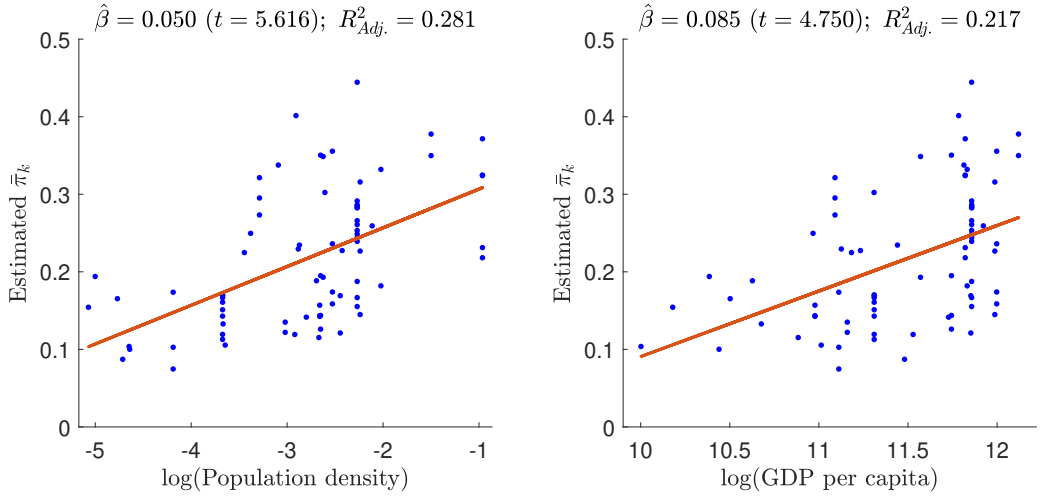


Figure IA.6: Smart entrant fraction versus population density and number of brokers This figure presents scatter plots of $\bar{\pi}_k$, time-series means of the estimated smart entrant fraction $\bar{\pi}_{kt}$ against two explanatory variables: (1) the logarithm of population density of branch k's 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.

