ISSN 0956-8549-815

Resolving the Excessive Trading Puzzle: An Integrated Approach Based on Surveys and Transactions

By

Hongqi Liu

Cameron Peng

Wei A. Xiong

Wei Xiong

DISCUSSION PAPER NO 815

PAUL WOOLLEY CENTRE WORKING PAPER No 69

September 2020

Resolving the Excessive Trading Puzzle: An Integrated Approach Based on Surveys and Transactions*

Hongqi Liu, Cameron Peng, Wei A. Xiong, and Wei Xiong

July 2020

Abstract

The literature has provided over a dozen explanations for the widely documented excessive trading puzzle of retail investors trading so much that it hurts their performance. It is difficult to use transaction data to differentiate these explanations as they share similar predictions by design. To confront this challenge, we design and administer a nationwide survey to elicit investors' responses to an exhaustive list of trading motives. By merging survey responses with account-level transaction data, we validate survey responses with actual trading behaviors and compare the power of survey-based and transaction-based trading motives. A horse race among survey-based trading motives suggests that perceived information advantage and gambling preference dominate other explanations. Moreover, other popular arguments, such as neglect of trading costs, do not contribute to excessive trading.

*Hongqi Liu: Chinese University of Hong Kong, Shenzhen (email: hongqiliu@cuhk.edu.cn); Cameron Peng: London School of Economics and Political Science (email: c.peng9@lse.ac.uk); Wei A. Xiong: Shenzhen Stock Exchange (email: wxiong@szse.cn); Wei Xiong: Princeton University and NBER (email: wxiong@princeton.edu). We thank Jingxuan Chen and Zi Ye for able research assistance. For helpful comments, we thank Nick Barberis, Hui Chen, Thummim Cho, Bing Han, Rawley Heimer, Xing Huang, Lawrence Jin, Christian Julliard, Terry Odean, Søren Leth-Petersen, Andrei Shleifer, and seminar participants at Baruch, CEPR Household Finance Workshop, Chapman, CUHK Shenzhen, Imperial College, INSEAD, LSE, NBER Behavioral Finance Meeting, Princeton, and Tilburg. Cameron Peng acknowledges financial support from the Paul Woolley Centre at LSE. Wei A. Xiong acknowledges support from the National Natural Science Foundation of China (Project Number 71703104).

The field of behavioral economics has advanced significantly over the last few decades by using keen insights from psychology to explain many anomalies in individuals' economic and financial decision making.[1] A byproduct of such rapid development, however, is that researchers often face multiple behavioral biases—perhaps too many—for explaining each of these anomalies. For example, consider the excessive trading puzzle, which documents that retail investors appear to be trading *too much*: they perform poorly relative to the market index before fees, transaction costs make their performance even worse, and those who trade the most often perform the worst (Odean 1999; Barber and Odean 2000). Motivated by these puzzling facts, the literature has proposed a number of behavioral explanations, for example, overconfidence, realization utility, gambling preference, sensation seeking, social interaction, and low financial literacy, beyond standard arguments such as portfolio rebalancing and liquidity needs (see Table 1 for a complete list). This large number of behavioral explanations is not satisfying: it is unlikely that all these explanations are equally important, and it is also possible that certain explanations may be subsumed by others. To further develop this field, it is important to consolidate the multiple explanations for each anomaly to develop a unified conceptual framework—one that is based on a small number of biases and explains a wide range of individual behaviors.

The consolidation task is challenging because many of the existing explanations, by design, share similar predictions on a targeted anomaly. While some explanations may offer different predictions on more subtle dimensions, the power from testing these subtle predictions is often constrained by the availability of administrative data. It is even harder to compare multiple explanations at the same time, as constructing a large number of empirical proxies is often difficult, if not implausible, within a single dataset. The recent literature, for example, Greenwood and Shleifer (2014), Choi and Robertson (2019), and Chinco, Hartzmark and Sussman (2019), has turned to survey-based approaches by having investors self-examine and report the drivers of their trading and investment decisions. Survey-based approaches permit collection of information on multiple explanations, which information can then facilitate a horse race. However, there are also some common concerns about the use of survey data in economic analysis, specifically, that respondents may not truthfully report their answers and even if they do, their subjective answers may not translate into real actions (Bertrand and Mullainathan (2001) and Cochrane (2011)).

[1] See DellaVigna (2009), Barber and Odean (2013), and Barberis (2018) for recent literature reviews.

In this paper, we adopt a new approach to address the excessive trading puzzle by combining surveys with transactions. This integrated approach enables us to overcome the challenges posed by the existing approaches that are based on either administrative data or surveys alone. First, the use of surveys allows us to elicit investor responses to a large set of trading motives, making it possible to directly compare competing explanations for excessive trading. For certain explanations, such as perceived information advantage, it is inherently difficult to construct their empirical proxies from administrative data. However, surveys allow researchers to elicit responses to these subtle trading motives through investors' introspection and self-examination. To our knowledge, this is the first attempt to measure and compare such a wide range of explanations for excessive trading. Second, by merging survey responses with transaction data at the individual level, we directly verify that survey responses are largely consistent with the actual trading patterns they are designed to capture. This consistency provides justification for the use of surveys, not only to our analysis of the excessive trading puzzle, but to other studies as well.

Specifically, we designed and administered a nation-wide survey in China through the Investor Education Center at the Shenzhen Stock Exchange, with the respondents randomized across regions and brokers. The survey asked a series of multiple-choice questions related to financial literacy, return expectations, and, most importantly, an exhaustive list of trading motives. The survey took place in September 2018; more than 10,000 investors responded.

To understand what drives the variation in trading intensity across investors, we merge the survey responses with account-level transaction data from the Shenzhen Stock Exchange. This step gives rise to a unique advantage of our setting: we are able to link an investor's survey responses with her actual trading behavior and examine their consistency. We provide four pieces of evidence to show that subjective survey responses are consistent with real actions: 1) survey-based measures of gambling preference explain the tendency to buy lottery-like stocks, 2) survey-based measures of extrapolation explain the tendency to buy stocks with positive recent returns, 3) investors who are more risk-averse according to the survey hold less volatile stocks, and 4) investors with higher return expectations increase their stock holdings by more.

After this consistency check, we formally examine the explanatory power of survey-based trading motives for excessive trading. As a baseline exercise, we first run a series of cross-sectional regressions of turnover on each trading motive *alone*. These regressions confirm that many of the previous explanations for excessive trading also hold true in our sample. We then include all survey-based trading motives as regressors to compare their explanatory power in a horse race. Together, these two sets of exercises reveal a number of novel findings.

First, two trading motives stand out in the horse race as the dominant drivers of excessive trading: gambling preference and perceived information advantage. Their explanatory power is sizable: while the standard deviation of monthly turnover rate in our sample is 123%, gambling preference can explain up to 21% and perceived information advantage can explain up to 24%. These two motives contribute to an annualized transaction fee of 0.6% and 0.7%, respectively, implying substantial investment consequences borne by investors who display either or both of these trading motives.

Second, in further support of these two channels, we find that survey-based gamblers trade smaller, high-beta, more volatile, and more positively skewed stocks. However, the stocks they buy do not subsequently outperform, suggesting that gambling does not lead to better returns. Furthermore, investors with *perceived* information advantage do *not* deliver better performance in their trading, suggesting that they are *overconfident* about their own information.

Third, for several trading motives, their coefficients turn from large and significant in the baseline to small and insignificant in the horse race. For instance, we have constructed two measures of sensation seeking, one for novelty seeking and the other for volatility seeking. While both measures exhibit positive and significant explanatory power in univariate regressions, their explanatory power is largely subsumed by other trading motives in the horse race. In comparison, the explanatory power for both gambling preference and perceived information advantage is robust across various specifications. This apples-to-apples comparison among a large set of behavioral biases allows us to narrow down to a few that are the most important.

Fourth, in both the baseline regressions and the horse race, we report a number of "null" results. Contrary to popular accounts, low financial literacy, social interaction, and neglect of trading costs do not appear to contribute to more trading. Perhaps the most consistent, yet surprising set of results concerns neglect of trading costs. While we have constructed three different measures, none of them explain turnover. Furthermore, in a randomized experiment, we give half of the respondents a "nudge" by having them read a message with pictures illustrating how excessive trading hurts their investment performance due to transaction costs. The treatment group, however, does not exhibit any difference in turnover after the "nudge," leading to a further questioning of the role of neglect of trading costs in driving excessive trading.

Our analysis above highlights how surveys can help consolidate the large set of behavioral explanations for excessive trading. However, suppose that, through some magical way, we were able to gather measures of the same list of trading motives using transaction data. How should we choose between the survey-based measures and transaction-based measures? We address this question in the context of gambling preference. Following the approach used by Kumar (2009), we measure an investor's gambling behavior in transaction data as the propensity to buy lottery-like stocks. Compared to the survey-based gambling preference, the transaction-based gambling behavior quadruples in its explanatory power for turnover. However, this greater explanatory power comes at a cost: when regressing it on other survey-based trading motives, it is not only explained by gambling preference but also correlated with a number of other trading motives.

This contrast nicely highlights the pros and cons of these two approaches. On the one hand, when carefully designed, surveys can directly target a specific trading motive without being confounded by other trading motives. However, as discussed by Bertrand and Mullainathan (2001), survey responses are subject to measurement noise at the individual level and are thus less powerful. On the other hand, although transaction-based measures are less subject to measurement noise, they may simultaneously capture multiple trading motives and are less reliable in isolating a single economic mechanism.

As reviewed by Barber and Odean (2013), there is extensive literature that analyzes the excessive trading puzzle from both the theoretical and empirical sides. Our paper differs from these prior studies in its scope and approach. While most of the existing papers focus on one or two trading motives, we simultaneously examine many mechanisms by directly surveying investors. This sets up the first way to consolidate the behavioral bias "zoo." Moreover, we connect the two prevailing approaches of studying decision making by integrating *subjective* survey responses with *objective* transaction data. Our integrated approach offers a more powerful tool to consolidate the large number of behavioral mechanisms offered by the literature.

Several studies, for example, Dorn and Huberman (2005), Glaser and Weber (2007), and Dorn and Sengmueller (2009), have also combined survey data with administrative data, albeit in smaller scope, to study the excessive trading puzzle.[2] Each of these studies elicits responses about one or

[2] Specifically, Dorn and Huberman (2005) focus on risk aversion and perceived financial knowledge, Glaser and Weber (2007) examine two forms of overconfidence, overplacement and miscalibration, and Dorn and Sengmueller (2009) study sensation seeking.

two trading motives and then examines their explanatory power for the respondents' trading or portfolio choices. In the absence of a horse race among different mechanisms, significant effects associated with survey responses to one mechanism may be a reflection of other mechanisms, as in the case of sensation seeking in our analysis. Furthermore, by systematically comparing survey responses and transaction data, our analysis is able to demonstrate that, while survey responses may be noisy at the individual level, they are consistent with actual trading behavior at the aggregate level. In this regard, our paper shares a theme similar to that of Giglio et al. (2019), which studies the relationship between portfolio decisions and return expectations by combining survey expectations with mutual fund holdings data from Vanguard, and that of Epper et al. (2020), which uses a survey approach to measure individuals' time discount rate and examine its relationship with their wealth accumulation over time. However, our paper is different in all other dimensions, such as research questions, survey designs, and transaction data.

The rest of the paper is organized as follows. In Section 1, we explain the survey design and report some stylized facts about Chinese investors from the survey. In Section 2, we validate survey responses using actual trading data and compare survey-based trading motives in a horse race. In Section 3, we provide additional evidence on three selected trading motives. In Section 4, we directly compare survey-based and transaction-based measures. We conclude in Section 5. We also report detailed information about the survey and additional analysis in an Online Appendix.

1. The Survey

In this section, we first elaborate on our survey design and then explain the procedure for survey distribution and data collection. Finally, we summarize some basic facts about the trading motives of Chinese investors based on the survey.

1.1. Survey Design

We designed the survey to test and differentiate a large set of trading motives developed by the literature. Table 1 provides a summary of all the trading motives we consider. A trading motive may take several forms. For instance, overconfidence comes in at least two forms: overplacement, that is, people have overly rosy views of their abilities relative to others, and miscalibration of uncertainty, that is, people are too confident in the accuracy of their beliefs. The survey included at least one question for each form of overconfidence, as detailed in the Online Appendix.

We do not take any prior stand on the relationships across these trading motives, some of which share overlapping theoretical underpinnings, while others may even build on opposite premises. Instead, our research strategy is to use survey questions to directly seek the perspectives of a pool of investors about each of the motives and then compare the explanatory power of these survey responses with their actual trading behaviors. Some respondents may agree with a particular trading motive, but we can determine that this motive is a relevant driver of observed trading only if these respondents also trade more than others. Furthermore, by putting these motives in a horse race, we can let the data determine whether one motive may be subsumed by others. It is possible that when we put several trading motives in the same regression, one of the motives may become insignificant even though it is significant by itself.

We note two common limitations of surveys. First, survey responses are subjective: they capture how people consciously perceive themselves to be making investment decisions. In the language of Adam Smith, respondents are effectively asked to act as the "impartial spectators" to evaluate the reasons and drivers behind their own decisions (Grampp 1948). A common criticism of subjective surveys in economic analysis is the so-called "as if" critique: respondents may not consciously perceive a factor to be important, but they still behave as if it were (Friedman 1953). However, as argued by Choi and Robertson (2020), subjective perceptions are still useful for many reasons: they shed light on the true decision process, they help differentiate competing theories, and they have predictive power for implications of debiasing mechanisms on individuals' future behaviors. It is also inherently interesting to know about people's subjective reasoning. We add that subjective perceptions are also relevant for nudge interventions: if a nudge is targeting a bias that people are not even aware of, it is unlikely that the intervention would successfully produce the desired outcome (DellaVigna and Linos 2020).

The second limitation of surveys is that survey responses are noisy. As explained by Bertrand and Mullainathan (2001), measurement error can arise in the form of white noise, but it could also be attributed to other factors such as wording, scaling (of the answer options), mental effort, social desirability, the lack of opinions, and cognitive dissonance. If measurement errors are white noise, they would create an attenuation bias and result in underdetection of relevant factors. Then, the factors we find relevant would be even more important in practice. These concerns lead Bertrand

and Mullainathan (2001) to conclude that subjective survey responses are useful as explanatory variables for “explaining differences in behavior across individuals,” which is precisely the approach we take in this paper. However, we go beyond simply following their recommendation: in the initial design, we actively addressed the possible bias sources as best we could, as described below.

