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

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

April 2020

Abstract

The behavioral finance literature has provided over a dozen explanations for the so-called excessive trading puzzle – retail investors trade a lot even though more trading 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 nation-wide survey among retail investors to elicit their 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 measures of trading motives. A horse race among survey-based trading motives suggests that overconfidence in having information advantage and gambling preference quantitatively dominate other explanations. Moreover, other popular arguments such as neglect of trading cost 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. We are grateful to Lawrence Jin, and seminar participants at Baruch, Chapman, CUHK Shenzhen, Imperial College, LSE, and Princeton for helpful feedback. 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 made significant advancement over the last few decades by bringing sharp insights from psychology to explain many anomalies in individuals' economic and financial decision makings.[1] A byproduct of such rapid development, however, is that we often face multiple behavioral biases – perhaps too many – for explaining each of these anomalies. For example, consider the excessive trading puzzle, which suggests that retail investors appear to be trading *too much*: they perform poorly relative to the market index before fees, transaction cost makes their performance even worse, and those who trade the most often perform the worst (Odean 1999; Barber and Odean 2000 and 2013). Motivated by these puzzling observations, the literature has proposed a number of behavioral explanations, e.g., overconfidence, realization utility, gambling preference, sensation seeking, social interaction, and low financial literacy,[2] beyond standard arguments such as portfolio rebalancing and liquidity needs. The large set of behavioral explanations we face is not satisfying: it is unlikely that all 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 so that we can eventually develop a unified conceptual framework based on a small number of biases to explain a wide range of individual behaviors.

This task of consolidation 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. In response to this challenge, the recent literature has turned to survey-based approaches by having investors self-examine and report the drivers of their trading and investment decisions, e.g., Greenwood and Shleifer (2014), Choi and Robertson (2019), and Chinco, Hartzmark and Sussman (2019). Survey-based approaches can quickly collect information on multiple explanations and therefore have the advantage of permitting horse races. However, there are also common concerns about the use of subjective survey data in economic analysis – that respondents may not report truth in their answers and that, even if they do, their

[1] See DellaVigna (2009) and Barberis (2018) for recent reviews of the literature.

[2] A more comprehensive review of the literature can be found in Table 1.

answers may not translate into real actions, e.g., Bertrand and Mullainathan (2001) and Cochrane (2011 and 2017).

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 faced 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 have a serious comparison among competing explanations for excessive trading. For certain explanations such as belief in having information advantage and influence of social interaction, it is inherently difficult to construct their empirical proxies from administrative data, but surveys allow researchers to collect 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 transactions at the individual level, we are able to directly verify that survey responses are largely consistent with the actual trading patterns they are designed to capture. This consistency provides further justification – not only to our analysis of the excessive trading puzzle but also to other studies that are based on surveys and experiments – for the use of surveys.

More specifically, we design and administer a nation-wide survey in China through the Investor Education Center at the Shenzhen Stock Exchange. Respondents are randomized across regions and brokers, and are incentivized with monetary rewards. The survey asks 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 and gathered responses from more than 10,000 investors.

To understand what drives the variation of trading volume across investors, we merge survey responses with account-level transaction data at the Shenzhen Stock Exchange. This gives rise to a unique advantage of our setting: we are able to link an investor's survey responses her actual trading behavior and examine their consistency. We provide four pieces of evidence to show that survey responses are consistent with actions: 1) survey-based measures of gambling preference predict the tendency to buy lottery-like stocks, 2) survey-based measures of extrapolation predict the tendency to buy stocks with prices that have recently gone up, 3) investors who are more risk averse according to the survey hold less volatile stocks, and 4) investors who report to hold higher expected returns about the stock market increase their stock holdings more.

After validating survey responses, we examine the explanatory power of survey-based trading motives for excessive trading. As a baseline, we first run cross-sectional regressions by regressing turnover on each trading motive in a univariate framework. This set of exercises confirms that some of the previous explanations for excessive trading also hold true in our setting. In the next set of exercises, we include all survey-based trading motives as regressors to examine their relative importance 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 to quantitatively dominate others: gambling preference and the belief of having information advantage. Their explanatory power is sizable: while the standard deviation of monthly turnover across all investors in our sample is 126%, gambling preference can explain up to 21% and the belief of having information advantage can explain up to 24%. They contribute to an annualized transaction fee of 0.6% and 0.7%, respectively, implying substantial investment consequences borne by retail investors who display either or both of the two trading motives.

Second, we provide further evidence in support of these two channels. In particular, we find that, survey-based gamblers trade smaller, high-beta, more volatile, and more positively skewed stocks. However, the stocks they buy subsequently perform worse, suggesting that gambling does not improve their performance. Similarly, we find that investors who *report* to have an information advantage do *not* deliver better performance in their trading. This suggests that their belief in having information advantage is unwarranted: they are *over*-confident 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 are significantly positive in univariate regressions, in the horse race their explanatory power is largely subsumed by other trading motives. In comparison, for both gambling preference and overconfidence in having information advantage, their coefficients are essentially unchanged across specifications. Therefore, by having an apples-to-apples comparison across a large set of behavioral biases, we can narrow down to the few that are the most important.

Fourth, in both the baseline regressions and the horse race, we report a number of "null" results. In our setting, low financial literacy, social interaction, and neglect of trading cost do not appear to contribute to excessive trading. Perhaps the most consistent, yet surprising set of results concerns neglect of trading cost. While we have constructed three different measures, none of them explains turnover with the "correct" sign: the coefficients are either insignificant or marginally significant with the opposite sign. 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 cost. The treatment group, however, do not exhibit any difference in turnover after the "nudge", further questioning the role of neglect of trading cost in driving excessive trading.

Our analysis above highlights how surveys could help consolidate the large set of behavioral explanations for excessive trading. However, for a given explanation, how does the survey-based measure compare to the corresponding transaction-based measure, provided that the latter can be constructed from administrative data? We study this comparison in the context of gambling preference. Following the approach used by Kumar (2009), we measure an investor's gambling *behavior*, a transaction-based measure of gambling preference, by examining the lottery-like features of the stocks she tends to buy. 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 associated with gambling *preference* but also correlated with alternative trading motives. This contrast nicely highlights the pros and cons of these two different approaches. On the one hand, when carefully designed, surveys can directly measure a specific trading motive, but, as discussed by Bertrand and Mullainathan (2001), they are subject to measurement noise at the individual level and are thus less powerful. On the other hand, although transaction-based measures can be more powerful in explaining the observed investor behavior, they may simultaneously capture multiple trading motives and are less reliable in isolating a particular economic mechanism. By combining these two methods together, our integrated approach offers a more powerful tool to consolidate the large number of behavioral biases and mechanisms offered by the behavioral economics and finance literatures.