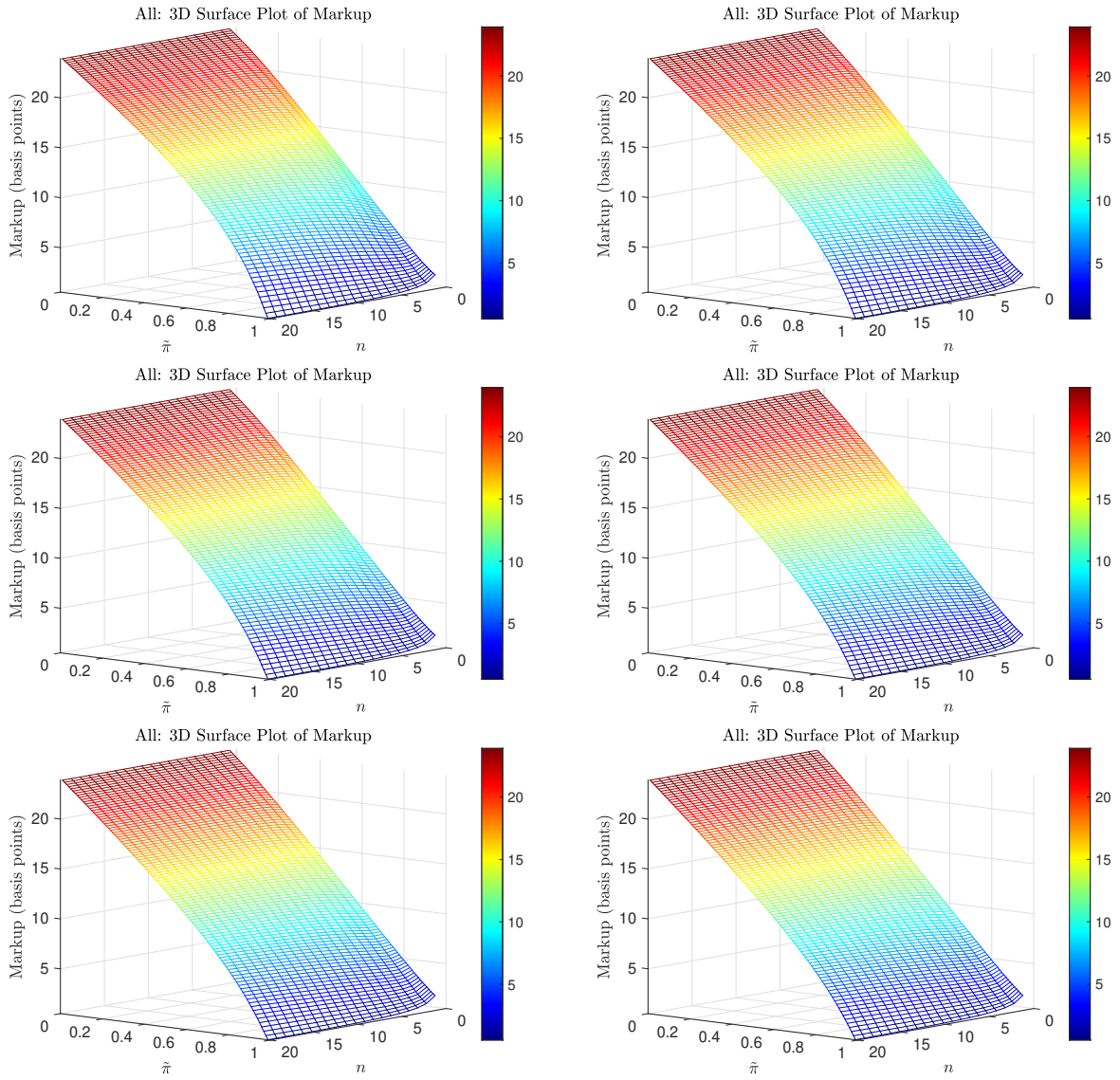


Figure IA.7: Competition effect on average markup
The left plots display average markup levels across all investors at $T = 0, 2, 4$ as initial belief $\tilde{\pi}$ and number of competitors n changes. The right plots display average markup levels across all investors at $T = 1, 3, 5$ as initial belief $\tilde{\pi}$ and number of competitors n changes.