At a general level, we faced a significant tradeoff between “being rigorous” and “being intuitive” in the design of survey questions. To be fully rigorous in investigating trading motives, the corresponding survey questions needed to comprehensively capture all their aspects. For instance, to fully grasp realization utility requires calibrating a utility function that captures not only different attitudes between gains and losses but also the shape of the utility function in the gain/loss region. Such a design would make some of the questions exceedingly long and unavoidably include academic jargon, which is difficult for respondents to comprehend and could consequently reduce their incentive to complete the survey. Moreover, the psychology literature documents an attribute substitution bias, whereby participants may not respond to complicated questions but rather answer a related question that is easier to respond to (Kahneman and Frederick, 2002). In light of these concerns, we used the “being intuitive” design to make the phrasing as intuitive as possible to laypeople. This prevented the survey length from exploding and ensured that respondents could immediately understand the questions and were willing and able to provide truthful and intuitive responses.

Because biases could arise due to poor phrasing, we adopted a jargon-free protocol. We phrased the questions as accurately as possible when describing the underlying concept while ensuring that they remain comprehensible to the average respondent. To confirm that respondents could immediately understand each question, we ran a series of pilot tests among the general population on a Chinese version of the Mechanical Turk and solicited their feedback on the survey design. The overwhelming majority of respondents found the questions easy to understand. Another concern, particularly relevant to eliciting “biases,” is that respondents may want to look good to others and avoid admitting doing anything “stupid” or “wrong.” This concern arises naturally in interview-based surveys, where the respondents directly interact with the interviewer. As we explain later, because we conducted our survey online, respondents had less of a need to appear “socially desirable.” Moreover, we carefully phrased the questions to be objective and avoided making any inference about a certain behavior being right or wrong. For instance, for questions related to overconfidence, instead of asking respondents “How overconfident do you think you are?” we asked them to self-assess their investment performance and compare it to their actual performance.

Biases could also arise due to the scaling and presentation of the answer options. We designed all questions to be multiple choice so that respondents did not have to fill in an answer themselves. The qualitative questions fell into two types. The first type—“agreement”—asked respondents whether they agree or disagree with a statement that describes a particular trading motive. Answer options included: “strongly agree,” “agree,” “neutral,” “disagree,” “strongly disagree,” “do not know,” and “decline to answer.” The second type—“frequency”— asked respondents how often they consider a particular motive when they trade. Answer options included: “always,” “often,” “sometimes,” “rarely,” “never,” “do not know,” and “decline to answer.” The inclusion of two options, “do not know” and “decline to answer,” further reduced the biases created by a respondent’s lack of opinions. We also sought quantitative answers for certain trading motives (e.g., estimates of transaction fees to measure neglect of trading costs). In such cases, we provided several options, each covering a specific value range. The standardization of answer options ensured that the bias resulting from the design of answer options is small and consistent across all the questions.

Post-survey, we design our empirical strategy with the aforementioned measurement issues in mind. First, we validate survey responses with actual trading behavior. The strong consistency we find between survey responses and transaction data confirms that the survey questions were well designed and captured the targeted mechanisms. Second, we encode all survey-based trading motives into dummy variables, which minimizes the variation of measurement errors across the survey-based trading motives and facilitates an apples-to-apples comparison.

Note that although we ask respondents to assess whether a trading motive matters to their trading or how often they consider a certain motive, we do *not* ask them to evaluate the importance of any motive to their frequency of trading—our subject of interest— relative to other motives. This is different from the approach taken by Choi and Robertson (2019), which asks correspondents themselves to evaluate and compare the relevance of different theories in describing their decision making. In contrast, we retain this task for ourselves, which we perform by regressing individual-level turnover on a set of trading motives indicated by survey responses.

This analysis is made possible by our capacity to trace a respondent's survey responses to her trading record.

The final survey contained four main parts. The first part contained eight questions measuring financial literacy. These questions included the classic "big three" questions, as well as several other widely used questions to measure financial literacy (Lusardi and Mitchell (2007, 2011)). At the end of this section, we also asked respondents to self-assess how many questions they answered correctly. This allows us to construct a measure for overconfidence based on financial literacy. The second part represented the core of the survey, where we asked respondents to answer a series of questions related to various trading motives. We postpone a more detailed discussion about this part to Section 1.3. The third part asked their basic demographic characteristics, including name, gender, date of birth, province, city, education, income, net worth, phone number, brokerage firm, and broker branch. While many of these variables serve as control variables in subsequent analysis, they also provide crucial identifying information for us to be able to locate each correspondent in the transaction database. Finally, for a randomly selected group of respondents (the treatment group), we also included a fourth "nudge" section. We explain the "nudge" below and discuss the results in more detail in Section 3.3.

1.2. Data

We administered the survey through the Investor Education Center of the Shenzhen Stock Exchange (SZSE). As part of its regular operations, the Investor Education Center annually surveys domestic retail investors to assess their financial literacy and trading motives. In 2018, we began to collaborate with the center to redesign the survey with the aforementioned research question in mind. Our target sample was 10,000 investors, which size provides sufficient statistical power but was feasible to implement. To ensure that the survey sample was nationally representative, we randomized across branch offices of China's ten largest brokers. Specifically, we selected 500 branch offices across 29 provinces (and regions) and required each branch office to collect at least 20 valid responses. The number of branch offices allocated to each province (region) was proportional to the total trading volume from that province (region) in 2017.

The survey took place in September 2018, and respondents were given two weeks to complete the survey.[3] A valid response had to be completed within 30 minutes. Respondents could open the survey using their personal computers or their smartphones.[4] We collected an initial sample of 12,856 respondents. Table 2 reports the distribution of respondents across brokers and provinces. By design, respondents were evenly distributed across the ten brokers, with only slight variation. In terms of geographic variation, areas that are more financially developed (e.g., Guangdong, Zhejiang, Jiangsu, and Shanghai) are more represented in our sample.

Table 3 reports a more detailed summary of the sample's demographic characteristics. Overall, the sample is balanced in gender and highly educated: more than half of the respondents had a college or higher degree. Respondents were primarily middle-aged: almost half of the sample were aged 30–50. They were also quite wealthy: the median annual income was around 200,000 RMB and the median household net worth was around 500,000 RMB, both of which far exceed the national median. Overall, our sample represents a relatively well-educated, wealthy set of retail investors, which means that any results we find may not be simply interpreted as an average effect. Instead, to the extent that rich and sophisticated investors are less affected by behavioral biases in their portfolio decision making, our results may serve as a lower bound.

Finally, while we feel confident that the use of monetary incentives and the brand names of our respective institutions should on average invite high-quality responses, we nevertheless cannot avoid having a few respondents who quickly clicked through the survey without spending much time on the questions, especially given the survey's large scale. We eliminate these responses by examining the total amount of time spent on the survey. Figure 1 plots the distribution: it took a median investor about eight minutes to complete the survey, and 95% of respondents finished within 20 minutes. However, we find that respondents who spent less than three minutes on the survey experienced a sharp drop in their financial literacy score, suggesting that they may have

[3] The distribution of the survey proceeded as follows. The SZSE Center first distributed the link to the survey to each broker's headquarters. The headquarters then distributed it to the pre-selected branches, where local client managers redistributed the survey to their clients (investors), likely via phone calls or WeChat messages. Once an investor had completed the survey, the client manager recorded her name, phone number, and the name of the branch. This information was then sent back to us for verification purposes.

[4] To boost the response rate, we included the logos of both the SZSE and the Shenzhen Finance Institute on the front page of the survey. We also explicitly included a confidentiality agreement to make respondents feel more secure about their answers. Finally, we used monetary rewards as incentives. Specifically, among those who completed the survey, 20 would be randomly selected to receive a gift card worth 500 RMB (around 80 USD) and 1000 would receive a gift card worth 50 RMB (around 8 USD).

shirked during the survey. In subsequent analysis, we dropped these observations, which reduced our sample size to 11,268.

1.3. Survey Results

Financial literacy

Table 4 reports the summary statistics for the eight questions on financial literacy. In addition to the classic "big three" questions on interest rates, inflation, and diversification, as in Lusardi and Mitchell (2014), we also include five other questions that capture additional dimensions of financial (or investment) literacy.[5] Panel A shows that, out of all eight questions, seven have a correct rate above 75%. The only exception is the question about the relationship between interest rates and bond prices. Panel B shows that more than 80% of respondents correctly answered at least six questions. In fact, one-third of them were correct on all eight questions. Panel B shows the distribution of self-assessed scores, which is similar to that of the actual scores. Overall, investors in our sample display a high level of financial literacy.[6]

Overconfidence

Overconfidence is an important concept in behavioral finance and has been adopted by various models to explain a wide range of anomalies in financial markets, including excessive trading, use of leverage, price momentum and reversals, and asset bubbles, e.g., Kyle and Wang (1997), Daniel, Hirshleifer and Subramanyam (1998), Odean (1998), Gervais and Odean (2001), Scheinkman and Xiong (2003), and Barber et al. (2019). The literature also suggests that overconfidence may present in several closely related, albeit distinct, forms: overplacement of ability, miscalibration of uncertainty, and overprecision of information. We designed questions to capture each of these forms.

---

[5] These questions are related to the concept of risks and volatility (Question 4), the definitions of shareholders, the price-to-earnings ratio, and mutual funds (Question 5, 7, and 8), and the relationship between interest rates and bond prices (Question 6).

[6] Lusardi and Mitchell (2014) show that among eight countries including Germany, the Netherlands, and the United States, the fraction of respondents who correctly answer all "big three" questions ranges from 3% (Russia) to 57% (Germany). In contrast, 70.4% of investors correctly answer all "big three" questions in our survey. One possible reason for this difference is that their surveys typically draw respondents from the general population, whereas ours draws from investors already participating in the stock market.

Overplacement of one's own ability is perhaps the most direct form of overconfidence. We construct two measures of this form, one by the difference between self-assessed and actual performance in 2017 and the other by the difference between self-assessed and actual literacy scores.[7] In Table 5, Panel A reports the summary statistics for both measures. In constructing overplacement of performance*,* self-assessed performance is one's self-reported rank of her investment performance among all investors in 2017; actual performance is measured by the actual rank in the population. At this point, we have not yet merged survey responses with transaction data, so Panel A only reports the distribution of self-assessed performance and suggests that the respondents are rather optimistic about their performance: almost two-thirds of them believe that their performance is better than average, while only a quarter believe that their performance is below average. Panel A also reports the second measure, overplacement of literacy. Overall, respondents do *not* overestimate their level of financial literacy. This is perhaps not that surprising given the sample's overall high level of financial literacy.

Overconfidence may also show up as miscalibration of uncertainty, as suggested by Alpert and Raiffa (1982).[8] We include a similar measure of miscalibration by the difference between the estimates of upside returns and downside returns. This measure is based on two questions in which we ask respondents to estimate how much the stock market will go up (down) with 10% probability within the next year; the difference between these two estimates results in an 80% confidence interval. As reported by Panel A of Table 5, while a rational benchmark (based on historical market volatility) suggests that the upside and downside returns should exhibit a difference of 76%, the majority of the respondents report a much narrower range.

Overconfidence may also show up as overprecision about one's own information. We will describe this measure later when we discuss information-related questions.

Extrapolation

The behavioral finance literature has also emphasized the tendency of investors to extrapolate past returns as a key driver of stock return predictability, e.g., Barberis, Shleifer and Vishny (1998),

[7] A similar measure is also used by Dorn and Huberman (2005) and Barber et al. (2019) to measure perceived financial knowledge.

[8] Ben-David, Graham and Harvey (2013) show that 80% confidence intervals provided by firm executives for the subsequent year's stock market return only cover 36% of the realizations, and they use the surveyed confidence interval to measure the executives' overconfidence.

Barberis et al. (2015), and Jin and Sui (2019), and excessive trading, e.g., Hong and Stein (1999) and Barberis et al. (2018). In Table 5, Panel B reports the summary statistics for two questions concerning whether investors form expectations about future returns based on past returns. These two questions elicit investors' extrapolative beliefs in two scenarios. In the first scenario, a stock's price keeps going up, and in the second scenario, a stock's price keeps going down. Respondents are then asked whether they believe the stock's price will rise or fall in the future. In both scenarios, more respondents believe in price continuation than reversal, suggesting that Chinese investors on average exhibit extrapolative beliefs.

Neglect of trading costs

Barber and Odean (2000) and Barber et al. (2009) show that trading causes retail investors in the United States and Taiwan to underperform relative to the overall market, and more than 60% of their underperformance is directly due to commissions and transaction taxes. While overconfidence and other behavioral biases may cause investors to trade despite trading costs, these findings also suggest the possibility that those investors who trade a lot may have neglected the various fees and taxes associated with trading. As it is common for financial regulators across the world to use Tobin taxes to curb speculative trading, the possible neglect of trading costs by investors undermines the effectiveness of such financial policies.

An investor's neglect of trading costs stems from at least two possible sources. The first source is simply underestimation: investors systematically believe the fee is lower than it actually is due to a lack of financial sophistication. The second source is a lack of salience (Bordalo et al. 2012): even if investors do have full knowledge about trading costs, it still matters very little to their trading because the amount associated with each transaction is negligible.[9]

To capture these two forms of neglect of trading costs, we constructed three different measures. Panel C of Table 5 reports the summary statistics. First, we directly asked investors to estimate the total transaction costs associated with a round-trip buy and sell at 10,000 RMB. The results show that respondents significantly underestimated trading costs: while on average, such a round-trip transaction should incur a fee of 15 to 26 RMB, depending on the fee rate charged by the particular

[9] Several papers show that manipulating the salience of a stock's purchase price affects the level of the disposition effect (e.g. Frydman and Rangel 2014; Birru 2015; Frydman and Wang 2019). Other papers find that manipulating the salience of taxes affects consumer responsiveness to taxes (e.g., Chetty, Looney and Kroft 2009; Taubinsky and Rees-Jones 2017).

broker, almost 70% of the respondents reported an estimate below the lower bound. The second question asked how often an investor considers transaction costs when trading stocks. Similarly, more than half of the respondents said that they never or rarely do so. The third question targeted the implicit cost of the bid-ask spread by asking whether the respondent agrees that bid-ask spread is a form of trading cost. Around 60% of respondents agreed while 23% disagreed. Overall, there is strong evidence that retail investors in China underestimate or neglect trading costs.

If neglect of trading costs is due to (a lack of) salience, then presenting transaction costs in a more salient manner or more frequently reminding investors of the costs may lead them to trade less. To test this hypothesis, we gave a random half of respondents a "nudge" and compare their turnover with that of other investors before and after the survey. For the treated group, we increased the salience of trading costs by presenting them in annualized terms and reminding the investors about the negative impact of excessive trading to their overall returns. We discuss these results later in Section 3.3.

Gambling preference

Barberis and Huang (2008) show that the cumulative prospect theory of Tversky and Kahneman (1992) can generate a preference for gambling stocks, meaning stocks with positively skewed returns. In particular, this gambling preference is driven by prospect theory's probability weighting component, through which investors over-weight the likelihood of tail events.[10] To the extent that gambling stocks change over time due to fluctuations of volatility and tail distribution, gambling preference may also contribute to excessive trading by leading some investors to chase gambling stocks and thus trade with other investors (Barber and Odean 2000).

