Taming the Bias Zoo[*]

Hongqi Liu[a], Cameron Peng[b,*], Wei A. Xiong[c], and Wei Xiong[d]

[a] Chinese University of Hong Kong, Shenzhen

[b] London School of Economics and Political Science (LSE)

[c] Shenzhen Stock Exchange

[d] Princeton University

Abstract

The success of behavioral economics has led to a new challenge—many biases offering observationally similar predictions for a targeted financial anomaly. To tame this bias zoo, we combine subjective survey responses with observational data to propose a new approach, one that is robust to question-specific biases introduced through surveys. We illustrate this approach by administering a nationwide survey of Chinese retail investors to elicit their trading motives. In cross-sectional regressions of respondents' actual turnover on survey-based trading motives, perceived information advantage and gambling preference dominate other motives, even though they are not the most prevalent biases simply based on survey responses.

JEL Classification Code: G11, G41

Keywords: Bias zoo, Excessive trading, Gambling preference, Perceived information advantage

Over the last few decades, behavioral economists have used keen insights from psychology to explain many anomalies in individuals' financial decision making.[1] As a result of these successes, there are now multiple behavioral biases proposed to explain each anomaly, and the set of proposed behavioral explanations often varies from one anomaly to another. This "zoo" of biases is not satisfying: quantitatively, it is unlikely that they are equally important; and qualitatively, it is possible that a seemingly relevant bias may be just a manifestation of a different yet more fundamental bias. For the field of behavioral economics to eventually arrive at a unified conceptual framework—one that is based on a small set of biases to explain a wide range of phenomena—it is necessary to consolidate the many biases proposed for each anomaly.[2]

This consolidation task is challenging because existing explanations, by design, share similar or identical predictions for a targeted anomaly. Therefore, key moments in observational data may not differentiate one from the others. Some explanations may offer unique predictions along more subtle dimensions, but testing such predictions often requires particular data that are difficult to collect. Comparing the relative importance of different explanations is even more demanding as it requires empirical proxies of different explanations in the same data sample.

Choi and Robertson (2020) adopt a survey-based approach to directly compare many factors that may affect investment decisions. Specifically, they administer a survey to elicit individual responses to an exhaustive list of economic mechanisms ranging from expectations and risk concerns to biases and transactional factors. Survey responses make it possible to rank the relevance of these factors. Despite the appeal, surveys also raise some methodological concerns due to their subjective nature: respondents may not truthfully report their answers and, even when they do, their subjective responses are noisy and may be influenced by the wording and framing of the questions (Bertrand and Mullainathan, 2001). Such question-specific biases may distort the ranking of survey responses and lead to spurious conclusions.

In this paper, we propose a new approach to consolidate the bias zoo. We design and administer a survey to elicit individual responses to an exhaustive list of behavioral biases. However, we depart from a purely survey-based approach by using subjective survey responses to explain respondents' *actual* investment behaviors. This integrated approach enables us to

---

[1] See Barber and Odean (2013), Hirshleifer (2015), and Barberis (2018) for recent literature reviews.

[2] This effort can be thought of as a response to the "lack-of-discipline critique" about behavioral finance that was common in the 1990s, which said that because people may depart from full rationality in various ways, it is too easy to pick biases for a given anomaly by flipping through the pages of a psychology textbook (Fama, 1998).

overcome some of the challenges faced by the existing approaches that are based on either observational data or survey responses alone.

The first advantage of our approach is that, by collecting people's attitudes towards a variety of economic mechanisms, it allows us to conduct an apples-to-apples comparison of their explanatory power for the targeted anomaly. This advantage is similar to the purely survey-based approach and overcomes the difficulty faced by using observational data alone when differentiating multiple explanations that are observationally equivalent. Second, because the targeted anomaly is measured using field data rather than survey responses, our subject of interest is immune from noise introduced through surveys. This avoids biases that arise when survey responses are used for both dependent variables and independent variables, in which case correlated measurement errors on both sides of the regression can significantly bias the coefficients (Bertrand and Mullainathan, 2001). Third, by regressing observational outcomes on survey responses at the individual level, the regression coefficients are not affected by question-specific biases arising from misunderstanding or prejudice. That is, a common bias in survey responses favoring a certain factor—potentially due to the survey's choice of words or framing of questions—does not affect the cross-sectional explanatory power of that factor, even though the common bias may distort the ranking of the factor relative to other factors purely based on survey responses.

We illustrate this integrated approach with an attempt to resolve the so-called "excessive trading puzzle." Initially documented by Odean (1999) and Barber and Odean (2000) for U.S. retail investors and later found to be prevalent across many markets, the puzzle is characterized by three robust facts about retail investor behavior: 1) they perform poorly relative to the market index before fees, 2) transaction costs make their performance even worse, and 3) those who trade more often perform worse. The literature has proposed several behavioral explanations, for example, overconfidence, realization utility, gambling preference, sensation seeking, social interaction, and low financial literacy, on top of standard arguments such as portfolio rebalancing and liquidity needs (see Table 1 for a complete list of the explanations along with the references).

We ran a nationwide survey among Chinese retail investors, with respondents randomized across regions and brokers. As of the end of 2018, the Chinese equity market stood as the second largest in the world. As highlighted by Allen et al. (2020), one of the most striking features of the Chinese stock market is the coexistence of low returns and high trading volume, with more than

80% of total trading volume coming from retail investors. To understand this phenomenon, our survey asked a series of questions related to financial literacy and return expectations, and, most importantly, included an exhaustive list of behavioral biases and motives as potential explanations of excessive trading. Implicit in our empirical design is the assumption that trading motives are persistent over time and explain cross-sectional differences in trading behavior. The survey took place in September 2018 and received responses from more than 10,000 investors. We then merged these responses with account-level transaction data from the Shenzhen Stock Exchange.

We conduct three sets of exercises that elucidate our approach. In the first set of exercises, we show that, by and large, there is robust statistical consistency between what people *say* and what they *do*: survey responses are largely in line with the actual trading patterns they are designed to capture. For the four trading motives that can be directly matched with our observational data, we find the following: 1) survey-based measures of gambling preference explain the tendency to buy lottery-like stocks, 2) survey-based measures of extrapolation explain the tendency to buy stocks with positive recent returns, 3) survey-based measures of risk aversion explain the holding of stocks with greater volatility, and 4) survey-based return expectations explain changes in stock holdings. By demonstrating this consistency for a wide range of question, we provide external support for prior studies that test finance theories based on surveys alone.

In our second set of exercises, we illustrate how our approach can be used to tame the bias zoo. In the first step, we first run a series of cross-sectional regressions of turnover on each trading motive *alone*. These regressions confirm that many of the previous explanations for excessive trading also hold true in our sample. In the second step, we include all survey-based trading motives as regressors to compare their explanatory power in a horse race.

Our findings from these exercises are three-fold. First, two trading motives stand out in the horse race as the dominant drivers of excessive trading: gambling preference and perceived information advantage. Explanatory power is sizable for both trading motives: while the standard deviation of the monthly turnover rate in our sample is 123%, gambling preference can explain up to 21% and perceived information advantage can explain up to 24%. Despite their substantial cross-sectional explanatory power, these two motives are only supported by 37% and 18% of the respondents, respectively, much lower than several other motives in the survey with supporting rates over 60%. This contrast highlights the important difference between the cross-sectional explanatory power of survey responses and the simple ranking of survey responses, possibly due

4

to the presence of question-specific biases in the survey. Furthermore, these two motives contribute to an annualized transaction fee of 0.6% and 0.7%, respectively, implying substantial investment consequences borne by investors who display either or both of these trading motives.

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

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

In the third and last set of exercises, we compare the effectiveness of transaction-based and survey-based measures of trading motives, by constructing two measures of gambling preference. While the transaction-based measure appears to have greater explanatory power for turnover, it is also correlated with several other trading motives. We therefore conclude by discussing the pros and cons of these two approaches. On the one hand, when carefully designed, surveys can directly target a specific trading motive without being confounded by other trading motives. However, survey responses are subject to measurement noise at the individual level and are thus less powerful. On the other hand, although transaction-based measures are less subject to measurement noise, they may simultaneously capture multiple trading motives and are less reliable in isolating a single economic mechanism.

Our paper contributes to the growing literature that uses surveys to construct economic variables that are otherwise difficult to measure. Several studies, for example, Dorn and Huberman (2005), Glaser and Weber (2007), and Dorn and Sengmueller (2009), have previously combined survey data with observational data to study the excessive trading puzzle. They focus on one or two behavioral biases or trading motives: risk aversion and perceived financial knowledge in Dorn and Huberman (2005), two forms of overconfidence (overplacement and miscalibration) in Glaser and Weber (2007), and sensation seeking in Dorn and Sengmueller (2009). Our study expands the idea of combining survey responses with observational data by running a horse race of multiple trading motives. In the absence of such a horse race, significant effects associated with one motive may simply reflect other motives, as in the case of sensation seeking in our analysis. Furthermore, by showing consistency between survey-based trading motives and observed trading behaviors, we provide external validation to the survey responses in our sample.

Another strand of the literature, for example, Greenwood and Shleifer (2014), Barberis et al. (2018), and Giglio et al. (2020; 2021), uses survey-based expectations to analyze people's belief dynamics. Similar to our paper, Giglio et al. (2020; 2021) combine survey expectations with mutual fund holdings data to validate the consistency between survey expectations and actual investments. In other related studies, Chinco, Hartzmark, and Sussman (2020) use surveys to uncover subjective perception of consumption risk in investors' portfolio choice decisions, while Epper et al. (2020) use experiments to measure the time discount rate and examine its relationship to wealth accumulation over time. These studies again tend to focus on a single variable or bias. In this regard, our paper is most closely related to Choi and Robertson (2020), who also use survey responses to compare a large number of potentially relevant factors for investment decisions. Employing a different framework by combining survey responses with observational data, our study not only provides external validation to survey responses but also overcomes question-specific biases, which may distort the simple ranking of survey responses.

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

1. The Survey

In this section, we first discuss the survey design to further elaborate, from an econometric point of view, the advantages of our approach and the concerns that may arise in our framework. We then explain the procedure for survey distribution and data collection. Finally, we summarize some basic facts from the survey.

1.1. Survey Design

We designed the survey to test and differentiate a large set of trading motives developed by the literature. Table 1 provides a summary of all the trading motives we consider. A trading motive may take several forms. For instance, overconfidence comes in at least three forms. The first is overplacement, which means that people have overly rosy views of their abilities relative to others. The second is miscalibration of uncertainty, which means people are too confident in the accuracy of their beliefs. The third is perceived information advantage, which means that people believe they have superior information over others. The survey included at least one question for each form of overconfidence, as detailed in Sections 1 and 2 of the Online Appendix.

In our research design, we first survey a pool of investors about their trading motives and then compare the different motives' explanatory power for an actual trading behavior. Specifically, we consider a standard linear model to relate investor turnover y to a list of trading motives $x_1, \dots, x_K$:

$$y^i = \beta_0 + \beta_1 \tilde{x}_1^i + \cdots + \beta_K \tilde{x}_K^i + \varepsilon_i, \tag{1}$$

where i indexes individuals. Surveys allow us to collect noisy measures of the trading motives $\{\tilde{x}_k^i\}$, where $\tilde{x}_k^i = x_k^i + u_k^i$ with u_k^i representing the measurement error of variable x_k induced through the survey. In our approach, we do not rank the trading motives by the values of their noisy measures $\{\tilde{x}_k^i\}$, but instead by their cross-sectional explanatory power $\{\beta_k\}$ for the observed turnover. Many respondents may agree with a particular trading motive, but we can confirm its relevance *only if* we also observe that these respondents trade more than other respondents.

A first advantage of our approach is that by directly observing dependent variable y from transaction data, we can avoid spurious coefficients due to mismeasurement in dependent variables. To see why, suppose that instead we use survey-based measure $\tilde{y}^i$, where $\tilde{y}^i = y^i + \delta^i$ and δ^i reflects the survey-induced measurement error in y^i. In our analysis, this corresponds to using self-

reported turnover rather than actual turnover. When δ^i is white noise and uncorrelated with x_k^i, the OLS coefficients will not be biased. However, as discussed by Bertrand and Mullainathan (2001), if measurement error δ^i is correlated with x_k, which is highly likely, OLS coefficients can be severely biased. For example, it may be more difficult for overconfident investors to recall bad trading experience in the past, resulting in a negative bias (δ^i) in the self-reported turnover rate. If we use overconfidence as an explanatory variable x_k, coefficient β_k can be substantially biased downward due to the negative correlation between x_k and δ.

A second advantage of our approach is that question-specific biases in the measurement of x_k will not bias the OLS coefficients. Suppose that $u_k^i = u_k + \eta_k^i$, where $u_k \neq 0$ and η_k^i is pure white noise. For instance, if a trading motive is poorly phrased and subsequently invites misperception or prejudice, there is a question-specific bias u_k against that motive among all survey respondents. This bias reduces the mean of the survey responses $\tilde{x}_k^i$ and thus may distort the ranking of the motive relative to other motives in the purely survey-based approach used by Choi and Robertson (2020).[3] In contrast, in the cross-sectional regression (1), the question-specific bias u_k will not bias the OLS estimate of β_k, as the bias will be absorbed by the intercept. Therefore, when a question is poorly worded and generates, on average, less-favorable responses from the respondents, it will not bias the OLS estimation as long as the downward bias is common to all respondents and does not interact with individual characteristics that we do not control for.

Although our approach is not subject to question-specific measurement bias, other measurement issues may still arise. We now discuss their implications and our solutions.

White noise

We start with η_k^i, the white noise component in the measurement error. This component will produce an attenuation bias in the estimate of the regression coefficients β_k. The magnitude of this bias may differ across motives, depending on the variance-covariance matrix of the K explanatory variables and the variance of each type of white noise. For instance, a larger variance of white noise contributes a greater attenuation bias, which leads to the common concern that insignificant

---

[3] More precisely, Choi and Robertson (2020) asked respondents in their survey to rank competing mechanisms specifically for a given decision variable, that is, the y variable in equation (1). Therefore, an alternative way to interpret their survey responses is that the responses may already capture the respondents' own estimates of the beta coefficients, β_k. In our view, this interpretation may further complicate the task assigned to the respondents and result in other sources of bias due to the more complex inference process.

results from horse races may simply reflect a lot of white noise. To the extent that white noise in measurement errors makes it more difficult to detect significant factors, any significant factor from our analysis is even more important in practice.

Wording, scaling, and mental effort

Measurement errors could also arise due to the wording of questions and scaling of answer options. For example, people give rather different answers to the following two questions: "Do you think the United States should *forbid* public speeches against democracy?" and "Do you think that the United States should *allow* public speeches against democracy?" Similarly, when the scaling of answer options changes, subjects may report their answers differently as they might be anchored by the choice of options. Lack of mental effort typically makes these issues worse, as subjects may not read the questions in detail and choose answers that appear first or last in the list of options. As discussed above, when wording or scaling induces a question-specific bias, it will be absorbed by the intercept and will not bias the OLS coefficients. When the bias is individual-specific and more prevalent in certain demographic groups, then individual characteristics should be properly accounted for. In our main regressions, we control for an exhaustive list of demographic variables.

To mitigate biases induced by wording, we adopt a jargon-free protocol. We phrased the questions as accurately as possible when describing the underlying concept while ensuring that they remained comprehensible to the average respondent. To confirm that respondents could immediately understand each question, we ran a series of pilot tests among the general population on a Chinese version of Mechanical Turk and solicited their feedback on the survey design. The overwhelming majority of respondents found the questions easy to understand. This also ensures that subjects typically did not find it mentally burdensome to complete the survey.

To deal with biases induced by scaling, we designed all questions to be multiple choice with a standardized menu of answer options. There are two types of qualitative questions. The first type—"agreement"—asked respondents whether they agree or disagree with a statement that describes a particular motive driving trading decisions. Answer options included: "strongly agree," "agree," "neutral," "disagree," "strongly disagree," "do not know," and "decline to answer." The second type—"frequency"— asked respondents how often they consider a particular motive when they trade. Answer options included: "always," "often," "sometimes," "rarely," "never," "do not

know," and "decline to answer." We also sought quantitative answers for certain trading motives (e.g., estimates of transaction fees to measure neglect of trading costs). In such cases, we provided several options, each covering a specific value range. The standardization of answer options ensured that any bias resulting from the design of answer options should be small and consistent across all the questions.

Attitudes

Survey questions typically elicit people' attitudes toward a certain description or statement. However, a clear attitude may not exist, and if forced to give an answer, people may just randomly pick an answer, causing further noise in measurement. To deal with this "no-attitude" issue, we include two answer options—"do not know" and "decline to answer"—so that respondents do not feel compelled to give an answer when they do not have a clear one in mind.

Social desirability

Another concern, particularly relevant to eliciting biases and mistakes, is that respondents may want to look good in front of others and avoid giving answers that may sound stupid or wrong. This concern arises naturally in interview-based surveys, in which respondents directly interact with the interviewer. Since our survey was conducted online, our respondents had less of a need to appear socially desirable. Moreover, we carefully phrased the questions to be objective and avoided making any inference about a certain behavior being right or wrong. For instance, to elicit a measure of overconfidence, instead of asking "How overconfident do you think you are?" we asked respondents to only self-assess their investment performance. Later, we would compare it to their actual performance to get our measures of overconfidence.

Others

We discuss three final considerations in our survey design. First, survey responses are subjective: they capture how people consciously perceive themselves to be making investment decisions.[4] A common criticism of subjective surveys in economic analysis is the "as if" critique: respondents may not consciously perceive a factor to be important, but they still behave as if it

---

[4] In the language of Adam Smith, respondents are effectively asked to act as the "impartial spectators" to evaluate the reasons and drivers behind their own decisions (Grampp, 1948).

were (Friedman, 1953). However, subjective perceptions are still useful for many reasons: they shed light on the true decision making process, they help differentiate competing theories, and they have predictive power for implications of debiasing mechanisms on individuals' future behaviors (Choi and Robertson, 2020). It is also inherently interesting to know about people's subjective reasoning. We add that subjective perceptions are also relevant for nudge interventions: if a nudge is targeting a bias that people are not even aware of, it is unlikely that the intervention would successfully produce the desired outcome (DellaVigna and Linos, 2020).

Second, at a general level, there is a significant tradeoff between "being rigorous" and "being intuitive" in the design of survey questions. To be fully rigorous in investigating trading motives, the corresponding survey questions needed to comprehensively capture all their aspects. For instance, to fully grasp realization utility requires calibrating a utility function that captures not only the different attitudes between gains and losses but also the shape of the utility function in different regions. Such a design would make the survey exceedingly long and unavoidably include academic jargon, which, as discussed above, would immediately raise issues on wording and mental efforts. The psychology literature also documents an attribute substitution bias, whereby participants may not respond to complicated questions but rather answer a related question that is easier to respond to (Kahneman and Frederick, 2002). In light of these concerns, we used the "being intuitive" design to make the phrasing as intuitive as possible to laypeople.