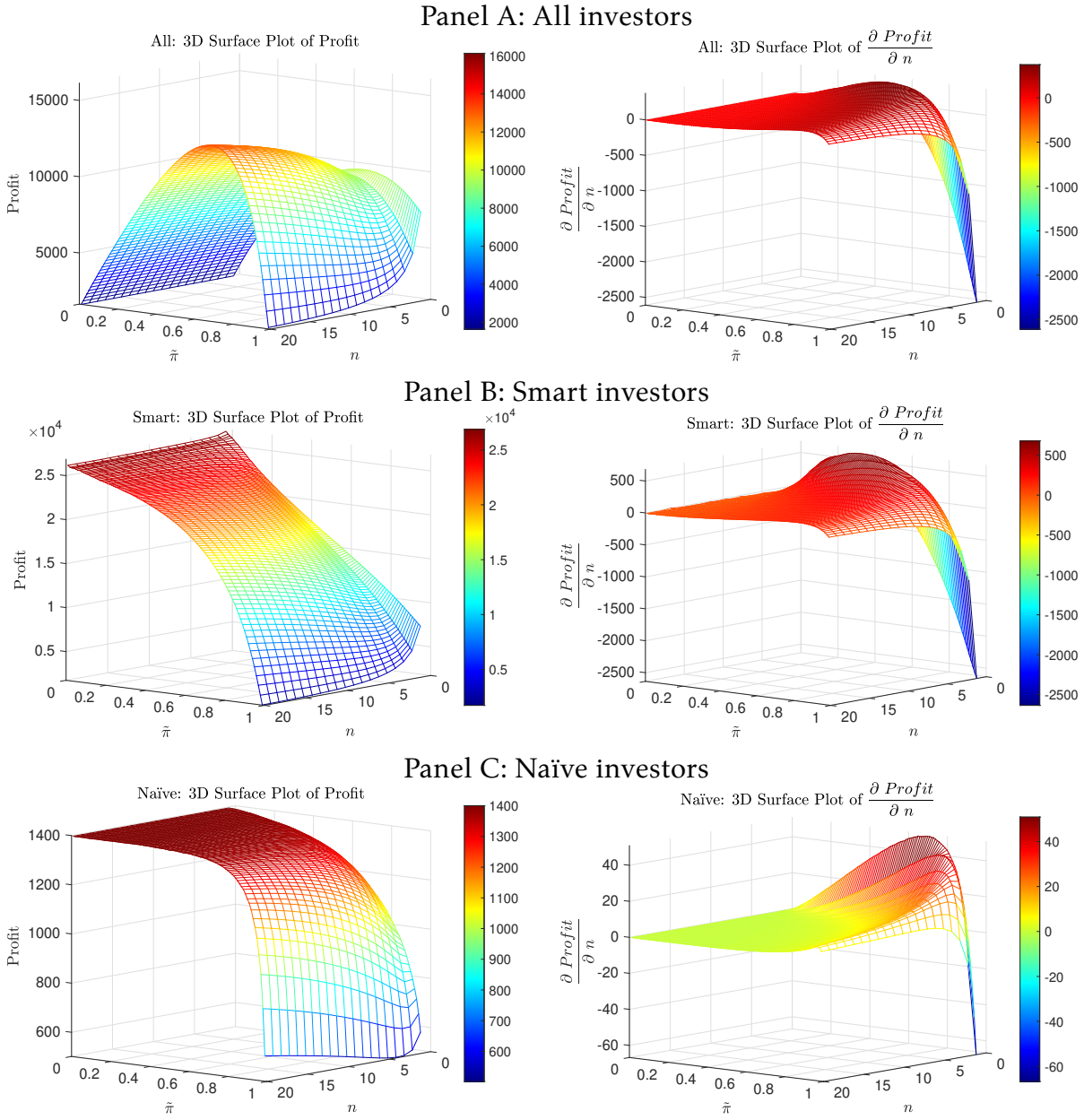


Figure IA.8: Competition effect on average profit per account
The left plot in each panel displays lifetime net profit per account as initial belief $\tilde{\pi}$ and number of competitors n changes. The right plot in each panel shows the corresponding partial derivative with respect to n. The top-left panel demonstrates non-monotonicity over $\tilde{\pi}$ because smart investors mechanically trade more frequently than naive ones; in our simulation, greater trading by smart investors raises total broker profits.