In Table 6, Panel A shows the responses on the two questions about gambling preference. The first question asked whether the respondent aims to select a few blockbuster stocks with the intention of getting rich quickly. The second question asked whether the respondent consciously perceives trading stocks as buying lotteries in that they are willing to exchange small losses for the small probability of a big gain. Overall, for each question, about one-third of the respondents agreed or strongly agreed with the statements. In what follows, we differentiate these two questions

[10] Kumar (2009) and Boyer, Mitton and Vorkink (2010) provide empirical evidence that supports the presence of such gambling preference.

by labeling the first one as representing "blockbusters" and the second one as representing "lotteries."

In phrasing these two questions, we had the following design in mind: the "blockbusters" question focuses on the salient upside and deliberately tones down the fact that "blockbusters" are rare. Therefore, investors who agree with this statement are the ones drawn to the large upside without necessarily assessing its small probability. In the language of prospect theory, these investors tend to over-weight small probabilities. In contrast, the "lotteries" question contains a direct description of lotteries by explicitly stating that large payoffs rarely happen. Therefore, the two questions not only help identify the gamblers among the respondents, but also help differentiate their assessments of the tail probabilities. As we will show, the "blockbusters" question has substantially stronger explanatory power for investor trading.[11]

Realization utility

Shefrin and Statman (1985), Odean (1999), Grinblatt and Keloharju (2001), and Grinblatt and Han (2005) argue that trading can arise as a result of the widely observed disposition effect. To provide a robust explanation to the disposition effect, Barberis and Xiong (2009, 2012) and Ingersoll and Jin (2013) propose a theory of realization utility, which posits that trading causes investors to realize enjoyment from selling winning stocks and pains from liquidating losing stocks.[12]

In Table 6, Panel B reports the summary statistics for the two questions on realization utility. Similar to the questions on extrapolative beliefs, these two questions ask respondents to make investment decisions under two hypothetical scenarios. In the first scenario, the respondent is given a stock whose price has gone up since purchase and is then asked which of the two actions would make her happier: selling the stock or holding on to it. In the second scenario, the respondent instead faces a stock whose price has gone down since purchase and is asked which action would be more painful. According to realization utility, selling winners is more pleasing than holding winners while selling losers is more painful than holding losers. Survey responses for the two questions are mixed. In the first question, consistent with realization utility, more respondents say

[11] An alternative explanation for the difference between these two questions is that the "blockbusters" question helps to identify the "impatient" gamblers. As the literature does not offer any link between trading volume and the discount rate, we attribute the question's better explanatory power to incorrect probability assessment rather than to impatience.

[12] Frydman et al. (2014) provide neural evidence to support realization utility in financial decision making.

selling winners makes them happier. In the second question, however, more respondents report that holding on to losers is more painful than selling losers. In what follows, we differentiate these two questions by labeling the first question as realization utility for winners and the second question as realization utility for losers.

Sensation seeking

Grinblatt and Keloharju (2009) argue that sensation seeking, a measurable psychological trait linked to gambling, risky driving, drug abuse, and a host of other behaviors, is an important motivation for trading. Dorn and Sengmueller (2009) provide supportive evidence that sensation seeking drives the retail investors' trading. Brown et al. (2018) further argue that sensation seeking may even affect the trading of hedge fund managers. We have designed two questions to capture two distinct dimensions of sensation seeking: novelty seeking, which says that people derive utility from doing something new, and volatility seeking, which says that people derive utility from doing something risky. In Table 6, Panel C reports the summary statistics for these two questions. Overall, answers to these two questions exhibit a similar distribution, but the respondents in general do not exhibit a strong tendency for sensation seeking.

Information

Economists have long argued that access to private information is a key reason for investors to trade in financial markets. However, the classic no-trade theorem posits that when all investors are rational and share the same prior beliefs, asymmetric information cannot cause them to trade due to the concern of adverse selection (Milgrom and Stokey (1982)). Instead, theories of financial market trading with asymmetric information, e.g., Grossman and Stiglitz (1980) and Kyle (1985), typically involve the presence of noise traders, who may trade at losses, so that rational traders may trade despite the potential concern of adverse selection.

Are retail investors in China rational investors with a genuine information advantage or noise traders who believe they hold superior information even though they do not? We included two questions in the survey to elicit a respondent's perception of their information. The first question measures one's belief in having an information advantage by asking how often she believes she knows stocks better than other investors. A positive response to this question may be associated with a genuine information advantage, but it could also reflect a misperceived information advantage due to overconfidence. This latter possibility potentially reflects a tendency to exaggerate one's own information but not the information of others. Various theoretical models have used this tendency, e.g., Kyle and Wang (1997), Odean (1998), and Scheinkman and Xiong (2003), to specify investor overconfidence, which is the third form of overconfidence that we mentioned earlier. In our empirical analysis, we can differentiate a genuine information advantage from a perceived information advantage by examining whether the respondent actually delivers better trading performance.

The second question measures one's fear about potential adverse selection concerns by asking how often she worries that others know stocks better than herself. This question potentially measures dismissiveness about others' information, a form of investor bias that offers distinct implications from overconfidence for equilibrium prices and trading volume (Eyser, Rabin and Vayanos 2019). Panel A of Table 7 shows that about 18% of the respondents say that they often or always believe they have an information advantage, while 47% of the respondents never or rarely believe that they face an information disadvantage.

Social interaction

Shiller (1984) argues that investing in speculative assets is a social activity because investors enjoy discussing investments and gossiping about others' investment successes or failures. As a result, social influences would affect investors' trading behavior.[13] We designed two questions to capture social interactions, one about the influence from family, friends, and other acquaintances, and the other about the influence from investment advisors. Panel B of Table 7 shows that while around 14% of the respondents say that they are often or always influenced by their family, friends, or other acquaintances, only 8% say their investment advisors often or always have an influence on their trading.

Other trading motives

In Table 7, Panel C reports the responses on the two questions related to liquidity needs and rebalancing motives. Overall, only about 11% of the respondents say portfolio rebalancing often

[13] Hong, Kubik and Stein (2004) provide evidence that stock market participation is influenced by social interaction. Han, Hirshleifer and Walden (2019) develop a model to show that social interaction exacerbates excessive trading among investors.

or always affects their trading, whereas about 17% say liquidity needs often or always affect their trading. Consistent with prior literature, retail investors do not appear to be considering these rational trading motives in their day-to-day trading activities.

Panel D of Table 7 reports three standard questions for measuring risk aversion. Following Lusardi and Mitchell (2011), we elicit investors' risk attitude by asking whether they would be willing to give up their current stable jobs for other jobs with higher expected income but also higher uncertainty in three hypothetical scenarios. While about 34% of the investors were unwilling to take the job with the smallest risk, 26% of the investors were willing to take the riskiest job.

Comparison with U.S. investors

While our study primarily focuses on Chinese retail investors, it is of general interest to know how U.S. retail investors—who are often believed to be more sophisticated than their Chinese counterparts—would respond to our survey. We translated the original survey into English with slight modifications (tailored to American investors) and ran the survey on Mechanical Turk among a small sample of 400 U.S. retail investors. On the one hand, we find that U.S. investors care more about trading costs, rely more on investment advisors, and are more alert to being at an information disadvantage. These differences may be attributed to the institutional environment of the U.S. stock market: higher transaction fees charged by brokers, the popularity of investment advisors, and a highly institutionalized investor base. On the other hand, contrary to conventional wisdom, U.S. retail investors exhibit stronger biases on several fronts: they are more subject to realization utility, display a stronger preference for gambling, and are more prone to sensation seeking. A more detailed discussion about these differences is included in the Online Appendix.

2. A Horse Race Based on Survey Responses

In this section, we use survey responses to differentiate various explanations for the excessive trading puzzle. We start by merging the respondents' survey responses with their transaction data in Section 2.1. In Section 2.2, we address some of the common concerns associated with surveys by showing that survey responses are consistent with actual trading behavior. In Section 2.3, we examine all trading motives separately. Finally, in Section 2.4, we run a horse race among all survey-based trading motives.

2.1. Merging Surveys with Transactions

In the third part of our survey, we asked respondents to provide information on various demographic variables, including name, date of birth, broker name, and branch name. This allows us to uniquely identify a substantial fraction of the respondents in the transaction database of the Shenzhen Stock Exchange. Specifically, out of the 11,268 respondents that remain in our sample, we are able to uniquely identify 6,013 investors.[14] Our transaction data cover January 2018 through June 2019; our survey date of September 2018 is nicely in the middle of this time frame. We further require an investor to have held at least one stock in the Shenzhen Stock Exchange during the two-year window before the survey.[15] This further reduces the sample size to 4,671, which is our main sample.

Table 8 compares the average characteristics between the main sample and the population of Chinese investors, where the population's characteristics are obtained using the centralized database at the Shenzhen Stock Exchange. While over 70% of the investor population is male, the gender ratio is much more balanced in our main sample, which has 54% male investors. Consistent with our previous discussion, our main sample covers slightly younger, more-educated investors. In terms of trading, investors in our main sample tend to have larger accounts, slightly lower turnover rates, and better investment performance.

To make different trading motives comparable, we encode all the measures of trading motives into dummy variables. A detailed description of the construction of these dummy variables can be found in the Online Appendix. In a nutshell, for the agreement type of questions, we code "strongly agree" and "agree" as 1 and other answers as 0; for the frequency type of questions, we code "always" and "often" as 1 and other answers as 0; for quantitative questions, we typically use zero as the cut-off value.[16] Table 9 reports the summary statistics of these dummy variables and their

---

[14] In the Online Appendix, we report the distribution of this subset of correspondents across various demographic variables and show that it is almost identical to that of the original sample.

[15] An investor may be invited to our survey without any stockholding in the Shenzhen Stock Exchange due to various reasons: she could hold mutual funds or ETFs, or she could hold stocks listed in the Shanghai Stock Exchange.

[16] The only exception is when we code the question of dismissiveness, where we code "never" or "rarely" as 1 and others as 0.

pairwise correlations. Note that for the multiple questions targeting the same trading motive, their pairwise correlation, highlighted in bold in Table 9, is generally high, which suggests that their responses are internally consistent.

2.2. Validating Survey Responses

There are several widely held concerns about the use of survey responses in testing economic hypotheses. First, respondents may not take the survey seriously and may not truthfully report what they really think or believe. Second, even if their responses are truthful, they may not act in a way that is consistent with their responses. Indeed, because most existing papers are limited to the use of either survey data or transaction data, the literature is still missing a systematic test of the external validity of survey responses of investors.[17]

Ideally, we would like to validate responses to all the questions in the survey, but this is neither efficient nor plausible. For instance, although the survey has several questions regarding sources of information and the influence of social interaction, it is difficult, if not impossible, to infer these aspects from transaction data without any additional administrative data and/or making strong assumptions. Given these limitations, we validate survey responses only for questions with a natural empirical counterpart that can be directly constructed from the transaction data. This set of questions concerns extrapolation, gambling preference, risk aversion, and return expectation. In addition to having straightforward implications about trading behavior, these questions also span a wide range of trading motives—belief formation, preferences, and return expectations. For brevity, we report the results related to gambling preference and extrapolative beliefs in the main text, and include other results and more details in the Online Appendix.

Gambling preference

We start by measuring gambling behavior from transaction data. Gambling preference motivates investors to buy assets with positively skewed returns. While it seems straightforward

---

[17] Several earlier examples of such validation exercises are worth noting. Using survey and administrative data from Denmark and Sweden, respectively, Koijen at al. (2015) and Kreiner et al. (2015) show that, while survey-based consumption is noisy at the individual level, it is consistent with actual consumption measured from administrative data. More recently, Giglio et al. (2019) examine the relationship between survey expectations and mutual fund holdings and find that survey expectations are consistent with respondents' mutual fund holdings. Compared to these earlier papers that study consumption and expectation, our main interest is to validate whether survey-based trading motives reflect investors' actual trading behavior.

to measure gambling behavior based on return skewness, the literature, for example, Kumar (2009), argues that return skewness is difficult to compute and is not a metric sufficiently intuitive to investors. Instead, salient stock characteristics such as realizations of extreme returns would attract investors with a gambling preference. This argument is particularly compelling as it connects well with our earlier discussion about gambling preference originating from investors' overweighting of tail outcomes (Barberis and Huang (2008)). Motivated by this argument, we take advantage of a unique regulation in the Chinese stock market: the daily price limits rule. This rule states that daily stock returns of individual stocks cannot exceed 10%. We use the total count of up-limit hits (i.e., the number of days with prices hitting the up-limit) in a preceding period to proxy for a stock's positive return skewness. As hitting the daily up-limit puts a stock in the headlines of the stock exchange, this event is highly salient and attracts attention from investors. Thus, we measure an investor's gambling behavior by the volume-weighted count of up-limit hits based on all the stocks she bought over either a month or a quarter.

Table 10 reports the results when regressing transaction-based gambling behavior on survey-based gambling preference. Panel A uses the total count of up-limit hits over the preceding one-month horizon, while Panel B uses one quarter as the horizon. Recall that we included two survey questions regarding gambling preference, one about the desire to pick blockbusters to get rich and the other about a conscious perception of stocks being lottery-like. Indeed, responses to the first question significantly explain gambling behavior in transaction data with a positive sign. On average, the stocks they purchase have a larger count of up-limit hits by around 0.1 (0.2) times in the preceding month (quarter), and this relationship holds in both the pre-survey and post-survey periods. Interestingly, responses to the second question do not explain gambling behavior. We document a similar pattern about their explanatory power on turnover in Sections 2.3 and 2.4.

Extrapolation

Next, we validate that survey-based measures of extrapolative beliefs are consistent with actual extrapolative behavior. Similar to before, we measure extrapolative behavior as the volume-weighted past return among all the stocks bought by an investor. Table 11 reports the results when regressing transaction-based extrapolative behavior on survey-based extrapolative beliefs, where, in measuring extrapolative behavior, Panel A uses past one-month return and Panel B uses past one-quarter return. Indeed, investors who report having extrapolative beliefs exhibit stronger extrapolative behavior: on average, the stocks they purchase experience 1% higher returns in the preceding month and more than 2% higher returns in the preceding quarter, and this holds in both pre-survey and post-survey samples. The two measures of extrapolation have equally strong explanatory power for extrapolative behavior.

Risk aversion and survey expectations

We perform two additional exercises to validate survey-based measures of risk aversion and return expectations, using a method similar to before. First, we find that, consistent with Dorn and Huberman (2005), survey-based measures of risk aversion are negatively associated with holding more volatile stocks. Second, we also find that, consistent with Giglio et al. (2019), survey-based expectations about future stock market returns are positively associated with an increase in stock holdings, but the magnitude, as noted by Giglio et al. (2019), is relatively small.