As reviewed by Barber and Odean (2013), there is an extensive literature that analyzes the excessive trading puzzle from both theoretical and empirical sides. We will systematically introduce these mechanisms and the related studies in Section 1.3. Our paper differs from these prior studies in its scope and its approach. While most of the existent papers focus on one or two trading motives, we examine a large number of mechanisms at the same time, by directly measuring these motives through investors' own perspectives rather than indirectly inferring from administrative data. This horse race allows us to not only confirm or reject certain mechanisms but also to speak to each mechanism's relative importance.

Several studies have also combined survey data with administrative data to study the excessive trading puzzle, e.g., Dorn and Huberman (2005), Glaser and Weber (2007), and Dorn and Sengmueller (2009). Each of these studies elicits responses about one or two trading motives and then examines whether survey responses can explain the respondents' trading or portfolio choices.[3] In the absence of a horse race among different mechanisms, significant effects associated with survey response 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 similar theme as Giglio et al. (2019), which studies the relationship between portfolio decision and return expectations by combining survey expectations with mutual fund holdings data at Vanguard, and Epper et al. (2020), which uses a survey approach to measure individuals' time discount rate and examines its relationship with the individuals' 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 retail investors from the survey. In Section 2, we validate survey responses using actual trading behavior and compare survey-based trading motives in a horse race. In Section 3, we provide additional evidence on a selected number of trading motives.

[3] Specifically, Dorn and Huberman (2005) focus on risk aversion and perceived financial knowledge, Glaser and Weber (2007) examine two forms of overconfidence, over-placement and miscalibration, and Dorn and Sengmueller (2009) aim at sensation seeking.

In Section 4, we discuss the pros and cons of survey-based and transaction-based measures. We conclude in Section 5.

1. The Survey

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

1.1. Survey Design

The survey is designed to test and differentiate among a large set of trading motives that provide theoretical foundations for many existing theories of trading volume. A summary of all the trading motives we consider can be found in Table 1. For each motive, we phrase the corresponding question(s) to map as closely as possible to the underlying concept, and we do so often by going back to the original paper that proposes the particular motive. A trading motive may take several forms of representation. For instance, overconfidence comes in at least two forms: *over-placement*, i.e., people have overly rosy views of their abilities relative to other people, and *miscalibration of uncertainty*, i.e., people are too confident in the accuracy of their beliefs. In such cases, we include at least one question for each form. A detailed description of how we design each question can be found in the Appendix.

Ideally, for each trading motive, we would like to design the corresponding survey question(s) to capture all aspects of the motive, but we are also concerned that a long and complex survey may confuse respondents or discourage them from answering the questions truthfully. To ensure the quality of survey responses, we design all questions to be multiple-choice so that respondents do not have to fill in an answer themselves. We include two types of qualitative questions. The first ("agreement") type asks respondents whether they agree or disagree with a statement that describes a particular trading motive. Answer options include: "strongly agree", "agree", "neutral", "disagree", "strongly disagree", "do not know", and "decline to answer". The second ("frequency") type asks respondents how often they consider a particular motive when they trade. Answer options include: "always", "often", "sometimes", "rarely", "never", "do not know", and "decline to answer". We also hope to obtain quantitative answers for certain trading motives (e.g., estimates of transaction fees to measure neglect of trading cost). In such cases, we provide several options each covering a specific range of value.

Inherent in any horse race is the concern that some variables are better measured than others and differences in coefficients simply reflect differences in measurement errors. To address this concern *ex-ante*, we follow a jargon-free protocol and standardize the phrasing of all the questions and answer options throughout the survey. To make sure that the survey is comprehensible to even the most naïve subset of investors, we run a series of pilot tests among the general population and solicit their feedback on the survey design. The overwhelming majority of the respondents find the survey easy to understand. To address this concern *ex-post*, we first validate these survey responses with actual trading behavior. In the analysis of the excessive trading puzzle, we further encode all survey-based trading motives into dummy variables. These efforts should help minimize the variation of measurement errors across the survey-based trading motives and better facilitate an apples-to-apples comparison.

It is worth noting that, while we ask the 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 that 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). In their survey, they ask correspondents how well a theory describes the way they make decisions on, for instance, what fraction of their portfolio to invest in equities and whether to own any stocks in their portfolios. In other words, they ask investors themselves to evaluate and compare the relevance of different theories in describing their decision making. In contrast, we do not delegate this task to the respondents but keep it to ourselves – later by regressing individual-level turnover on a variety of different trading motives. This is made possible by the fact that we are able to trace a respondent's survey responses to her actual trading behavior.

Moreover, our empirical strategy addresses a number of methodological concerns raised by Bertrand and Mullainathan (2001) about the use of subjective survey data in economic analysis. They argue that survey responses are noisy at the individual level due to various factors – e.g., white noise, phrasing and ordering of the questions, and cognitive dissonance – which can significantly contaminate the inference process. For instance, differences in responses across *time* to the same question may capture time fixed effects (e.g., overall market sentiment), whereas differences in responses across *questions* may be attributed to the phrasing and ordering of the survey questions. As a result, they conclude that changes in survey responses "do not appear useful in explaining changes in behavior" and recommend that survey responses are useful as explanatory variables for "explaining differences in behavior across individuals". This is precisely the approach we take in this paper. We return to issues related to measurement errors in Section 4.