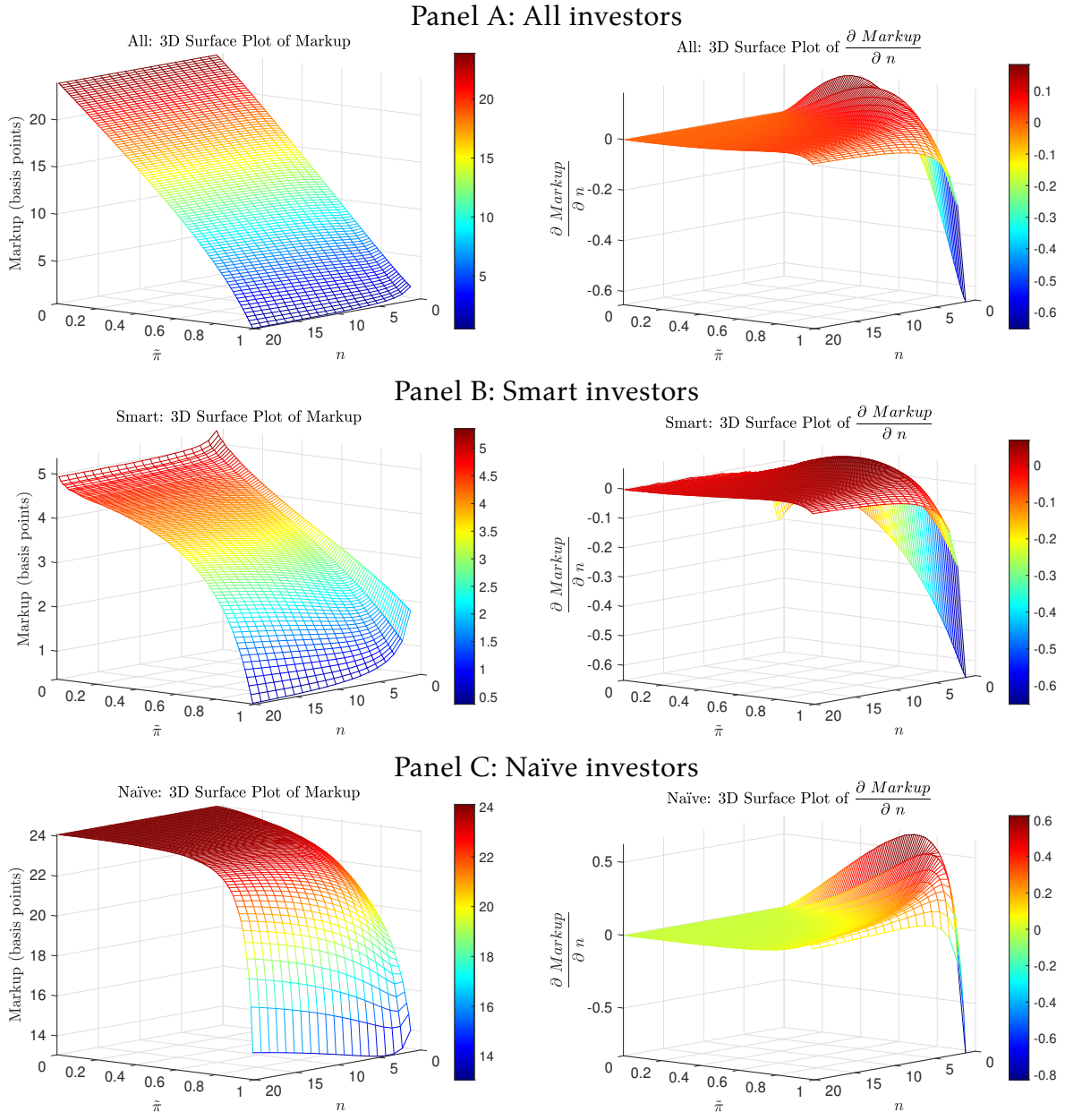


Figure IA.9: Competition effect on average markup at $T = 5$
The left plot in each panel displays average markup levels at $T = 5$ as initial belief $\tilde{\pi}$ and number of competitors n changes. The right plot in each panel shows the corresponding partial derivative with respect to n.