Finally, we note that throughout the validation exercises, although the coefficient between the survey response and trading behavior is highly significant, the *R*-squared is generally small. For instance, in Table 10, across all specifications, the *t*-statistic for gambling preference (blockbusters) remains around 4, but the *R*-squared is consistently below 1%. This suggests that although survey responses are in aggregate consistent with behavior, much of the variation in trading behavior is left unexplained. This could be due to measurement errors or white noise in survey responses, or other factors simultaneously driving trading behavior. We will discuss this important issue further in Section 4.

2.3. Baseline Results on Turnover

After validating survey responses, we proceed to examine the relationship between survey-based trading motives and turnover. We primarily focus on using survey responses to explain post-survey turnover.[18] Table 12 reports the summary statistics of their monthly turnover and portfolio returns in the post-survey sample from October 2018 through June 2019, which is the nine-month

[18] If we measure turnover at the time of or before the survey, then the exercise is subject to the concern that some common shocks may have affected both survey responses and trading behavior. For instance, a positive shock to one's recent return may lead her to report a higher self-assessed performance—resulting in more overplacement of performance—and to trade more.

window after the survey. When needed, however, we also extend the window to cover the nine months before the survey, spanning our full sample from January 2018 through June 2019.

Table 12 shows that excessive trading is pronounced among Chinese retail investors. First, they trade intensively: the median monthly turnover rate in our sample is almost one, suggesting that they fully reshuffle their portfolios almost once every month.[19] Second, their performance is poor: while the monthly return of the Shenzhen Composite Index is about 0.6% from October 2018 through June 2019, the median net return in our sample is only 0.0%. Third, those who trade more perform worse: the correlation between turnover and *raw* returns is −0.07 while the correlation between turnover and *net* returns is −0.16. These negative correlations are statistically significant and confirm the key findings of Odean (1999) and Barber and Odean (2000).

Table 13 presents the baseline results, where in each column we regress turnover on a particular survey-based trading motive. Most regressions are univariate, except for a few instances where we need to control for some additional characteristics.

Columns (1) to (3) report the results on three measures of overconfidence—overplacement of performance, overplacement of literacy, and miscalibration of uncertainty. Out of these three measures of overconfidence, the only one that is significantly and positively related to turnover is overplacement of performance: in column (1), conditional on having the same past performance, investors who self-report having higher performance tend to trade more subsequently. Column (1) also shows that past performance positively predicts future turnover. In column (2), financial literacy positively predicts future turnover. This finding is in sharp contrast to a widely held view that excessive trading may be driven by the lack of financial knowledge. Therefore, improving investors' financial literacy, a policy often advocated in emerging economies such as China, may not be effective in reducing excessive trading. Furthermore, column (2) shows that overplacement of literacy does not predict future turnover. In column (3), miscalibration of uncertainty does not significantly predict future turnover. This set of results is broadly consistent with Glaser and Weber (2007), who find that overplacement predicts more trading, but miscalibration does not.

Columns (4) to (6) report the results on neglect of trading costs. Surprisingly, for all three measures we have constructed, none of them significantly predict future turnover with the predicted sign: in columns (4) and (5), the coefficients are close to zero and insignificant; in

[19] In comparison, the average monthly turnover rate of investors on the Shenzhen Stock Exchange was around 129.5% during the same period.

column (6), investors who do not understand the bid-ask spread as a form of trading cost trade *less*. The result in column (4) is particularly puzzling because the measure is constructed directly using the estimate of fees in a round-trip transaction and should clearly identify those investors who underestimate trading costs.[20] The fact that we cannot find any supporting evidence despite having constructed three measures for neglect of trading costs gives us pause about its role in explaining investor trading. We will return to this issue with more analysis in Section 3.3.

Columns (7) and (8) report the results on extrapolative beliefs. For the two measures of extrapolation of positive and negative returns, we do not find a strong relationship between extrapolative beliefs and turnover. One possibility is that extrapolation generates trading only in a bullish market (Barberis et al. 2018; Liao, Peng and Zhu 2020), but the period we examine is relatively quiet—the market increased by just a few percentage points. Another possibility is that extrapolation alone cannot explain volume and must be combined with some additional forces to generate a trading frenzy (Liao, Peng and Zhu 2020). We leave these issues to future research.

Columns (9) and (10) report the results on gambling preference. We find that, consistent with the conjecture in Barber and Odean (2000) and the implications of Barberis and Huang (2008), investors who are subject to gambling preference trade significantly more. Again, the question about "blockbusters" is much more powerful than the "lotteries" question. This is consistent with the pattern in Table 10, which shows gambling behavior can be explained by answers to the "blockbusters" question but not by answers to the "lotteries" question.

Columns (11) and (12) report the results on realization utility and show asymmetry. The first measure—the one that proxies for taking pleasure in selling winners—positively predicts future turnover, whereas the second measure—the one that proxies for feeling pain when selling losers—does not predict future turnover. This pattern is consistent with the implications of realization utility (Barberis and Xiong 2012), as investors who exhibit realization utility are more willing to let go of stocks once they exceed their purchase price and to hold on to stocks after their prices fall from the purchase prices.

Columns (13) and (14) report the results on sensation seeking. Both the "novelty-seeking" and the "volatility-seeking" measures positively predict future turnover with a large coefficient. These

[20] Transaction fees are standard and almost homogeneous across different brokers. While some variation across brokers still remains, in our construction we use a rather conservative bound to identify those who underestimate trading costs. In addition, we control for differences in fees across brokers with branch fixed effect.

results are consistent with the findings by Grinblatt and Keloharju (2009) and Dorn and Sengmueller (2009) that investors most prone to sensation seeking trade more frequently.

Columns (15) and (16) report the results on perceived information advantage and dismissiveness of others' information. Column (15) suggests that those who believe in having an information advantage tend to trade more, whereas column (16) suggests that those who dismiss others' information do *not* trade more. As we discussed earlier, the first measure captures a particular form of overconfidence as perceived information advantage,[21] as modelled by Kyle and Wang (1997), Odean (1998), and Scheinkman and Xiong (2003), while the second measure captures the dismissiveness modelled by Eyster, Rabin and Vanayos (2019). Thus, these results suggest that perceived information advantage leads to high volume, while dismissiveness of others' information does not.

Finally, columns (17) and (18) concern two measures of social influence, one from family and friends and the other from investment advisors. Interestingly, investors who are more influenced by their family, friends, and investment advisors tend to trade *less*, not more. This pattern does not lend support to the aforementioned literature that argues that social interaction contributes to the spread of investor sentiment and excessive trading. Columns (19) and (20) show that rational trading motives such as portfolio rebalancing needs and liquidity needs can only explain a small part of the variation in turnover across investors.

In sum, Table 13 confirms several of the existing explanations for trading volume, specifically, overplacement of performance, gambling preference, sensation seeking (for both novelty and volatility), realization utility, and perceived information advantage. Table 13 also highlights a number of "null" results that cast doubt on several prominent explanations of excessive trading, specifically, lack of financial literacy, neglect of trading costs, dismissiveness about others' information, and social interaction.

2.4. Horse Race Results on Turnover

While the baseline results confirm several of the previous explanations for trading volume, it remains unclear whether their explanatory power will survive once they are all included in the same regression. Table 14 presents the full regression results. In addition to including all the

[21] Note that this interpretation assumes that those who claim to have an information advantage do not have one in reality. We will verify this interpretation later in Section 3.2.

survey-based trading motives, we also include: 1) basic demographic characteristics such as gender, income, net worth, and education; 2) return expectations to control for differences in optimism and pessimism; and 3) recent performance to control for "mood."[22] Table 14 reveals a number of notable observations.

First, two trading motives that stand out in the horse race: gambling preference ("blockbusters") and overconfidence in the form of perceived information advantage. Both coefficients are quantitatively large and significant at the 1% level. The finding of overconfidence as a key driver of turnover nicely supports the large volume of prior studies in the behavioral finance literature emphasizing the roles of overconfidence. Even more interesting, our finding highlights that a particular form of overconfidence—through perceived information advantage—rather than other forms, such as overplacement of literacy and miscalibration of uncertainty, is most relevant in explaining trading. This form of overconfidence also confirms the specification adopted by Kyle and Wang (1997), Odean (1998), and Scheinkman and Xiong (2003) in modeling investor overconfidence in financial markets.

Our finding of gambling preference as a key driver of investor trading is surprising given that the literature tends to treat gambling preference as an important mechanism for understanding demand for lottery-like stocks but not for excessive trading. Our finding suggests that gambling preference may also lead investors to trade more. A possible mechanism works as follows. As individual stocks fluctuate in their volatility and tail distribution, the set of lottery-like stocks changes over time. Consequently, investors subject to gambling preference chase one lottery-like stock after another, leading to large trading volume.

Note that in Table 9, the correlation coefficients between perceived information advantage and the two measures of gambling preference fall between −0.09 and −0.06. The small correlation suggests that overconfidence and gambling preference are likely two independent traits that contribute to trading volume through two distinct channels. We will present additional evidence to support these trading motives as key drivers of excessive trading in Section 3.

Second, several trading motives that are significant in the baseline regressions become insignificant or only marginally significant in the horse race. They include financial literacy, sensation seeking for novelty, sensation seeking for volatility, social influence, and advisor

[22] We also have a specification that includes branch fixed effects to control for clustering at the branch level. Results are essentially unchanged and reported in the Online Appendix.

influence. The results for the two sensation seeking measures are particularly striking: while both measures are highly significant in univariate regressions, their significance largely disappears after controlling for other factors, suggesting that their explanatory power is subsumed by other factors. This contrast nicely highlights the advantage of our direct comparison of different mechanisms.

Finally, consistent with the finding of Barber and Odean (2001), we also report a significant gender effect: on average, the monthly turnover of male investors is 21% higher than female investors. Barber and Odean (2001) attribute this difference to overconfidence: men trade more because they are more overconfident. Interestingly, the gender effect in Table 14 persists even after controlling for various forms of overconfidence, suggesting the gender effect may go beyond overconfidence. We leave that for future research to explore.

2.5. Robustness and Subsample Analysis

As robustness checks, we report the results from alternative regression specifications in the Online Appendix, including bootstrapped standard errors, adding branch fixed effects as control variables, a larger sample that includes investors that have not traded for more than two years before the survey, and a small sample that only includes investors who are active around the survey. We also consider several alternative measures of turnover, including: an equal-weighted version of turnover as opposed to the value-weighted one we use throughout the paper, and a version of turnover measured in the nine-month window before the survey, as opposed to the nine-month window after the survey. Throughout all these specifications, gambling preference and perceived information advantage remain the most powerful drivers of excessive trading.

We also perform two sets of subsample analysis and report the results in the Online Appendix. In the first one, we split the full sample based on account size and compare the behaviors of small and large investors. Overall, consistent with the notion that small investors are more affected by behavioral biases, we find that the results are slightly stronger among small investors. In the second subsample, we split the full sample based on the fraction of wealth invested in the stock market. In both subsamples, gambling preference and perceived information advantage largely remain the important factors. However, for investors whose wealth is more invested in the stock market, portfolio rebalancing needs become a more pronounced factor to their frequent trading.

To conclude this section, we discuss two limitations of our horse race. First, it is possible that the importance of each mechanism is time-varying, and, without a panel of survey responses, we can only capture a snapshot of their relative importance. For instance, realization utility may contribute to excessive trading more in a market boom than in a market downturn (Barberis and Xiong 2012, Liao, Peng and Zhu 2020). However, we show, in the Online Appendix, that the explanatory power of each motive remains stable during the 18-month window around the survey, suggesting relatively persistent importance in the time-series. Second, and relatedly, it is also possible that some retail investors learn to debias themselves from past mistakes, and the importance of certain mechanisms may decay over time (Seru, Shumway and Stoffman 2010). While our cross-sectional setting does not allow us to directly speak to the issue of learning, we note that some recent evidence suggests that retail investors do not appear to learn from their prior mistakes (e.g., Anagol, Balasubramaniam and Ramadorai 2019).

3. Additional Evidence on Different Mechanisms

In this section, we conduct additional analysis to further reinforce the trading motives highlighted in Section 2. Sections 3.1 and 3.2 further analyze the two positive results, gambling preference and perceived information advantage, respectively. Section 3.3 focuses on one "null" result: neglect of trading costs.

3.1. Gambling Preference

So far, we have coded the survey responses into dummy variables that take on values of 0 and 1, but this may reduce their explanatory power. To address this concern, Table 15 reports a more detailed summary of trading characteristics when investors are sorted into five groups based on their answers to the "blockbusters" question. While this single-sorting approach ignores the correlations of gambling preference with other trading motives, it provides a more granular look at the explanatory power of gambling preference.[23]

Panel A shows the distribution of turnover for each of the five groups. There is a monotonically increasing pattern across the five groups that differs in the extent that investors agreed with the gambling preference question. This monotonic pattern is present not just in the mean and the median of the monthly turnover rate, but also across various percentiles in the

[23] Note that the coefficient of gambling preference is virtually unchanged from the univariate regression in Table 13 to the horse race in Table 14, suggesting that the effect is not affected by other trading motives.

distribution, indicating that this pattern is not driven by outliers. On average, the difference between "strongly agree" and "strongly disagree" is about 21%, suggesting sizable economic significance—a monthly turnover rate of 21% translates into an annualized transaction fee of 0.6%.

Is the trading associated with gambling preference excessive? Panel B reports portfolio returns for the five groups of investors and shows that this is the case: the five groups exhibit similar raw returns before fees. In fact, the "strongly agree" group on average earns −0.35% lower monthly returns than the "strongly disagree," albeit the difference is not statistically significant. Together, the lack of superior performance and the large transaction costs suggest their trading is excessive.

In Panel C, we examine the characteristics of stocks purchased by the five groups of investors. Investors with a survey-based gambling preference tend to buy stocks that are smaller, have a larger market beta, and have larger counts of daily up-limit hits, and higher past volatility and past returns. These stocks also perform worse subsequently, confirming that investors with a gambling preference trade in the wrong direction and their trading is excessive.

3.2. Perceived Information Advantage

We now further analyze perceived information advantage in Table 16, again by sorting investors into five groups based on their answers to the question asking how often they think they have an information advantage over others. Panel A presents the monthly turnover rate of these groups. Similar to before, investors who "always" think they have an information advantage exhibit higher turnover than those who "never" think so for almost all the distribution percentiles we look at. The magnitude is also similar: the difference in monthly turnover rate between the "always" and the "never" groups is about 24%, implying an annual transaction fee of 0.7%.

Is the perceived information advantage supported by superior performance in portfolio returns? Panel B suggests that this is not the case: the five groups exhibit similar performance before fees, indicating that those who report having an information advantage do not outperform others in selecting better stocks. Accounting for trading fees would make their net performance clearly worse. Thus, the perceived information advantage reflects a form of overconfidence rather than better information.