The survey contains four main parts. The first part contains eight questions measuring financial literacy. These questions include the classic "big three" questions, e.g., Lusardi and Mitchell (2007, 2011), as well as several other widely used questions to measure financial literacy. At the end of this section, we also ask respondents to self-assess how many questions they have answered correctly. This allows us to construct a measure for overconfidence based on financial literacy. The second part represents the core of the survey, where we ask 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 asks about their basic demographic characteristics, including name, gender, date of birth, province, city, education, income, wealth, 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 locate each correspondent in the transaction database. Finally, for a randomly selected group of respondents (the treatment group), we also include 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 operation, the Investor Education Center surveys domestic retail investors on an annual basis 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, a size that provides sufficient statistical power while remains 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 a total of two weeks to complete the survey.[4] A valid response must be completed within 30 minutes. Respondents could open the survey using their personal computers or on their smartphones.[5] We collected an initial sample of 12,856 respondents. Table 2 reports the distribution of respondents across brokers and provinces. As designed, the respondents are 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 have a college or higher degree. Respondents are primarily middle-aged: almost half of them are between 30 to 50. They are also quite wealthy: the median annual income is around 200,000 RMB and the median household wealth is around 500,000 RMB, both far exceeding the national median. Overall, our sample represents a relatively sophisticated, 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

[4] The distribution of the survey proceeded in the following way. The SZSE center first distributed the link to the survey to each broker's parent office. After receiving the link, the parent office then distributed it to the pre-selected branches, where the local client manager then redistributed the survey to their clients (investors). While we do not observe direct conversations between client managers and investors, we suspect much of the communication happened via phone calls and WeChat messages. Once an investor had completed the survey, the client manager recorded down her name, phone number, and the name of the branch. This information was then sent back to us for verification purposes.

[5] To boost response rate, we put up the logos of both SZSE and 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 have completed the survey, 20 would be randomly selected to receive a gift card worth 500 RMB (around 80 USD) and 1000 to receive a gift card worth 50 RMB (around 8 USD).

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 takes a median investor about 8 minutes to complete the survey, and 95% of respondents finish within 20 minutes. However, we find that respondents who spend less than 3 minutes on the survey experience a sharp drop in their financial literacy score, suggesting that they may have shirked. In subsequent analysis, we dropped these observations, which reduces 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.[6] Panel A shows that, out of all the eight questions, seven of them 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 have correctly answered at least six questions. In fact, one-third of them are correct on all eight questions. Panel B shows the distribution of self-assessed scores, which is similar to that of the actual scores. Therefore, investors in our sample display a high level of financial literacy.[7]

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 has also suggested that overconfidence may

[6] 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).

[7] Lusardi and Mitchell (2014) show that among eight countries including Germany, Netherlands, and U.S., 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 is among investors already participating in the stock market.

adopt several closely related, albeit distinct, forms: over-placement of ability, miscalibration of uncertainty, and over-precision of information. We have designed questions to capture each of these forms.

Over-placement 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. A similar measure is also used by Dorn and Huberman (2005) and Barber et al. (2019) to measure perceived financial knowledge. In Table 5, Panel A reports the summary statistics for both measures. In constructing *over-placement* 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. Since we have not yet merged survey responses with transaction data, 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, called *over-placement* of *literacy*. Overall, respondents do *not* overestimate their level of financial sophistication. 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). 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 use the surveyed confidence interval to measure the executives' overconfidence. 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 where we ask investors to estimate how much the stock market will go up (down) with 10% probability within the next year; the difference between these two estimates gives the 80% confidence interval. As reported by Panel A of Table 5, while a rational benchmark 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 over-precision about one's own information. We will describe this measure slightly later when we discuss information related questions.

Extrapolation

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

Barber and Odean (2000) and Barber, Lee, Liu and Odean (2009) show that trading causes retail investors in the U.S. and Taiwan to underperform relative to the overall market and more than 60% of their under-performance is directly due to commissions and transaction taxes. While overconfidence and other behavioral biases may cause investors to trade despite the trading cost, these findings also suggest the possibility that those investors who trade a lot may have neglected the various fees and taxes associated with trading. There are at least two possible sources for neglect of trading cost. The first one is simply due to underestimation – investors systematically underestimate the fee rate due to their lack of financial sophistication. The second one is due to (lack of) "salience" (Bordalo et al. 2012): even if investors do have the full knowledge about trading cost, it still matters very little to their trading because the amount associated with each single transaction is small and negligible.[8]

To capture these two forms of neglect of trading cost, we have constructed three different measures. Panel C of Table 5 reports the summary statistics. First, we directly ask investors to estimate the total transaction cost associated with a round-trip buy and sell at 10,000 RMB. The

[8] 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).

results show that respondents significantly underestimate trading cost: 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 broker, almost 70% of the respondents report an estimate below the lower bound. The second question asks how often an investor considers transaction cost when she trades stocks. Similarly, more than half of the respondents say that they never or rarely do so. The third question targets 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 agree while 23% disagree. Overall, there is strong evidence that retail investors in China underestimate or neglect trading cost.

If neglect of trading cost is due to (the lack of) "salience", then presenting transaction cost in a more salient manner or more frequently reminding investors of the cost may lead investors to trade less. To test this hypothesis, we give a random half of respondents a "nudge" and compare their turnover to other investors before and after the survey. For the treated group, we increase the salience of trading cost by presenting it in annualized terms and reminding them 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 lead investors to have a preference for gambling stocks, i.e., 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. Kumar (2009) and Boyer, Mitton, and Vorkink (2010) provide empirical evidence that supports the presence of such gambling preference. These existing studies tend to focus on the implication of gambling preference for stock selection. To the extent that gambling stocks change over time due to fluctuations of volatility and return distribution in the tails, 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 asks whether the respondent aims to select the few blockbusters stocks so that he or she could get rich quickly. This question deliberately tones down the fact that picking a blockbuster is a small probability event. In contrast, the second question contains a more objective description by asking whether the respondent views trading stocks as buying lotteries in that they are willing to exchange small losses for the small probability of a big gain. These two questions reveal not only the respondent's gambling preference but also her assessment of tail probability. According to the cumulative prospect theory, investors may over-weight small probability events such as choosing future blockbusters. Thus, loosely speaking, the first question identifies "behavioral" gamblers – those who overweight the small probability of a big gain, while the second question identifies "rational" gamblers – those who recognize that a big gain is a rare event. Overall, for each question, about one third of the respondents agree or strongly agree with that statement. In what follows, we differentiate these two questions by labeling the first one as gambling preference *with* probability weighting and the second one as *without* probability weighting.

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. In order 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. Frydman et al. (2014) provide evidence from neural data to support the relevance of realization utility in financial decision making.