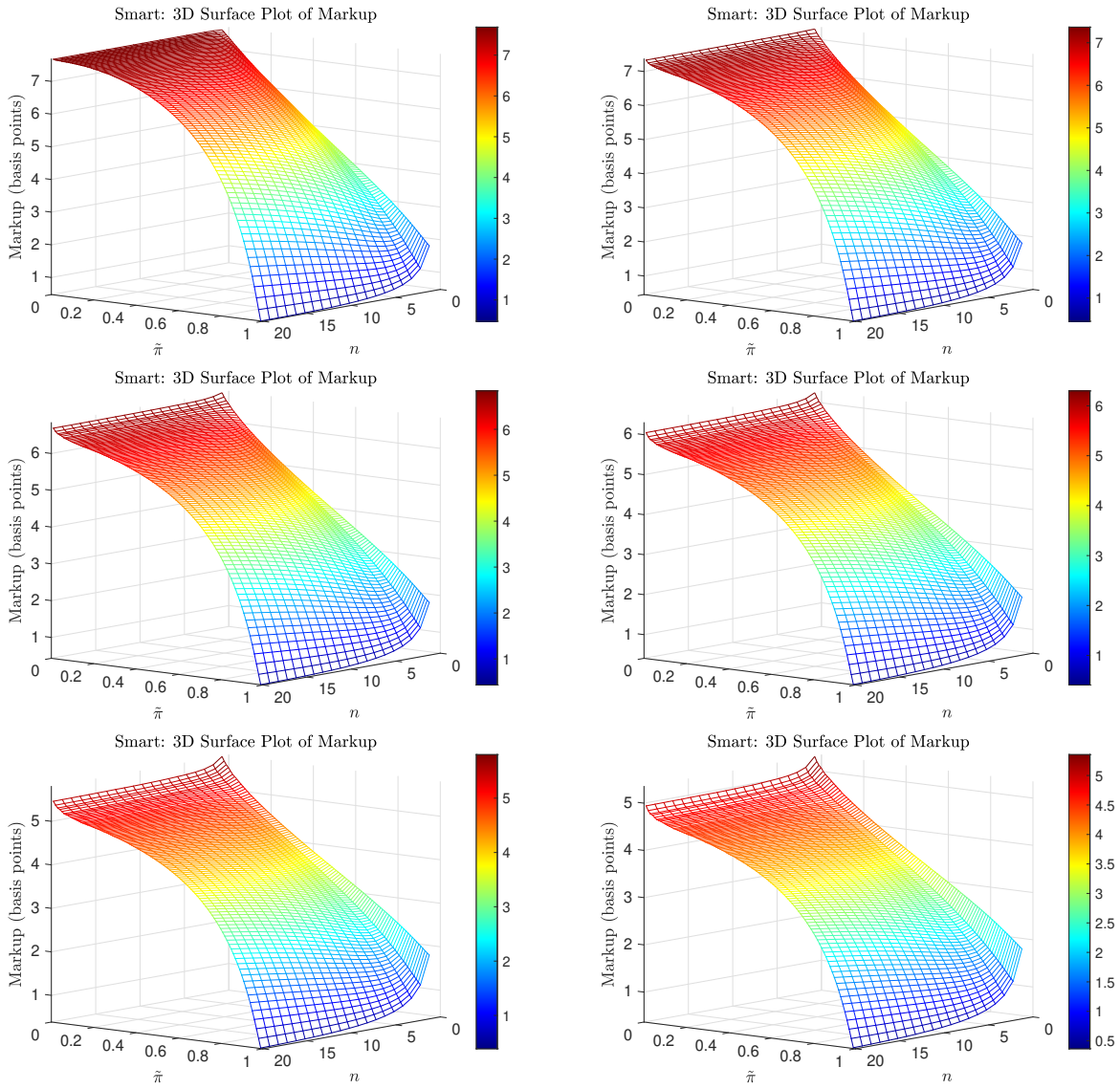


Figure IA.10: Competition effect on smart investors' average markup
The left plots display average markup levels across smart investors at $T = 0, 2, 4$ as initial belief $\tilde{\pi}$ and number of competitors n changes. The right plots display average markup levels across smart investors at $T = 1, 3, 5$ as perceived prior $\tilde{\pi}$ and number of competitors n changes.