3.3. Neglect of Trading Costs

In both the baseline and the horse race, none of the survey variables for neglect of trading costs can explain turnover in the right direction. This contradicts the popular view that Chinese retail investors trade so much because they neglect trading costs. The regression results reported in Tables 13 and 14 even suggest an opposite pattern in one of the measures that investors with more awareness of trading costs trade more. This pattern, however, may reflect a reverse selection that investors who trade more incur more total costs and are more aware of their existence. To further isolate the effect of awareness of trading costs, we have also implemented a randomized experiment.

Among all 500 brokerage branches we distributed the survey to, we randomly selected 250 branches to include an additional "nudge." The "nudge" asked the respondent to read a short article that highlighted the negative consequences of excessive trading. As shown in Figure 2, the article contained a detailed calculation of how much investors lose from frequent trading along with a quote from Warren Buffett advising investors to buy and hold. Instead of presenting trading costs as a fraction of total transaction value, we made it more salient by presenting the annualized fee rate for a frequent trader. We also included a "validation" question after the article by asking the respondent to calculate the total trading costs of a given level of turnover. Answers to this question help identify those who have actually read the article and therefore been treated.

We study the effect of this "nudge" in a difference-in-difference framework, and the results are reported in Table 17. Column (1) shows that the interaction term is small and insignificant, suggesting that the treatment and control groups exhibit similar turnover rates one month after the survey. We repeat this exercise in columns (2) and (3) by expanding the window to three months and six months before and after the survey, and the interaction term remains insignificant. Overall, these results suggest that the nudge had no effect on reducing trading. One might argue that the "nudge" was not sufficiently strong and the treated group may not have read the article carefully. However, we identify an investor as treated only if she was in the treated group and answered the "validation" question correctly.

Taken together, our analysis suggests that neglect of trading costs is not a key driver of excessive trading. This finding has an important policy implication. Policy makers across the world, including China's stock market regulator, the China Securities Regulatory Commission (CSRC), frequently use Tobin taxes as a policy tool to curb speculative trading in stock markets. To the extent that investors may engage in excessive trading despite their awareness of the trading costs, our finding casts doubt on the effectiveness of Tobin taxes.[24]

4. Comparing Survey-Based and Transaction-Based Measures

In our analysis so far, we have taken survey responses as direct measures of trading motives and use them to study why investors trade so much. These survey-based measures have some clear advantages over transaction-based measures. First, well-designed surveys provide relatively clean measures of trading motives. Second, survey responses allow researchers to measure a large set of trading motives from the perspectives of the respondents at the same time, including those that are hard to measure from administrative data. There are also various concerns about survey data. The primary concern, the one we have already addressed through various validation exercises, is that survey responses may not capture actual trading behavior. A second concern is that survey responses are noisy—perhaps on average respondents do answer truthfully, but their responses at the individual level may be noisy. This is a concern that also arises in our setting. For instance, in Table 10, while the relationship between survey-based gambling preference and transaction-based gambling behavior is statistically significant, the R-squared is rather small across all specifications.

The concern about noise in survey responses motivates a follow-up question: do transaction-based behavioral measures have stronger power than survey-based measures? We now address this question by comparing survey-based and transaction-based measures of gambling behavior. Table 18 reports the results when we sort investors into different groups based on their gambling behavior directly measured from transaction data in the pre-survey sample period. This transaction-based measure turns out to be much more powerful in explaining turnover in the post-survey sample: the difference in the monthly turnover rate between the top and bottom groups is 97%, quadrupling the magnitude of 21% reported in Table 15 based on the survey-based measure of gambling behavior. In addition, the difference in other trading characteristics between the top and bottom groups is also larger in magnitude than the respective value reported in Table 15.

If this transaction-based measure of gambling behavior is so powerful, why is it not used directly instead of relying on the survey-based measure? To address this question, we regress the

[24] There is mixed evidence on the effects of Tobin taxes in reducing speculative trading and price volatility. See Song and Xiong (2018) for a detailed review of the CSRC's policy interventions in the stock market and Deng, Liu and Wei (2018) and Cai et al. (2019) for studies of effects of increasing the stamp tax for stock trading in China.

transaction-based measure of gambling behavior on all survey-based trading motives and report the results in Table 19. It is reassuring to see that the survey-based measure of gambling preference is indeed the most powerful explanatory variable in this regression. However, a number of other survey-based trading motives are also significantly correlated with the transaction-based measure of gambling behavior. For instance, investors with perceived information advantage also gamble more. Therefore, although the transaction-based measure of gambling behavior is more powerful in explaining trading, this measure is partially correlated with other trading motives and its explanatory power may not come solely from gambling preference.[25]

Taken together, our comparison shows a trade-off between survey-based and transaction-based measures of trading motives. Survey-based measures have stronger power from the economic perspective of having qualitative tests of different trading motives, even though they may contain more noise and thus have weaker power from the statistical perspective of explaining cross-individual variation of trading. Transaction-based measures have stronger statistical power, but they may reflect multiple mechanisms, and their economic interpretations are thus not as sharp as survey-based measures.

5. Conclusion

We design and administer a nation-wide survey to study why investors trade so much, by directly comparing a large set of explanations for trading volume. The key innovation in our approach is to combine survey responses and transaction data, allowing us to not only validate survey responses but also to compare survey-based and transaction-based approaches.

Based on this integrated approach, we highlight a number of new findings. First, we find systematic evidence that survey responses are consistent with actual trading behavior. Second, overconfidence (in having an information advantage) and gambling preference dominate other trading motives in explaining excessive trading. Third, popular arguments such as neglect of trading costs, low financial literacy, and social interaction do not contribute to excessive trading. Finally, by analyzing the pros and cons of survey-based and transaction-based approaches, we

[25] The transaction-based measure of gambling behavior may also contain effects from other omitted variables. For example, one possible omitted variable is investor attention—investors who pay more attention to the stock market are more likely to be drawn to lottery-like stocks as they appear more often in the news. While these investors may exhibit gambling-like behavior, their frequent trading is explained by their attention to the stock market.

argue that our integrated approach can address the concerns faced by each of these approaches alone.

References

Alpert, M., & Raiffa, H. (1982). A Progress Report on the Training of Probability Assessors.Anagol, Santosh, Vimal Balasubramaniam, and Tarun Ramadorai, 2019, Learning from Noise: Evidence from India's IPO lotteries, *Working Paper*.

Banerjee, Abhijit V., 1992, A Simple Model of Herd Behavior, *Quarterly Journal of Economics* 107, 797–817.

Barber, Brad M., Xing Huang, Kwangmin Ko, and Terrance Odean, 2019, Leveraging Overconfidence, *Working Paper*.

Barber, Brad M., Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2009, Just How Much Do Individual Investors Lose by Trading? *Review of Financial Studies* 22, 609–632.

Barber, Brad M., and Terrance Odean, 2000, Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors, *Journal of Finance* 55, 773–806.

Barber, Brad M., and Terrance Odean, 2013, The Behavior of Individual Investors, in George Constantinides ed.: *Handbook of the Economics of Finance* (Elsevier).

Barberis, Nicholas, 2018, Psychology-Based Models of Asset Prices and Trading Volume, *Handbook of Behavioral Economics: Applications and Foundations 1* (North-Holland).

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-CAPM: An Extrapolative Capital Asset Pricing Model, *Journal of Financial Economics* 115, 1–24.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2018, Extrapolation and Bubbles, *Journal of Financial Economics* 129, 203–227.

Barberis, Nicholas, and Ming Huang, 2008, Stocks as Lotteries: The Implications of Probability Weighting for Security Prices, *American Economic Review* 98, 2066–2100.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A Model of Investor Sentiment, *Journal of Financial Economics* 49, 307–343.

Barberis, Nicholas, and Wei Xiong, 2009, What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation, *Journal of Finance* 64, 751–784.

Barberis, Nicholas, and Wei Xiong, 2012, Realization Utility, *Journal of Financial Economics* 104, 251–271.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey, 2013, Managerial Miscalibration, *Quarterly Journal of Economics* 128, 1547–1584.

Benos, Alexandros V., 1998, Aggressiveness and Survival of Overconfident Traders, *Journal of Financial Markets* 1, 353–383.

Bertrand, Marianne, and Sendhil Mullainathan, 2001, Do People Mean What They Say? Implications for Subjective Survey Data, *American Economic Review* 91, 67–72.

Birru, Justin, 2015, Confusion of Confusions: A Test of the Disposition Effect and Momentum, *Review of Financial Studies* 28, 1849–1873.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2012, Salience Theory of Choice Under Risk, *Quarterly Journal of Economics* 127, 1243–1285.

Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected Idiosyncratic Skewness, *Review of Financial Studies* 23, 169–202.

Brown, Stephen, Yan Lu, Sugata Ray, and Melvyn Teo, 2018, Sensation Seeking and Hedge Funds, *Journal of Finance* 73, 2871–2914.

Cai, Jinghan, Jibao He, Wenxi Jiang, and Wei Xiong, 2019, The Whack-A-Mole Game: Tobin Taxes and Trading Frenzy, *Working Paper*.

Chetty, Raj, Adam Looney, and Kory Kroft, 2009, Salience and Taxation: Theory and Evidence, *American Economic Review* 99, 1145–1177.

Chinco, Alex, Samuel M. Hartzmark, and Abigail B. Sussman, 2019, Risk-Factor Irrelevance, *Working Paper*.

Choi, James J., and Adriana Z. Robertson, 2019, What Matters to Individual Investors? Evidence from the Horse's Mouth, *Working Paper*.

Cochrane, John H., 2011, Presidential Address: Discount Rates, *Journal of Finance* 66, 1047–1108.

Cochrane, John H., 2017, Macro-Finance, *Review of Finance* 21, 945–985.

Da, Zhi, Xing Huang, and Lawrence J. Jin, 2019, Extrapolative Beliefs in the Cross-Section: What Can We Learn from the Crowds? *Working Paper*, 1–54.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor Psychology and Security Market Under- and Overreactions, *Journal of Finance* 53, 1839–1885.

Dellavigna, Stefano, 2009, Psychology and Economics: Evidence from the Field, *Journal of Economic Literature* 47, 315–372.

DellaVigna, Stefano, and Elizabeth Linos, 2020, RCTs to Scale: Comprehensive Evidence from Two Nudge Units, *Working Paper*.

Deng, Yongheng, Xin Liu, and Shang-Jin Wei, 2018, One Fundamental and Two Taxes: When Does a Tobin Tax Reduce Financial Price Volatility?, *Journal of Financial Economics* 130, 663–692.

Dorn, Daniel, and Gur Huberman, 2005, Talk and Action: What Individual Investors Say and What They Do, *Review of Finance* 9, 437–481.

Dorn, Daniel, and Paul Sengmueller, 2009, Trading as Entertainment? *Management Science* 55, 591–603.

Epper, Thomas, Ernst Fehr, Helga Fehr-Duda, Claus Thustrup Kreiner, David Dreyer Lassen, Søren Leth-Petersen, and Gregers Nytoft Rasmussen, 2020, Time Discounting and Wealth Inequality, *American Economic Review* 110, 1177–1205.

Eyster, Erik, Matthew Rabin, and Dimitri Vayanos, 2019, Financial Markets Where Traders Neglect the Informational Content of Prices, *Journal of Finance* 74, 371–399.

Friedman, Milton, and L.J. Savage, 1948, The Utility Analysis of Choices Involving Risk, *Journal of Political Economy* 56, 279–304.

Friedman, Milton, 1953, The Methodology of Positive Economics, *Essays in Positive Economics* 3, 145–178.

Frydman, Cary, Nicholas Barberis, Colin Camerer, Peter Bossaerts, and Antonio Rangel, 2014, Using Neural Data to Test a Theory of Investor Behavior: An Application to Realization Utility, *Journal of Finance* 69, 907–946.

Frydman, Cary, and Antonio Rangel, 2014, Debiasing the Disposition Effect by Reducing the Saliency of Information about a Stock's Purchase Price, *Journal of Economic Behavior and Organization* 107, 541–552.

Frydman, Cary, and Baolian Wang, 2020, The Impact of Salience on Investor Behavior: Evidence from a Natural Experiment, *Journal of Finance* 75, 229–276.

Gao, Xiaohui, and Tse-Chun Lin, 2014, Do Individual Investors Treat Trading as a Fun and Exciting Gambling Activity? Evidence from Repeated Natural Experiments, *Review of Financial Studies* 28, 2128–2166.

Gervais, Simon, and Terrance Odean, 2001, Learning to Be Overconfident, *Review of Financial Studies* 14, 1–27.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Steve Utkus, 2019, Five Facts about Beliefs and Portfolios, *Working Paper*.

Glaser, Markus, Markus Nöth, and Martin Weber, 2004, Behavioral Finance, *Blackwell Handbook of Judgement and Decision Making*, 527–546.

Glaser, Markus, and Martin Weber, 2007, Overconfidence and Trading Volume, *Geneva Risk and Insurance Review* 32, 1–36.

Graham, John R, Campbell R Harvey, and Hai Huang, 2009, Investor Competence, Trading Frequency, and Home Bias, *Management Science* 55, 1094–1106.

Grampp, William D, 1948, Adam Smith and the Economic Man, *Journal of Political Economy* 56, No. 4: 315–336.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of Returns and Expected Returns, *Review of Financial Studies* 27, 714–746.

Grinblatt, Mark, and Bing Han, 2005, Prospect Theory, Mental Accounting, and Momentum, *Journal of Financial Economics* 78, 311–339.

Grinblatt, Mark, and Matti Keloharju, 2001, How Distance, Language, and Culture Influence Stockholdings and Trades, *Journal of Finance* 56, 1053–1073.

Grinblatt, Mark, and Matti Keloharju, 2009, Sensation Seeking, Overconfidence, and Trading Activity, *Journal of Finance* LXIV, 549–578.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa, 2011, IQ and Stock Market Participation, *Journal of Finance* 66, 2121–2164.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70, 393–408.

Han, Bing, David Hirshleifer, and Johan Walden, 2019, Social Transmission Bias and Investor Behavior, *Working Paper*.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, 2004, Social Interaction and Stock-Market Participation, *Journal of Finance* 59, 137–163.

Hong, Harrison, José Scheinkman, and Wei Xiong, 2008, Advisors and Asset Prices: A Model of the Origins of Bubbles, *Journal of Financial Economics* 89, 268–287.

Hong, Harrison, and Jeremy C. Stein, 1999, A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, *Journal of Finance* 54, 2143–2184.

Ingersoll, Jr., Jonathan E., and Lawrence J. Jin, 2013, Realization Utility with Reference-Dependent Preferences, *Review of Financial Studies* 26, 723–767.

Jin, Lawrence J., and Pengfei Sui, 2019, Asset Pricing with Return Extrapolation, *Working Paper*, 1–59.

Kahneman, Daniel, and Shane Frederick, 2002, Representativeness Revisited: Attribute Substitution in Intuitive Judgment, *Heuristics and Biases: The Psychology of Intuitive Judgment* 49: 81.