Third, post-survey, we designed our empirical strategy with the aforementioned measurement issues in mind. First, we validated survey responses with actual trading behavior and found strong consistency between survey responses and transaction data. This provides further validation of our survey design. Second, we encoded all survey-based trading motives into dummy variables. This standardization minimizes the variation of measurement errors across all survey-based trading motives and facilitates an apples-to-apples comparison.

The final survey contained four main parts. The first part contained eight questions measuring financial literacy. These questions included the classic "big three" questions as well as several other widely used questions to measure financial literacy (Lusardi and Mitchell, 2007; 2011). At the end of this section, we also asked respondents to assess how many questions they answered correctly. This allowed us to construct a measure for overconfidence based on financial literacy. The second part represented the core of the survey, in which we asked respondents to answer a

series of questions related to various trading motives. We postpone a more detailed discussion of this part to Section 1.3. The third part asked about basic demographic characteristics, including name, gender, date of birth, province, city, education, income, net worth, phone number, brokerage firm, and broker branch. While many of these variables serve as control variables in subsequent analysis, they also provide crucial identifying information that enabled us to locate each correspondent in the transaction database. Finally, for a randomly selected group of respondents (the treatment group), we also included a fourth "nudge" section. We explain the "nudge" and discuss the results in more detail in Section 2.8.

1.2. Data

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

The survey took place in September 2018, and respondents were given two weeks to complete the survey.[5] A valid response had to be completed within 30 minutes. Respondents could open the survey using their personal computers or their smartphones.[6] We collected an initial sample of 12,856 respondents. We report the distribution of respondents across brokers and provinces in Table A1 of the Online Appendix. By design, respondents are evenly distributed across the ten

---

[5] The SZSE Center first distributed the survey link to each broker's headquarter. The headquarter then redistributed it to the preselected branches, after which local client managers would send the survey to their clients (investors), likely via phone calls or WeChat messages. Once an investor had completed the survey, the manager would record the investor's name, phone number, and branch name. This information was then sent back to us for verification purposes.

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

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 2 reports a more detailed summary of the sample's demographic characteristics. Overall, the sample is balanced in gender and highly educated: more than half of the respondents had a college or higher degree. Respondents were primarily middle-aged: about half of the sample were aged between 30 and 50. They were also quite wealthy: the median annual income was around 200,000 RMB and the median household net worth was around 500,000 RMB. Overall, our sample represents a relatively well-educated, wealthy set of retail investors, and this means that any results we find should 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 investment decision making, our results may serve as a lower bound.

Finally, we eliminate respondents who shirked by examining the total amount of time spent on the survey. We show, in Figure A2 of the Online Appendix, that it took a median respondent about eight minutes to complete the survey and 95% of respondents finished within 20 minutes. Respondents who spent less than three minutes on the survey experienced a sharp drop in their financial literacy score, suggesting that they may have shirked during the survey. In subsequent analysis, we dropped these observations, which reduced our sample size to 11,268.

1.3. Survey Results

We now summarize the questions covered in the survey and their responses. Our empirical strategy is not to rank trading motives by their supportive rates but to compare their explanatory power for actual turnover in a cross-sectional regression. Nevertheless, it is useful to have an overall picture of the survey responses.

Financial literacy

Table 3 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 (Lusardi and Mitchell, 2014), we also included five other questions to capture additional dimensions of financial

(or investment) literacy.[7] Panel A shows that, out of all eight questions, seven have a correct rate above 75%. The only exception is the question about the relationship between interest rates and bond prices. Panel B shows that more than 80% of the respondents correctly answered at least six questions. In fact, one-third of them were correct on all eight questions. Panel B also shows the distribution of self-assessed scores, which is similar to that of the actual scores. Overall, investors in our sample display a high level of financial literacy.[8]

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.[9] The literature also suggests that overconfidence may be present in several closely related, albeit distinct, forms: overplacement of ability, miscalibration of uncertainty, and overprecision of information. We designed questions to capture each of these forms.

Overplacement of one's own ability is perhaps the most direct form of overconfidence. We constructed two measures of this form, one by the difference between self-assessed and actual performance and the other by the difference between self-assessed and actual literacy scores.[10] In Table 4, Panel A reports their summary statistics. In constructing overplacement of performance, self-assessed performance is measured by the self-reported rank of investment performance among all investors in 2017 while actual performance is measured by the actual rank in the population. At this point, we have not yet merged survey responses with transaction data, so Panel A only reports the distribution of self-assessed performance. The patterns suggest that respondents are rather optimistic about their performance: almost two-thirds believe that their performance is better than average while only a quarter believe that it is below average. Panel A also reports the second

---

[7] 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).

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

[9] See, for example, Kyle and Wang (1997), Daniel, Hirshleifer, and Subrahmanyam (1998; 2001), Odean (1998), Gervais and Odean (2001), Scheinkman and Xiong (2003), Glaser and Weber (2007), and Barber et al. (2020).

[10] Dorn and Huberman (2005) and Barber et al. (2020) use a similar measure for perceived financial knowledge.

measure, overplacement of literacy. Overall, respondents do *not* overestimate their financial literacy, which is perhaps not 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).[11] We measure miscalibration of uncertainty by the difference between two estimates: one for upside returns and the other for downside returns. This is based on two questions asking respondents to estimate how much the stock market will go up or down next year with 10% probability. The difference between the two estimates gives an 80% confidence interval. The rational benchmark (based on historical market volatility) suggests that this difference should be 76%, but Panel A of Table 4 shows that most respondents report a much narrower range.

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

Extrapolation

The behavioral finance literature has also emphasized the tendency of investors to extrapolate past returns as a key driver of stock return predictability and excessive trading.[12] In Table 4, 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 rising, and in the second scenario, a stock's price keeps falling. Respondents are asked whether they believe the stock's price will rise or fall in the future. In both scenarios, more respondents believe in price continuation than reversal, suggesting that Chinese investors on average exhibit extrapolative beliefs.

Neglect of trading costs

Barber and Odean (2000) and Barber et al. (2009) show that trading causes retail investors in the United States and Taiwan to underperform relative to the overall market, with more than 60% of underperformance directly due to commissions and transaction taxes. These findings suggest that investors who trade a lot may have neglected the various fees and taxes associated with trading.

---

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

[12] See, for example, Barberis, Shleifer, and Vishny (1998), Hong and Stein (1999), Barberis et al. (2015, 2018), and Jin and Sui (2019).

As it is common for financial regulators across the world to use Tobin taxes to curb speculative trading, neglect of trading costs could undermine the effectiveness of such financial policies.

Neglect of trading costs stems from at least two possible sources. The first one is underestimation: investors systematically believe the fee is lower than it actually is, possibly due to lack of financial knowledge. The second one is a lack of attention: even when investors have full knowledge about trading costs, it may still matter very little to their trading because the amount associated with each transaction is small and the concept of trading costs may not come to mind at the time of trading.[13]

To capture these two forms of neglect of trading costs, we constructed three different measures. Panel C of Table 4 reports the summary statistics. First, we directly asked investors to estimate the total transaction costs associated with a round-trip buy and sell at 10,000 RMB. The results show that respondents significantly underestimated trading costs: while, on average, such a round-trip transaction should incur a fee of 15 to 26 RMB, almost 70% of the respondents reported an estimate below the lower bound. The second question asked how often an investor considers transaction costs when trading stocks. Similarly, more than half of the respondents said that they never or rarely do so. The third question targeted the implicit cost of the bid-ask spread by asking whether the respondent agrees that bid-ask spread is a form of trading cost. Around 60% of respondents agreed while 23% disagreed. Overall, there is strong evidence that retail investors in China underestimate or neglect trading costs.

If neglect of trading costs is due to (a lack of) attention, then presenting transaction costs in a more salient manner or reminding investors of these costs more frequently may lead them to trade less. To test this hypothesis, we gave half of the respondents a "nudge": we increase the salience of trading costs by presenting them in annualized terms and reminding investors about the negative impact of excessive trading to overall returns. We discuss these results later in Section 2.8.

Gambling preference

Barberis and Huang (2008) show that the cumulative prospect theory of Tversky and Kahneman (1992) can generate a preference for gambling stocks, meaning stocks with positively

---

[13] 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, 2020). 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, 2018).

skewed returns. Bordalo, Gennaioli, and Shleifer (2012) suggest that salience may cause investors to exaggerate the probability of salient payoffs, also leading to a preference for gambling stocks.[14] Barber and Odean (2000) argue that if gambling stocks change over time due to fluctuations of volatility and tail distribution, gambling preference may also contribute to excessive trading by leading some investors to chase gambling stocks and thus trade with other investors.

In Table 5, Panel A shows the responses on the two questions about gambling preference. The first question asked whether the respondent aims to select a few blockbuster stocks in order to get rich quickly. The second question asked whether the respondent consciously perceives trading stocks as buying lotteries in that they are willing to exchange small losses for the small probability of a big gain. About one-third of the respondents agreed or strongly agreed with each statement. In what follows, we differentiate these two questions by labeling the first one as representing "blockbusters" and the second one as representing "lotteries."

In phrasing these two questions, we had the following design in mind: the "blockbusters" question focuses on the salient upside and deliberately tones down the fact that "blockbusters" are rare. Therefore, investors who agree with this statement are the ones drawn to the large upside without necessarily assessing its small probability. In the language of prospect theory, these investors tend to over-weight small probabilities. In contrast, the "lotteries" question contains a direct description of lotteries by explicitly stating that large payoffs rarely happen. Therefore, the two questions not only help identify the gamblers among the respondents, but also help differentiate their assessments of the tail probabilities. As we will show, the "blockbusters" question has substantially stronger explanatory power for investor trading.[15]

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. Barberis and Xiong (2009, 2012) and Ingersoll and Jin (2013) formalize theories of realization utility, in which realization utility results in realization utility and contributes to excessive

---

[14] Kumar (2009) and Boyer, Mitton, and Vorkink (2010) provide evidence that supports the presence of such gambling preference. Barberis, Jin, and Wang (2020) study how prospect theory can explain stock market anomalies.

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

trading.[16] In Table 5, 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 bring more personal happiness: selling the stock or holding on to it. In the second scenario, the respondent instead faces a stock whose price has gone down since purchase and is asked which action would be more painful. According to realization utility, selling winners is more pleasing than holding winners while selling losers is more painful than holding losers. Survey responses for the two questions are mixed. In the first question, consistent with realization utility, more respondents say selling winners makes them happier. In the second question, however, more respondents report that holding on to losers is more painful than selling losers. In what follows, we differentiate these two questions by labeling the first question as realization utility for winners and the second question as realization utility for losers.

Sensation seeking

Sensation seeking, a measurable psychological trait linked to gambling, risky driving, drug abuse, and a host of other behaviors, is shown to be an important motivation for trading (Grinblatt and Keloharju, 2009; Dorn and Sengmueller, 2009).[17] We 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 5, Panel C reports the summary statistics for these two questions. Overall, answers to these two questions exhibit a similar distribution, but the respondents in general do not exhibit a strong tendency for sensation seeking.

Information

Economists have long argued that access to private information is a key reason for investors to trade in financial markets. However, the classic no-trade theorem posits that when all investors are rational and share the same prior beliefs, asymmetric information cannot cause them to trade due to the concern of adverse selection (Milgrom and Stokey, 1982). Instead, theories of financial

---

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

[17] Brown et al. (2018) further argue that sensation seeking may even affect the trading of hedge fund managers.

market trading with asymmetric information, for example, Grossman and Stiglitz (1980) and Kyle (1985), typically involve the presence of noise traders, who may trade at losses, so that rational traders may trade despite the potential concern of adverse selection.

Are retail investors in China rational investors with a genuine information advantage or noise traders who believe they hold superior information even though they do not? We included two questions in the survey to elicit a respondent's perception of their information advantage. The first question measures one's belief in having an information advantage by asking how often they believes they know stocks better than other investors. A positive response to this question may be associated with a genuine information advantage, but it could also reflect a misperceived information advantage due to overconfidence. This latter possibility potentially reflects a tendency to exaggerate one's own information but not the information of others. Various theoretical models have used this tendency to specify investor overconfidence, the third form of overconfidence that we mentioned earlier.[18] Later, we differentiate a genuine information advantage from a misperceived one by examining whether the respondent actually performs better.

The second question measures one's potential adverse selection concerns by asking how often they worry that others know stocks better than themselves. This question measures dismissiveness of others' information, a form of investor bias that offers distinct implications from overconfidence for equilibrium prices and trading volume (Eyster, Rabin, and Vayanos, 2019). Panel A of Table 6 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. Despite the relatively small fraction of respondents who indicate a perceived information advantage, they indeed trade more than others, as we will show later.

Social interaction

Shiller (1984) argues that investing in speculative assets is a social activity because investors enjoy discussing investments and gossiping about others' investment successes or failures. As a result, social influences could lead to excessive trading.[19] We designed two questions to capture

---

[18] For example, Kyle and Wang (1997), Odean (1998), Gervais and Odean (2001), and Scheinkman and Xiong (2003) all model overconfidence as stemming from a perceived information advantage.

[19] Hong, Kubik, and Stein (2004) provide evidence that stock market participation is influenced by social interaction. Han, Hirshleifer, and Walden (2020) develop a model to show that social interaction exacerbates excessive trading among investors.

social interactions, one about the influence of family, friends, and other acquaintances, and the other about the influence of investment advisors. Panel B of Table 6 shows that around 14% of the respondents say that they are often or always influenced by family, friends, or other acquaintances while 8% say their investment advisors often or always have an influence on their trading.

Other trading motives

In Table 6, Panel C reports the responses on the two questions related to liquidity needs and rebalancing motives. Overall, only about 11% of the respondents say portfolio rebalancing often 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 6 reports three standard questions for measuring risk aversion. We elicit investors' risk attitude by asking whether they would be willing to give up their current stable jobs for other jobs with higher expected income but also higher uncertainty in three hypothetical scenarios. While about 34% of the investors were unwilling to take the job with the smallest risk, 26% of the investors were willing to take the riskiest job.[20]

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. We demonstrate the external validity of survey responses by showing their ability to capture actual trading behaviors. We examine the explanatory power of each trading motive alone for turnover, followed by a horse race among all survey-based trading motives. We also provide some robustness checks and additional evidence for several key motives at the end of the section.

2.1. Merging Surveys with Transactions

In the third part of our survey, we asked respondents to provide information on various demographic variables, including name, date of birth, broker name, and branch name. This allows

---

[20] We offer a comparison between Chinese and U.S. investors in Table A3 of the Online Appendix.

us to uniquely identify a substantial fraction of the respondents in the transaction database of the Shenzhen Stock Exchange. Specifically, out of the 11,268 respondents that remain in our sample, we can uniquely identify 6,013 investors. Our transaction data cover January 2018 through June 2019, which nicely staddles our survey date of September 2018. We further require an investor to have held at least one stock in the Shenzhen Stock Exchange during the two-year window before the survey.[21] This further reduces the sample size to 4,671, which is our *main sample*. Table 2 shows that investor characteristics are comparable between all the respondents and those in the main sample, suggesting that the merging process does not induce further biases.

Is the excessive trading prevalent among investors in our sample? Table 7 reports the summary statistics of the monthly turnover and portfolio return for the post-survey sample from October 2018 through June 2019, that is, the nine-month window after the survey. When needed, however, we also extend the window to cover the nine months before the survey. Table 7 confirms the existence of excessive trading. First, investors trade intensively: the median monthly turnover rate 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 was about 0.6% from October 2018 through June 2019, the median net return in our sample is 0.0%. Third, those who trade more perform worse: the correlation between turnover and raw returns is −0.07 while the correlation between turnover and net returns is −0.16. These negative correlations are statistically significant and confirm the key findings of Odean (1999) and Barber and Odean (2000).

2.2. Encoding Survey-based Variables

To make different survey-based variables comparable, we encode them into dummy variables. A detailed description of the construction of these dummy variables is in Table A2 of the Online Appendix. In a nutshell, for the "agreement" type of questions, we code "strongly agree" and "agree" as 1 and other answers as 0; for the "frequency" type of questions, we code "always" and "often" as 1 and other answers as 0; for quantitative questions, we typically use zero as the cut-off value.[22] Table 8 reports the summary statistics of these dummy variables and their pairwise

---

[21] An investor may have been invited to take our survey without any stockholding in the Shenzhen Stock Exchange due to various reasons: holding mutual funds or ETFs, or holding stocks listed on the Shanghai Stock Exchange.

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

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

A high supporting rate in the survey for a certain trading motive does *not* necessarily mean that this motive is a key determinant of excess trading, due to the potential presence of question-specific biases induced by the survey. We filter out such biases by examining the cross-sectional explanatory power of survey responses for actual turnover in the aforementioned regression framework. Only when variation in the survey responses of a given motive explains the cross-sectional variation in turnover can we conclude that the motive is relevant to excessive trading. Column (1) of Table 8 shows the degree to which each trading motive is supported by the respondents in our survey. Several motives, such as overplacement of performance, miscalibration, and underestimation of transaction costs, have strong supporting rates above 60%. Interestingly, as we will show, these motives do not have the strongest explanatory power for turnover in the cross-section, possibly because these survey questions are easier for the respondents to understand. In contrast, some other motives such as gambling preference for blockbusters and perceived information advantage have substantially stronger explanatory power, despite their lower rankings indicated by the values in column (1).

2.3. Validating Survey Responses

There are several widely held concerns about the use of survey responses in testing economic hypotheses. First, respondents may not take the survey seriously and may not truthfully report what they really think or believe. Second, even if their responses are truthful, they may not act in a way that is consistent with their responses. Indeed, because most existing papers are limited to the use of either survey data or transaction data only, the literature is still missing a systematic test of the external validity of subjective survey responses from investors.[23]

---

[23] Several earlier examples of such validation exercises are worth noting. Using survey and administrative data from Denmark and Sweden, respectively, Koijen, Van Nieuwerburgh, and Vestman (2014) and Kreiner, Lassen, and Leth-Petersen (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.