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 make her more painful. According to realization utility, selling winners is more pleasing than holding on to them while selling losers is more painful. Survey responses for the two questions are mixed. In the first question, consistent with realization utility, more respondents say selling winners makes them happier. In the second question, however, more respondents report that holding on to losers is more painful. In what follows, we differentiate these two questions by labeling the first one as realization utility for *winners* and the second one as 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 trading of retail investors. 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 correspondents in general do not exhibit a strong tendency of 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, e.g., Milgrom and Stokey (1982), 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. 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 genuine information advantage or noise traders who believe they hold superior information even though they do not? We have included two questions in the survey to elicit a respondent's perception of her information. The first question measures one's belief in having information advantage by asking how often they believe they know stocks better than other investors. A positive response to this question may be associated with genuine information advantage, but it could also reflect 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 genuine information advantage from perceived information advantage by examining whether the respondent actually delivers better trading performance.

The second question measures one's fear or alert of potential adverse selection concerns by asking how often they worry that others know stocks better than themselves do. 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' successes or failures in investing. As a result, investors' trading behavior would be influenced by social movements. 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.

We have 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 correspondents say portfolio rebalancing often 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 that we use to measure risk aversion. Following Lusardi and Mitchell (2011), we elicit investors' risk attitude by asking if 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 are unwilling to take the job with the slightest risk, 26% of the investors are 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 translate the original survey into English with slight modifications (tailored to American investors) and run the survey on Amazon MTurk among a small sample of 400 U.S. retail investors. On the one hand, we find that U.S. investors care more about trading cost, 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, 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 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 ask 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 left in our sample, we are able to uniquely identify 6,013 investors.[9] Our transaction data sample covers from January 2018 to June 2019, which nicely splits around the time of our survey in September 2018. We further require an investor to hold at least one stock in the Shenzhen Stock Exchange during the two-year window before the survey.[10] This further reduces the sample size to 4,671 – 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 with 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 a larger account, a slightly lower turnover rate, and better performance.

To make different trading motives comparable with each other, we encode all the measures of trading motives into dummy variables. A detailed description about the construction of these dummy variables can be found in the 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; and for quantitative questions, we typically use zero as the cut-off value. Table 9 reports the summary statistics of these dummy variables and their pairwise correlations. It is worth noting that for the multiple questions targeted at the same trading motive, their pairwise correlation is generally high, which suggests that their responses are internally consistent.

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

[10] An investor may hold non-stock position in the sample due to various reasons: they could be holding mutual funds or ETFs, or they could be holding stocks trading at the Shanghai Stock Exchange, etc.

2.2. Validating Survey Responses

There are several widely held concerns about the use of survey response in testing economic hypotheses. First, respondents may not take the survey seriously and 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 existent papers are limited to the use of either survey data or transaction data, a *systematic* test of the external validity of survey responses of investors is still missing from the literature.[11]

Ideally, we would like to validate responses to all of the questions in the survey, but this is neither efficient nor plausible. For instance, while the survey has several questions on 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 an empirical counterpart that can be 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, they also span a wide range of trading motives – belief formation, preferences, and return expectations. For brevity, in the main part of this paper, we are primarily concerned with gambling preference and extrapolative beliefs. We briefly discuss other results and include more details in the 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 to measure gambling behavior based on return skewness, the literature, e.g., 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

[11] 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. (2020) 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.

with gambling preference. This argument is particularly compelling as it is well connected with our earlier discussion that gambling preference is originated from an investor's overweighting of tail outcomes, e.g., Barberis and Huang (2008). Realizations of extreme returns would likely stimulate an investor with gambling preference to extrapolate extreme returns into the future.

Motivated by this argument, we take advantage of a unique regulation in the Chinese stock market: the daily price limits rule. This rule imposes that daily stock returns of individual stocks cannot exceed 10%, and we use the total count of up-limit hits (i.e., the number of days with prices hitting the upper-price limit) in a preceding period to proxy for a stock's positive return skewness. As hitting the daily upper-price 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 include two survey questions of gambling preference, one without reminding the respondent that large stock returns have small probabilities and thus capturing gambling preference *with* probability weighting, while the other specifically reminding her so and thus capturing gambling preference *without* probability weighting. Interestingly, we find that survey responses to the first question have a significant, positive correlation with gambling behavior in transaction data. In other words, those who *report* to have gambling preference (with probability weighting) exhibit stronger gambling *behavior*. 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. In contrast, the relationship between gambling preference (without probability weighting) and gambling behavior is much weaker, suggesting that gamblers are precisely those who incorrectly assess the tail probabilities of large stock returns.

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 to have 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. More details about these exercises can be found in the Appendix.

Finally, we note that, throughout the validation exercises, while 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 (with probability weighting) remains around 4, but the R-squared is consistently below 1%. This suggests that, while survey responses are in aggregate consistent with behavior, much of the variation in trading behavior left unexplained. This could be due to measurement errors or white noise in survey responses, or a result of other factors driving trading behavior at the same time. We will further discuss this important issue later 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.[12] Table 12 reports the summary statistics of their monthly turnover and portfolio returns in the post-survey sample from October 2018 to June 2019, a 9-month window after the survey. When needed, however, we also extend the window to cover the 9 months before the survey, spanning our full sample from January 2018 to June 2019.

Table 12 shows that excessive trading is pronounced among Chinese retail investors. First, they trade a lot: the median monthly turnover rate in our sample is almost one, suggesting that they fully reshuffle their portfolios almost once every month. Second, their performance is poor: while the monthly return of the Shenzhen Composite Index is about 0.6% from October 2018 to June 2019, the median net return in our sample is only 0.1%. 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 – over-placement of performance, over-placement 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 over-placement of performance: in Column (1), conditional on having the same past performance, investors who self-report to have 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, further improving investors' financial literacy, a policy often advocated in emerging economies such as China, may not be effective in reducing their excessive trading. Furthermore, Column (2) shows that over-placement of literacy does not predict future turnover. In Column (3), miscalibration

[12] 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 over-placement of performance – and to trade more.

does not significantly predict future turnover. This set of results is broadly consistent with Glaser and Weber (2007), who find that over-placement predicts more trading, but miscalibration does not.

Columns (4) to (6) report the results on neglect of trading cost. Surprisingly, for all the three measures we have constructed, none of them significantly predicts future turnover with the correct sign: in Columns (4) and (5), the coefficients are close to zero and insignificant; in 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 cost.[13] The fact that we cannot find any supporting evidence despite having constructed three measures for neglect of trading cost gives us pause about its role in explaining investor trading. We will come back to this issue with more analysis in Section 3.3.