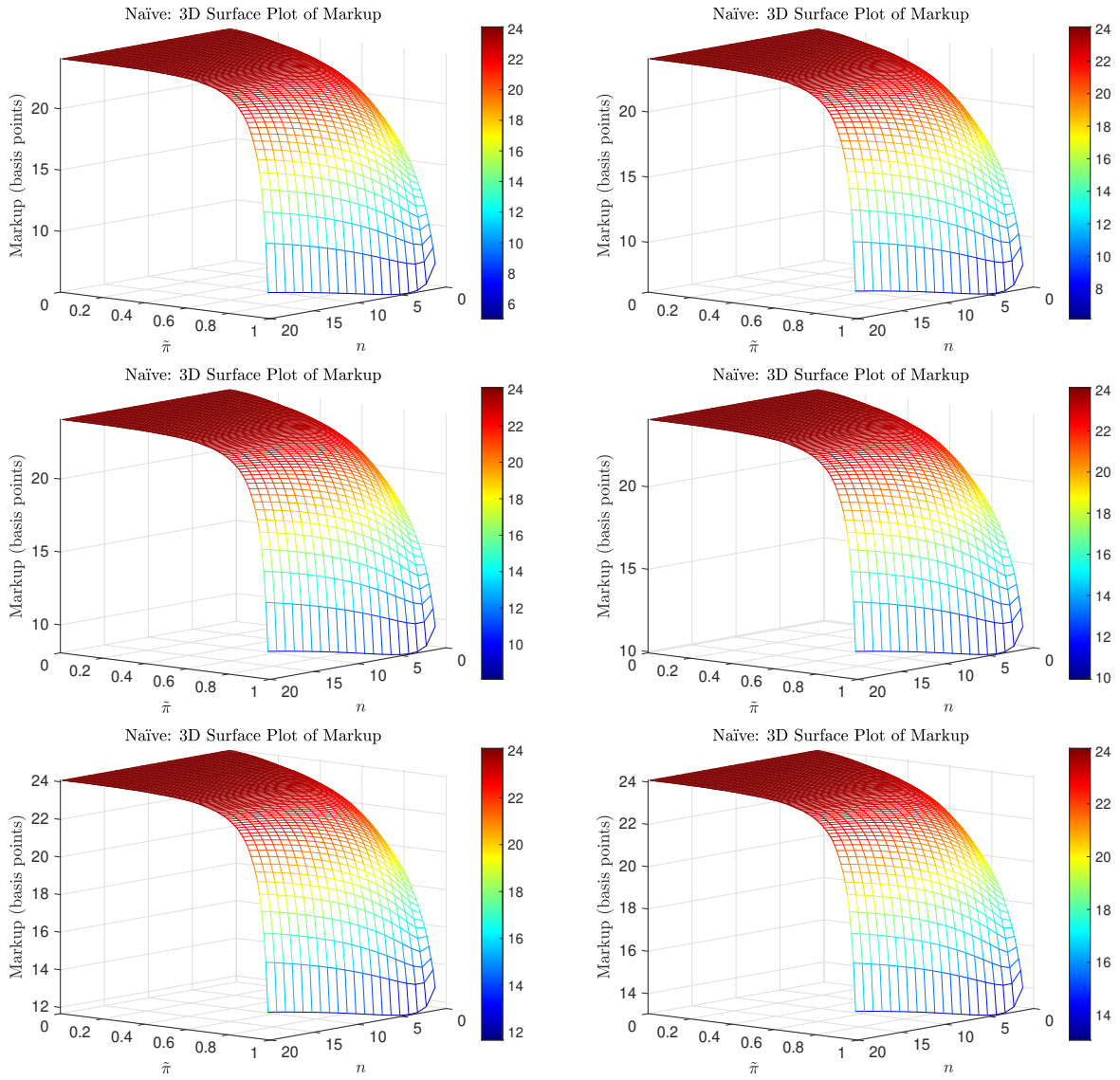
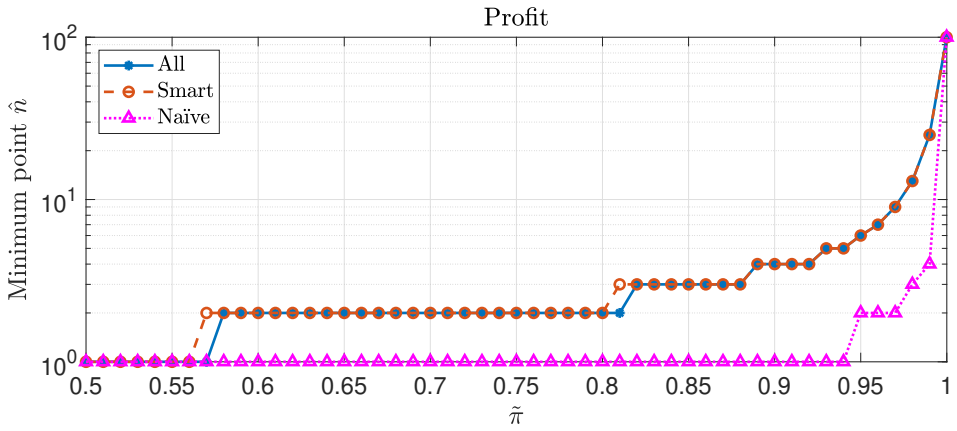
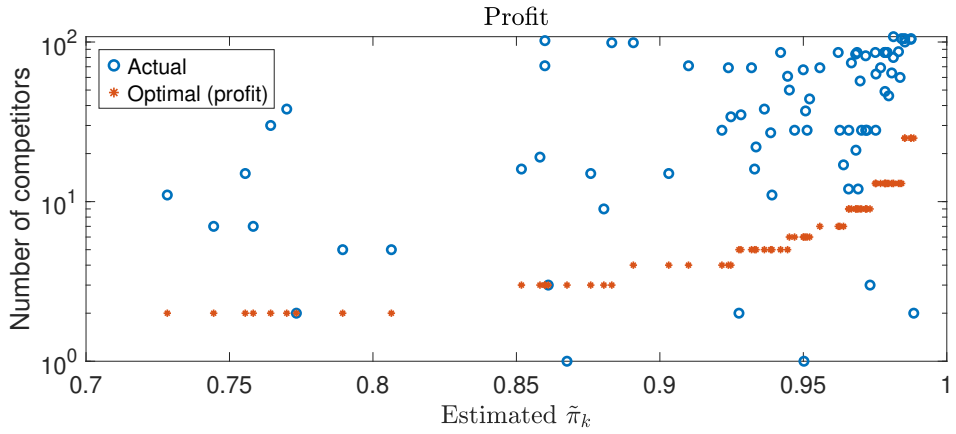