Ideally, we would like to validate responses to all the questions asked in the survey, but this is not plausible. For instance, although the survey has several questions regarding the sources of information and the influence of social interactions, it is difficult, if not impossible, to infer these aspects from observational data without collecting additional data and making strong assumptions. Given these limitations, we validate survey responses for a set of four questions with natural empirical counterparts that can be directly constructed from transactions. These questions concern extrapolation, gambling preference, risk aversion, and return expectation. In addition to having straightforward implications about trading behaviors, these questions span a wide range of trading motives: belief formation, preferences, and return expectations. For brevity, we focus on gambling preference in the main text. We briefly talk about other validation exercises with their details included in Section 8 of the Online Appendix.[24]

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, for example, Kumar (2009), argues that return skewness is difficult to compute and is not a metric sufficiently intuitive to investors. Instead, salient stock characteristics such as realizations of extreme returns would attract investors with a gambling preference. This argument is particularly compelling as it connects well with our earlier discussion of gambling preference driven by investors' over-weighting of tail outcomes (Barberis and Huang, 2008; Bordalo, Gennaioli, and Shleifer, 2012). Motivated by this argument, we take advantage of a unique regulation in the Chinese stock market: the daily price limits rule. This rule states that daily stock returns of individual stocks cannot exceed 10%. We use the total count of up-limit hits (i.e., the number of days with prices hitting the up limit) in a preceding period to proxy for a stock's positive return skewness. As hitting the daily up limit puts a stock in the headlines of the stock exchange, this event is highly salient and attracts attention from investors. Thus, we measure an investor's gambling behavior by the volume-weighted count of up-limit hits over either a month or a quarter based on all the stocks they added to the portfolio.

---

[24] Note that while we demonstrate consistency between survey responses and trading behaviors, we do not claim that the targeted trading behavior is solely captured by the designed question. Indeed, as we will show later in Section 3, one type of observed behavior (such as purchase of gambling stocks) can be driven by multiple motives. Therefore, the purpose of our validation exercise is simply to demonstrate the relevance and usefulness of survey responses.

Table 9 reports the results when regressing transaction-based gambling behavior on survey-based gambling preference. Panel A uses the total count of up-limit hits over the preceding one-month horizon, while Panel B uses one quarter as the horizon. Recall that we included two survey questions regarding gambling preference, one about the desire to pick blockbusters to get rich and the other about a conscious perception of stocks being lottery-like. Indeed, responses to the first question significantly explain gambling behavior with a positive sign. On average, the stocks they purchase have a larger count of up-limit hits by around 0.1 (0.2) times in the preceding month (quarter), and this relationship holds in both the pre-survey and post-survey periods. Interestingly, responses to the second question do not explain gambling behavior. We document a similar pattern about their explanatory power on turnover later.

Extrapolation, risk aversion and survey expectations

We perform three additional exercises to validate survey-based measures of extrapolative beliefs, risk aversion, and return expectations, using a method similar to before. The results are reported in Tables A5 to A7 of the Online Appendix. First, investors who report having extrapolative beliefs exhibit stronger extrapolative behavior: on average, the stocks they purchase experience 1% higher returns in the preceding month and more than 2% higher returns in the preceding quarter, and this holds in both pre-survey and post-survey samples. Second, consistent with Dorn and Huberman (2005), survey-based measures of risk aversion are negatively associated with holding more-volatile stocks. Third, consistent with Giglio et al. (2020), 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. (2020), is relatively small.

Finally, we note that throughout the validation exercises, although the coefficient between the survey response and the targeted trading behavior is highly significant, the R-squared is generally small. For instance, in Table 9, across all specifications, the t-statistic for gambling preference (blockbusters) remains around 4, but the R-squared is consistently below 2.5%. This suggests that although survey responses are consistent with the targeted behavior, much of its variation is left unexplained. This low R-squared could be due to measurement errors in survey responses, but it could also be that the behavior itself is driven simultaneously by multiple factors. We will discuss this important issue further in Section 3.

2.4. 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.[25] Table 10 presents the baseline results, where in each column we regress turnover on a particular survey-based trading motive. Most regressions are univariate, except for a few instances where we need to control for some additional characteristics.

Columns (1) to (3) report the results on three measures of overconfidence: overplacement of performance, overplacement of literacy, and miscalibration of uncertainty. Out of these three measures of overconfidence, the only one that is significantly and positively related to turnover is overplacement of performance: in column (1), conditional on having the same past performance, investors who self-report having higher performance tend to trade more subsequently. In column (3), miscalibration of uncertainty does not significantly predict future turnover. These results are consistent with Glaser and Weber (2007), who also show overplacement predicts excess trading while miscalibration of uncertainty does not.

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 view that excessive trading may be driven by a lack of financial knowledge. Therefore, improving investors' financial literacy, a policy often advocated in emerging economies such as China, may not be effective in reducing excessive trading. Furthermore, column (2) shows that overplacement of literacy does not predict future turnover.

Columns (4) to (6) report the results on neglect of trading costs. Surprisingly, for all three measures we have constructed, none of them significantly predict future turnover with the predicted sign: in columns (4) and (5), the coefficients are close to zero and insignificant; in 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 using direct estimates of transaction fees in a round-trip trade and should clearly identify those underestimating

---

[25] 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 one to report a higher self-assessed performance—resulting in more overplacement of performance—and to trade more.

trading costs.[26] That we cannot find any supporting evidence despite having constructed three measures for neglect of trading costs gives us pause about its role in explaining investor trading. It is possible that investors who underestimate trading costs are more naïve about financial markets, and such naivety results in passivity in trading and low volume. It is also possible that this pattern reflects a reverse selection that investors who trade more incur more total costs and are more aware of their existence. We further examine these issues in Section 2.7.

Columns (7) and (8) report the results on extrapolative beliefs. For the two measures of extrapolation of positive and negative returns, we do not find a strong relationship between extrapolative beliefs and turnover. One possibility is that extrapolation generates trading only in a bullish market (Barberis et al., 2018; Liao, Peng and Zhu, 2021), but the period we examine is relatively quiet—the market increased by just a few percentage points during the 9-month window. Another possibility is that extrapolation alone cannot explain volume and must be combined with some additional forces to generate a trading frenzy (Liao, Peng, and Zhu, 2021).

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) and Bordalo, Gennaioli and Shleifer (2012), investors who are subject to gambling preference trade significantly more. Again, the question about "blockbusters" is much more powerful than the "lotteries" question. This is consistent with the pattern in Table 9, which shows gambling behavior can be explained by answers to the "blockbusters" question but not by answers to the "lotteries" question.

Columns (11) and (12) report the results on realization utility and show 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 pain when selling losers—does not predict future turnover. This pattern is consistent with the implications of realization utility (Barberis and Xiong, 2012), as investors who exhibit realization utility are more willing to let go of stocks in gains and to hold on to stocks in losses.

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

---

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

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

Columns (15) and (16) report the results on perceived information advantage and dismissiveness of others' information. Column (15) shows that those who believe in themselves having an information advantage trade more, whereas column (16) shows that those who are more dismissive do *not* trade more. As we discussed earlier, the first measure can capture a particular form of overconfidence if we can show these investors do not deliver better returns; indeed, we show this later Section 2.7. The second measure captures the dismissiveness modelled by Eyster, Rabin, and Vayanos (2019). Thus, we find supportive evidence for perceived information advantage in explaining excessive trading, but not for dismissiveness.

Finally, columns (17) and (18) concern two measures of social influence. Interestingly, investors who are more influenced by their family, friends, and investment advisors tend to trade *less*, not more. This pattern does not lend support to the aforementioned literature that argues that social interaction contributes to the spread of investor sentiment and excessive trading.[27] Columns (19) and (20) show that rational trading motives such as portfolio rebalancing needs and liquidity needs can only explain a small part of the variation in turnover across investors.

In sum, Table 10 confirms several of the existing explanations for trading volume: for example, overplacement of performance, gambling preference, sensation seeking, realization utility, and perceived information advantage. Table 10 also shows a number of "null" results for some prominent explanations of excessive trading: for example, lack of financial literacy, neglect of trading costs, dismissiveness, and social interaction.

2.5. Horse Race Results on Turnover

Although the baseline results confirm several of the previous explanations for trading volume, it remains unclear whether their explanatory power will survive once they are all included in the same regression. Such a horse race has not been run before. Table 11 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, net worth, and education; 2) return

---

[27] However, we note that recent models of social interactions such as Han, Hirshleifer, and Walden (2020) are inherently conditional: social interactions lead to more trading when the market is going up and people are making money. Our tests rely on a period of quiet market reactions and therefore does not test these models directly.

expectations to control for differences in optimism and pessimism; and 3) recent performance to control for "mood."[28] Table 11 reveals a number of notable observations.

First, two trading motives stand out in the horse race: gambling preference ("blockbusters") and overconfidence in the form of perceived information advantage. Both coefficients are quantitatively large and significant at the 1% level. The finding of overconfidence as a key driver of turnover supports the large volume of prior studies emphasizing the roles of overconfidence. Even more interesting, our finding highlights that a particular form of overconfidence through perceived information advantage—rather than other forms such as overplacement of literacy and miscalibration of uncertainty—is most relevant in explaining trading. This form of overconfidence also confirms the specification adopted by Kyle and Wang (1997), Daniel, Hirshleifer, and Subrahmanyam (1998, 2001), Odean (1998), Gervais and Odean (2001), 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. Earlier literature tends to treat gambling preference as an important mechanism for understanding demand for lottery-like stocks but hasn't fully established its link with excessive trading. Our finding suggests that gambling preference may also lead investors to trade more. Barber and Odean (2000) conjecture a mechanism that works as follows. As individual stocks fluctuate in their volatility and tail distribution, the set of lottery-like stocks changes over time. Consequently, investors subject to gambling preference chase one lottery-like stock after another, leading to large trading volume.

We note several interesting aspects of our findings of perceived information advantage and gambling preference as the most powerful factors in explaining turnover. First, note that in Table 8, perceived information advantage is supported by only 18% of the respondents and gambling preference (for blockbusters) by 37%. Both are substantially lower than some other factors with over 60% supporting rates. Therefore, although the two motives affect a small fraction of the population, their explanatory power is greater. This contrast also echoes our earlier discussion that question-specific biases may make it challenging to rely on the simple ranking of survey responses to compare the importance of different trading motives in explaining actual turnover.

Second, in Table 8, the correlation coefficient between perceived information advantage and gambling preference for blockbusters is −0.06. The small correlation suggests that overconfidence

---

[28] 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 Table A8 of the Online Appendix.

and gambling preference contribute to trading volume through two orthogonal channels. Below, we present additional evidence to support these trading motives as key drivers of excessive trading.

Third, several trading motives that are significant in the baseline regressions become insignificant or 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, suggesting that their explanatory power is subsumed by other factors. The contrast between sensation seeking and gambling preference is also worth noting, given the literature sometimes mixes the two. Sensation seeking suggests that investors like to gamble because they derive utility from gambling activities independent of the final payoffs while gambling preference suggests that the appeal of gambles is ultimately driven by the potential of a large payoff. Our analysis suggests that, while sensation seeking and gambling preference are correlated, gambling preference is the more relevant factor for the observed trading.

Finally, consistent with the finding of Barber and Odean (2001), we 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 11 persists even after controlling for various forms of overconfidence, suggesting the gender effect may go beyond overconfidence.

2.6. Robustness and Subsample Analysis

As robustness checks, we report the results from alternative regressions in Section 9 of the Online Appendix, including specifications in which we bootstrap standard errors, add branch fixed effects as control variables, use a larger sample that includes investors that have not traded for more than two years before the survey, and use a small sample that only includes investors who are active around the time of the survey. We also consider alternative measures of turnover, including: an equal-weighted version as opposed to the value-weighted one we use throughout the paper; and a version measured in the nine-month window before the survey, as opposed to the nine-month window after the survey. Throughout all these specifications, gambling preference and perceived information advantage remain the most powerful factors for explaining turnover.

We also perform two sets of subsample analysis and report the results in Section 9 of the Online Appendix. In the first one, we split the full sample based on account size and compare the behaviors of small and large investors. Overall, consistent with the notion that small investors are more affected by behavioral biases, we find that the results are slightly stronger among small investors. In the second subsample, we split the full sample based on the fraction of wealth invested in the stock market. In both subsamples, gambling preference and perceived information advantage remain significant factors. However, for investors whose wealth is more invested in the stock market, portfolio rebalancing needs become a more pronounced factor to their trading.

We discuss two limitations of our horse race. First, it is possible that the importance of each mechanism is time-varying. Without a panel of survey responses, we can only capture a snapshot of their relative importance. For instance, realization utility (Barberis and Xiong, 2012; Liao, Peng, and Zhu, 2021) and social interactions (Han, Hirshleifer, and Walden, 2020) may contribute to excessive trading more in a market boom than in a market downturn. However, we show, in Table A17 of the Online Appendix, that the explanatory power of each motive remains stable during the 9-month window before the survey, suggesting relatively persistent importance in the time series. Second, and relatedly, it is also possible that some retail investors learn to debias themselves from past mistakes, and the importance of certain mechanisms may decay over time (Seru, Shumway, and Stoffman, 2010). While our cross-sectional setting does not allow us to directly speak to the issue of learning, we note that some recent evidence suggests that retail investors do not appear to learn from their prior mistakes (e.g., Anagol, Balasubramaniam, and Ramadorai, 2019).

2.7. Additional Evidence of Excessive Trading

Trading is not necessarily excessive if more trading is associated with better returns. We further examine the portfolio returns based on investors' responses to the gambling preference and perceived information advantage questions to show that indeed the associated trading is excessive.

Panel A of Table 12 sorts investors into five groups based on their answers to the "blockbusters" question and reports each group's monthly turnover and portfolio return. While this single-sorting approach ignores the correlations of gambling preference with other trading motives, it provides a more granular look at the explanatory power of gambling preference.[29] For

---

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

turnover, there is a monotonically increasing pattern from the least gambling-prone to the most gambling-prone group. 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 outliers. On average, the difference between the "strongly agree" group and the "strongly disagree" group 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? The result shows that this is the case: the five groups exhibit similar raw returns before fees. In fact, the "strongly agree" group on average earns −0.35% lower monthly returns than the "strongly disagree" group, albeit the difference is not statistically significant. Together, the lack of superior performance and the large transaction costs suggest their trading is excessive.

We also examine the characteristics of stocks purchased by the five groups of investors in Table A18 of the Online Appendix. Investors with a survey-based gambling preference tend to buy stocks that are smaller, have a larger market beta, and have larger counts of daily up-limit hits, and higher past volatility and past returns. These stocks also perform worse subsequently, confirming that investors with a gambling preference trade in the wrong direction and their trading is excessive.

Panel B of Table 12 performs a similar exercise by sorting investors into five groups based on their answers to the "information advantage." 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 the monthly turnover rate between the "always" and the "never" groups is about 24%, implying an annual transaction fee of 0.7%.

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

2.8. Neglect of Trading Costs

In both the baseline and the horse race, none of the survey variables for neglect of trading costs can explain turnover in the right direction. This contradicts the popular view that Chinese retail investors trade so much because they neglect trading costs. In Table 10 and Table 11, some measures even suggest an opposite pattern that investors with more awareness of trading costs trade more. To further isolate the effect of awareness of trading costs, we have also implemented a randomized experiment.

Among all the 500 brokerage branches we distributed the survey to, we randomly selected 250 branches to include an additional "nudge." The "nudge" asked the respondent to read a one-page article that highlighted the negative consequences of excessive trading. As shown in Figure A3 in the Online Appendix, the article contained a detailed calculation of how much investors lose from frequent trading along with a quote from Warren Buffett advising investors to buy and hold. Instead of presenting trading costs as a fraction of total transaction value, we made it more salient by presenting the annualized fee rate for a frequent trader. We also included a "validation" question after the article by asking the respondent to compute the total trading cost for a given level of turnover. Answers to this question help identify those who have actually read the article and therefore were treated. We study the effect of this "nudge" in a difference-in-difference framework and report the results in Table A19 of the Online Appendix. Overall, the nudge had no effect on reducing trading. One might argue that the "nudge" was not sufficiently strong, and the treated group may not have read the article carefully. However, we identify an investor as treated only if they were in the treated group and answered the "validation" question correctly.

Taken together, our analysis suggests that neglect of trading costs is not a key driver of excessive trading. This finding has an important policy implication. Policy makers across the world, including China's stock market regulator, the China Securities Regulatory Commission (CSRC), frequently use Tobin taxes as a policy tool to curb speculative trading in stock markets. To the extent that investors may engage in excessive trading despite their awareness of the trading costs, our finding casts doubt on the effectiveness of Tobin taxes.[30]

3. Comparing Survey-Based and Transaction-Based Measures

---

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

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 survey questions can measure trading motives in a way closer to textbook definitions. Second, survey responses allow researchers to measure a large set of trading motives from the perspectives of the respondents at the same time, including those that are hard to measure from administrative data. However, there are also various concerns about survey data. The primary concern, the one we have already addressed through various validation exercises, is that survey responses may not capture actual trading behavior. A second concern is that survey responses are noisy: perhaps respondents, on average, 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 9, 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.

This concern about noise in survey responses raises a follow-up question: do transaction-based behavioral measures have stronger power than survey-based measures? We address this question by comparing survey-based and transaction-based measures of gambling behavior. Table 13 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 12 based on the survey-based measure of gambling behavior.

While the transaction-based measure of gambling behavior appears to work well, we demonstrate that this measure may capture multiple forces at the same time. We regress the transaction-based measure of gambling behavior on all survey-based trading motives and report the results in Table 14. 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.[31]

Taken together, our results show a trade-off between survey-based and transaction-based measures of trading motives. Survey-based measures have stronger power from the economic perspective of qualitatively testing different trading motives, even though they may contain more noise and thus have weaker statistical power. Transaction-based measures have stronger statistical power, but one measure may reflect multiple mechanisms and the economic interpretations are thus not as sharp as survey-based measures.

4. Conclusion

We propose a new approach to consolidate the zoo of behavioral biases by examining the cross-sectional explanatory power of survey-based measures of different biases for certain observed investor behavior. This approach allows researchers to examine multiple mechanisms at the same time and addresses some econometric issues that arise in a purely survey-based approach. We illustrate this approach by designing and administering a nationwide survey to study why investors trade so much. We highlight a number of new findings. First, survey responses are consistent with actual trading behaviors. Second, overconfidence (in having an information advantage) and gambling preference dominate other trading motives in explaining observed turnover, despite their relatively low supporting rates in the survey. Third, other explanations such as neglect of trading costs and low financial literacy do not contribute to excessive trading. Finally, by analyzing the pros and cons of survey-based and transaction-based approaches, we argue that our integrated approach can help mitigate the concerns faced by each of these approaches alone.

The analysis in this paper has focused on one of the fundamental puzzles in financial markets: why do retail investors trade so much? A similar exercise can be carried out to compare competing mechanisms for other anomalies. It is possible that a few mechanisms are most relevant for explaining multiple anomalies, but it is also possible that different mechanisms are driving

---

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

different anomalies. In this regard, our exercise takes a necessary first step to eventually tame the bias zoo.