Kelley, Eric K., and Paul C. Tetlock, 2013, How Wise Are Crowds? Insights from Retail Orders and Stock Returns, *Journal of Finance* 68, 1229–1265.

Kelly, Morgan, and Cormac O Grada, 2000, Market Contagion: Evidence from the Panics of 1854 and 1857, *American Economic Review* 90, 1110–1124.

Koijen, Ralph, Stijn Van Nieuwerburgh, and Roine Vestman, 2015, Judging the Quality of Survey Data by Comparison with "Truth" as Measured by Administrative Records Evidence from Sweden, in *Improving the Measurement of Consumer Expenditures*, pp. 308–346, University of Chicago Press.

Kreiner, Claus Thustrup, David Dreyer Lassen, and Søren Leth-Petersen, 2015, Measuring the Accuracy of Survey Responses Using Administrative Register Data: Evidence from Denmark, in *Improving the Measurement of Household Consumption Expenditures*, pp. 289–307, University of Chicago Press.

Kumar, Alok, 2009, Who Gambles in the Stock Market? *Journal of Finance* 64, 1889–1933.

Kyle, Albert S., 1985, Continuous Auctions and Insider Trading, *Econometrica* 53, 1315–1335.

Kyle, Albert S., and F. Albert Wang, 1997, Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test? *Journal of Finance* 52, 2073–2090.

Liao, Jingchi, Cameron Peng, and Ning Zhu, 2020, Price and Volume Dynamics in Bubbles, *Working Paper*.

Lusardi, Annamaria, and Olivia S. Mitchell, 2007, Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth, *Journal of Monetary Economics* 54, 205–224.

Lusardi, Annamaria, and Olivia S. Mitchell, 2011, Financial Literacy around the World: An Overview, *Journal of Pension Economics and Finance* 10, 497–508.

Lusardi, Annamaria, and Olivia S. Mitchell, 2014, The Economic Importance of Financial Literacy: Theory and Evidence, *Journal of Economic Literature* 52, 5–44.

Markowitz, Harry, 1952, Portfolio Selection, *Journal of Finance* 7, 77–91.

Milgrom, Paul, and Nancy Stokey, 1982, Information, Trade and Common Knowledge, *Journal of Economic Theory* 26, 17–27.

Odean, Terrance, 1998, Volume, Volatility, Price, and Profit When All Traders Are Above Average, *Journal of Finance* 53, 1887–1934.

Odean, Terrance, 1999, Do Investors Trade Too Much? *American Economic Review* 89, 1279–1298.

Pool, Veronika K., Noah Stoffman, and Scott E. Yonker, 2015, The People in Your Neighborhood: Social Interactions and Mutual Fund Portfolios, *Journal of Finance* 70, 2679–2732.

Scheinkman, José A., and Wei Xiong, 2003, Overconfidence and Speculative Bubbles, *Journal of Political Economy* 111, 1183–1219.

Seru, Amit, Tyler Shumway, and Noah Stoffman, 2010, Learning by Trading, *Review of Financial Studies* 23, 705-739.

Shefrin, Hersh, and Meir Statman, 1985, The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence, *Journal of Finance* 40, 777–790.

Shefrin, Hersh, and Meir Statman, 2000, Behavioral Portfolio Theory, *Journal of Financial and Quantitative Analysis* 35, 127–151.

Shiller, Robert J., 1984, Stock Prices and Social Dynamics, *Brookings Papers on Economic Activity* 2, 457–510.

Shiller, Robert J., 1989, Comovements in Stock Prices and Comovements in Dividends, *Journal of Finance* 44, 719–729.

Shiller, Robert J., 2000, Measuring Bubble Expectations and Investor Confidence, *Journal of Psychology and Financial Markets* 1, 49–60.

Song, Zheng (Michael), and Wei Xiong, 2018, Risks in China's Financial System, *Annual Review of Financial Economics* 10, 261–286.

Taubinsky, Dmitry, and Alex Rees-Jones, 2018, Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment, *Review of Economic Studies* 85, 2462–2496.

Tversky, Amos, and Daniel Kahneman, 1992, Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty* 5, 297–323.

Van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie, 2011, Financial Literacy and Stock Market Participation, *Journal of Financial Economics* 101, 449–472.

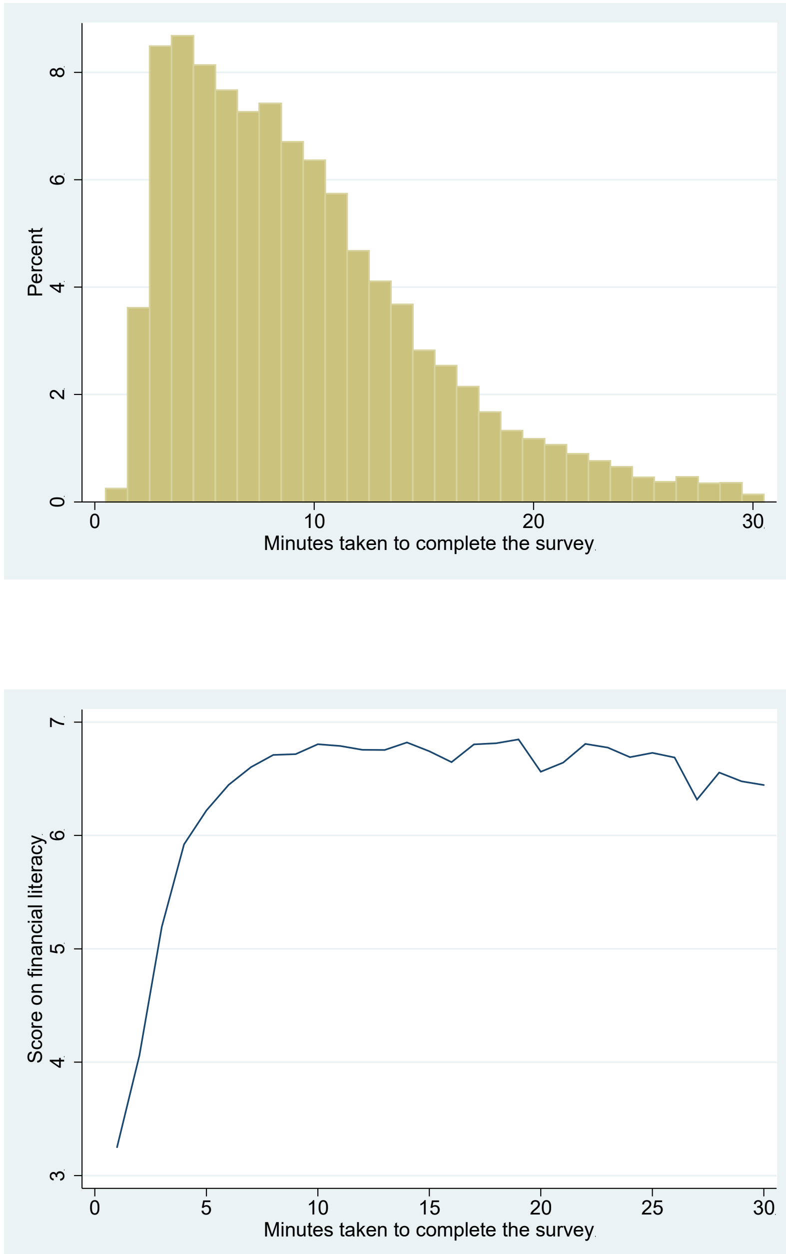

Figure 1: Relationship Between Financial Literacy Score and the Time Taken to Complete the Survey

Figure 2: The Treatment that Nudges Investors to Reduce Trading Due To Transaction Costs

Note: This figure shows the message with pictures that a random half of respondents read in the survey.

Theory	Forms of Representation	Papers
Overconfidence	1. overplacement 2. miscalibration of uncertainty	Odean (1998); Benos (1998); Glaser and Weber (2007); Dorn and Huberman (2005); Graham, Harvey and Huang (2009); Ben-David, Graham and Harvey (2013)
Extrapolation	1. upward trend to continue 2. downward trend to continue	Barberis et al. (2018); Jin and Sui (2019); Da et al. (2019); Liao, Peng and Zhu (2020)
Neglect of trading costs	1. underestimation of transaction fees 2. knowledge about the bid-ask spread 3. salience of transaction fees	Barber and Odean (2000); Barber, Lee, Liu and Odean (2009); Bordalo et al. (2012)
Gambling preferences	1. overweight small probability 2. understand small probability	Friedman and Savage (1948); Markowitz (1952); Shiller (1989, 2000); Barber and Odean (2000); Shefrin and Statman (2000); Barberis and Huang (2008); Kumar (2009); Barber et al. (2008)
Realization utility	1. utility from realizing gains 2. disutility from realizing losses	Barberis and Xiong (2009, 2012); Ingersoll and Jin (2013); Frydman et al. (2014)
Sensation seeking	1. novelty seeking 2. volatility seeking	Grinblatt and Keloharju (2009); Dorn and Sengmueller (2009); Gao and Lin (2014)
Private information	1. belief in having information advantage 2. fear of being at information disadvantage	Kyle (1985); Grossman and Stiglitz (1980); Gervais and Odean (2001); Scheinkman and Xiong (2003)
Social/advisor influence	1. advisor influence 2. social influence	Shiller (1989); Banerjee (1992); Kelly and Grada (2000); Hong, Kubik and Stein (2004a, 2004b); Hong, Scheinkman, and Xiong (2008); Pool, Stoffman and Yonker (2015)
Financial/investment literacy	1. compounding; 2. inflation; 3. diversification; 4. asset risk; 5. definition of stocks; 6. definition of bonds; 7. the PE ratio; 8. definition of mutual funds.	Van Rooij, Lusardi and Alessie (2011); Grinblatt, Keloharju and Linnainmaa (2011)
Liquidity and rebalance needs		Kyle (1985)

Table 1: Summary of Theories on Trading Volume

Panel A: By Broker	Observations	Percentage
Guotai Junan Securities	1,519	11.8%
CITIC Securities	1,410	11.0%
Haitong Securities	1,390	10.8%
China Merchants Securities	1,372	10.7%
Huatai Securities	1,350	10.5%
Guosen Securities	1,252	9.8%
China Securities	1,203	9.4%
Shenwan Hongyuan Securities	1,169	9.1%
GF Securities	1,111	8.7%
China Galaxy Securities	1,051	8.2%
Panel B: By Province/Region		
Guangdong	1,674	13.1%
Zhejiang	1,201	9.4%
Jiangsu	1,138	8.9%
Shanghai	1,135	8.9%
Hubei	629	4.9%
Beijing	622	4.9%
Fujian	600	4.7%
Hunan	572	4.5%
Shandong	542	4.2%
Henan	531	4.1%
Sichuan	530	4.1%
Anhui	463	3.6%
Jiangxi	388	3.0%
Hebei	385	3.0%
Liaoning	331	2.6%
Chongqing	284	2.2%
Heilongjiang	250	2.0%
Guangxi	230	1.8%
Shanxi	222	1.7%
Shaanxi	198	1.5%
Others	931	7.2%
Total	12,856	100%

Table 2: Distribution of Survey Respondents Across Brokers and Provinces

Note: This table shows the distributions of survey respondents across brokerage firms (Panel A) and across province/regions (Panel B).

Gender	Survey Respondents	Investor Population	Income (RMB)	Survey Respondents
Male	54.0%	71.7%	< 20K	3.8%
Female	46.0%	28.3%	20K to 100K	17.2%
			100K to 200K	29.5%
Education			200K to 500K	29.5%
Middle school or below	8.6%	7.3%	500K to 1M	12.6%
High school	15.6%	24.7%	1M to 2M	4.2%
Professional school	21.9%	26.0%	2M to 10M	2.1%
College	44.9%	23.6%	10M and above	1.2%
Graduate school and above	9.2%	3.4%		
			Net worth (RMB)	
Age			< 20K	4.8%
20 to 30	27.8%	21.3%	20K to 100K	12.3%
30 to 40	29.1%	27.4%	100K to 500K	27.5%
40 to 50	19.9%	24.5%	500K to 1M	22.3%
50 to 60	14.8%	15.1%	1M to 2M	21.9%
>60	8.5%	11.7%	2M to 10M	6.5%
			10M and above	4.8%

Table 3: Distribution of Survey Respondents Across Different Demographic Groups

Note: This table compares the demographics between survey respondents and the investor population. For the investor population, information on gender, education, and age is obtained from the centralized database at the Shenzhen Stock Exchange, and information on income and net worth is missing.

Panel A: Correct Rate by Question

Question	Correct rate
1. Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?	88.4%
2. Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much will you be able to buy with the money in this account?	91.5%
3. Do you agree with the following statement? Buying an individual stock is usually less risky than buying a stock mutual fund.	86.2%
4. Normally, which asset displays the highest fluctuation over time?	95.2%
5. Which of the following statements is correct? If somebody buys a stock of firm B in the stock market….	76.3%
6. Normally, when the market interest rate falls, the price of an existing bond will ….	54.7%
7. What is the P/E ratio?	75.8%
8. Which of the following statements about mutual funds is correct?	90.3%

Panel B: Distribution of Financial Literacy Scores

Score	Actual	Self-assessed
0	0.4%	0.6%
1	0.7%	0.7%
2	1.7%	1.8%
3	2.3%	4.6%
4	5.1%	6.9%
5	8.9%	13.0%
6	17.9%	16.2%
7	30.1%	17.7%
8	33.0%	32.7%
N/A	0.0%	5.8%

Table 4: Survey Responses on Questions on Financial Literacy

Note: This table shows the summary statistics of investors' responses to questions on financial literacy. In Panel A, we show the correct rate by question. In Panel B, we compare their actual and self-assessed performances, where actual performance is measured by the total number of questions answered correctly and self-assessed performance by the total number of questions one reports to have answered correctly.

Panel A: Overconfidence											
1. What fraction of retail investors do you think earned higher returns than you in 2017?	< 10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	80–90%	> 90%	N/A
	11.8%	13.8%	15.8%	13.5%	12.4%	10.4%	5.8%	3.8%	2.2%	3.4%	7.2%
2. Actual score–Self-assessed score	< –4	–4	–3	–2	–1	0	1	2	3	4	> 4
	0.8%	1.8%	5.4%	11.4%	19.7%	35.1%	17.7%	5.6%	1.7%	0.6%	0.4%
3. Upside return–Downside return	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	>50%
	32.7%	14.9%	9.2%	6.9%	5.2%	5.2%	4.3%	3.4%	3.1%	2.5%	12.7%

Panel B: Extrapolation						
1. After a stock's price keeps rising for a while, I usually believe that the price will rise even further in the future.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	4.8%	26.9%	39.3%	22.8%	1.3%	5.0%
2. After a stock's price keeps falling for a while, I usually believe that the price will fall even further in the future.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	4.4%	29.1%	41.9%	18.2%	1.3%	5.3%

Panel C: Neglect of Trading Costs								
1. Estimating the cost of a round-trip buy and sell at the value of 10,000 RMB	0–5	5–10	10–15	15–20	20–25	25–30	30–35	>35
	17.3%	27.7%	23.6%	12.8%	8.4%	3.7%	2.1%	5.5%
2. How often do you consider transaction costs when you trade?	Never	Rarely	Sometimes	Often	Always	N/A		
	14.6%	37.7%	27.0%	13.8%	4.6%	2.5%		
3. The bid-ask spread is one form of transaction cost (The bid-ask spread is the difference between the lowest ask price and the highest bid price).	Agree	Disagree	Don't Understand	Don't Know	N/A			
	59.8%	23.1%	8.5%	7.2%	1.4%			

Table 5: Survey Responses on Questions on Beliefs

Note: This table tabulates the distribution of investors' answers to questions related to overconfidence (Q10, Q11, Q13, Q14), extrapolation (Q26, Q27), and neglect of trading costs (Q15, Q16, Q17).

