Understanding Retail Investors: Evidence from China

Charles M. Jones, Columbia University
Donghui Shi, Fudan University
Xiaoyan Zhang, Tsinghua University
Xinran Zhang, Central University of Finance & Economics

October 2021
Why Chinese Retail Investors?

Trading Volume

- Retail <100K
- Retail (100K, 500K)
- Retail (500K, 3M)
- Retail (3M, 10M)
- Retail >10M
- INST
- CORP
Why Chinese Retail Investors?

Trading Volume vs Holdings Value

- Retail <100K
- Retail (100K, 500K)
- Retail (500K, 3M)
- Retail (3M, 10M)
- Retail >10M
- INST
- CORP
Chinese Retail Investors: A 20-Year Debate

- Investors:
  - Investing vs. Gambling?

- Capital Market:
  - Financing platforms vs. Casinos?

- Market Efficiency
  - Information
  - Liquidity
Research Questions

This study examine three fundamental questions of retail investors in Chinese stock market with account level data.

- How retail investors contribute to price discovery?
  - Can they predict future returns?

- What drives retail investors trading behaviors?
  - Information? Liquidity? Behavioral?

- How retail investors affect liquidity?
Previous Literature

- **US retail investors literature**
  - Kaniel, Saar, and Titman (2008 JF), Kelley and Tetlock (2013 JF)
  - Boehmer, Jones, Zhang, and Zhang (2021 JF)
  - Welch (2021 JF), Eaton, Green, Roseman, and Wu (2021 WP)
  - Barber, Huang, Odean, and Schwarz (2021 WP)

- **Chinese retail investors literature**
  - An, Lou, and Shi (2021 WP)
  - Liu, Peng, Xiong, and Xiong (2021 JFE), Liao, Peng, and Zhu (2021 RFS)
Contributions

- Our study, with its large coverage of the market for a recent sample period, is one of the most thorough and comprehensive studies for Chinese retail investors.

- We separate retail investors into groups based on account sizes and provide unique, direct evidence on investor heterogeneity in trading and return predictability.

- We examine different hypotheses for the return prediction patterns for different retail investor groups, and provide clear evidence on the sources of the retail investors.

- We provide new evidence on how retail investors affect liquidity.
Possible Concerns

- Too descriptive? Just summary statistics?
  - We do provide important stylized facts for the Chinese stock market.
  - We also provide thorough analyses of sources of return predictive patterns and liquidity provision.

- Are these patterns already found in U.S.?
  - The literature on U.S. retail investors have mixed results on our research questions.
  - Set-ups in the U.S. are quite different from those in China.
  - Understanding current status of Chinese retail investors is important for both academics and regulators.
## Data

- Account level data from a major exchange
  - Sample period: January 2016 to June 2019
  - Number of accounts: Over 53 million accounts

<table>
<thead>
<tr>
<th>Account value (CNY)</th>
<th>Number of Accounts (‘000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT1 &lt;100K</td>
<td>31,410</td>
</tr>
<tr>
<td>RT2 (100K,500K)</td>
<td>15,282</td>
</tr>
<tr>
<td>RT3 (500K,3M)</td>
<td>5,827</td>
</tr>
<tr>
<td>RT4 (3M,10M)</td>
<td>735</td>
</tr>
<tr>
<td>RT5 &gt;10M</td>
<td>235</td>
</tr>
<tr>
<td>INST</td>
<td>40</td>
</tr>
<tr>
<td>CORP</td>
<td>47</td>
</tr>
</tbody>
</table>
Figure 1.