References

Allen, F., Qian, J., Shan, C., Zhu, J., 2020. The development of the Chinese stock market. In: Amstad, M., Sun G., Xiong W. (Eds.), Handbook of China's Financial System, Princeton University Press, pp. 283–313.

Alpert, M., Raiffa, H., 1982. A progress report on the training of probability assessors. In Daniel Kahneman, Paul Slovic & Amos Tversky (Eds.), Judgment Under Uncertainty: Heuristics and Biases, Cambridge University Press, pp. 294–305.

Anagol, S., Balasubramaniam, V., Ramadorai T., 2020. Learning from noise: evidence from India's IPO lotteries. Journal of Financial Economics, forthcoming.

Barber, B.M., Huang X., Ko K.J., Odean, T., 2020. Leveraging overconfidence. Unpublished working paper.

Barber, B.M., Lee, Y.T, Liu, Y.J., Odean T., 2009. Just how much do individual investors lose by trading? Review of Financial Studies 22, 609–632.

Barber, B.M., Odean T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. Journal of Finance 55, 773–806.

Barber, B.M., Odean T., 2001. Boys will be boys: gender, overconfidence, and common stock investment. Quarterly Journal of Economics 116, 261-292.

Barber, B.M., Odean, T., 2013. The behavior of individual investors. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance, Vol. 2, Elsevier, pp. 1533-1570.

Barberis, N., 2018. Psychology-based models of asset prices and trading volume. In: Bernheim, B.D., DellaVigna, S., Laibson, D. (Eds.), Handbook of Behavioral Economics: Applications and Foundations 1, Vol. 1, North-Holland, pp. 79-175.

Barberis, N., Greenwood R., Jin L.J., Shleifer, A., 2015. X-CAPM: an extrapolative capital asset pricing model. Journal of Financial Economics 115, 1–24.

Barberis, N., Greenwood, R., Jin, L.J., Shleifer, A., 2018. Extrapolation and bubbles. Journal of Financial Economics 129, 203–227.

Barberis, N., Jin, L., Wang, B., 2020. Prospect theory and stock market anomalies. Journal of Finance, forthcoming.

Barberis, N., Huang, M., 2008. Stocks as lotteries: the implications of probability weighting for security prices. American Economic Review 98, 2066–2100.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.

Barberis, N., Xiong, W., 2009. What drives the disposition effect? An analysis of a long-standing preference-based explanation. Journal of Finance 64, 751–784.

Barberis, N., Xiong, W., 2012. Realization utility. Journal of Financial Economics 104, 251–271.

Ben-David, I., Graham, J.R., Harvey, C.R., 2013. Managerial miscalibration. Quarterly Journal of Economics 128, 1547–1584.

Benos, A.V., 1998. Aggressiveness and survival of overconfident traders. Journal of Financial Markets 1, 353–383.

Bertrand, M., Mullainathan, S., 2001. Do people mean what they say? Implications for subjective survey data. American Economic Review 91, 67–72.

Birru, J., 2015. Confusion of confusions: a test of the disposition effect and momentum. Review of Financial Studies 28, 1849–1873.

Bordalo, Pedro, Gennaioli, N., Shleifer, A., 2012. Salience theory of choice under risk. Quarterly Journal of Economics 127, 1243–1285.

Boyer, B., Mitton, T., Vorkink, K., 2010. Expected idiosyncratic skewness. Review of Financial Studies 23, 169–202.

Brown, S., Lu, Y., Ray, S., Teo, M., 2018. Sensation seeking and hedge funds. Journal of Finance 73, 2871–2914.

Cai, J., He J., Jiang, W., Xiong, W., 2019. The Whack-a-mole game: Tobin taxes and trading frenzy. Review of Financial Studies, forthcoming.

Chetty, R., Looney, A., Kroft, K., 2009. Salience and taxation: theory and evidence. American Economic Review 99, 1145–1177.

Chinco, A., Hartzmark, S.M., Sussman, A.B., 2020. Risk-factor irrelevance. Unpublished working paper.

Choi, J.J., Robertson, A.Z., 2020. What matters to individual investors? Evidence from the horse's mouth. Journal of Finance 75, 1965–2020.

Da, Z., Huang, X., Jin, L.J., 2020. Extrapolative beliefs in the cross-section: what can we learn from the crowds? Journal of Financial Economics, forthcoming.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53, 1839–1885.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 2001. Overconfidence, arbitrage, and equilibrium asset pricing. Journal of Finance 56, 921–965.

DellaVigna, S., Linos, E., 2020. RCTs to scale: comprehensive evidence from two nudge units. Unpublished working paper.

Deng, Y., Liu, X., Wei, S.J., 2018. One fundamental and two taxes: when does a Tobin tax reduce financial price volatility? Journal of Financial Economics 130, 663–692.

Dorn, D., Huberman, G., 2005. Talk and action: what individual investors say and what they do. Review of Finance 9, 437–481.

Dorn, D., Sengmueller, P., 2009. Trading as entertainment? Management Science 55, 591–603.

Epper, T., Fehr, E., Fehr-Duda, H., Kreiner, C.T., Lassen, D.D., Leth-Petersen, S. and Rasmussen, G.N., 2020. Time discounting and wealth inequality. American Economic Review 110, 1177–1205.

Eyster, E., Rabin, M., Vayanos, D., 2019. Financial markets where traders neglect the informational content of prices. Journal of Finance 74, 371–399.

Fama, E.F., 1998. Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics 49, 283-306.

Friedman, M., 1953. The methodology of positive economics. Essays in Positive Economics 3, 145–178.

Frydman, C., Barberis, N., Camerer, C., Bossaerts, P., Rangel, A., 2014. Using neural data to test a theory of investor behavior: an application to realization utility. Journal of Finance 69, 907–946.

Frydman, C., Rangel, A., 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, C., Wang, B., 2020. The impact of salience on investor behavior: evidence from a natural experiment. Journal of Finance 75, 229–276.

Gao, X., Lin, T.C., 2015. 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, S., Odean, T., 2001. Learning to be overconfident. Review of Financial Studies 14, 1–27.

Giglio, S., Maggiori, M., Stroebel, J., Utkus, S., 2020. Five facts about beliefs and portfolios. American Economic Review, forthcoming.

Giglio, S., Maggiori, M., Stroebel, J. and Utkus, S., 2021. The joint dynamics of investor beliefs and trading during the COVID-19 crash. Proceedings of the National Academy of Sciences, 118(4).

Glaser, M., Weber, M., 2007. Overconfidence and trading volume. Geneva Risk and Insurance Review 32, 1–36.

Graham, J.R, Harvey, C.R, Huang, H., 2009. Investor competence, trading frequency, and home bias. Management Science 55, 1094–1106.

Grampp, W.D, 1948. Adam Smith and the economic man. Journal of Political Economy 56, No. 4: 315–336.

Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. Review of Financial Studies 27, 714–746.

Grinblatt, M., Han, B., 2005. Prospect theory, mental accounting, and momentum. Journal of Financial Economics 78, 311–339.

Grinblatt, M., Keloharju, M., 2001. How distance, language, and culture influence stockholdings and trades. Journal of Finance 56, 1053–1073.

Grinblatt, M., Keloharju, M., 2009. Sensation seeking, overconfidence, and trading activity. Journal of Finance 64(2), 549–578.

Grinblatt, M., Keloharju, M., Linnainmaa, J., 2011. IQ and stock market participation. Journal of Finance 66, 2121–2164.

Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70, 393–408.

Han, B., Hirshleifer, D., Walden, J., 2020. Social transmission bias and investor behavior. Journal of Financial and Quantitative Analysis, forthcoming.

Hirshleifer, D., 2015. Behavioral finance. Annual Review of Financial Economics, 7, 133–159.

Hong, H., Kubik, J.D, Stein, J.C, 2004. Social interaction and stock-market participation. Journal of Finance 59, 137–163.

Hong, H., Scheinkman, J., Xiong, W., 2008. Advisors and asset prices: a model of the origins of bubbles. Journal of Financial Economics 89, 268–287.

Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143–2184.

Ingersoll, J.E., Jin, L.J., 2013. Realization utility with reference-dependent preferences. Review of Financial Studies 26, 723–767.

Jin, L.J., Sui, P., 2019. Asset pricing with return extrapolation. Unpublished working paper.

Kahneman, D., Frederick, S., 2002. Representativeness revisited: attribute substitution in intuitive judgment. Heuristics and Biases: The Psychology of Intuitive Judgment 49: 81.

Kelly, M., O Grada, C., 2000. Market contagion: evidence from the panics of 1854 and 1857. American Economic Review 90, 1110–1124.

Koijen, R., Van Nieuwerburgh, S., Vestman, R., 2014. Judging the quality of survey data by comparison with "truth" as measured by administrative records evidence from Sweden. In: Carroll, C.D., Crossley, T.F., Sabelhaus, J. (Eds.), Improving the Measurement of Consumer Expenditures, University of Chicago Press, pp. 308–346.

Kumar, A., 2009. Who gambles in the stock market? Journal of Finance 64, 1889–1933.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Kyle, A.S., Wang, F.A., 1997. Speculation duopoly with agreement to disagree: can overconfidence survive the market test? Journal of Finance 52, 2073–2090.

Liao, J., Peng, C., Zhu, N., 2021. Extrapolative bubbles and trading volume. Unpublished working paper.

Lusardi, A., Mitchell, O.S., 2007. Baby boomer retirement security: the roles of planning, financial literacy, and housing wealth. Journal of Monetary Economics 54, 205–224.

Lusardi, A., Mitchell, O.S., 2011. Financial literacy around the world: an overview. Journal of Pension Economics and Finance 10, 497–508.

Lusardi, A., Mitchell, O.S., 2014. The economic importance of financial literacy: theory and evidence. Journal of Economic Literature 52, 5–44.

Milgrom, P., Stokey, N., 1982. Information, trade and common knowledge. Journal of Economic Theory 26, 17–27.

Odean, T., 1998. Volume, volatility, price, and profit when all traders are above average. Journal of Finance 53, 1887–1934.

Odean, T., 1999. Do investors trade too much? American Economic Review 89, 1279–1298.

Pool, V.K., Stoffman, N., Yonker, S.E., 2015. The people in your neighborhood: social interactions and mutual fund portfolios. Journal of Finance 70, 2679–2732.

Scheinkman, J.A., Xiong, W., 2003. Overconfidence and speculative bubbles. Journal of Political Economy 111, 1183–1219.

Seru, A., Shumway, T., Stoffman, N., 2010. Learning by trading. Review of Financial Studies 23, 705-739.

Shefrin, H., Statman, M., 1985. The disposition to sell winners too early and ride losers too long: theory and evidence. Journal of Finance 40, 777–790.

Shefrin, H., Statman, M., 2000. Behavioral portfolio theory. Journal of Financial and Quantitative Analysis 35, 127–151.

Shiller, R.J., 1984. Stock prices and social dynamics. Brookings Papers on Economic Activity, 457–498.

Shiller, R.J., 1989. Comovements in stock prices and comovements in dividends. Journal of Finance 44, 719–729.

Song, Z., Xiong, W., 2018. Risks in china's financial system. Annual Review of Financial Economics 10, 261–286.

Taubinsky, D., Rees-Jones, A., 2018. Attention variation and welfare: theory and evidence from a tax salience experiment. Review of Economic Studies 85, 2462–2496.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

Van Rooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation. Journal of Financial Economics 101, 449–472.

Theory	Forms of Representation	Papers
Overconfidence	1. overplacement 2. miscalibration of uncertainty	Daniel, Hirshleifer, and Subrahmanyam (1998; 2001); 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, Huang, and Jin (2020); Liao, Peng, and Zhu (2021)
Neglect of trading costs	1. underestimation of transaction fees 2. knowledge about the bid-ask spread 3. salience of transaction fees	Barber and Odean (2000); Barber et al. (2009); Bordalo, Gennaioli, and Shleifer (2012)
Gambling preferences	1. blockbuster 2. lotteries	Barber and Odean (2000); Shefrin and Statman (2000); Barberis and Huang (2008); Kumar (2009); Barber et al. (2008); Bordalo, Gennaioli, and Shleifer (2012)
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 (2015)
Information	1. perceived information advantage 2. dismissiveness (of others' information)	Kyle and Wang (1997); Odean (1998); Gervais and Odean (2001); Scheinkman and Xiong (2003); Eyster, Rabin, and Vayanos (2019)
Social and advisor influences	1. advisor influence 2. social influence	Shiller (1989); Kelly and Grada (2000); Hong, Kubik, and Stein (2004); Hong, Scheinkman, and Xiong (2008); Pool, Stoffman, and Yonker (2015); Han, Hirshleifer, and Walden (2020)
Financial or investment literacy	1. compounding 2. inflation 3. diversification 4. asset risk 5. definition of stocks 6. definition of bonds 7. the PE ratio 8. definition of mutual funds	Van Rooij, Lusardi, and Alessie (2011); Lusardi and Mitchell (2007; 2011; 2014); Grinblatt, Keloharju, and Linnainmaa (2011)

Table 1: Summary of Theories on Trading Volume

Gender	Population	All Respondents	Main Sample
Male	71.70%	54.00%	54.40%
Female	28.30%	46.00%	45.60%
Education			
Middle school or blow	7.30%	8.60%	8.10%
High school	24.70%	15.60%	18.35%
Professional school	26.00%	21.90%	24.83%
College	23.60%	44.90%	40.95%
Graduate school and above	3.40%	9.20%	7.77%
Others	14.80%	0.00%	0.00%
Age			
< 30	21.30%	27.80%	26.10%
30 to 40	27.40%	29.10%	27.40%
40 to 50	24.50%	19.90%	22.40%
50 to 60	15.10%	14.80%	16.00%
> 60	11.70%	8.50%	8.10%
Investment age (in years)			
< 2	10.00%	NA	21.20%
2 to 6	29.80%	NA	26.20%
6 to 10	18.00%	NA	17.40%
> 10	42.20%	NA	35.10%

Income (RMB)	Population	All Respondents	Main Sample
< 20K	NA	3.80%	2.08%
20K to 100K	NA	17.20%	16.42%
100K to 200K	NA	29.50%	30.08%
200K to 500K	NA	29.50%	30.16%
500K to 1M	NA	12.60%	13.74%
1M to 2M	NA	4.20%	4.47%
2M to 10M	NA	2.10%	2.46%
10M and above	NA	1.20%	0.58%
Net worth (RMB)			
< 20K	NA	4.80%	2.57%
20K to 100K	NA	12.30%	9.59%
100K to 500K	NA	27.50%	25.50%
500K to 1M	NA	22.30%	23.91%
1M to 2M	NA	21.90%	25.16%
2M to 10M	NA	6.50%	8.05%
10M and above	NA	4.80%	5.22%
Trading characteristics in 2017			
Maximum value of investment (in thousand RMB)	639	NA	1,250
Annual turnover rate	9.4	NA	8.3
Annual raw return rate	-3.90%	NA	-1.20%

Table 2 Summary Statistics for the Investor Population, Survey Respondents, and Investors in the Main Sample

Note: This table shows the summary statistics for the investor population, all respondents, and the main sample. The population's characteristics are obtained from the centralized database at the Shenzhen Stock Exchange. Survey respondents are the 12,856 investors who have completed the survey. The main sample includes 4,671 survey respondents that: 1) can be identified in the Shenzhen Stock Exchange centralized database, and 2) held at least one SZSE stock during the two-year window before the survey. Gender, education, and age are from either the SZSE centralized database or the survey answers. Investment age and trading characteristics in 2017 are calculated from trading records and thus missing for the survey respondents who cannot be identified in the SZSE centralized database. Income and net worth are obtained from the survey and thus missing for the investor population. See Table A2 in the Online Appendix for more details about variable definitions.

Panel A: Correct Rate by Question

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

Panel B: Distribution of Financial Literacy Scores

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

Table 3 Survey Responses on Questions on Financial Literacy

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

Panel A: Overconfidence

1. What fraction of retail investors do you think earned higher returns than you in 2017?

< 10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	80–90%	> 90%	N/A
11.8%	13.8%	15.8%	13.5%	12.4%	10.4%	5.8%	3.8%	2.2%	3.4%	7.2%

2. Actual score–Self-assessed score

< −4	−4	−3	−2	−1	0	1	2	3	4	> 4
0.8%	1.8%	5.4%	11.4%	19.7%	35.1%	17.7%	5.6%	1.7%	0.6%	0.4%

3. Upside return–Downside return

0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	> 50%
32.7%	14.9%	9.2%	6.9%	5.2%	5.2%	4.3%	3.4%	3.1%	2.5%	12.7%

Panel B: Extrapolation

1. After a stock's price keeps rising for a while, I usually believe that the price will rise even further in the future.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
4.8%	26.9%	39.3%	22.8%	1.3%	5.0%

2. After a stock's price keeps falling for a while, I usually believe that the price will fall even further in the future.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
4.4%	29.1%	41.9%	18.2%	1.3%	5.3%

Panel C: Neglect of Trading Costs

1. Estimating the cost of a round-trip buy and sell at the value of 10,000 RMB

0–5	5–10	10–15	15–20	20–25	25–30	30–35	>35
17.3%	27.7%	23.6%	12.8%	8.4%	3.7%	2.1%	5.5%

2. How often do you consider transaction costs when you trade?

Never	Rarely	Sometimes	Often	Always	N/A
14.6%	37.7%	27.0%	13.8%	4.6%	2.5%

3. The bid-ask spread is one form of transaction cost (The bid-ask spread is the difference between the lowest ask price and the highest bid price).

Agree	Disagree	Don't Understand	Don't Know	N/A
59.8%	23.1%	8.5%	7.2%	1.4%

Table 4 Summary Statistics for Responses to Questions, Part I

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

Panel A: Gambling Preference

Blockbusters

1. When I trade stocks, I aim to select those stocks whose price would rise sharply in a short period of time so that I can make a lot of money quickly.

Strongly Agree 10.4%	Agree 25.4%	Neutral 33.9%	Disagree 23.0%	Strongly Disagree 4.6%	N/A 2.7%

Lotteries

2. When I trade stocks, I often think of them as lotteries; I am willing to accept small losses in exchange for the possibility of a big upside.

Strongly Agree 5.5%	Agree 24.9%	Neutral 27.2%	Disagree 32.5%	Strongly Disagree 7.3%	N/A 2.7%

Panel B: Realization Utility

Winners

1. Normally, if the price of a stock in your portfolio rose substantially since you bought it, which of these two actions would make you feel happier: holding on to the stock or selling that stock?

Sell 37.2%	Same 23.7%	Hold 25.3%	No Feeling 9.2%	N/A 4.5%

Losers

2. Normally, if the price of a stock in your portfolio dropped substantially since you bought it, which of these two actions would make you feel more painful: holding on to the stock or selling that stock?