Columns (7) to (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 with the market going up by just a few percentage points. Another possibility is that extrapolation alone cannot explain volume and needs to be combined with some additional forces to generate a trading frenzy (Barberis et al. 2018; 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 overweight small probability trade significantly more. In contrast, those who acknowledge that stocks are like lotteries and picking the next blockbuster is a small probability event do *not* trade more. This contrast suggests that investors' assessment of tail probability plays a key role in explaining excessive trading. Also note that this result is also consistent with the

[13] Transaction fees are rather 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 cost. In addition, we control for difference in fees across brokers with branch fixed effect.

patterns in Table 10, where only investors who overweight small probability tend to buy lottery-like stocks; those who acknowledge stocks are like lotteries do not have such a tendency.

Columns (11) and (12) report the results on realization utility and shows an 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 painful in selling losers – does not predict future turnover. This pattern is consistent with the implications of realization utility (Barberis and Xiong 2012), as investors with realization utility are more willing to let go of stocks once they exceed the purchase prices 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 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 dismissive 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,[14] 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 arguing that social interaction contributes to the spread of investor sentiment and excessive trading. Columns (19) and (20) show

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

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.

To sum up, Table 13 confirms several of the previous explanations for trading volume: over-placement of performance, gambling preference, sensation seeking, and perceived information advantage. Which of these explanations are most relevant? Are some of the explanations subsumed by others? Addressing these issues requires putting all of them in a horse race, which we pursue below. Table 13 also highlights a number of "null" results that cast doubt on several prominent explanations of excessive trading: lack of financial literacy, neglect of trading cost, 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 respective 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 survey-based trading motives, we also include 1) basic demographic characteristics such as gender, income, wealth, and education, 2) return expectations to control differences in optimism and pessimism, and 3) recent performance to control for "mood".[15] Compared to Table 13, Table 14 reveals a number of notable observations.

First, 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 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. This contrast nicely highlights the advantage of our setting that allows for direct comparison across different mechanisms.

Second, two trading motives that stand out in the horse race: gambling preference (with probability weighting) and overconfidence in the form of perceived information advantage. Both

---

[15] 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 Appendix.

coefficients are quantitatively large and significant at 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 interestingly, our finding highlights that a particular form of overconfidence – through perceived information advantage – rather than other forms, such as over-placement of performance or 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 associate gambling preference as an important mechanism for understanding investor demand for lottery-like stocks. Our finding suggests that gambling preference may also lead investors to trade lottery-like stocks with the fluctuations of volatility and tail distribution of individual stocks. We will present additional evidence to support these two highlighted trading motives as key drivers of excessive trading in Section 3.

It is worth nothing that, in Table 5, 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 two independent traits that contribute to trading volume through two distinct channels. Therefore, investors who are subject to both biases are particularly prone to excessive trading.

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 it for future research to explore.

2.5. Robustness and Subsample Analysis.

As robustness checks, we repeat the same regression specification used in Table 14 and report the results in the Appendix. These alternative specifications include: 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 9-month window before the survey, as opposed to the 9-month window after the survey. Throughout all these specifications, gambling preference and perceived information advantage remain the most powerful drivers for excessive trading.

We also perform two sets of subsample analysis and report these results in the 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 one, we split the full sample based on the fraction of wealth invested in the stock market. In both subsamples, the coefficients for gambling preference and perceived information advantage remain similar in magnitude and statistically significant. However, among the subsample of investors who are more invested in the stock market, liquidity consideration becomes a more pronounced factor in explaining their turnover.

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 Appendix, that the explanatory power of each motive remains very stable during the 18-month window around the survey, suggesting relatively persistent importance in time-series. Second, and relatedly, it is also possible that some retail investors learn to de-bias themselves from past mistakes and the importance 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, building on the results in Section 2, we conduct additional analysis to further reinforce the highlighted trading motives. 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 cost.

3.1. Gambling Preference

We start by discussing the magnitude of the explanatory power of gambling preference for turnover. So far, we have coded the survey responses into dummy variables, 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 "gambling preference" question. While this single-sorting approach ignores the correlations with gambling preference with other trading motives, it provides a more granular look at the explanatory power of gambling preference. 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.

Panel A shows the distribution of turnover for each of the five groups. There is a nice, monotonically increasing pattern across the five groups that differ in the extent the investors agree with the gambling preference. 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 distribution, indicating that this pattern is not driven by some 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 -0.4% lower monthly returns than the "strongly disagree", albeit the difference is not statistically significant. The lack of superior performance and the large transaction cost together suggest their trading is excessive.

Finally, we examine the characteristics of stocks purchased by the five groups of investors in Panel C. Investors with survey-based gambling preference tend to buy stocks with positive skewness, larger counts of daily up-limit hits, higher past volatility and past returns, and smaller size, and larger market beta. These stocks also perform worse subsequently, confirming that investors with 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 on 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 "always" and "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 to have 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 Cost

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

Among all of 500 brokerage branches we distributed the survey to, we randomly selected 250 branches to include an additional "nudge". The "nudge" asks the respondent to read a short article that highlights the negative consequences of excessive trading. As shown in Figure 2, the article contains a detailed calculation of how much investors lose from frequent trading, together with a quote from Warren Buffett advising investors to buy and hold. Instead of presenting trading cost as a fraction of total transaction value, we make it more salience by presenting the annualized fee rate for a frequent trader. We also include a "validation" question after the article by asking the respondent to calculate the total trading cost of a given level of turnover. The answer to this question helps to filter out 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 rate one month after the survey. We repeat this exercise in Columns (2) and (3) by expanding the window to 3 months and 6 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 find similar results among a subsample of investors who are identified as treated according to their answers to the "validation" question.