Panel A. Share Volume (%)
Retail Order Flows: Order Imbalance

For each type of investor $G$ trade stock $i$ on day $d$, the order imbalance measures are:

$$Oib(i, d, G) = \frac{\sum_{j \in G} Buyvol(i, d, j) - \sum_{j \in G} Sellvol(i, d, j)}{\sum_{j \in G} Buyvol(i, d, j) + \sum_{j \in G} Sellvol(i, d, j)}$$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>OibRT1 OibRT2 OibRT3 OibRT4 OibRT5 OibINST</td>
</tr>
<tr>
<td>OibRT1</td>
<td>-0.021</td>
<td>0.187</td>
<td></td>
</tr>
<tr>
<td>OibRT2</td>
<td>-0.011</td>
<td>0.171</td>
<td>0.802</td>
</tr>
<tr>
<td>OibRT3</td>
<td>-0.006</td>
<td>0.166</td>
<td>0.610 0.710</td>
</tr>
<tr>
<td>OibRT4</td>
<td>0.002</td>
<td>0.250</td>
<td>0.194 0.244 0.256</td>
</tr>
<tr>
<td>OibRT5</td>
<td>0.019</td>
<td>0.352</td>
<td>-0.151 -0.158 -0.163 -0.091</td>
</tr>
<tr>
<td>OibINST</td>
<td>-0.011</td>
<td>0.455</td>
<td>-0.315 -0.365 -0.380 -0.263 -0.188</td>
</tr>
<tr>
<td>OibCORP</td>
<td>-0.004</td>
<td>0.720</td>
<td>0.022 0.029 0.021 -0.007 -0.043 -0.044</td>
</tr>
</tbody>
</table>
Question 1. Can Retail Investors Predict Future Returns?

- Fama-MacBeth regression for each investor group:

\[ \text{Ret}(i, d) = a0(d) + a1(d)Oib(i, d - 1) + a2(d)'\text{Controls}(i, d - 1) + u1(i, d). \]

- If retail investors correctly predict future returns: \( a1 > 0 \)
- If retail investors incorrectly predict future returns: \( a1 < 0 \)
- If retail investors trading has nothing to do with future returns: 
  \( a1 = 0 \)
## Answer 1a. Over Next Day

### Can retail investors predict future stock returns?

<table>
<thead>
<tr>
<th>Dep.var</th>
<th>Ret</th>
<th>Ret</th>
<th>Ret</th>
<th>Ret</th>
<th>Ret</th>
<th>Ret</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oib.var</td>
<td>RT1</td>
<td>RT2</td>
<td>RT3</td>
<td>RT4</td>
<td>RT5</td>
<td>INST</td>
</tr>
<tr>
<td>Oib(-1)</td>
<td>-0.0093***</td>
<td>-0.0091***</td>
<td>-0.0065***</td>
<td>-0.0009***</td>
<td>0.0012***</td>
<td>0.0016***</td>
</tr>
<tr>
<td>Interquartile Ret</td>
<td>-0.21%</td>
<td>-0.17%</td>
<td>-0.11%</td>
<td>-0.02%</td>
<td>0.05%</td>
<td>0.10%</td>
</tr>
</tbody>
</table>

**Daily Interquartile Return Diff**

- RT1: -0.21%
- RT2: -0.17%
- RT3: -0.11%
- RT4: -0.02%
- RT5: 0.05%
- INST: 0.10%
Answer 1b. Over 12 Weeks

Predicting kth week ahead returns
Question 2a: What Drives Retail Investor’s Trading?

- Hypotheses for explaining the order flow dynamics and their return predictive power for future returns:
    - Proxy: previous day order flow
    - Proxy: previous returns
  - Behavior biases: Liu, Peng, Xiong and Xiong (2021)
    - Proxy for over-confidence: previous turnover
    - Proxy for gambling preferences: previous max daily returns
  - Information: Kelley and Tetlock (2013)
    - Proxy for public information: earnings surprise
### Question 2a: What Leads to the Predictive Pattern (Stage I)