Sell 22.9%	Same 28.0%	Hold 32.1%	No Feeling 12.2%	N/A 4.8%

Panel C: Sensation Seeking

Novelty

1. I feel excited about getting to know new stocks and new firms.

Strongly Agree 5.9%	Agree 20.3%	Neutral 43.9%	Disagree 21.0%	Strongly Disagree 3.2%	N/A 5.7%

Volatility

2. I feel excited about the stock market moving up and down.

Strongly Agree 5.4%	Agree 23.4%	Neutral 36.7%	Disagree 26.2%	Strongly Disagree 4.3%	N/A 4.1%

Table 5 Summary Statistics for Responses to Questions, Part II

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 8.7%	Rarely 27.9%	Sometimes 40.3%	Often 14.5%	Always 3.2%	N/A 5.4%

Dismissive of others' information

2. When you decide to trade a stock, how often do you worry that other investors know about the stock better than you do?

	Never 18.2%	Rarely 28.9%	Sometimes 32.3%	Often 12.6%	Always 2.5%	N/A 5.6%

Panel B: Social Interaction

Social influence

1. When you decide to trade a stock, how often are you influenced by your family members, friends, or other acquaintances?

	Never 11.6%	Rarely 31.2%	Sometimes 40.0%	Often 11.8%	Always 1.7%	N/A 3.8%

Advisor influence

2. When you decide to trade a stock, how often are you influenced by your investment advisors?

	Never 17.8%	Rarely 35.0%	Sometimes 35.8%	Often 7.2%	Always 1.2%	N/A 3.1%

Panel C: Others

Portfolio rebalancing needs

1. When you decide to trade a stock, how often is it that you need to rebalance your portfolio?

	Never 9.6%	Rarely 30.5%	Sometimes 44.5%	Often 9.5%	Always 1.7%	N/A 4.2%

Liquidity needs

2. When you decide to trade a stock, how often is it because you need money somewhere else?

	Never 7.0%	Rarely 25.9%	Sometimes 45.0%	Often 14.4%	Always 2.6%	N/A 5.1%

Panel D: Risk Aversion

1. Suppose you are the only income earner in the family, and you have a good job guaranteed to give you your current income every year for life. You are given the opportunity to take a new, equally good job. With a 50% chance it will double your income, and with a 50% chance, it will cut your income by 20%. Would you take the new job?

	Yes 51.6%	No 34.1%	Don't Know 11.3%	N/A 3.0%

2. Suppose the chances were 50% that it would double your income and 50% that it would cut it by 1/3. Would you take the new job?

	Yes 45.3%	No 37.5%	Don't Know 13.8%	N/A 3.4%

3. Suppose the chances were 50% that it would double your income and 50% that it would cut it by 1/2. Would you take the new job?

	Yes 26.0%	No 57.4%	Don't Know 13.2%	N/A 3.5%

Table 6 Summary Statistics for Responses to Questions, Part III

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).

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

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

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

Table 7 Summary Statistics of Turnover and Portfolio Returns

Note: Panel A shows the summary statistics of the monthly turnover, raw return, and net return for investors in the main sample between October 2018 and June 2019. The main sample includes 4,671 survey respondents that: 1) can be identified in the Shenzhen Stock Exchange centralized database, and 2) held at least one SZSE stock during the two-year window before the survey. Panel B shows the correlation coefficients among the three variables. See Table A2 in the Online Appendix for more details about variable definitions.

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

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

	Full sample (2018:01 – 2019:06)		Pre-survey (2018:01 – 2018:09)		Post-survey (2018:10 – 2019:06)	
Gambling preference, blockbusters	0.112***	0.109***	0.087***	0.086***	0.142***	0.139***
	(3.875)	(3.768)	(3.640)	(3.608)	(3.660)	(3.573)
Gambling preference, lotteries	0.038	0.019	0.025	0.018	0.051	0.029
	(1.257)	(0.653)	(1.013)	(0.727)	(1.237)	(0.698)
Male	-0.034	-0.033	-0.011	-0.01	-0.035	-0.034
	(-1.164)	(-1.140)	(-0.444)	(-0.403)	(-0.884)	(-0.866)
Controls	NO	YES	NO	YES	NO	YES
R2	0.004	0.023	0.004	0.017	0.004	0.020
N	4,145	4,145	3,435	3,435	3,550	3,550

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

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

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

Table 9 Validating Gambling Preferences Using Gambling Behavior

Note: This table studies the relationship between survey-based gambling preference and transaction-based gambling behavior. Gambling behavior is measured by the buy-volume (in RMB) weighted average of the past one-month (Panel A) or one-quarter (Panel B) number of up-limit hits based on the stocks an investor purchases in a given sample period. A purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase. Each panel presents OLS regression results based on three sample periods: full (January 2018 through June 2019), pre-survey (January 2018 through September 2018), and post-survey (October 2018 through June 2019). Gambling preference (blockbusters) equals one if an investor answers "Strongly agree" or "Agree" when asked if they aim to make a lot of money quickly through stock investment and zero otherwise. Gambling preference (lotteries) equals one if an investor answers "Strongly agree" or "Agree" when asked if they often think of stocks as lotteries and zero otherwise. See Table 5 for the exact phrasing of the survey questions. Control variables include age, gender, net worth, income, trading experience, account size, and education. T-statistics are based on robust standard errors and are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actual performance in 2017	4.104***							
	(5.332)							
Overplacement, performance	15.695***							
	(2.760)							
Financial literacy, dummy		11.922***						
		(3.127)						
Overplacement, literacy		1.729						
		(0.400)						
Miscalibration			1.116					
			(0.289)					
Underestimation of trading costs				-3.549				
				(-0.980)				
Do not consider trading costs					-2.143			
					(-0.548)			
Do not think bid-ask spread is a cost						-15.135***		
						(-4.254)		
Extrapolation, up							4.379	
							(1.110)	
Extrapolation, down								3.810
								(1.005)
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, blockbusters	10.924***							
	(2.878)							
Gambling preference, lotteries		2.750						
		(0.684)						
Realization utility, winners			7.188*					
			(1.874)					
Realization utility, losers				0.409				
				(0.093)				
Sensation seeking, novelty					10.184**			
					(2.270)			
Sensation seeking, volatility						11.984***		
						(2.885)		
Perceived information advantage							21.747***	
							(4.254)	
Dismissive of others' information								4.778
								(1.318)
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 rebalancing 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 10 Univariate Regression Results on Turnover

Note: In this table, we run univariate cross-sectional regressions of each investor's turnover (%) on survey-based trading motives. *T*-statistics are based on robust standard errors and are reported in parentheses. See Table A2 in the Online Appendix for more details about variable definitions.

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

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

Table 11: Regression Results Using the Full Set of Trading Motives

Note: In this table, we run a multivariate cross-sectional regression of each investor's turnover on all survey-based measures of trading motives. Control variables include age, gender, wealth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors and are reported in parentheses. See Table A2 in the Online Appendix for more details about variable definitions.

Panel A: Sort investors by their answers to the statement about gambling preference, blockbusters

	Monthly Turnover						Monthly Raw Returns	
	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 B: Sort investors by their perceived information advantages

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

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

Table 12 Additional Analysis of Gambling Preference, Blockbusters and Perceived Information Advantage

Note: In Panel A, 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 B, 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 each panel, we tabulate the summary statistics of monthly turnover ratios (monthly raw returns) for investors in each group. The last one or two rows report the differences between the bottom and top groups. When testing for the significance of the differences, we use robust standard errors.

Panel A: Monthly Turnover

Panel B: Characteristics of Stocks Bought

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

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

Table 13: Trading Characteristics for Investors Sorted on Transaction-Based Gambling Behavior

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

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

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

Table 14: Regressing Transaction-Based Gambling Behavior on Survey-Based Trading Motives

Note: In this table, we run a multivariate cross-sectional regression of each investor's transaction-based gambling behavior on survey-based trading motives. We construct a measure for transaction-based gambling behavior for each investor in two steps. First, for each of the nine months prior to the survey (January 2018 through September 2018), we first calculate the past one-month count of up-limit hits of the stock for each buy transaction and then take the transaction value weighted average across all buy orders. Second, we take the time-series average value weighted by monthly buy values. Control variables include age, gender, wealth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors and are reported in parentheses. See Table A2 in the Online Appendix for more details about variable definitions.

Online Appendix for

"Taming the Bias Zoo"

This appendix contains the following materials:

A1. *Survey questions (Chinese version)*

This section presents the Chinese version of our full survey. Figure A1 shows the cover page of the survey, and the full text of the survey follows.
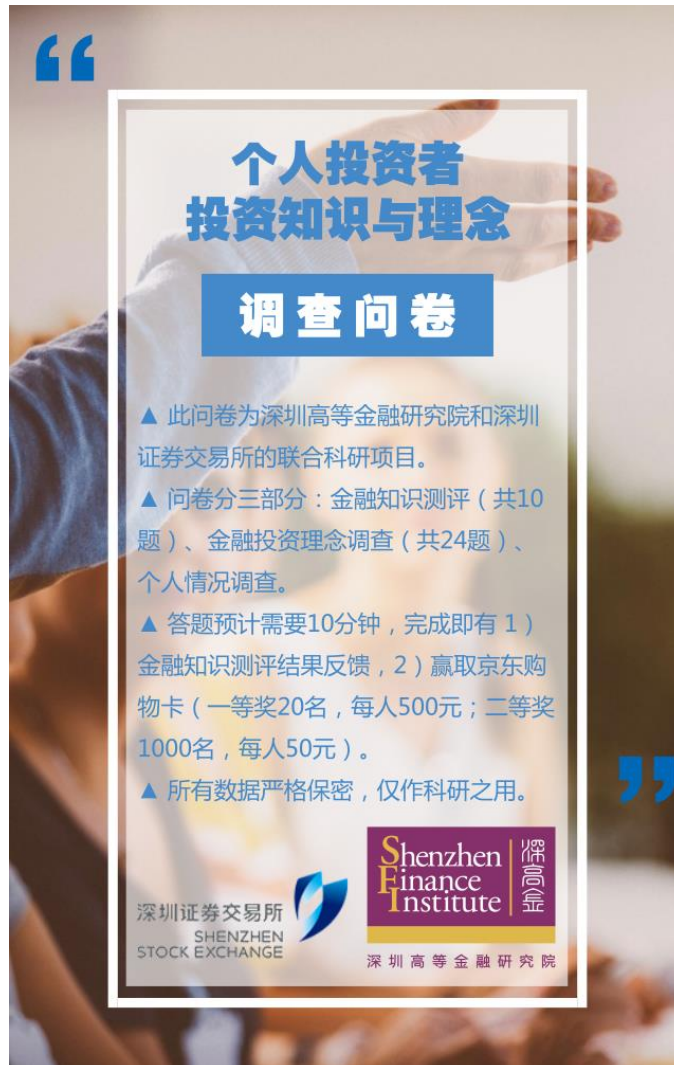
Figure A1: Cover Page of the Survey and Confidentiality Agreement

个人投资者投资知识与理念调查

1.　　以 1 到 7 来打分，1 分代表非常差，7 分代表非常好，请问您会给自己的金融知识水平打几分？
A.　　1-非常差
B.　　2-比较差
C.　　3-有点差
D.　　4-一般
E.　　5-有点好
F.　　6-比较好
G.　　7-非常好
H.　　拒绝回答

接下来的八道题，我们会对您的金融知识进行一个小测试，快来看看您能得几分吧（完成问卷即有结果反馈）！

2.　　假设您有 100 元的银行存款，银行的利息是每年 2%。如果您将钱一直存在银行里，5 年以后，您的银行账户里面会有多少钱？
A.　　多于 102 元
B.　　正好 102 元
C.　　少于 102 元
D.　　不知道
E.　　拒绝回答

3.　　假设您把钱存入银行，银行存款利息是每年 1%，物价水平预计每年上涨 2%。一年以后您能够买到的物品预计会
A.　　比现在多
B.　　和现在相等
C.　　比现在少
D.　　不知道
E.　　拒绝回答

4.　　请问您是否同意以下说法: 购买单支股票通常比购买股票型基金的风险更小
A.　　同意
B.　　不同意
C.　　不知道
D.　　拒绝回答

5.　　通常来说，以下哪种投资产品的收益波动最大？
A.　　银行存款
B.　　债券
C.　　股票

3

D. 不知道
E. 拒绝回答

6. 如果一个人购买了 B 公司的股票，以下正确的描述是
A. 他拥有 B 公司的一部分
B. 他借钱给 B 公司
C. 他对 B 公司债务承担无限责任
D. 不知道
E. 拒绝回答

7. 通常来说，当市场利率下降时，已发行债券的价格会
A. 上升
B. 不变
C. 下降
D. 不知道
E. 拒绝回答

8. 市盈率的定义是
A. 每股股价/每股收益
B. 每股股价/每股股息
C. 每股股价/每股账面价值
D. 不知道
E. 拒绝回答

9. 以下哪一项关于基金的描述是正确的？
A. 所有的基金都投资股票
B. 基金可以投资多种资产，例如股票和债券
C. 基金根据过往的业绩支付固定的收益
D. 不知道
E. 拒绝回答

10. 第 2 题至第 9 题这八道题中，您认为自己答对了几题？
A. 0 道
B. 1 道
C. 2 道
D. 3 道
E. 4 道
F. 5 道
G. 6 道
H. 7 道
I. 8 道

J.   不知道
K.   拒绝回答

11.   请问您认为有多少比例的股民在 2017 年的股市收益比您的高？
A.   低于 10%
B.   10% 到 20%
C.   20% 到 30%
D.   30% 到 40%
E.   40% 到 50%
F.   50% 到 60%
G.   60% 到 70%
H.   70% 到 80%
I.   80% 到 90%
J.   高于 90%
K.   不知道
L.   拒绝回答

12.   接下来，我们想请您对未来一年中国股市的走势做出预测。具体来说，在现在开始计算的未来**一年**内，您认为上证指数的累积收益会是多少？
为了方便您答题，以下我们给出了以 5% 为区间的多个选项，请您作出预测。
【下拉框中给出-50% 到 50% 的选项，以 5% 递增】
低于-50%;-50% 到 -45%; -45% 到 -40%; -40% 到 -35%; -35% 到 -30%; -30% 到 -25%; -25% 到 -20%; -20% 到 -15%; -15% 到 -10%; -10% 到 -5%; -5% 到 0%; 0% 到 5%; 5% 到 10%; 10% 到 15%; 15% 到 20%; 20% 到 25%; 25% 到 30%; 30% 到 35%; 35% 到 40%; 40% 到 45%; 45% 到 50%; 高于 50%;不知道; 拒绝回答

13.   接上题，我们请您预测上证指数上涨的空间。假设上证指数未来一年的累积收益有**一成**的概率能**高于**某个水平，您认为这个水平会是多少？
为了方便您答题，以下我们给出了以 5% 为区间的多个选项，请您作出预测。
【下拉框中给出-50% 到 50% 的选项，以 5% 递增】
低于-50%;-50% 到 -45%; -45% 到 -40%; -40% 到 -35%; -35% 到 -30%; -30% 到 -25%; -25% 到 -20%; -20% 到 -15%; -15% 到 -10%; -10% 到 -5%; -5% 到 0%; 0% 到 5%; 5% 到 10%; 10% 到 15%; 15% 到 20%; 20% 到 25%; 25% 到 30%; 30% 到 35%; 35% 到 40%; 40% 到 45%; 45% 到 50%; 高于 50%;不知道; 拒绝回答

14.   接上题，我们请您预测上证指数下跌的空间。假设上证指数未来一年的累积收益有**一成**的概率会**低于**某个水平，您认为这个水平会是多少？
为了方便您答题，以下我们给出了以 5% 为区间的多个选项，请您作出预测。

【下拉框中给出-50%到 50%的选项，以 5%递增】
低于-50%;-50% 到 -45%; -45% 到 -40%; -40% 到 -35%; -35% 到 -30%; -30% 到 -25%; -25%
到 -20%; -20% 到 -15%; -15% 到 -10%; -10% 到 -5%; -5% 到 0%; 0% 到 5%; 5% 到 10%;
10% 到 15%; 15% 到 20%; 20% 到 25%; 25% 到 30%; 30% 到 35%; 35% 到 40%; 40% 到
45%; 45% 到 50%; 高于 50%;不知道; 拒绝回答

15.　　在实际交易中，假设您以每股 100 元的价格购入了 100 股 A 股票，然后再以同样
的价格全部卖出。请问这两笔交易您需要支付的交易费用（包括券商佣金和相关税费）总
共是多少？ 为了方便您答题，以下我们给出了以 5 元为区间的多个选项，请选择您认为
正确的区间。
【下拉框中给出 0 到 50 元的选项，以 5 元递增】
0 - 5 元; 5 - 10 元; 10 - 15 元; 15 - 20 元; 20 - 25 元; 25 - 30 元; 30 - 35 元; 35 - 40 元; 40 - 45
元; 45 - 50 元; 50 元以上; 不知道; 拒绝回答

16.　　请问您在做出买卖股票的决定时，有多少时候考虑过交易费用？
A.　　从来没有
B.　　很少
C.　　有时候
D.　　经常
E.　　总是
F.　　不知道
G.　　拒绝回答

17.　　请问您是否同意股票市场上的买卖价差是一种交易成本？买卖价差是指某一个时间
点上，市场上限价委托单的最低卖价和最高买价之间的差值。
A.　　同意
B.　　不同意
C.　　不明白买卖价差的含义
D.　　不知道
E.　　拒绝回答

18.　　请问您是否同意以下说法？我平时买卖股票时，寄希望于能通过买到短期内连续上
涨的少数个股实现快速致富 。
A.　　非常同意
B.　　同意
C.　　中立
D.　　不同意
E.　　非常不同意

F. 不知道
G. 拒绝回答

19. 请问您是否同意以下说法？我平时买卖股票时，抱着一种买彩票的心态，愿意承担一定的亏损以换取小概率的高收益 。
A. 非常同意
B. 同意
C. 中立
D. 不同意
E. 非常不同意
F. 不知道
G. 拒绝回答

20. 通常情况下，当您持有的一支股票市场价显著**高于**您的平均购买价时，**卖出**或**继续持有**该股票这两种情形中，哪一种让您感到更快乐？
A. 卖出该股票
B. 一样快乐
C. 继续持有该股票
D. 没什么感觉
E. 不知道
F. 拒绝回答

21. 通常情况下，当您持有的一支股票市场价显著**低于**您的平均购买价时，**卖出**或**继续持有**该股票这两种情形中，哪一种让您感到更痛苦？
A. 卖出该股票
B. 一样痛苦
C. 继续持有该股票
D. 没什么感觉
E. 不知道
F. 拒绝回答