Taken together, our analysis suggests that investors in our sample engage in excessive trading despite their awareness of the substantial cost incurred by trading. This finding has important policy implications. Policy makers across the world, including China's stock market regulator, the China Securities Regulatory Commission (CSRC), often consider Tobin taxes as a policy tool to curb speculative trading in stock markets. To the extent that trading cost is not a key driver of excessive trading, our finding casts doubt on the effectiveness of Tobin taxes.[16]

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

[16] There is rather mixed evidence of 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 stamp tax for stock trading in China.

measures of trading motives. Second, survey responses allow researchers to measure a large set of trading motives at the same time, including those that are hard to measure from administrative data. There are also many 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 preference is so powerful, why don't we use it directly instead of relying on the survey-based measure? To address this question, we regress the 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.[17]

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 retail investors trade so much. The survey is designed to capture an exhaustive list of trading motives that are prevalent in the literature, and in doing so, we are able to offer serious comparison across a large set of explanations for trading volume. The key innovation in our approach is to combine survey data and transaction data, allowing us to not only validate survey responses, but also to offer a comparison between 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 information advantage) and gambling preference quantitatively dominate other trading motives in explaining frequent trading. Third, popular arguments such as neglect of trading cost, low financial literacy, and social interaction do not contribute to excessive trading. Finally, by discussing the pros and cons of survey-based and transaction-based approaches, we argue that our integrated approach can address the concerns each particular approach faces.

[17] 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, it is the attention to the stock market that explains their frequent trading.

References

Anagol, Santosh, Vimal Balasubramaniam, and Tarun Ramadorai, 2019, Learning from Noise: Evidence from India's IPO lotteries, *Working paper*.

Alpert, M., & Raiffa, H. (1982). A Progress Report on the Training of Probability Assessors.

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, 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.

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.

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.

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.

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.

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

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

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.

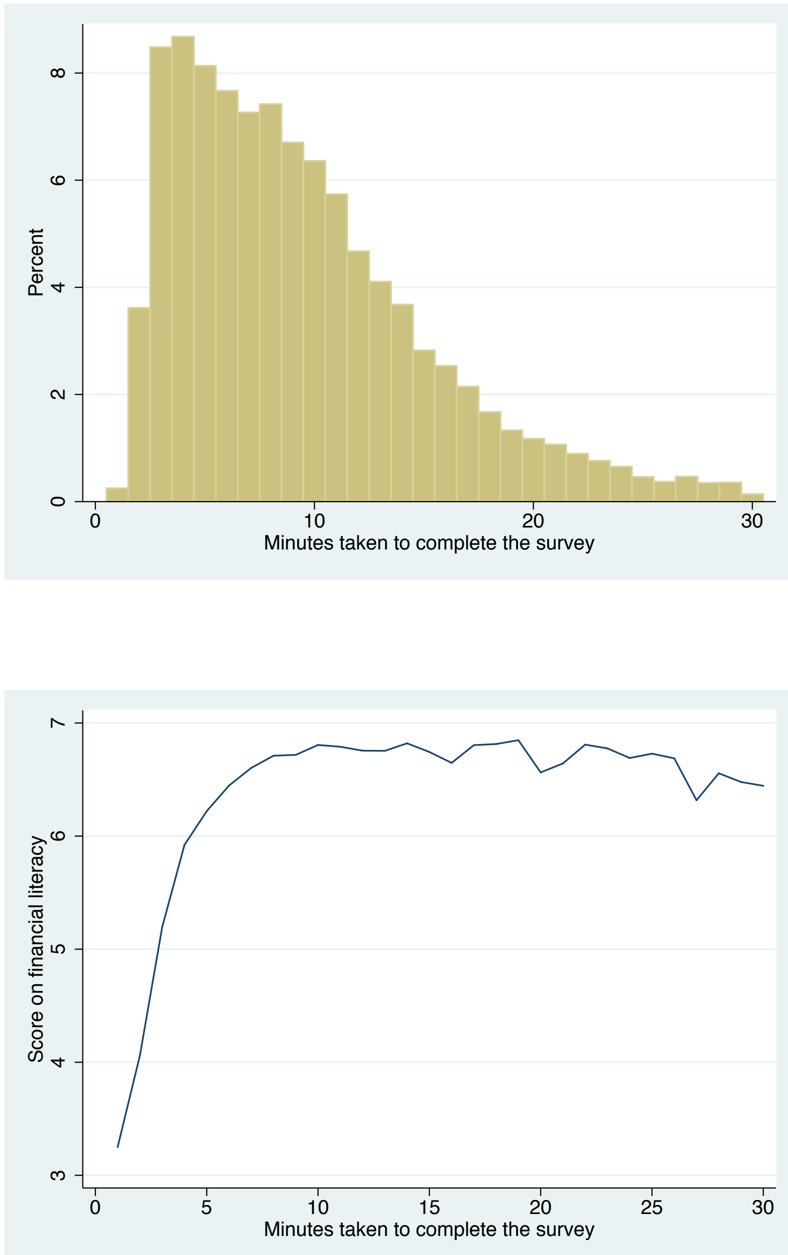

Figure 1: Relationship Between Financial Literacy Score and Minutes Taken to Complete the Survey

Figure 2: The Treatment that Nudges Investors to Reduce Trading due to Transaction Cost

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. over-placement 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 cost	1. transaction fees 2. bid-ask spread	Barber and Odean (2000), Barber, Lee, Liu and Odean (2009)
Gambling preferences	1. overweight small probability (behavioral) 2. understand small probability (rational)	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 for 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 literacy	1. numeracy; 2. inflation; 3. diversification; 4. assets' risk; 5. stock; 6. bond; 7. PE ratio; 8. mutual fund	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	11.80%
CITIC Securities	1,410	11.00%
Haitong Securities	1,390	10.80%
China Merchants Securities	1,372	10.70%
Huatai Securities	1,350	10.50%
Guosen Securities	1,252	9.80%
China Securities	1,203	9.40%
Shenwan Hongyuan Securities	1,169	9.10%
GF Securities	1,111	8.70%
China Galaxy Securities	1,051	8.20%
Panel B: By Province/Region		
Guangdong	1,674	13.10%
Zhejiang	1,201	9.40%
Jiangsu	1,138	8.90%
Shanghai	1,135	8.90%
Hubei	629	4.90%
Beijing	622	4.90%
Fujian	600	4.70%
Hunan	572	4.50%
Shandong	542	4.20%
Henan	531	4.10%
Sichuan	530	4.10%
Anhui	463	3.60%
Jiangxi	388	3.00%
Hebei	385	3.00%
Liaoning	331	2.60%
Chongqing	284	2.20%
Heilongjiang	250	2.00%
Guangxi	230	1.80%
Shanxi	222	1.70%
Shaanxi	198	1.50%
Others	931	7.20%
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	54.00%	71.70%	< 20K	3.80%
Female	46.00%	28.30%	20K to 100K	17.20%
			100K to 200K	29.50%
Education			200K to 500K	29.50%
Middle School or below	8.60%	7.30%	500K to 1M	12.60%
High School	15.60%	24.70%	1M to 2M	4.20%
Professional School	21.90%	26.00%	2M to 10M	2.10%
College	44.90%	23.60%	10M and above	1.20%
Graduate school and above	9.20%	3.40%		
Age			Wealth (RMB)	
20 to 30	27.80%	21.30%	< 20K	4.80%
30 to 40	29.10%	27.40%	20K to 100K	12.30%
40 to 50	19.90%	24.50%	100K to 500K	27.50%
50 to 60	14.80%	15.10%	500K to 1M	22.30%
>60	8.50%	11.70%	1M to 2M	21.90%
			2M to 10M	6.50%
			10M and above	4.80%