<table>
<thead>
<tr>
<th></th>
<th>OibRT1</th>
<th>OibRT2</th>
<th>OibRT3</th>
<th>OibRT4</th>
<th>OibRT5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oib(-1)</td>
<td>0.1867***</td>
<td>0.1965***</td>
<td>0.1711***</td>
<td>0.0499***</td>
<td>0.1037***</td>
</tr>
<tr>
<td>Ret(-1)</td>
<td>0.5131***</td>
<td>0.7244***</td>
<td>0.4469***</td>
<td>-0.2210***</td>
<td>-1.2907***</td>
</tr>
<tr>
<td>Ret(-6,-2)</td>
<td>-0.4186***</td>
<td>-0.2186***</td>
<td>-0.1047***</td>
<td>-0.0788***</td>
<td>-0.0513***</td>
</tr>
<tr>
<td>Ret(-27,-7)</td>
<td>-0.0325***</td>
<td>-0.0171***</td>
<td>-0.0205***</td>
<td>-0.0377***</td>
<td>-0.0231***</td>
</tr>
<tr>
<td>Overconf(-1)</td>
<td>0.0792***</td>
<td>0.0366***</td>
<td>0.0318***</td>
<td>0.0177*</td>
<td>-0.0590***</td>
</tr>
<tr>
<td>Gamble(-1)</td>
<td>0.0467***</td>
<td>0.1025***</td>
<td>0.1991***</td>
<td>0.2583***</td>
<td>-0.0863***</td>
</tr>
<tr>
<td>Adj.R2</td>
<td>7.08%</td>
<td>5.59%</td>
<td>3.88%</td>
<td>0.74%</td>
<td>2.02%</td>
</tr>
</tbody>
</table>
**Question 2a: What Leads to the Predictive Pattern (Stage I)**

<table>
<thead>
<tr>
<th></th>
<th>OibRT1</th>
<th>OibRT2</th>
<th>OibRT3</th>
<th>OibRT4</th>
<th>OibRT5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oib(-1)</td>
<td>0.1867***</td>
<td>0.1965***</td>
<td>0.1711***</td>
<td>0.0499***</td>
<td>0.1037***</td>
</tr>
<tr>
<td>Ret(-1)</td>
<td><strong>0.5131</strong>*</td>
<td><strong>0.7244</strong>*</td>
<td><strong>0.4469</strong>*</td>
<td>-0.2210***</td>
<td>-1.2907***</td>
</tr>
<tr>
<td>Ret(-6,-2)</td>
<td>-0.4186***</td>
<td>-0.2186***</td>
<td>-0.1047***</td>
<td>-0.0788***</td>
<td>-0.0513***</td>
</tr>
<tr>
<td>Ret(-27,-7)</td>
<td>-0.0325***</td>
<td>-0.0171***</td>
<td>-0.0205***</td>
<td>-0.0377***</td>
<td>-0.0231***</td>
</tr>
<tr>
<td>Overconf(-1)</td>
<td>0.0792***</td>
<td>0.0366***</td>
<td>0.0318***</td>
<td>0.0177*</td>
<td>-0.0590***</td>
</tr>
<tr>
<td>Gamble(-1)</td>
<td>0.0467***</td>
<td>0.1025***</td>
<td>0.1991***</td>
<td>0.2583***</td>
<td>-0.0863***</td>
</tr>
<tr>
<td>Adj.R2</td>
<td>7.08%</td>
<td>5.59%</td>
<td>3.88%</td>
<td>0.74%</td>
<td>2.02%</td>
</tr>
</tbody>
</table>
**Question 2a: What Leads to the Predictive Pattern (Stage I)**

<table>
<thead>
<tr>
<th></th>
<th>OibRT1</th>
<th>OibRT2</th>
<th>OibRT3</th>
<th>OibRT4</th>
<th>OibRT5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oib(-1)</td>
<td>0.1867***</td>
<td>0.1965***</td>
<td>0.1711***</td>
<td>0.0499***</td>
<td>0.1037***</td>
</tr>
<tr>
<td>Ret(-1)</td>
<td>0.5131***</td>
<td>0.7244***</td>
<td>0.4469***</td>
<td>-0.2210***</td>
<td>-1.2907***</td>
</tr>
<tr>
<td>Ret(-6,-2)</td>
<td>-0.4186***</td>
<td>-0.2186***</td>
<td>-0.1047***</td>
<td>-0.0788***</td>
<td>-0.0513***</td>
</tr>
<tr>
<td>Ret(-27,-7)</td>
<td>-0.0325***</td>
<td>-0.0171***</td>
<td>-0.0205***</td>
<td>-0.0377***</td>
<td>-0.0231***</td>
</tr>
<tr>
<td>Overconf(-1)</td>
<td>0.0792***</td>
<td>0.0366***</td>
<td>0.0318***</td>
<td>0.0177*</td>
<td><strong>-0.0590</strong>*</td>
</tr>
<tr>
<td>Gamble(-1)</td>
<td><strong>0.0467</strong>*</td>
<td><strong>0.1025</strong>*</td>
<td><strong>0.1991</strong>*</td>
<td><strong>0.2583</strong>*</td>
<td><strong>-0.0863</strong>*</td>
</tr>
<tr>
<td>Adj.R2</td>
<td>7.08%</td>
<td>5.59%</td>
<td>3.88%</td>
<td>0.74%</td>
<td>2.02%</td>
</tr>
</tbody>
</table>
**Question 2b: What Leads to the Predictive Pattern (Stage II)**