Panel A: Gambling Preference						
Blockbusters 1. When I trade stocks, I aim to select those stocks whose price would rise sharply in a short period of time so that I can make a lot of money quickly.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	10.4%	25.4%	33.9%	23.0%	4.6%	2.7%
Lotteries 2. When I trade stocks, I often think of them as lotteries: I am willing to accept small losses in exchange for the possibility of a big upside.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	5.5%	24.9%	27.2%	32.5%	7.3%	2.7%

Panel B: Realization Utility					
Winners 1. Normally, if the price of a stock in your portfolio rose substantially since you bought it, which of these two actions would make you feel happier: holding on to the stock, or selling that stock?	Sell	Same	Hold	No Feeling	N/A
	37.2%	23.7%	25.3%	9.2%	4.5%
Losers 2. Normally, if the price of a stock in your portfolio dropped substantially since you bought it, which of these two actions would make you feel more painful: holding on to the stock, or selling that stock?	Sell	Same	Hold	No Feeling	N/A
	22.9%	28.0%	32.1%	12.2%	4.8%

Panel C: Sensation Seeking						
Novelty 1. I feel excited about getting to know new stocks and new firms.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	5.9%	20.3%	43.9%	21.0%	3.2%	5.7%
Volatility 2. I feel excited about the stock market moving up and down.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	5.4%	23.4%	36.7%	26.2%	4.3%	4.1%

Table 6: Survey Responses on Questions Related to Preferences

Note: This table tabulates the distribution of investors' answers to questions related to gambling preference (Q18, Q19), realization utility (Q20, Q21), and sensation seeking (Q22, Q23).

Panel A: Information						
Perceived information advantage						
1. When you decide to trade a stock, how often do you believe that you know the stock better than others?	Never	Rarely	Sometimes	Often	Always	N/A
	8.7%	27.9%	40.3%	14.5%	3.2%	5.4%
Dismissive of others' information						
2. When you decide to trade a stock, how often do you worry that other investors know about the stock better than you do?	Never	Rarely	Sometimes	Often	Always	N/A
	18.2%	28.9%	32.3%	12.6%	2.5%	5.6%
Panel B: Social Interaction						
Social influence						
1. When you decide to trade a stock, how often are you influenced by your family members, friends, or other acquaintances?	Never	Rarely	Sometimes	Often	Always	N/A
	11.6%	31.2%	40.0%	11.8%	1.7%	3.8%
Advisor influence						
2. When you decide to trade a stock, how often are you influenced by your investment advisors?	Never	Rarely	Sometimes	Often	Always	N/A
	17.8%	35.0%	35.8%	7.2%	1.2%	3.1%
Panel C: Others						
Portfolio rebalance needs						
1. When you decide to trade a stock, how often is it that you need to rebalance your portfolio?	Never	Rarely	Sometimes	Often	Always	N/A
	9.6%	30.5%	44.5%	9.5%	1.7%	4.2%
Liquidity needs						
2. When you decide to trade a stock, how often is it because you need money somewhere else?	Never	Rarely	Sometimes	Often	Always	N/A
	7.0%	25.9%	45.0%	14.4%	2.6%	5.1%
Panel D: Risk Aversion						
1. Suppose you are the only income earner in the family, and you have a good job guaranteed to give you your current income every year for life. You are given the opportunity to take a new, equally good job. With a 50% chance it will double your income, and with a 50% chance, it will cut your income by 20%. Would you take the new job?		Yes	No	Don't Know	N/A	
		51.6%	34.1%	11.3%	3.0%	
2. Suppose the chances were 50% that it would double your income and 50% that it would cut it by 1/3. Would you take the new job?		Yes	No	Don't Know	N/A	
		45.3%	37.5%	13.8%	3.4%	
3. Suppose the chances were 50% that it would double your income and 50% that it would cut it by 1/2. Would you take the new job?		Yes	No	Don't Know	N/A	
		26.0%	57.4%	13.2%	3.5%	

Table 7: Survey Responses on Questions on Information and Other Trading Motives

Note: This table tabulates the distribution of investors' answers to questions related to information (Q24, Q25), social interaction (Q28, Q29), others (Q30, Q31), and risk aversion (Q32, Q33, Q34).

Gender	Main Sample	Population
Male	54.4%	71.7%
Female	45.6%	28.3%
Education		
Middle school or blow	5.1%	7.3%
High school	17.6%	24.7%
Professional school	24.4%	26.0%
College	38.5%	23.6%
Graduate school and above	6.1%	3.4%
Others	8.4%	14.8%
Age		
< 30	26.1%	21.3%
30 to 40	27.4%	27.4%
40 to 50	22.4%	24.5%
50 to 60	16.0%	15.1%
> 60	8.1%	11.7%
Investment age (in years)		
< 2	21.2%	10.0%
2 to 6	26.2%	29.8%
6 to 10	17.4%	18.0%
> 10	35.1%	42.2%
Trading characteristics in 2017		
Maximum value of investment (in thousand RMB)	1,250	639
Turnover	8.3	9.4
Raw return rate	-1.20%	-3.90%

Table 8: Summary Statistics for the Main Sample and the Population

Note: This table shows the summary statistics for investors in the main sample and the investor population. The main sample includes 4,671 survey respondents that: 1) can be identified in the Shenzhen Stock Exchange centralized database, and 2) hold at least one SZSE stock during the two-year window before the survey. The population's characteristics are obtained from the centralized database at the Shenzhen Stock Exchange. See the Online Appendix for more details about variable definitions.

	Variable	Mean	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	Overplacement, performance	0.67	1.00																				
2	Overplacement, literacy	0.24	0.03	1.00																			
3	Miscalibration	0.69	0.08	0.02	1.00																		
4	Underestimation of transaction costs	0.69	(0.02)	0.02	0.00	1.00																	
5	Do not consider transaction costs	0.53	0.03	(0.01)	0.01	0.11	1.00																
6	Do not think bid-ask spread is a cost	0.33	(0.01)	(0.01)	(0.05)	(0.08)	(0.06)	1.00															
7	Extrapolation, up	0.32	(0.01)	0.04	0.02	0.00	0.08	(0.09)	1.00														
8	Extrapolation, down	0.34	0.00	0.04	0.02	0.00	0.07	(0.10)	**0.62**	1.00													
9	Gambling preference, blockbusters	0.37	(0.01)	0.04	(0.01)	(0.02)	0.05	(0.09)	0.25	0.21	1.00												
10	Gambling preference, lotteries	0.30	(0.01)	0.04	0.01	0.02	0.07	(0.10)	0.24	0.21	**0.40**	1.00											
11	Realization utility, winners	0.36	(0.03)	0.02	0.05	0.07	0.01	(0.09)	(0.01)	0.05	0.04	0.07	1.00										
12	Realization utility, losers	0.22	0.01	0.02	(0.01)	0.03	0.04	(0.08)	0.06	0.06	0.04	0.04	**0.22**	1.00									
13	Sensation seeking, novelty	0.24	(0.03)	0.03	0.00	0.03	0.08	(0.12)	0.19	0.18	0.18	0.24	0.07	0.12	1.00								
14	Sensation seeking, volatility	0.29	0.00	0.04	0.03	0.03	0.05	(0.12)	0.22	0.23	0.22	0.26	0.09	0.13	**0.42**	1.00							
15	Perceived information advantage	0.18	0.06	0.07	0.01	(0.02)	(0.03)	(0.03)	(0.01)	0.01	(0.06)	(0.09)	(0.02)	(0.02)	0.01	0.02	1.00						
16	Dismissive of others' information	0.14	(0.02)	0.03	(0.03)	(0.05)	(0.11)	0.08	(0.03)	0.01	0.02	(0.01)	(0.01)	(0.04)	(0.02)	(0.03)	**0.14**	1.00					
17	Social influence	0.13	(0.01)	0.02	(0.04)	(0.04)	(0.02)	0.07	(0.01)	(0.01)	0.06	0.05	0.01	(0.02)	0.00	(0.03)	0.01	0.22	1.00				
18	Advisor influence	0.07	(0.01)	0.01	(0.01)	(0.02)	0.00	0.03	0.01	0.00	0.00	0.02	0.02	(0.02)	0.03	0.01	0.03	0.15	**0.32**	1.00			
19	Portfolio rebalance needs	0.17	0.01	0.02	(0.02)	(0.07)	(0.07)	0.07	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.02)	(0.01)	(0.02)	0.20	0.17	0.12	0.08	1.00		
20	Liquidity needs	0.10	0.00	0.03	(0.07)	(0.07)	(0.10)	0.08	(0.04)	(0.03)	0.05	(0.01)	(0.07)	(0.02)	(0.03)	(0.02)	0.09	0.22	0.21	0.10	**0.29**	1.00	
21	Risk aversion	0.34	0.02	(0.01)	0.01	0.01	0.00	0.06	0.02	0.02	0.00	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.01)	(0.02)	0.00	(0.03)	(0.05)	(0.01)	1.00

Table 9: Summary Statistics and Pair-wise Correlation Coefficients of Dummy Variables Based on Survey Responses

Note: This table shows the mean value of dummy variables based on survey responses and their pair-wise correlation coefficients. See the Online Appendix for more details about variable definitions. The bold fonts highlight correlation coefficients for survey responses that capture different aspects of the same mechanism.

Panel A: Volume-Weighted Past One-Month Count of Up-Limit Hits Based on Initial Buys

	Full sample (2018:01 – 2019:06)				Pre-survey (2018:01 – 2018:09)				Post-survey (2018:10 – 2019:06)			
Gambling preference, blockbusters	0.112***	0.109***			0.087***	0.086***			0.142***	0.139***		
	(3.875)	(3.768)			(3.640)	(3.608)			(3.660)	(3.573)		
Gambling preference, lotteries			0.038	0.019			0.025	0.018			0.051	0.029
			(1.257)	(0.653)			(1.013)	(0.727)			(1.237)	(0.698)
Male		-0.034		-0.033		-0.011		-0.01		-0.035		-0.034
		(-1.164)		(-1.140)		(-0.444)		(-0.403)		(-0.884)		(-0.866)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
$R2$	0.004	0.023	0.000	0.019	0.004	0.017	0.000	0.014	0.004	0.02	0.000	0.016
N	4,145	4,145	4,145	4,145	3,435	3,435	3,435	3,435	3,550	3,550	3,550	3,550

Panel B: Volume-Weighted Past One-Quarter Count of Up-Limit Hits Based on Initial Buys

	Full sample (2018:01 – 2019:06)				Pre-survey (2018:01 – 2018:09)				Post-survey (2018:10 – 2019:06)			
Gambling preference, blockbusters	0.209***	0.199***			0.174***	0.169***			0.256***	0.239***		
	(4.550)	(4.299)			(4.354)	(4.240)			(4.066)	(3.774)		
Gambling preference, lotteries			0.091*	0.055			0.103**	0.086**			0.071	0.024
			(1.897)	(1.144)			(2.389)	(1.994)			(1.107)	(0.373)
Male		-0.051		-0.049		-0.04		-0.039		-0.051		-0.05
		(-1.084)		(-1.051)		(-0.996)		(-0.949)		(-0.798)		(-0.784)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
$R2$	0.005	0.025	0.001	0.021	0.006	0.017	0.002	0.013	0.005	0.021	0.000	0.017
N	4,145	4,145	4,145	4,145	3,435	3,435	3,435	3,435	3,550	3,550	3,550	3,550

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 10: Validating Gambling Preferences Using Gambling Behavior

Note: This table studies the relationship between survey-based gambling preference and transaction-based gambling behavior. Gambling behavior is measured by the buy-volume (in RMB) weighted average of the past one-month (Panel A) or one-quarter (Panel B) number of up-limit hits based on the stocks an investor purchases in a given sample period. A purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase. Each panel presents OLS regression results based on three sample periods: full (January 2018 through June 2019), pre-survey (January 2018 through September 2018), and post-survey (October 2018 through June 2019). Gambling preference (blockbusters) equals one if an investor answers "Strongly agree" or "Agree" when asked if she aims to make a lot of money quickly through stock investment and zero otherwise. Gambling preference (lotteries) equals one if an investor answers "Strongly agree" or "Agree" when asked if she often thinks of stocks as lotteries and zero otherwise. See Table 6 for the exact phrasing of the survey questions. Control variables include age, gender, net worth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors and are reported in parentheses.

Panel A: Volume-Weighted Past One-Month Return Based on Initial Buys												
	Full sample (2018:01 – 2019:06)				Pre-survey (2018:01 – 2018:09)				Post-survey (2018:10 – 2019:06)			
Extrapolation, up	0.011**	0.011**			0.012***	0.013***			0.011*	0.011*		
	(2.170)	(2.134)			(2.689)	(2.902)			(1.668)	(1.704)		
Extrapolation, down			0.014***	0.013***			0.012***	0.012***			0.014**	0.014**
			(2.751)	(2.640)			(2.655)	(2.691)			-2.142	-2.142
Male		-0.014***		-0.014***		-0.012***		-0.012***		-0.014**		-0.014**
		(-2.854)		(-2.816)		(-2.740)		(-2.697)		(-2.284)		(-2.237)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
*R*2	0.001	0.017	0.002	0.018	0.002	0.016	0.002	0.016	0.001	0.017	0.001	0.017
N	4,142	4,142	4,142	4,142	3,432	3,432	3,432	3,432	3,550	3,550	3,550	3,550

Panel B: Volume-Weighted Past One-Quarter Return Based on Initial Buys												
	Full sample (2018:01 – 2019:06)				Pre-survey (2018:01 – 2018:09)				Post-survey (2018:10 – 2019:06)			
Extrapolation, up	0.020**	0.020**			0.019***	0.022***			0.026**	0.028***		
	(2.406)	(2.419)			(2.999)	(3.446)			(2.451)	(2.597)		
Extrapolation, down			0.021***	0.020**			0.020***	0.021***			0.021**	0.021**
			(2.615)	(2.532)			(3.112)	(3.316)			(2.032)	(2.091)
Male		-0.028***		-0.028***		-0.037***		-0.036***		-0.030***		-0.029***
		(-3.685)		(-3.638)		(-5.848)		(-5.801)		(-3.113)		(-3.031)
Controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
*R*2	0.001	0.023	0.002	0.023	0.003	0.033	0.003	0.033	0.002	0.021	0.001	0.02
N	4,136	4,136	4,136	4,136	3,428	3,428	3,428	3,428	3,544	3,544	3,544	3,544

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 11: Validating Extrapolative Belief Using Trend-Chasing Behavior

Note: This table studies the relationship between survey-based extrapolative beliefs and transaction-based trend-chasing behavior. Trend-chasing behavior is measured as the buy-volume (in RMB) weighted average of past one-month (Panel A) or one-quarter (Panel B) returns of stocks based on the stocks an investor purchases in a given sample period. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase. Each panel presents OLS regression results based on three sample periods: full (January 2018 through June 2019), pre-survey (January 2018 through September 2018), and post-survey (October 2018 through June 2019). Extrapolation-up (Extrapolation-down) equals one if an investor answers "Strongly agree" or "Agree" when asked if she believes stock price will rise (drop) even further in the future after it has risen (dropped) for a while. Otherwise, extrapolation-up (extrapolation-down) equals zero. See Table 5 for the exact phrasing of the survey questions. Control variables include age, gender, wealth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors are reported in parentheses.