22. 请问您是否同意以下说法？接触新的股票和公司让我感到兴奋。
A. 非常同意
B. 同意
C. 中立
D. 不同意
E. 非常不同意
F. 不知道

G.   拒绝回答

23.   请问您是否同意以下说法？我享受股票价格的上涨和下跌带来的刺激感。
A.   非常同意
B.   同意
C.   中立
D.   不同意
E.   非常不同意
F.   不知道
G.   拒绝回答

24.   请问您在做出买卖股票的决定时，有多少时候是因为您认为自己比其他投资者更了解这些股票？
A.   从来没有
B.   很少
C.   有时候
D.   经常
E.   总是
F.   不知道
G.   拒绝回答

25.   请问您在买卖股票的时候，有多少时候担心其他投资者知道一些您不知道的消息？
A.   从来没有
B.   很少
C.   有时
D.   经常
E.   总是
F.   不知道
G.   拒绝回答

26.   请问您是否同意以下叙述？当某支股票价格持续**上涨**一段时间后，我通常认为其价格在短期内还会继续**上涨**。
A.   非常同意
B.   同意
C.   中立
D.   不同意
E.   非常不同意
F.   不知道

G.　拒绝回答

27.　请问您是否同意以下叙述？当某支股票价格持续**下跌**一段时间后，我通常认为其价格在短期内还会继续**下跌**。
A.　非常同意
B.　同意
C.　中立
D.　不同意
E.　非常不同意
F.　不知道
G.　拒绝回答

28.　请问您在做出买卖股票的决定时，有多少时候是因为亲人，朋友或熟人的推荐或影响？
A.　从来没有
B.　很少
C.　有时候
D.　经常
E.　总是
F.　不知道
G.　拒绝回答

29.　请问您在做出买卖股票的决定时，有多少时候是因为投资顾问（如证券公司客户经理，银行理财经理，炒股专家等）的推荐或影响？
A.　从来没有
B.　很少
C.　有时候
D.　经常
E.　总是
F.　不知道
G.　拒绝回答

30.　请问您在做出买卖股票的决定时，有多少时候是因为需要调整投资组合？
A.　从来没有
B.　很少
C.　有时候
D.　经常
E.　总是

F.　不知道
G.　拒绝回答

31.　请问您在做出卖出股票的决定时，有多少时候是因为其他地方需要用钱？
A.　从来没有
B.　很少
C.　有时候
D.　经常
E.　总是
F.　不知道
G.　拒绝回答

32.　假设您是家庭中唯一的收入来源，您有一个很好的工作可以保证今后每年给您稳定不变的收入。现在您有机会换一个同样好的新工作，这份工作有50%的可能将您的收入增加**一倍**，同时也有50%的可能将收入减少**五分之一**。您会换新工作吗？
A.　会
B.　不会
C.　不知道
D.　拒绝回答

33.　接上题，假设新工作有50%的可能将您的收入增加**一倍**，50%的可能将您的收入减少**三分之一**。您会换新工作吗？
A.　会
B.　不会
C.　不知道
D.　拒绝回答

34.　接上题，假设新工作有50%的可能将您的收入增加**一倍**，50%的可能将您的收入减少**一半**。您会换新工作吗？
A.　会
B.　不会
C.　不知道
D.　拒绝回答

【针对实验组】
35.　非常感谢您完成答卷！快来看看巴菲特给个人投资者的建议：

数据显示，A股市场中百分之八十以上的交易量来自个人投资者，大量个人投资者存在频繁交易的行为。然而研究发现：交易频率越高的投资者，收益往往越差。

为什么会这样呢？

那是因为频繁交易有两大负面影响：
1.频繁交易会产生高额的交易费用。
2.股市的短期走势是无法预测的。个人投资者由于缺乏经验和信息，把握不好买入卖出的时机，容易高买低卖，造成亏损。

交易费用详解：

假设：证券公司收取的佣金是万三

如果你以相同的价格买卖股票一次，那么：

则交易费用占初始投资额的百分比＝
买入佣金0.03%＋卖出佣金0.03%
＋印花税0.1%＝0.16%

我明白了！
如果我每月进行一次买入再卖出的操作，交易费用就占到我初始投资的1.92%。如果每周一次，则交易费用高达8.32%！

对的！你知道吗？过去15年间，沪深300指数的年平均收益为8.96%。

这样呀！这意味着什么呢？

这意味着，
一个和市场收益持平的投资者，
如果每月交易一次，
则交易费用占总收益的五分之一。
如果每周交易一次，
则交易费用几乎和总收益持平。
而且！很多投资者的收益是低于市场收益的，那么频繁交易的后果很有可能就是亏损。

股神巴菲特给个人投资者的建议

获得更好投资回报的秘诀就是：
买进一只股票，然后忘掉它。
如果试图频繁交易股票，
那么将赚不到钱。
如果你经常买卖股票，
它将以交易佣金和税收的形式
拿走很大一部分回报。所以，
最好是买好股票并持有很久。

咱们来看看您是否理解了图片中的内容吧！在上图的例子中，请问如果证券公司的佣金为万三，您每周进行一次买入股票，再以相同价格全部卖出的交易的话，一年之后，交易费用会占到您的初始投资额的多少呢？

A.　　0%
B.　　0.16%
C.　　1.92%
D.　　8.32%
E.　　不知道

36.　　上一题的答案是8.32%！交易费用不可忽视，请尽量避免频繁交易。

感谢您完成关于投资知识及理念的题目，下面请帮助我们了解一下您的个人情况吧！完成这一部分之后，会有您的答题成绩，更可参与抽奖。一等奖20名，每人500元，二等奖1000名，每人50元。若中奖，我们会将相应金额的京东电子购物卡通过短信发送至您的手机，请注意查收。请继续吧！若此时放弃，则答卷作废。

请问您开户的券商是？
华泰证券;中信证券;国泰君安;广发证券;银河证券;海通证券;招商证券;申万宏源;国信证券;中信建投

37.　　请问您开户的营业部是？

38.　　请问您的出生于哪一年？
39.　　请问您的出生于几月？
40.　　请问您的出生于几号？
41.　　请问您的性别是
A.　　男
B.　　女

42.　　请问您的最高学历是？
A.　　没有正式入学
B.　　小学
C.　　初中
D.　　高中/中专/技校
E.　　大专/成人大专/夜大/电大
F.　　大学本科
G.　　研究生及以上

43.　　请问您的家庭净资产是（总资产-总负债）多少？
A.　　2万元以下
B.　　2万-10万元（不含10万）

C. 10万元-50万元（不含50万）
D. 50万元-100万元（不含100万）
E. 100万元-500万元（不含500万）
F. 500万-1000万元（不含1000万）
G. 1000万元及以上

44. 请问您的家庭年收入是多少？
A. 2万元以下
B. 2万-10万元（不含10万）
C. 10万-20万（不含20万）
D. 20万-50万（不含50万）
E. 50万-100万（不含100万）
F. 100万-200万（不含200万）
G. 200万-1000万（不含1000万）
H. 1000万元及以上

45. 请问您的出生省份是？（注：在选择出生省市时，以现有行政划分为准。如直辖市重庆市在1997年以前隶属四川省。若张三在1988年出生于该市，则应按照现有行政区划选择重庆市，而非四川省）

46. 请问您的出生城市是？（注：在选择出生省市时，以现有行政划分为准。如直辖市重庆市在1997年以前隶属四川省。若张三在1988年出生于该市，则应按照现有行政区划选择重庆市，而非四川省）

这是最后一道题啦，非常感谢您完成问卷。请在下一页中填入您的姓名和手机号码参与抽奖以及查看答卷成绩。若中奖，我们会将京东电子购物卡通过短信发送至您的手机，请注意查收。若此时放弃，则答卷作废

47. 请问您的姓名是？
48. 请问您的手机号是？

A2. *Survey questions (English version)*

We translated the original Chinese survey into English, with slight modifications made to accommodate the different institutional details of the U.S. market. This translated version was used for a smaller survey of U.S. investors on Amazon Mechanical Turk. The English version of the survey is attached below.

1. How do you assess the level of your financial literacy?
A. Very low
B. Low
C. Slightly low
D. Fair
E. Slightly high
F. High
G. Very high
H. Decline to answer

2. Normally, which of the following assets displays the highest fluctuation over time?
A. Savings accounts
B. Bonds
C. Stocks
D. Do not know
E. Decline to answer

3. Which of the following statements is correct? If somebody buys a stock of firm B in the stock market:
A. He owns a part of firm B
B. He has lent money to firm B
C. He is fully liable for firm B's total debts
D. Do not know
E. Decline to answer

4. Normally, when the market interest rate falls, the price of an existing bond will:
A. Go up
B. Remain the same
C. Go down
D. Do not know
E. Decline to answer

5. What is the P/E ratio?
A. Price per share/Earnings per share
B. Price per share/Dividend per share
C. Price per share/Book value per share
D. Do not know
E. Decline to answer

6. Which of the following statements about mutual funds is correct?
A. All mutual funds invest in stocks
B. Mutual funds can invest in multiple asset classes, such as stocks and bonds
C. Mutual funds pay fixed returns based on their past performance
D. Do not know
E. Decline to answer

7. From Question 2 to 9, how many questions do you think you have answered correctly?
A. 0
B. 1
C. 2
D. 3
E. 4
F. 5
G. 6
H. 7
I. 8
J. Do not know
K. Decline to answer

8. What fraction of retail investors do you think earned higher returns than you in 2018?
A. Lower than 10%
B. 10% to 20%
C. 20% to 30%
D. 30% to 40%
E. 40% to 50%
F. 50% to 60%
G. 60% to 70%
H. 70% to 80%
I. 80% to 90%
J. Higher than 90%
K. Do not know
L. Decline to answer

9. We now ask you to make predictions about the U.S. stock market over the next year. What do you think is the expected return for the S&P 500 Index over the next year?

10. As a follow-up question, we ask you to predict the upside potential for the S&P 500 Index. Assume that with 10% chance the return of the S&P 500 Index over the next year will be higher than a certain level. What do you think this level will be?

11. As a follow-up question, we ask you to predict the downside potential for the S&P 500 index. Assume that with 10% chance the return of the S&P 500 Index over the next year will be lower than a certain level. What do you think this level will be?

12. Suppose you purchase 100 shares of stock A at $100 per share and then sell all the shares at the same price. Based on your trading experience, what's your estimated total transaction cost (including the commission fees) in $?
A. 0-5
B. 5-10
C. 10-15
D. 15-20
E. 20-25
F. 25-30
G. 30-35
H. 35-40
I. 40-45
J. 45-50
K. 50 or more
L. Do not know
M. Decline to answer

13. How often do you consider transaction costs when you trade stocks?
A. Never
B. Rarely
C. Sometimes
D. Often
E. Always
F. Do not know
G. Decline to answer

14. Do you agree the following statement? 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.)
A. Agree
B. Disagree
C. Do not understand what the bid-ask spread is
D. Do not know
E. Decline to answer

15. Do you agree with the following statement? 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.
A. Strongly agree
B. Agree
C. Neutral
D. Disagree
E. Strongly disagree
F. Do not know
G. Decline to answer

16. Do you agree with the following statement? 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.
A. Strongly agree
B. Agree
C. Neutral
D. Disagree
E. Strongly disagree
F. Do not know
G. Decline to answer

17. 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?
A. Holding on to the stock
B. The same
C. Selling the stock
D. I am indifferent to price changes
E. Do not know
F. Decline to answer

18. 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?
A. Holding on to the stock
B. The same
C. Selling the stock
D. I am indifferent to price changes
E. Do not know
F. Decline to answer

19. Do you agree with the following statement? I feel excited about getting to know new stocks and new firms.
A. Strongly agree
B. Agree

C. Neutral
D. Disagree
E. Strongly disagree
F. Do not know
G. Decline to answer

20. Do you agree with the following statement? I feel excited about the stock market moving up and down.
A. Strongly agree
B. Agree
C. Neutral
D. Disagree
E. Strongly disagree
F. Do not know
G. Decline to answer

21. When you decide to trade a stock, how often do you believe that you know the stock better than others?
A. Never
B. Rarely
C. Sometimes
D. Often
E. Always
F. Do not know
G. Decline to answer

22. When you decide to trade a stock, how often do you worry that other investors know about the stock better than you do?
A. Never
B. Rarely
C. Sometimes
D. Often
E. Always
F. Do not know
G. Decline to answer

23. Do you agree with the following statement? After a stock's price keeps rising for a while, I usually believe that the price will rise even further in the future.
A. Strongly agree
B. Agree
C. Neutral
D. Disagree
E. Strongly disagree
F. Do not know
G. Decline to answer

24. Do you agree with the following statement? After a stock's price keeps dropping for a while, I usually believe that the price will drop even further in the future.
A. Strongly agree
B. Agree
C. Neutral
D. Disagree
E. Strongly disagree
F. Do not know
G. Decline to answer

25. When you decide to trade a stock, how often are you influenced by your family members, friends, or other acquaintances?
A. Never
B. Rarely
C. Sometimes
D. Often
E. Always
F. Do not know
G. Decline to answer

26. When you decide to trade a stock, how often are you influenced by your investment advisors?
A. Never
B. Rarely
C. Sometimes
D. Often
E. Always
F. Do not know
G. Decline to answer

27. When you decide to trade a stock, how often is it that you need to rebalance your portfolio?
A. Never
B. Rarely
C. Sometimes
D. Often
E. Always
F. Do not know
G. Decline to answer

28. When you decide to trade a stock, how often is it that you need money somewhere else?
A. Never
B. Rarely
C. Sometimes
D. Often
E. Always

F. Do not know
G. Decline to answer

29. 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?
A. Yes
B. No
C. Do not know
D. Decline to answer

30. 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?
A. Yes
B. No
C. Do not know
D. Decline to answer

31. 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?
A. Yes
B. No
C. Do not know
D. Decline to answer

A3. *Distribution of respondents across brokers and provinces*

Table A1 reports the distribution of respondents across brokers and provinces. By design, respondents were evenly distributed across the ten brokers, with only slight variation. In terms of geographic variation, areas that are more financially developed (e.g., Guangdong, Zhejiang, Jiangsu, and Shanghai) are more represented in our sample.

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

Table A1 Distribution of Survey Respondents Across Brokers and Provinces

A4. *Relationship between financial literacy score and time spent to complete the survey*

Figure A2 plots the distribution of the total amount of time spent on the survey: it took a median respondent about eight minutes to complete the survey, and 95% of respondents finished within 20 minutes. However, we find that respondents who spent less than three minutes on the survey experienced a sharp drop in their financial literacy score, suggesting that they may have shirked during the survey. In the main analysis of the paper, we dropped these observations, which reduced our sample size to 11,268.
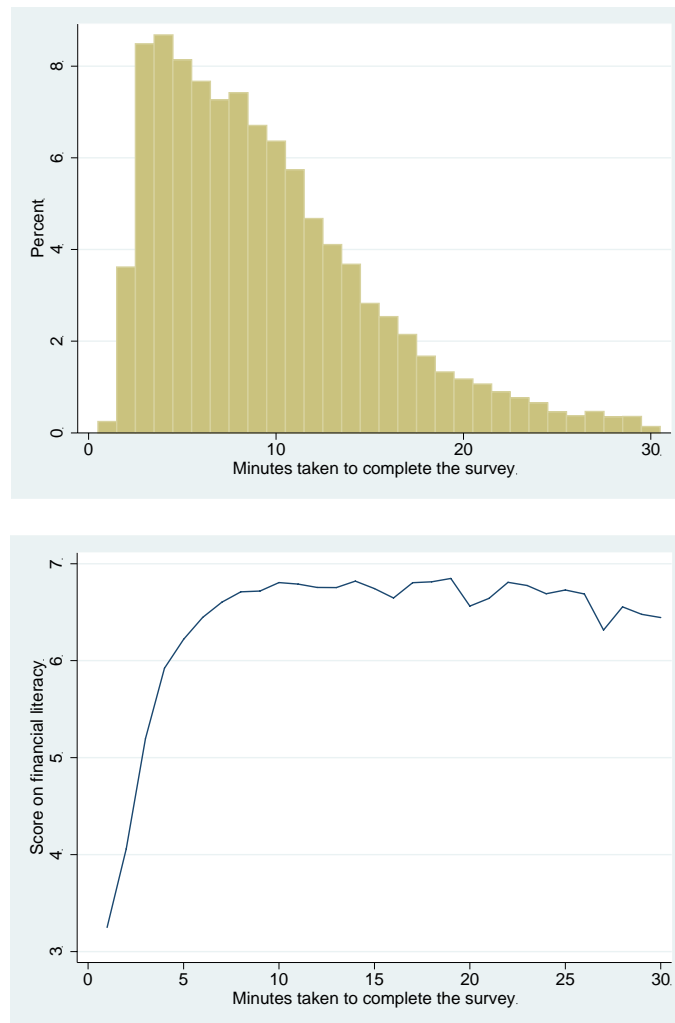
Figure A2 Relationship Between Financial Literacy Score and the Time Taken to Complete the Survey

A5. Construction of variables

Table A2 provides a detailed description about the construction of each variable we use in the paper.