Figure IA.11: Competition effect on naïve investors' average markup
The left plots display average markup levels across naïve investors at $T = 0, 2, 4$ as initial belief $\tilde{\pi}$ and number of competitors n changes. The right plots display average markup levels across naïve investors at $T = 1, 3, 5$ as perceived prior $\tilde{\pi}$ and number of competitors n changes.

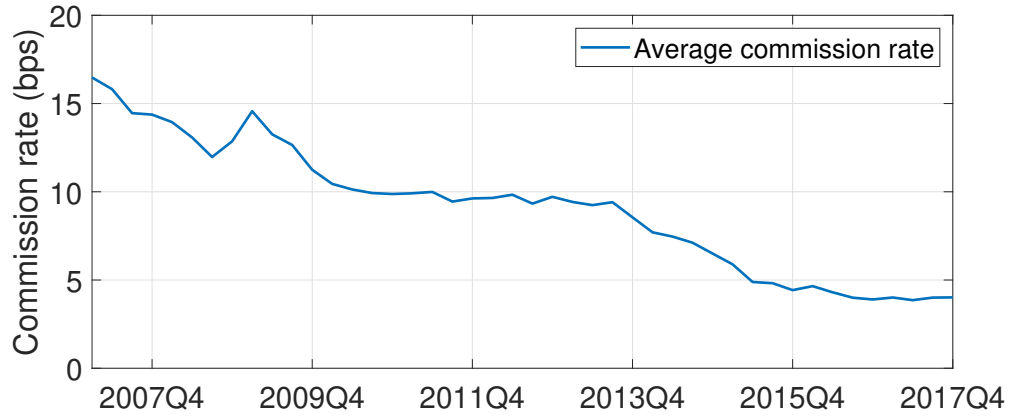


(a) Optimal competition level for smart and naive investors

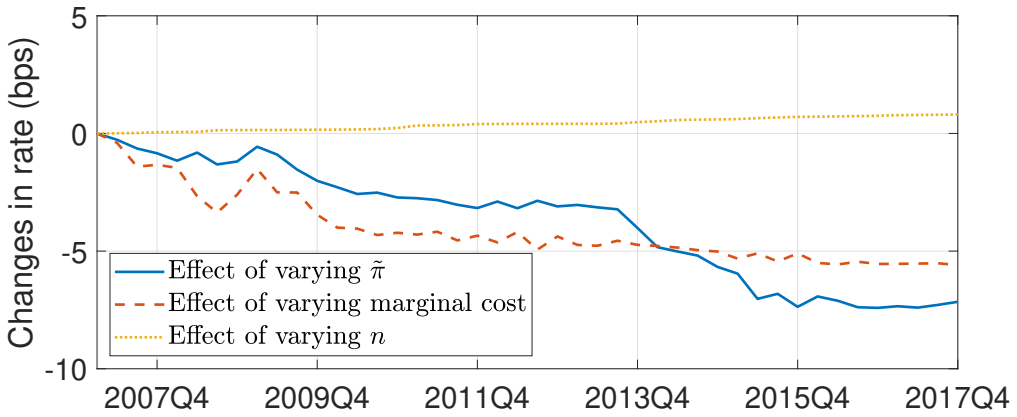


(b) Optimal competition for all investors v.s. actual competition

Figure IA.12: Actual and optimal competition levels across branches by profit
Panel (a) of this figure shows the optimal number of competitors that minimize the average lifetime profit for different values of $\tilde{\pi}$ for all investors, smart investors, and naive investors. For each branch k, the optimal number of competitors are the minimum point $\hat{n}$ when the associated $\tilde{\pi}$ is closest to the $\tilde{\pi}_k$ that minimizes the average lifetime profit. Panel (b) of this figure shows the actual and optimal number of competitors that minimize the average lifetime profit for all investors across all branches. We use a logarithmic scale for the y-axis to improve clarity. The x-axis represents the perceived prior $\tilde{\pi}_k$ for branch k calculated as the time-series mean of the estimated $\tilde{\pi}_{kt}$. The actual number of competitors is calculated as the time-series average of end-of-quarter figures.