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 wealth 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 score

Score	Actual	Self-assessed
0	0.40%	0.60%
1	0.70%	0.70%
2	1.70%	1.80%
3	2.30%	4.60%
4	5.10%	6.90%
5	8.90%	13.00%
6	17.90%	16.20%
7	30.10%	17.70%
8	33.00%	32.70%
N/A	0.00%	5.80%

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 believes 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
	11.80%	13.80%	15.80%	13.50%	12.40%	10.40%	5.80%	3.80%	2.20%	3.40%	7.20%
2. Actual score–Self-assessed score	<-4	-4	-3	-2	-1	0	1	2	3	4	>4
	0.80%	1.80%	5.40%	11.40%	19.70%	35.10%	17.70%	5.60%	1.70%	0.60%	0.40%
3. Upside return–Downside return	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	>50%
	32.70%	14.90%	9.20%	6.90%	5.20%	5.20%	4.30%	3.40%	3.10%	2.50%	12.70%

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
	4.80%	26.90%	39.30%	22.80%	1.30%	5.00%
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
	4.40%	29.10%	41.90%	18.20%	1.30%	5.30%

Panel C: Neglect of trading cost								
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
	17.30%	27.70%	23.60%	12.80%	8.40%	3.70%	2.10%	5.50%
2. How often do you consider transaction cost when you trade?	Never	Rarely	Sometimes	Often	Always	N/A		
	14.60%	37.70%	27.00%	13.80%	4.60%	2.50%		
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			
	59.80%	23.10%	8.50%	7.20%	1.40%			

Table 5: Survey Responses on Questions on Beliefs

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

Panel A: Gambling preference						
Gambling preference, with probability weighting 1. When I trade stocks, I aim to select those stocks whose price would rise sharply in a short period time so that I can make a lot of money quickly.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	10.40%	25.40%	33.90%	23.00%	4.60%	2.70%
Gambling preference, without probability weighting 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
	5.50%	24.90%	27.20%	32.50%	7.30%	2.70%
Panel B: Realization utility						
Realization utility, winner 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	
	37.20%	23.70%	25.30%	9.20%	4.50%	
Realization utility, loser 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	
	22.90%	28.00%	32.10%	12.20%	4.80%	
Panel C: Sensation seeking						
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
	5.90%	20.30%	43.90%	21.00%	3.20%	5.70%
Sensation seeking, volatility 2. I feel excited about the stock market moving up and down.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
	5.40%	23.40%	36.70%	26.20%	4.30%	4.10%

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
	8.70%	27.90%	40.30%	14.50%	3.20%	5.40%
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
	18.20%	28.90%	32.30%	12.60%	2.50%	5.60%
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
	11.60%	31.20%	40.00%	11.80%	1.70%	3.80%
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
	17.80%	35.00%	35.80%	7.20%	1.20%	3.10%
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
	9.60%	30.50%	44.50%	9.50%	1.70%	4.20%
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
	7.00%	25.90%	45.00%	14.40%	2.60%	5.10%
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	
		51.60%	34.10%	11.30%	3.00%	
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	
		45.30%	37.50%	13.80%	3.40%	
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	
		26.00%	57.40%	13.20%	3.50%	

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).

	Main Sample	Population
Gender		
Male	54.40%	71.70%
Female	45.60%	28.30%
Education		
Middle School or blow	5.10%	7.30%
High School	17.60%	24.70%
Professional School	24.40%	26.00%
College	38.50%	23.60%
Graduate school and above	6.10%	3.40%
Others	8.40%	14.80%
Age		
< 30	26.10%	21.30%
30 to 40	27.40%	27.40%
40 to 50	22.40%	24.50%
50 to 60	16.00%	15.10%
> 60	8.10%	11.70%
Investment age (in years)		
< 2	21.20%	10.00%
2-6	26.20%	29.80%
6-10	17.40%	18.00%
> 10	35.10%	42.20%
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 the main sample and the investor population. The main sample include 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 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	Over-placement, performance	0.67	1.00																				
2	Over-placement, literacy	0.24	0.03	1.00																			
3	Miscalibration	0.69	0.08	0.02	1.00																		
4	Underestimation of transaction cost	0.69	(0.02)	0.02	0.00	1.00																	
5	Do not consider transaction cost	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, with prob. weighting	0.37	(0.01)	0.04	(0.01)	(0.02)	0.05	(0.09)	0.25	0.21	1.00												
10	Gambling preference, without prob. weighting	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, winner	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, loser	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	Affected by family and friends	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	Affected by investment advisors	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	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	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 of Dummy Variables Based on Survey Responses

Note: This table shows the mean value (Column 3) of dummy variables based on survey responses and their pair-wise correlation coefficients (Columns 4 to 24). See the Appendix for more details about variable definitions.