Variable	Definition
Over-placement, performance	1: self-assessed performance rank among the investor population in 2017 is higher than her actual rank 0: otherwise
Over-placement, literacy	1: self-assessed score in the financial literacy test is higher than her actual score 0: otherwise
Miscalibration of uncertainty	Investors are asked 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 of an investor's forecast for the market's next year return. 1: confidence interval is narrower than 76%, the estimated confidence interval from the historical data 0: otherwise
Underestimation of transaction cost	Investors are asked to estimate the total transaction cost from buying 10,000 RMB worth of stock and then selling everything at the same value. 1: answer falls below the estimated true cost (15 to 26 RMB) 0: otherwise
Do not consider transaction cost	1: answers "Rarely" or "Never" when asked how often she considers transaction costs when trading a stock 0: otherwise
Do not think bid-ask spread is a cost	1: answers "Disagree", "Do not understand what the bid-ask spread is", "Do not know" or "Decline to answer" when asked if she agrees that the bid-ask spread is one form of transaction cost 0: otherwise
Extrapolation, up	1: answers "Strongly agree" or "Agree" when asked if she believes stock price will rise even further in the future after it keeps rising for a while 0: otherwise
Extrapolation, down	1: answers "Strongly agree" or "Agree" when asked if she believes the stock price will drop even further in the future after it keeps dropping for a while 0: otherwise
Gambling preference, blockbusters	1: answers "Strongly agree" or "Agree" when asked if she aims to make a lot of money quickly through stock investment

0: otherwise

Gambling preference, lotteries	1: answers "Strongly agree" or "Agree" when asked if she often think of stocks as lotteries 0: otherwise
Realization utility, winners	1: answers "Selling the stock" when asked if the price of a stock in her portfolio rose substantially since she bought it, which of these two actions make her fell happier: holding on to the stock, or selling that stock? 0: otherwise
Realization utility, losers	1: answers "Selling the stock" when asked if the price of a stock in her portfolio dropped substantially since she bought it, which of these two actions would make her feel more painful: holding on to the stock, or selling that stock? 0: otherwise
Sensation seeking, novelty	1: answers "Strongly agree" or "Agree" when asked if she feels excited about getting to know new stocks and new firms 0: otherwise
Sensation seeking, volatility	1: answers "Strongly agree" or "Agree" when asked if she feels excited about the stock market moving up and down 0: otherwise
Perceived information advantage	1: answers "Often" or "Always" when asked how often she believes that she knows the stock better than others when she decides to trade a stock 0: otherwise
Dismissive of others' information	1: answers "Rarely" or "Never" when asked how often she worries that other investors know about the stock better than she does when she decides to trade a stock 0: otherwise
Social influence	1: answers "Often" or "Always" when asked how often she was influenced by her family members, friends, or other acquaintances when she decides to trade a stock 0: otherwise
Advisor influence	1: answers "Often" or "Always" when asked how often she was influenced by her investment advisors when she decides to trade a stock 0: otherwise
Portfolio rebalancing needs	1: answers "Often" or "Always" when asked how often it is because she need to rebalance her portfolio when she decides to trade a stock 0: otherwise

Liquidity needs	1: answers "Often" or "Always" when asked how often it is because she need to rebalance her portfolio when she decides to trade a stock 0: otherwise
Risk aversion	Investors are asked 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. 1: answers "No" in all three questions 0: otherwise
Expected future one-year market return	Investors' prediction for the future one-year market returns, ranging from 1 (lower than −50%) to 22 (higher than 50%).
Financial literacy, dummy	1: scores 7 or higher out 8 questions designed to test her financial literacy 0: otherwise
Other variables	**Definition**
Turnover	For each month, we calculate an investor's turnover ratio as the total value of buy and sell orders divided by the maximum value of investment in that month. We then take the time series average of monthly turnover ratios weighted by monthly maximum value of investment.
Raw return	For each month, we calculate an investor's raw return as raw profit divided by the maximum value of investment in that month. We then take the time series average of monthly raw returns weighted by monthly maximum value of investment. Monthly raw profit = ending portfolio balance + value of sell orders + dividends − beginning portfolio balance − value of buy orders
Net return	For each month, we calculate an investor's net return as net profit divided by the maximum value of investment in that month. We then take the time series average of monthly net returns weighted by monthly maximum value of investment. Monthly net profit = ending portfolio balance + value of sell orders + dividends − beginning portfolio balance − value of buy orders − comission fees − stamp duty
Volume-weighted Past One-month Count of Up-limit Hits Based on Initial Buys	For each month, we first calculate the past one-month # of up-limit hits of the stock for each initial buy transaction and then take the transaction value weighted average across all initial buy orders. We then take the time-series average value weighted by monthly initial buy values. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase.

Volume-weighted Past One-quarter Count of Up-limit Hits Based on Initial Buys	For each month, we first calculate the past one-quarter # of up-limit hits of the stock for each initial buy transaction and then take the transaction value weighted average across all initial buy orders. We then take the time-series average value weighted by monthly initial buy values. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase.
Volume-weighted Past One-month Return Based on Initial Buys	We first calculate the past one-month return of the stock for each initial buy order and then take the average for all initial buy orders in a month weighted by the transaction value. We then calculate the time-series average weighted by the total value of monthly initial buy orders. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase.
Volume-weighted Past One-quarter Return Based on Initial Buys	We first calculate the past one-quarter return of the stock for each initial buy order and then take the average for all initial buy orders in a month weighted by the transaction value. We then calculate the time-series average weighted by the total value of monthly initial buy orders. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase.
Volume-weighted Past One-month Return Volatility Based on Initial Buys	We first calculate the past one-month return volatility of the stock for each initial buy order and then take the average for all initial buy orders in a month weighted by the transaction value. We then calculate the time-series average weighted by the total value of monthly initial buy orders. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase.
Volume-weighted Past One-quarter Return Volatility Based on Initial Buys	We first calculate the past one-quarter return volatility of the stock for each initial buy order and then take the average for all initial buy orders in a month weighted by the transaction value. We then calculate the time-series average weighted by the total value of monthly initial buy orders. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase.
After	The dummy equals one for the period from October 2018 onward and zero otherwise.
Treated	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 include educational content about how frequent trading negatively affect their investment performance. The dummy equals one if an investor is in the treated group and correctly answered the follow-up question designed to test if the respondent understands the content of the message. The dummy equals zero if an investor is in the control group.

Table A2 Definition of Variables

A6. Comparison between Chinese and U.S. investors

To study the differences between Chinese and US investors, we run the English version of our survey on Mechanical Turk with a small sample of 400 investors. The results are reported in Table A3. American investors are more "sophisticated" in three aspects: they consider transaction cost more often when they trade, they are less dismissive of others' information advantage, and they have a closer relationship with their advisors. At the same time, on several fronts, American retail investors are more prone to behavioral biases than their Chinese counterparts: they exhibit stronger gambling preferences, are more prone to realization utility and sensation seeking, and display slightly stronger extrapolative beliefs.

Survey question	China sample	US sample	Difference
# of responses	12,856	400	
Over-placement, literacy	0.25	0.20	0.04**
Miscalibration of uncertainty	0.67	0.88	-0.21***
Do not consider transaction cost	0.51	0.25	0.26***
Do not think bid-ask spread is a cost	0.42	0.40	0.02
Extrapolation, up	0.32	0.40	-0.08***
Extrapolation, down	0.33	0.35	-0.02
Gambling preference, blockbusters	0.36	0.52	-0.16***
Gambling preference, lotteries	0.30	0.52	-0.22***
Realization utility, winners	0.35	0.55	-0.20***
Realization utility, losers	0.23	0.41	-0.18***
Sensation seeking, novelty	0.26	0.74	-0.48***
Sensation seeking, volatility	0.28	0.65	-0.37***
Perceived information advantage	0.18	0.15	0.03
Dismissive of others' information	0.45	0.25	0.20***
Social influence	0.14	0.12	.02
Advisor influence	0.08	0.31	-0.22***
Portfolio rebalancing needs	0.18	0.21	-0.03
Liquidity	0.12	0.19	-0.06***
Risk aversion	0.19	0.29	-0.09***

Table A3 Differences in Survey Responses between Chinese and American Investors

A7. *Comparison between the pre-merged sample and the merged sample*

Table A4 reports the comparison between the pre-merged sample and the merged sample. The merged sample represents the 6,013 investors we are able to locate in the Shenzhen Stock Exchange's centralized database. Overall, the merged sample and the pre-merged sample are very similar in their survey responses. For certain responses such as risk aversion and return expectations, the difference is significant, but the magnitude is generally small. Overall, we don't find strong evidence that our merging process is biased towards certain investor groups.

Survey question	Merged sample	Pre-merged sample	Difference
Over-placement, literacy	0.24	0.26	-0.02**
Miscalibration of uncertainty	0.67	0.66	0.01
Underestimation of transaction cost	0.68	0.67	0.00
Do not consider transaction cost	0.52	0.50	0.02*
Do not think bid-ask spread is a cost	0.41	0.43	-0.01
Extrapolation, up	0.32	0.31	0.01
Extrapolation, down	0.34	0.32	0.02*
Gambling preference, with prob. weighting	0.37	0.36	0.00
Gambling preference, without prob. weighting	0.31	0.30	0.01
Realization utility, winner	0.36	0.35	0.01
Realization utility, loser	0.23	0.22	0.01
Sensation seeking, novelty	0.25	0.26	-0.01
Sensation seeking, volatility	0.28	0.28	0.01
Perceived information advantage	0.17	0.18	-0.01
Dismissive of others' information	0.46	0.45	0.01
Social influence	0.14	0.14	-0.01
Advisor influence	0.08	0.09	-0.01**
Portfolio rebalancing needs	0.17	0.19	-0.02***
Liquidity	0.11	0.14	-0.03***
Risk aversion	0.20	0.18	0.02***
Expected future one-year market return	9.92	9.25	0.67***

Table A4 Comparison between the Pre-merged Sample and the Merged Sample

A8. *Additional validation exercises*

In the main text of the paper, we report evidence from the validation exercises for gambling preferences. In this section, we present evidence from three other validation exercises: one for extrapolative beliefs, one for risk aversion, and one for return expectations.

We measure extrapolative behavior by the volume-weighted past return among all the stocks bought by an investor. Table A5 reports the results when regressing transaction-based extrapolative behavior on survey-based extrapolative beliefs, where, in measuring extrapolative behavior, Panel A uses the past one-month return and Panel B uses the past one-quarter return. Indeed, investors who report having extrapolative beliefs exhibit stronger extrapolative behavior: on average, the stocks they purchase experience 1% higher returns in the preceding month and more than 2% higher returns in the preceding quarter, and this holds in both pre-survey and post-survey samples. The two measures of extrapolation have equally strong explanatory power for extrapolative behavior.

Table A6 reports the relationship between risk aversion and the average return volatility of stocks in the portfolio. Overall, investors with a higher risk aversion tend to hold stocks with less volatility. This further validates that survey-based risk aversion is useful for explaining portfolio-level characteristics. The results are robust to alternative specifications of return volatility and different sample periods.

Table A7 reports the relationship between return expectations and changes in stock holdings. Consistent with the evidence from Giglio et al. (2020), we find that higher return expectations indeed lead to an increase in stock holdings. However, we also reaffirm their finding that the sensitivity of holding changes to return expectations is small in magnitude. Note that, unlike Giglio et al. (2020), which studies holdings of passive funds at Vanguard, we study holdings of individual stocks, which are arguably traded more actively.

29

Table A5 Validating Extrapolative Belief Using Trend-Chasing Behavior

Panel A: Volume-Weighted Past One-Month Return Based on Initial Buys

	Full sample (2018:01 – 2019:06)		Pre-survey (2018:01 – 2018:09)		Post-survey (2018:10 – 2019:06)	
Extrapolation, up	0.011**	0.011**	0.012***	0.013***	0.011*	0.011*
	(2.170)	(2.134)	(2.689)	(2.902)	(1.668)	(1.704)
Extrapolation, down	-0.014***	-0.014***	-0.012***	-0.012***	-0.014**	-0.014**
	(-2.854)	(-2.816)	(-2.740)	(-2.697)	(-2.142)	(-2.237)
Male		-0.028***				
		(-3.685)				
Controls	NO	YES	NO	YES	NO	YES
R2	0.001	0.017	0.002	0.016	0.001	0.017
N	4,142	4,142	3,432	3,432	3,550	3,550

Panel B: Volume-Weighted Past One-Quarter Return Based on Initial Buys

	Full sample (2018:01 – 2019:06)		Pre-survey (2018:01 – 2018:09)		Post-survey (2018:10 – 2019:06)	
Extrapolation, up	0.020**	0.020**	0.019***	0.022***	0.026**	0.028***
	(2.406)	(2.419)	(2.999)	(3.446)	(2.451)	(2.597)
Extrapolation, down	0.021***	0.020**	0.020***	0.021***	0.021**	0.021**
	(2.615)	(2.532)	(3.112)	(3.316)	(2.032)	(2.091)
Male	-0.028***	-0.028***	-0.037***	-0.036***	-0.030***	-0.029***
	(-3.685)	(-3.638)	(-5.848)	(-5.801)	(-3.113)	(-3.031)
Controls	NO	YES	NO	YES	NO	YES
R2	0.001	0.023	0.003	0.033	0.002	0.021
N	4,136	4,136	3,428	3,428	3,544	3,544

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

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

30

Panel A: Average Past One-month Volatility Based on Stock Holdings

	Full sample		Pre-survey		Post-survey	
	(2018:01 − 2019:06)		(2018:01 − 2018:09)		(2018:10 − 2019:06)	
Risk aversion	-0.068***	-0.054**	-0.041	-0.033	-0.084***	-0.069**
	(-2.932)	(-2.327)	(-1.342)	(-1.075)	(-3.071)	(-2.515)
Male		0.043**		0.052**		0.029
		(2.185)		(2.125)		(1.035)
Controls	NO	YES	NO	YES	NO	YES
R2	0.002	0.027	0.001	0.016	0.002	0.016
N	4,381	4,381	3,807	3,807	3,956	3,956

Panel B: Average Past One-quarter Volatility Based on Stock Holdings

	Full sample		Pre-survey		Post-survey	
	(2018:01 − 2019:06)		(2018:01 − 2018:09)		(2018:10 − 2019:06)	
Risk aversion	-0.065***	-0.050**	-0.059*	-0.048	-0.077***	-0.062**
	(-2.614)	(-2.001)	(-1.771)	(-1.433)	(-2.662)	(-2.125)
Male		0.046**		0.062***		0.021
		(2.200)		(2.306)		(0.674)
Controls	NO	YES	NO	YES	NO	YES
R2	0.002	0.029	0.001	0.018	0.001	0.014
N	4,381	4,381	3,807	3,807	3,956	3,956

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

Table A6 Validation of Risk Aversion

Note: This table studies the relationship between investors' risk attitude and return volatilities of the stocks they buy. The dependent variables are buy-volume (in RMB) weighted average of past one-month (Panel A) or one-quarter (Panel B) return volatilities of stocks an investor purchases during various sample periods. The return volatility is defined as the standard deviation of daily returns. A stock purchase is considered as an initial buy if the investor holds zero share of the stock before the purchase. Each panel presents results of an OLS regression based on three sample periods: full (January 2018 through June 2019), pre-survey (January 2018 through September 2018), and post-survey (October 2018 through June 2019). The key independent variables are dummies that indicate if an investor is strongly risk averse or not. It equals one if an investor answers "No" when asked if she would be willing to quit her current stable job for a higher expected pay but higher risk one in all three hypothetical scenarios. Otherwise, the dummy equals zero. Control variables include age, net worth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors are reported in parentheses.

Panel A: Change in Stock Holdings During the 6-month Window Around the Survey

	Full sample		Pre-survey		Post-survey	
	(2018:07 – 2018:12)		(2018:07 – 2018:09)		(2018:10 – 2018:12)	
Return Expectation	0.006*	0.007*	0.005**	0.005***	0.003	0.003
	(1.781)	(1.836)	(2.363)	(2.603)	(1.391)	(1.496)
Male		0.113***		0.032		0.037*
		(3.209)		(1.596)		(3.209)
Controls	NO	YES	NO	YES	NO	YES
R2	0.001	0.053	0.002	0.042	0.001	0.021
N	3,122	3,122	3,174	3,174	3,159	3,159

Panel B: Change in Stock Holdings During the 18-month Window Around the Survey

	Full sample		Pre-survey		Post-survey	
	(2018:01 – 2019:06)		(2018:01 – 2018:09)		(2018:10 – 2019:06)	
Return Expectation	0.030**	0.029**	0.015***	0.012**	-0.005	-0.005
	(2.084)	(2.106)	(2.614)	(2.186)	(-0.631)	(-0.572)
Male		0.359***		0.046		0.162**
		(2.760)		(0.820)		(1.969)
Controls	NO	YES	NO	YES	NO	YES
R2	0.001	0.099	0.002	0.077	0.000	0.041
N	3,187	3,187	3,143	3,143	3,225	3,225

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

Table A7 Validation of Survey Expectations

Note: This table studies the relationship between investors' one-year market return forecasts and portfolio changes. The dependent variables are percentage changes in investors' total stock holdings during various sample periods. Panel A looks at investors' portfolio changes in the 6-month window around the survey, and Panel B examines a longer 18-month window around the survey. The key independent variables are investors' forecasts of one-year ahead market return. Its discrete values range from 1 (corresponds to less than -50%) to 21 (corresponds to 45% to 50%). Each panel presents results for an OLS regression based on three sample periods: full (January 2018 through June 2019), pre-survey (January 2018 through September 2018), and post-survey (October 2018 through June 2019). Control variables include age, net worth, income, trading experience, account size, and education. *T*-statistics are based on robust standard errors are reported in parentheses.

A9. Robustness checks for horse-race results

In this section, we present additional evidence for the horse race by considering alternative empirical specifications and subsample analyses.

A9.1. Including branch fixed effects

Table A8 reports the results for the horse race when we include branch fixed effects. As the table shows, virtually all the results are unchanged, and gambling preference and perceived information advantage remain the most powerful explanatory variables for trading. One caveat is that the significance of gambling preference drops from significant at the 1% level to 5%, and the only trading motive that is significant at the 1% level in this specification is perceived information advantage.

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual rank of 2017 performance	4.347***	Gambling preference, blockbusters	10.570**
	(5.085)		(2.428)
Over-placement, performance	12.718**	Gambling preference, lotteries	-1.882
	(2.120)		(-0.402)
Financial literacy, dummy	7.161*	Sensation seeking, novelty	8.506*
	(1.687)		(1.651)
Over-placement, literacy	-5.285	Sensation seeking, volatility	5.671
	(-1.183)		(1.207)
Miscalibration of uncertainty	-3.009	Perceived information advantage	14.511***
	(-0.723)		(2.642)
Do not consider transaction cost	-5.980	Dismissive of others' information	4.016
	(-1.506)		(1.059)
Underestimation of transaction cost	-2.111	Social influence	-7.979
	(-0.513)		(-1.499)
Do not think bid-ask spread is a cost	-9.364**	Advisor influence	-12.094
	(-2.438)		(-1.632)
Extrapolation, up	0.306	Portfolio rebalancing needs	8.706
	(0.059)		(1.558)
Extrapolation, down	-2.718	Liquidity	-6.778
	(-0.555)		(-1.111)
Realization utility, winners	5.039	Risk aversion	-2.313
	(1.246)		(-0.506)

Realization utility, losers	-2.224	Expected future one-year market return	0.514
	(-0.490)		(1.311)
Gender: male	22.262***	Branch fixed effects	YES
	(5.998)	N	4,648
Controls	YES	R^2	0.177

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

Table A8 Horse Race with Branch Fixed Effects

A9.2. *Bootstrapped standard errors*

Table A9 reports the results for the horse race using bootstrapped standard errors. Significance is virtually unchanged.