Panel A: Summary Statistics							
	Min	P25	Median	P75	Max	Mean	Std Dev
Turnover	0.0%	4.7%	35.5%	109.8%	650.6%	84.8%	123.4%
Raw returns	-12.6%	-1.4%	0.0%	2.0%	10.0%	0.0%	3.5%
Net returns	-12.9%	-1.6%	0.0%	1.8%	9.6%	-0.2%	3.6%

Panel B: Correlation Matrix			
	Turnover	Raw returns	Net returns
Turnover	1		
Raw returns	-0.07***	1	
Net returns	-0.16***	0.99***	1

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 12: Summary Statistics of Turnover and Portfolio Returns

Note: Panel A shows the summary statistics of monthly turnover, raw return, and net return for investors in the main sample between October 2018 and June 2019. The main sample includes 4,671 survey respondents that: 1) can be identified in the Shenzhen Stock Exchange centralized database, and 2) hold at least one SZSE stock during the two-year window before the survey. Panel B shows the correlation coefficients among the three variables. See the Online Appendix for more details about variable definitions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actual performance in 2017	4.104***							
	(5.332)							
Overplacement, performance	15.695***							
	(2.760)							
Financial literacy, dummy		11.922***						
		(3.127)						
Overplacement, literacy		1.729						
		(0.400)						
Miscalibration			1.116					
			(0.289)					
Underestimation of trading costs				-3.549				
				(-0.980)				
Do not consider trading costs					-2.143			
					(-0.548)			
Do not think bid-ask spread is a cost						-15.135***		
						(-4.254)		
Extrapolation, up							4.379	
							(1.110)	
Extrapolation, down								3.810
								(1.005)
*R*2	0.007	0.002	0.000	0.000	0.000	0.004	0.000	0.000
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Gambling preference, blockbusters	10.924***							
	(2.878)							
Gambling preference, lotteries		2.750						
		(0.684)						
Realization utility, winners			7.188*					
			(1.874)					
Realization utility, losers				0.409				
				(0.093)				
Sensation seeking, novelty					10.184**			
					(2.270)			
Sensation seeking, volatility						11.984***		
						(2.885)		
Perceived information advantage							21.747***	
							(4.254)	
Dismissive of others' information								4.778
								(1.318)
*R*2	0.002	0.000	0.001	0.000	0.001	0.002	0.005	0.000
	(17)	(18)	(19)	(20)				
Social influence	-15.647***							
	(-3.317)							
Advisor influence		-16.469**						
		(-2.708)						
Portfolio rebalance needs			12.652**					
			(2.423)					
Liquidity needs				-9.974*				
				(-1.853)				
*R*2	0.002	0.001	0.001	0.001				

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 13: Univariate Regression Results on Turnover

Note: In this table, we run univariate cross-sectional regressions of each investor's turnover (%) on survey-based trading motives. *T*-statistics are based on robust standard errors and are reported in parentheses. See the Online Appendix for more details about variable definitions.

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual performance in 2017	4.198***	Gambling preference, blockbusters	11.764***
	(5.219)		(2.920)
Overplacement, performance	11.549**	Gambling preference, lotteries	-1.159
	(2.063)		(-0.263)
Financial literacy, dummy	7.065*	Sensation seeking, novelty	6.598
	(1.800)		(1.360)
Overplacement, literacy	-2.621	Sensation seeking, volatility	3.632
	(-0.625)		(0.824)
Miscalibration of uncertainty	-2.989	Perceived information advantage	15.660***
	(-0.764)		(2.988)
Do not consider trading costs	-3.989	Dismissive of others' information	2.942
	(-1.071)		(0.805)
Underestimation of trading costs	-4.029	Social influence	-7.839
	(-1.052)		(-1.616)
Do not think bid-ask spread is a cost	-9.456***	Advisor influence	-12.089*
	(-2.650)		(-1.943)
Extrapolation, up	-1.255	Portfolio rebalance needs	12.571**
	(-0.254)		(2.280)
Extrapolation, down	-1.208	Liquidity needs	-7.651
	(-0.262)		(-1.335)
Realization utility, winners	7.049*	Risk Aversion	-2.943
	(1.848)		(-0.692)
Realization utility, losers	-2.321	Expected 1-year market return	0.709*
	(-0.538)		(1.901)
Gender: male	21.488***	Controls	YES
	(6.124)	N	4,648
		R2	0.089

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 14: Regression Results Using the Full Set of Trading Motives

Note: In this table, we run a multivariate cross-sectional regression of each investor's turnover on all survey-based measures of trading motives. Control variables include age, gender, wealth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors and are reported in parentheses. See the Online Appendix for more details about variable definitions.

	Panel A: Monthly Turnover (October 2018 – June 2019)						Panel B: Monthly Raw Returns (October 2018 – June 2019)	
	P10	P25	P75	P90	Median	Mean	Median	Mean
1. Strongly disagree	0%	4%	99%	206%	25%	74%	0.19%	0.15%
2. Disagree	0%	3%	100%	222%	31%	77%	0.00%	0.04%
3. Neutral	0%	5%	112%	238%	33%	84%	0.01%	0.11%
4. Agree	0%	7%	117%	248%	42%	90%	0.03%	-0.04%
5. Strongly agree	0%	5%	119%	274%	42%	95%	0.00%	-0.20%
5 – 1	**0%**	**0%**	**20%**	**68%**	**17%**	**21%****	**-0.19%**	**-0.35%**
Annual transaction fee (5 – 1)	0.00%	0.00%	0.60%	1.96%	0.51%	0.63%		

	Panel C: Characteristics of Stocks Bought (October 2018 – June 2019)						
	Past 30-day # of Up-limit Hits	Past 30-day Return Volatility (%)	Past 30-day Return (%)	Size (Billion RMB)	Beta	B/M	Future 30-day Return (%)
1. Strongly disagree	0.60	3.25	9.71	43.73	0.93	0.62	-0.03
2. Disagree	0.75	3.39	11.58	35.21	0.96	0.62	-0.87
3. Neutral	0.83	3.49	11.94	26.92	0.99	0.61	-1.53
4. Agree	0.89	3.56	12.45	26.29	1.00	0.61	-1.36
5. Strongly agree	0.92	3.55	12.74	26.65	1.02	0.62	-1.77
5 − 1	**0.32*****	**0.30*****	**3.03****	**-17.08****	**0.09*****	**0.00**	**-1.74****

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 15: Additional Analysis of Gambling Preference, Blockbusters

Note: We sort investors into five groups based on their answers to the question, "Do you agree with the following statement? When I trade stocks, I often wish to select those stocks whose price would rise sharply in a short period time so that I can make a lot of money quickly." In Panel A (B), we tabulate the summary statistics of monthly turnover ratios (monthly raw returns) for investors in each group. In Panel C, we tabulate the equal-weighted average of various characteristics of stocks bought by investors in each group. In each panel, the last one or two rows report the differences between the bottom and top groups. When testing for the significance of the differences, we use robust standard errors.

	Panel A: Monthly Turnover (October 2018 – June 2019)						Panel B: Monthly Raw Returns (October 2018 – June 2019)	
	P10	P25	P75	P90	Median	Mean	Median	Mean
1. Never	0%	4%	102%	232%	30%	76%	0.10%	0.12%
2. Rarely	0%	3%	100%	218%	32%	76%	0.07%	0.06%
3. Sometimes	0%	5%	109%	244%	34%	86%	0.00%	0.08%
4. Often	0%	11%	139%	286%	46%	103%	0.00%	-0.13%
5. Always	0%	10%	139%	253%	44%	100%	0.00%	-0.01%
5 − 1	**0%**	**6%**	**37%**	**21%**	**14%****	**24%****	**-0.10%**	**-0.13%**
Annual transaction fee (5 − 1)	0.00%	0.18%	1.11%	0.63%	0.42%	0.72%		

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 16: Additional Analysis of Perceived Information Advantage

Note: We sort investors into five groups based on their answers to the question, "When you decide to trade a stock, how often do you believe that you know the stock better than others?". In Panel A (B), we tabulate the summary statistics of monthly turnover ratios (monthly raw returns) for investors in each group. In each panel, the last one or two rows report the differences between the bottom and top groups. When testing for the significance of the differences, we use robust standard errors.

Turnover Around the Survey (%)			
	1-month window	3-month window	6-month window
	(1)	(2)	(3)
After*Treated	0.672	-5.971	-4.417
	(0.119)	(-0.944)	(-0.675)
Treated	-0.219	4.153	0.583
	(-0.053)	(0.911)	(0.130)
After	-2.858	-1.012	16.144***
	(-0.956)	(-0.305)	(4.612)
Controls	YES	YES	YES
$R2$	0.056	0.058	0.056
N	6,628	6,628	6,628

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 17: Comparing Turnover Before and After the Survey for the Control and Treatment Groups

Note: Before distributing the survey, we randomly assigned 500 targeted branches of brokerage firms into treated and control groups. Investors in the two groups received questionnaires that were otherwise identical except for one difference: the questionnaire for the treated group included a "nudge" that highlighted the negative consequences of excessive trading. In this table, we study the effect of the "nudge" on investors' trading frequencies using difference-in-difference tests. The dependent variables from Columns (1) to (3) are investors' average monthly turnover rates in the one, three, and six months before and after the survey. The dummy, Treated, equals one if an investor is in the treated group and correctly answered the follow-up question designed to test if the respondent understands the content of the message. The dummy, Treated, equals zero if an investor is in the control group. The dummy, After, equals one for the periods after the survey month and zero for the periods before or in the survey month (September 2018). Control variables include age, gender, wealth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors and are reported in parentheses. See the Online Appendix for more details about variable definitions.

	Panel A: Monthly Turnover		Panel B: Characteristics of Stocks Bought						
	Mean	Median	Past 30-day # of Up-limit Hits	Past 30-day Return Volatility (%)	Past 30-day Return (%)	Size (Billion RMB)	Beta	B/M	Future 30-day Return (%)
1 (lowest)	60.37	29.43	0.70	3.30	10.65	36.46	0.94	0.66	-0.91
2	80.76	38.69	0.67	3.36	10.28	35.14	0.95	0.62	-0.91
3	71.91	29.49	0.80	3.41	11.18	29.79	0.99	0.61	-0.81
4	92.69	43.92	0.74	3.48	10.13	23.37	1.04	0.58	-0.88
5 (highest)	157.29	98.45	1.12	3.78	14.63	20.13	1.02	0.59	-2.02
5−1	**96.92*** **	**69.02*** **	**0.42*** **	**0.48*** **	**3.97*** **	**-16.34*** **	**0.09*** **	**-0.07*** **	**-1.11** **

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 18: Trading Characteristics for Investors Sorted on Transaction-Based Gambling Behavior

Note: We construct a measure for transaction-based gambling behavior for each investor in two steps. First, for each of the nine months prior to the survey (January 2018 through September 2018), we first calculate the past one-month count of up-limit hits of the stock for each buy transaction and then take the transaction value weighted average across all buy orders. Then, we take the time-series average value weighted by monthly buy values. We then sort investors into five groups according to transaction-based gambling behavior and compared their behaviors after the survey, from October 2018 through June 2019. In Panel A (B), we tabulate the summary statistics of monthly turnover ratios (characteristics of stocks bought) for investors in each group. In the last row of each panel, we report the differences between the bottom and top groups. When testing for the significance of the differences, standard errors are adjusted for heteroscedasticity.

Dependent Variable: Volume-Weighted Past One-Month Count of Up-Limit Hits Based on Initial Buys (January 2018 – September 2018)			
Actual performance in 2017	-0.009**	Gambling preference, blockbusters	0.071***
	(-2.533)		(3.598)
Overplacement, performance	0.002	Gambling preference, lotteries	-0.011
	(0.071)		(-0.482)
Financial literacy, dummy	-0.031	Sensation seeking, novelty	-0.032
	(-1.478)		(-1.518)
Overplacement, literacy	-0.014	Sensation seeking, volatility	0.022
	(-0.633)		(1.030)
Miscalibration of uncertainty	0.017	Perceived information advantage	0.049**
	(0.942)		(2.097)
Do not consider trading costs	0.040**	Dismissive of others' information	-0.001
	(2.221)		(-0.031)
Underestimation of trading costs	-0.005	Social influence	-0.005
	(-0.276)		(-0.178)
Do not think bid-ask spread is a cost	-0.043**	Advisor influence	0.025
	(-2.436)		(0.647)
Extrapolation, up	0.003	Portfolio rebalance needs	-0.039*
	(0.133)		(-1.741)
Extrapolation, down	-0.001	Liquidity needs	0.021
	(-0.045)		(0.679)
Realization utility, winners	0.015	Risk Aversion	0.004
	(0.843)		(0.205)
Realization utility, losers	0.009	Expected 1-year market return	0.000
	(0.409)		(0.266)
Gender: male	0.011	Controls	YES
	(0.623)	N	3,528
		R2	0.031

t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 19: Regressing Transaction-Based Gambling Behavior on Survey-Based Trading Motives

Note: In this table, we run multivariate cross-sectional regressions of each investor's transaction-based gambling behavior on survey-based measures of trading motives based. We construct a measure for transaction-based gambling behavior for each investor in two steps. First, for each of the nine months prior to the survey (January 2018 through September 2018), we first calculate the past one-month count of up-limit hits of the stock for each buy transaction and then take the transaction value weighted average across all buy orders. Second, we take the time-series average value weighted by monthly buy values. Control variables include age, gender, wealth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors and are reported in parentheses. See the Online Appendix for more details about variable definitions.