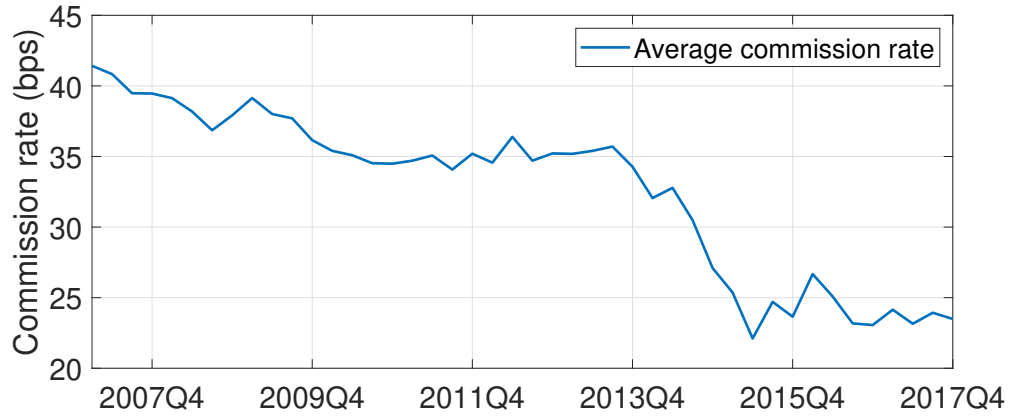
(a) Average commission rate from simulation



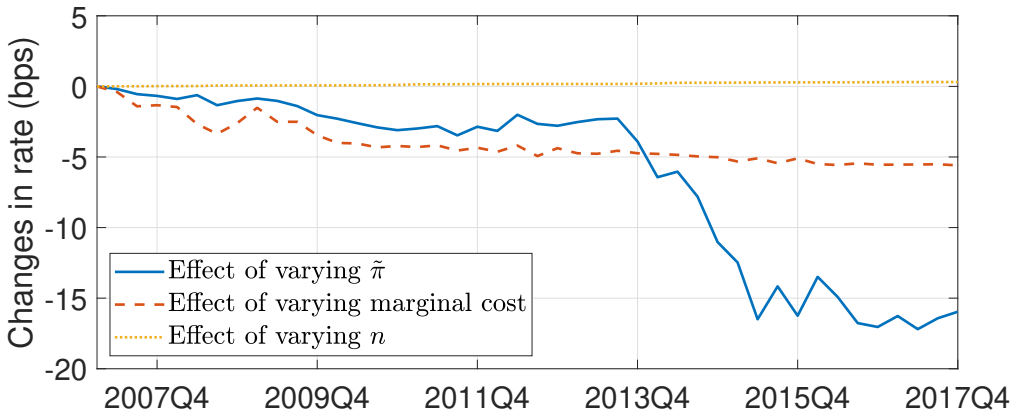
(b) Effects of varying investor composition, marginal cost, and number of brokers

Figure IA.13: Decomposition of the downward trend in commission fees (smart investors)

Panel (a) of this figure simulates the average commission rate for new entrants over the sample period, using the estimated parameters reported in Table 7. Panel (b) decomposes this decline into contributions from changes in investor composition, marginal cost, and the number of brokers. Each component in Panel (b) is simulated by varying one factor at a time while holding the others fixed at their values estimated for the first quarter of 2007.



(a) Average commission rate from simulation



(b) Effects of varying investor composition, marginal cost, and number of brokers

Figure IA.14: Decomposition of the downward trend in commission fees (naïve investors)

Panel (a) of this figure simulates the average commission rate for new entrants over the sample period, using the estimated parameters reported in Table 7. Panel (b) decomposes this decline into contributions from changes in investor composition, marginal cost, and the number of brokers. Each component in Panel (b) is simulated by varying one factor at a time while holding the others fixed at their values estimated for the first quarter of 2007.