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual performance in 2017	4.198***	Gambling preference, blockbusters	11.764***
	(5.314)		(2.808)
Overplacement, performance	11.549**	Gambling preference, lotteries	-1.159
	(2.018)		(-0.265)
Financial literacy, dummy	7.065*	Sensation seeking, novelty	6.598
	(1.829)		(1.356)
Overplacement, literacy	-2.621	Sensation seeking, volatility	3.632
	(-0.643)		(0.824)
Miscalibration of uncertainty	-2.989	Perceived information advantage	15.660***
	(-0.827)		(2.973)
Do not consider trading costs	-3.989	Dismissive of others' information	2.942
	(-1.168)		(0.822)
Underestimation of trading costs	-4.029	Social influence	-7.839
	(-0.987)		(-1.496)
Do not think bid-ask spread is a cost	-9.456***	Advisor influence	-12.089*
	(-2.667)		(-1.896)
Extrapolation, up	-1.255	Portfolio rebalancing needs	12.571**
	(-0.256)		(2.285)
Extrapolation, down	-1.208	Liquidity needs	-7.651
	(-0.253)		(-1.247)
Realization utility, winners	7.049*	Risk Aversion	-2.943
	(1.824)		(-0.743)
Realization utility, losers	-2.321	Expected 1-year market return	0.709**
	(-0.468)		(2.274)
Gender: male	21.488***	Controls	YES
	(6.236)	N	4,648
		R^2	0.089

A9.3.   Subsample analysis

Table A10 reports the horse-race results for small investors, that is, investors with a maximum account balance of less than 100K RMB by the end of 2017. Again, gambling preference and perceived information advantage remain the most powerful explanatory variables for trading, and their coefficients increase in magnitude. This suggests that these factors are more powerful among small investors. Furthermore, over-placement based on performance becomes insignificant.

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual rank of 2017 performance	3.486***	Gambling preference, blockbusters	11.266***
	(4.238)		(2.695)
Over-placement, performance	7.020	Gambling preference, lotteries	-1.097
	(1.202)		(-0.240)
Financial literacy, dummy	6.339	Sensation seeking, novelty	9.929**
	(1.564)		(1.974)
Over-placement, literacy	-3.591	Sensation seeking, volatility	0.677
	(-0.828)		(0.149)
Miscalibration of uncertainty	-2.434	Perceived information advantage	20.687***
	(-0.603)		(3.705)
Do not consider transaction costs	-4.714	Dismissive of others' information	3.099
	(-1.224)		(0.819)
Underestimation of transaction costs	-3.770	Social influence	-9.454**
	(-0.946)		(-1.973)
Do not think bid-ask spread is a cost	-7.993**	Advisor influence	-10.549*
	(-2.175)		(-1.666)
Extrapolation, up	-2.014	Portfolio rebalancing needs	8.959
	(-0.391)		(1.598)
Extrapolation, down	0.055	Liquidity needs	-6.621
	(0.011)		(-1.127)
Realization utility, winners	8.877**	Risk aversion	-4.704
	(2.253)		(-1.066)
Realization utility, losers	-3.351	Expected future one-year market return	0.771**
	(-0.752)		(1.981)
Gender: male	23.338***	Controls	YES
	(6.481)	N	4,241
		R^2	0.094

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

Table A10 Horse Race for Small Investors

Table A11 reports the horse-race results for large investors, that is, investors with a maximum account balance greater than 100K RMB by the end of 2017. Unlike in the sample of small investors, in this sample of large investors, the most significant explanatory variables are overconfidence based on performance and portfolio rebalancing needs.

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual rank of 2017 performance	10.499***	Gambling preference, blockbusters	23.300
	(3.017)		(1.517)
Over-placement, performance	52.447**	Gambling preference, lotteries	-6.862
	(2.589)		(-0.372)
Financial literacy, dummy	26.847*	Sensation seeking, novelty	-26.370
	(1.722)		(-1.335)
Over-placement, literacy	11.502	Sensation seeking, volatility	32.999*
	(0.586)		(1.734)
Miscalibration of uncertainty	-9.376	Perceived information advantage	-29.456*
	(-0.563)		(-1.873)
Do not consider transaction cost	-5.314	Dismissive of others' information	6.487
	(-0.370)		(0.495)
Underestimation of transaction cost	-2.900	Social influence	17.155
	(-0.219)		(0.619)
Do not think bid-ask spread is a cost	-29.042*	Advisor influence	-12.954
	(-1.795)		(-0.413)
Extrapolation, up	0.645	Portfolio rebalancing needs	45.272**
	(0.033)		(2.044)
Extrapolation, down	-12.854	Liquidity needs	-37.934
	(-0.684)		(-1.453)
Realization utility, winners	-11.929	Risk aversion	7.232
	(-0.749)		(0.486)
Realization utility, losers	2.456	Expected future one-year market return	0.812
	(0.134)		(0.567)
Gender, male	-4.092	Controls	YES
	(0.278)	N	407
		R^2	0.165

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

Table A11 Horse Race for Large Investors

We next split the main sample into two halves based on how much of an investor's net worth is invested in the stock market. For each investor, we first calculate the ratio of her entire portfolio value at the Shenzhen Stock Exchange to her reported net worth. This, however, does not take into account her investment at the Shanghai Stock Exchange or mutual fund holdings. We then sort all investors into two groups of equal size, where investors with the ratio above (below) the median are considered more (less) invested in the stock market. Table A12 reports the horse race results

for the more invested investors. A notable observation is that, in this subsample, portfolio rebalancing becomes a more important trading motive. This is consistent with these investors having to move money in and out of the stock market more frequently. Another observation is that, due to a smaller sample size, perceived information advantage is no longer significant, although the magnitude remains large. Gambling preference remains significant.

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 − June 2019)			
Actual rank of 2017 performance	6.330***	Gambling preference, blockbusters	14.076**
	(4.937)		(2.264)
Over-placement, performance	16.820**	Gambling preference, lotteries	-3.279
	(1.990)		(-0.476)
Financial literacy, dummy	7.378	Sensation seeking, novelty	6.600
	(1.207)		(0.853)
Over-placement, literacy	-5.092	Sensation seeking, volatility	-2.396
	(-0.777)		(-0.343)
Miscalibration of uncertainty	-2.498	Perceived information advantage	10.652
	(-0.407)		(1.319)
Do not consider transaction cost	2.893	Dismissive of others' information	3.024
	(0.502)		(0.517)
Underestimation of transaction cost	-6.834	Social influence	-1.485
	(-1.116)		(-0.178)
Do not think bid-ask spread is a cost	-12.834**	Advisor influence	-4.972
	(-2.277)		(-0.432)
Extrapolation, up	1.858	Portfolio rebalancing needs	30.106***
	(0.234)		(3.228)
Extrapolation, down	-5.378	Liquidity needs	-16.607*
	(-0.737)		(-1.850)
Realization utility, winners	9.048	Risk aversion	-4.485
	(1.499)		(-0.706)
Realization utility, losers	3.985	Expected future one-year market return	0.661
	(0.559)		(1.191)
Gender, male	28.625***	Controls	YES
	(5.070)	N	2,321
		$R2$	0.075

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

Table A12 Horse Race for Highly Invested Investors

Table A13 reports the horse race results for the less invested investors. In this subsample, perceived information advantage is more important, while gambling preference remains significant. However, portfolio rebalancing becomes an insignificant factor in explaining trading.

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual rank of 2017 performance	2.038**	Gambling preference, blockbusters	10.746**
	(2.010)		(2.145)
Over-placement, performance	2.517	Gambling preference, lotteries	-0.619
	(0.342)		(-0.114)
Financial literacy, dummy	6.151	Sensation seeking, novelty	7.160
	(1.231)		(1.251)
Over-placement, literacy	0.289	Sensation seeking, volatility	7.705
	(0.054)		(1.412)
Miscalibration of uncertainty	-2.247	Perceived information advantage	17.560***
	(-0.452)		(2.607)
Do not consider transaction cost	-12.097**	Dismissive of others' information	3.420
	(-2.576)		(0.770)
Underestimation of transaction cost	-1.534	Social influence	-15.371***
	(-0.329)		(-2.925)
Do not think bid-ask spread is a cost	-7.940*	Advisor influence	-15.509**
	(-1.816)		(-2.522)
Extrapolation, up	-6.465	Portfolio rebalancing needs	-2.057
	(-1.095)		(-0.333)
Extrapolation, down	4.402	Liquidity needs	0.166
	(0.758)		(0.023)
Realization utility, winners	4.430	Risk aversion	-0.416
	(0.944)		(-0.077)
Realization utility, losers	-9.229*	Expected future one-year market return	0.837*
	(-1.805)		(1.697)
Gender, male	14.740***	Controls	YES
	(3.449)	N	2,327
		$R2$	0.106

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

Table A13 Horse Race for Less Invested Investors

A9.4. Other robustness checks

Finally, we consider four additional robustness checks. First, we consider an alternative measure of turnover, where, instead of value-weighting monthly turnover using monthly balance, we equal-weight monthly turnover. Results are reported in Table A14. Second, we consider including investors with zero holdings at the Shenzhen Stock Exchange, and their turnovers are coded as zeros. Results are reported in Table A15. Third, we consider restricting to "active" investors, who were actively trading prior to the survey. Results are reported in Table A16. Fourth, we consider measuring turnover using transactions from January 2018 through September 2018,

and results are presented in Table A17. In all four sets of analyses, gambling preference and perceived information advantage remain the most significant and powerful variables for trading.

Dependent Variable: Average Monthly Turnover Ratio (%), Equal-Weighted (October 2018 – June 2019)			
Actual rank of 2017 performance	2.928***	Gambling preference, blockbusters	10.715***
	(4.125)		(2.994)
Over-placement, performance	9.420*	Gambling preference, lotteries	-3.053
	(1.890)		(-0.791)
Financial literacy, dummy	5.658*	Sensation seeking, novelty	6.868
	(1.650)		(1.589)
Over-placement, literacy	-3.750	Sensation seeking, volatility	4.250
	(-1.043)		(1.085)
Miscalibration of uncertainty	-2.915	Perceived information advantage	9.221**
	(-0.839)		(2.013)
Do not consider transaction cost	-3.380	Dismissive of others' information	2.805
	(-1.021)		(0.870)
Underestimation of transaction cost	-1.707	Social influence	-5.761
	(-0.505)		(-1.325)
Do not think bid-ask spread is a cost	-8.431***	Advisor influence	-7.642
	(-2.677)		(-1.367)
Extrapolation, up	-1.613	Portfolio rebalancing needs	10.590**
	(-0.368)		(2.212)
Extrapolation, down	-1.308	Liquidity needs	-7.077
	(-0.322)		(-1.433)
Realization utility, winners	5.651*	Risk aversion	-1.457
	(1.668)		(-0.381)
Realization utility, losers	-2.930	Expected future one-year market return	0.582*
	(-0.773)		(1.775)
Gender, male	18.187***	Controls	YES
	(5.889)	N	4,648
		$R2$	0.104

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

Table A14 Using Equal-weighted Turnover as Dependent Variables

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual rank of 2017 performance	3.102***	Gambling preference, blockbusters	12.040***
	(3.777)		(3.180)
Over-placement, performance	12.285**	Gambling preference, lotteries	-4.415
	(1.978)		(-1.073)
Financial literacy, dummy	1.566	Sensation seeking, novelty	6.264
	(1.466)		(1.381)
Over-placement, literacy	-6.528*	Sensation seeking, volatility	5.351
	(-1.690)		(1.282)
Miscalibration of uncertainty	-3.109	Perceived information advantage	14.163***
	(-0.845)		(2.801)
Do not consider transaction cost	-3.042	Dismissive of others' information	-5.870
	(-0.900)		(-1.220)
Underestimation of transaction cost	-3.029	Social influence	-10.383**
	(-0.848)		(-2.327)
Do not think bid-ask spread is a cost	-8.722***	Advisor influence	-7.451
	(-2.601)		(-1.264)
Extrapolation, up	-3.997	Portfolio rebalancing needs	10.300**
	(-0.859)		(2.009)
Extrapolation, down	0.911	Liquidity needs	-3.872
	(0.208)		(-0.732)
Realization utility, winners	6.896*	Risk aversion	-4.405
	(1.931)		(-1.106)
Realization utility, losers	-1.569	Expected future one-year market return	0.421
	(-0.385)		(1.194)
Gender, male	20.732***	Controls	YES
	(6.302)	$R2$	0.101
		N	4,937

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

Table A15 Horse Race When Including Accounts with Zero Holdings

Dependent Variable: Average Monthly Turnover Ratio (%) (October 2018 – June 2019)			
Actual rank of 2017 performance	4.989***	Gambling preference, blockbusters	12.148***
	(5.726)		(2.842)
Over-placement, performance	11.929*	Gambling preference, lotteries	-3.846
	(1.937)		(-0.831)
Financial literacy, dummy	8.265**	Sensation seeking, novelty	6.538
	(1.979)		(1.279)
Over-placement, literacy	-3.450	Sensation seeking, volatility	5.366
	(-0.779)		(1.155)
Miscalibration of uncertainty	-2.884	Perceived information advantage	15.954***
	(-0.697)		(2.889)
Do not consider transaction cost	-4.334	Dismissive of others' information	1.897
	(-1.107)		(0.487)
Underestimation of transaction cost	-3.570	Social influence	-8.870*
	(-0.878)		(-1.733)
Do not think bid-ask spread is a cost	-10.395***	Advisor influence	-11.379*
	(-2.765)		(-1.744)
Extrapolation, up	-1.378	Portfolio rebalancing needs	13.537**
	(-0.261)		(2.332)
Extrapolation, down	0.306	Liquidity needs	-5.889
	(0.062)		(-0.966)
Realization utility, winners	7.171*	Risk aversion	-3.349
	(1.763)		(-0.742)
Realization utility, losers	-4.156	Expected future one-year market return	0.740*
	(-0.908)		(1.877)
Gender, male	23.440***	Controls	YES
	(6.306)	$R2$	0.066
		N	4,398

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

Table A16 Horse Race Among Active Investors

Dependent Variable: Average Monthly Turnover Ratio (%)			
(January 2018 – September 2018)			
Actual rank of 2017 performance	6.476***	Gambling preference, blockbusters	10.812***
	(8.726)		(2.837)
Over-placement, performance	14.416***	Gambling preference, lotteries	-2.568
	(2.722)		(-0.624)
Financial literacy, dummy	4.260	Sensation seeking, novelty	-1.018
	(1.201)		(-0.237)
Over-placement, literacy	-4.472	Sensation seeking, volatility	6.988*
	(-1.190)		(1.717)
Miscalibration of uncertainty	-6.911*	Perceived information advantage	11.644**
	(-1.845)		(2.476)
Do not consider transaction cost	-1.993	Dismissive of others' information	5.925*
	(-0.578)		(1.744)
Underestimation of transaction cost	-4.171	Social influence	-13.520***
	(-1.146)		(-3.328)
Do not think bid-ask spread is a cost	-10.033***	Advisor influence	-8.331
	(-3.059)		(-1.563)
Extrapolation, up	-3.037	Portfolio rebalancing needs	8.235*
	(-0.715)		(1.654)
Extrapolation, down	5.305	Liquidity needs	-5.508
	(1.304)		(-1.006)
Realization utility, winners	4.057	Risk aversion	-2.366
	(1.143)		(-0.605)
Realization utility, losers	-2.898	Expected future one-year market return	0.587*
	(-0.709)		(1.713)
Gender, male	19.623***	Controls	YES
	(5.991)	$R2$	0.079
		N	4,648

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

Table A17 Horse Race Using Pre-Survey Turnover

A10. Gambling preferences and characteristics of stocks purchased

In the main text of the paper, we show that investors with a survey-based gambling preference tend to trade more and perform worse than others. In this section, we examine the characteristics of stocks purchased by different investors. Table A18 shows that investors with a survey-based gambling preference tend to buy stocks that are smaller, have a larger market beta, and have larger counts of daily up-limit hits, and higher past volatility and past returns. These stocks also perform worse subsequently, confirming that investors with a gambling preference trade in the wrong direction and their trading is excessive.

	(October 2018 – June 2019)						
	Past 30-day # of Up-limit Hits	Past 30-day Return Volatility (%)	Past 30-day Return (%)	Size (Billion RMB)	Beta	B/M	Future 30-day Return (%)
1. Strongly disagree	0.6	3.25	9.71	43.73	0.93	0.62	-0.03
2. Disagree	0.75	3.39	11.58	35.21	0.96	0.62	-0.87
3. Neutral	0.83	3.49	11.94	26.92	0.99	0.61	-1.53
4. Agree	0.89	3.56	12.45	26.29	1.00	0.61	-1.36
5. Strongly agree	0.92	3.55	12.74	26.65	1.02	0.62	-1.77
5 − 1	**0.32*****	**0.30*****	**3.03****	**-17.08****	**0.09*****	**0.00**	**-1.74****

Table A18 Trading Characteristics for Investors Sorted on Gambling Preference, Blockbusters

A11. A "Nudge" Experiment

Among all 500 brokerage branches we distributed the survey to, we randomly selected 250 branches to include an additional "nudge." The "nudge" asked the respondent to read a short article that highlighted the negative consequences of excessive trading. As shown in Figure A3, the article contained a detailed calculation of how much investors lose from frequent trading along with a quote from Warren Buffett advising investors to buy and hold. Instead of presenting trading costs as a fraction of total transaction value, we made it more salient by presenting the annualized fee rate for a frequent trader. We also included a "validation" question after the article by asking the respondent to calculate the total trading costs of a given level of turnover. Answers to this question help identify those who have actually read the article and therefore were treated.

Figure A3 The "Nudge" to Reduce Trading Due to Transaction Costs

We study the effect of this "nudge" in a difference-in-difference framework; Table A19 reports the results. Column (1) shows that the interaction term is small and insignificant, suggesting that the treatment and control groups exhibit similar turnover rates one month after the survey. We repeat this exercise in columns (2) and (3) by expanding the window to three months and six months before and after the survey, and the interaction term remains insignificant. Overall, these results suggest that the nudge had no effect on reducing trading. One might argue that the "nudge" was not sufficiently strong and the treated group may not have read the article carefully. However,

we identify an investor as treated only if they were in the treated group and answered the "validation" question correctly.

	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
*R*2	0.056	0.058	0.056
N	6,628	6,628	6,628

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

Table A19: Comparing Turnover Before and After the Survey for the Control and Treatment Groups

Note: Before distributing the survey, we randomly assigned 500 targeted branches of brokerage firms into treated and control groups. Investors in the two groups received questionnaires that were otherwise identical except for one difference: the questionnaire for the treated group included a "nudge" that highlighted the negative consequences of excessive trading. In this table, we study the effect of the "nudge" on investors' trading frequencies using difference-in-difference tests. The dependent variables from Columns (1) to (3) are investors' average monthly turnover rates in the one month, three months, and six months before and after the survey. The dummy, Treated, equals one if an investor is in the treated group and correctly answered the follow-up question designed to test whether the respondent understands the content of the message. The dummy, *Treated*, equals zero if an investor is in the control group. The dummy, *After*, equals one for the periods after the survey month and zero for the periods before or in the survey month (September 2018). Control variables include age, gender, wealth, income, trading experience, account size, and education. T-statistics are based on robust standard errors and are reported in parentheses. See the Table A2 for more details about variable definitions.