Panel A: Volume-weighted Past One-month Count of Up-limit Hits Based on Initial Buys												
	Full sample (2018:01 to 2019:06)				Pre-survey (2018:01 to 2018:09)				Post-survey (2018:10 to 2019:06)			
Gambling preference, with probability weighting	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, without probability weighting			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 to 2019:06)				Pre-survey (2018:01 to 2018:09)				Post-survey (2018:10 to 2019:06)			
Gambling preference, with probability weighting	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, without probability weighting			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 buy-volume (in RMB) weighted average of past one-month (Panel A) or one-quarter (Panel B) # 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 (Jan. 2018 to June 2019), pre-survey (Jan. 2018 to Sept.2018), and post-survey (Oct. 2018 to June 2019). The key independent variables are survey-based gambling preference. Gambling preference (with probability weighting) 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 (without probability weighting) equals one if an investor answers “Strongly agree” or “Agree” when asked if she often think of stocks as lotteries and zero otherwise. See Table 6 for the exact phrase of the survey questions. Control variables include age, gender, wealth, income, trading experience, account size, and education. T-statistics based on robust standard errors are reported in parentheses.

Panel A: Volume-weighted Past One-month Return Based on Initial Buys												
	Full sample (2018:01 to 2019:06)				Pre-survey (2018:01 to 2018:09)				Post-survey (2018:10 to 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
R2	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 to 2019:06)				Pre-survey (2018:01 to 2018:09)				Post-survey (2018:10 to 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
R2	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 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 (Jan. 2018 to June 2019), pre-survey (Jan. 2018 to Sept. 2018), and post-survey (Oct. 2018 to June 2019). The key independent variables survey-based extrapolative beliefs. 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 keeps rising (dropping) for a while. Otherwise, extrapolation-up (Extrapolation-down) equals zero. See Table 5 for the exact phrase of the survey questions. Control variables include age, gender, wealth, income, trading experience, account size, and education. T-statistics based on robust standard errors are reported in parentheses.

Panel A: Summary Statistics							
	Min	P25	Median	P75	Max	Mean	Std Dev
Turnover	0.00%	12.10%	46.60%	121.60%	650.60%	94.20%	125.70%
Raw returns	-12.60%	-1.80%	0.30%	2.20%	10.00%	-0.10%	3.80%
Net returns	-12.90%	-2.10%	0.10%	2.00%	9.60%	-0.30%	3.80%

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. The main sample include 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 Appendix for more details about variable definitions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actual performance in 2017	4.104***							
	(5.332)							
Over-placement, performance	15.695***							
	(2.760)							
Financial literacy, dummy		11.922***						
		(3.127)						
Over-placement, literacy		1.729						
		(0.400)						
Miscalibration			1.116					
			(0.289)					
Underestimation of trading cost				-3.549				
				(-0.980)				
Do not consider trading cost					-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)
R2	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, with probability weighting	10.924***							
	(2.878)							
Gambling preference, without probability weighting		2.750						
		(0.684)						
Realization utility, winner			7.188*					
			(1.874)					
Realization utility, loser				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)
R2	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)
R2	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 based on robust standard errors are reported in parentheses. See the Appendix for more details about variable definitions.

Dependent variable: Average Monthly Turnover Ratio (%) from 2018/10-2019/6			
Actual performance in 2017	4.198***	Gamble, with probability weighting	11.764***
	(5.219)		(2.920)
Over-placement, performance	11.549**	Gamble, without probability weighting	-1.159
	(2.063)		(-0.263)
Financial literacy, dummy	7.065*	Sensation, novelty	6.598
	(1.800)		(1.360)
Over-placement, literacy	-2.621	Sensation, volatility	3.632
	(-0.625)		(0.824)
Miscalibration	-2.989	Perceived information advantage	15.660***
	(-0.764)		(2.988)
Do not consider trading cost	-3.989	Dismissive of others' information	2.942
	(-1.071)		(0.805)
Underestimation of trading cost	-4.029	Social influence	-7.839
	(-1.052)		(-1.616)
Do not know bid-ask spread	-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, winner	7.049*	Risk Aversion	-2.943
	(1.848)		(-0.692)
Realization utility, loser	-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 based on robust standard errors are reported in parentheses. See the Appendix for more details about variable definitions.

	Panel A: Monthly Turnover (2018:10 to 2019:06)						Panel B: Monthly Raw Returns (2018:10 to 2019:06)	
	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 (2018:10 to 2019:06)						
	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	-1.74**

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

Table 15: Additional Analysis of Gambling Preference, with Probability Weighting

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 group. When testing for the significance of the differences, we use robust standard errors.

	Panel A: Monthly Turnover (2018:10 to 2019:06)						Panel B: Monthly Raw Returns (2018:10 to 2019:06)	
	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%		

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 group. 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

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

Note: Before distributing the survey, we randomly assign 500 targeted branches of brokerage firms into treated and control groups. Investors in the two groups receive questionnaires that are otherwise identical except for one difference: the questionnaire for the treated group includes a "nudge" that highlights 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 1-, 3-, and 6-month before and after the survey. The dummy, Treated, equals one if an investor is in the treated group and answers correctly the follow-up question designed to test if he/she 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 based on robust standard errors are reported in parentheses. See the 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	0.7	3.3	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	0.8	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**

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

Note: We construct a measure for transaction-based gambling behavior in two steps. First, for each of the nine months prior to the survey (2018/1-2018/9), 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. We then sort investors into five groups according to transaction-based gambling behavior and compared their behaviors post-survey, from 2018/10 to 2019/06. In Panel A, we tabulate the summary statistics of monthly turnover ratios (monthly raw returns) for investors in each group. In the last row of each panel, we report the differences between bottom and top group. 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, 2018:01-2018:09			
Actual performance in 2017	-0.009**	Gamble, with probability weighting	0.071***
	(-2.533)		(3.598)
Over-placement, performance	0.002	Gamble, without probability weighting	-0.011
	(0.071)		(-0.482)
Financial literacy, dummy	-0.031	Sensation, novelty	-0.032
	(-1.478)		(-1.518)
Over-placement, literacy	-0.014	Sensation, volatility	0.022
	(-0.633)		(1.030)
Over-precision	0.017	Belief in information advantage	0.049**
	(0.942)		(2.097)
Do not consider trading cost	0.040**	Dismiss information disadvantage	-0.001
	(2.221)		(-0.031)
Underestimation of trading cost	-0.005	Affected by family and friends	-0.005
	(-0.276)		(-0.178)
Do not know bid-ask spread	-0.043**	Affected by investment advisors	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, winner	0.015	Risk Aversion	0.004
	(0.843)		(0.205)
Realization utility, loser	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 in two steps. First, for each of the nine months prior to the survey (2018/1-2018/9), 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 based on robust standard errors are reported in parentheses. See the Appendix for more details about variable definitions.