<table>
<thead>
<tr>
<th>Interquartile return</th>
<th>RT1</th>
<th>RT2</th>
<th>RT3</th>
<th>RT4</th>
<th>RT5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oib(-1,Persistence)</td>
<td>-0.1179%</td>
<td>-0.0964%</td>
<td>-0.0598%</td>
<td>-0.0131%</td>
<td>0.0281%</td>
</tr>
<tr>
<td>Oib(-1,Liquidity)</td>
<td>-0.0287%</td>
<td>-0.0347%</td>
<td>-0.0257%</td>
<td>0.0099%</td>
<td>0.0095%</td>
</tr>
<tr>
<td>Oib(-1,Overconf)</td>
<td>-0.0162%</td>
<td>-0.0321%</td>
<td>-0.0379%</td>
<td>-0.0383%</td>
<td>0.0233%</td>
</tr>
<tr>
<td>Oib(-1,Gamble)</td>
<td>-0.0401%</td>
<td>-0.0347%</td>
<td>-0.0342%</td>
<td>-0.0355%</td>
<td>0.0448%</td>
</tr>
<tr>
<td>Oib(-1,Other)</td>
<td>-0.1782%</td>
<td>-0.1495%</td>
<td>-0.1009%</td>
<td>-0.0231%</td>
<td>0.0492%</td>
</tr>
</tbody>
</table>
Question 2c: What Leads to the Predictive Pattern – Public News

- Kelley and Tetlock (2013 JF), Boehmer et al. (2021 JF)
  “… retail orders might contain firm-level information…”

- Earnings announcement event, for each quarter:
  - Can retail investors anticipate future earnings surprises?
  - Can retail investors process contemporaneous earnings news?
  - Do predictive patterns of retail investors change on event days?
Answer 2c: What Leads to the Predictive Pattern – Public News

- Earnings announcement event, for each quarter:
  - Can retail investors anticipate future earnings surprises?
    - RT1-RT4 in the wrong direction, RT5 and INST in correct direction.

- Can retail investors process contemporaneous earnings news?
  - RT1-RT4 in the wrong direction, RT5 and INST in correct direction.

- Do predictive patterns of retail investors change on event days?
  - On event days, the predictive power becomes more negative for RT1-RT4, more positive for RT5 and INST.
Question 3. How Retail Investors Affect Future Market Liquidity?

- Black (1986) argues that noise traders are important participants for a well-functioning capital market, because they provide liquidity and lower transaction costs for other investors.
  - If all participants in the market are informed traders, then there probably would be no trades.

- Ozik, Sadka, and Shen (2021 JFQA)
  - “Retail provide liquidity during pandemic lockdown”

- Eaton, Green, Roseman and Wu (2021 WP)
  - “Robinhood investors reduce liquidity…”
Answer 3. Retail Trades vs. Future Liquidity?

- Liquidity measure: relative effective spread

\[ RES_{i,k} = 2|P_{ik} - M_{ik}|/M_{ik}. \]

<table>
<thead>
<tr>
<th>Dep.var</th>
<th>RES</th>
<th>RES</th>
<th>RES</th>
<th>RES</th>
<th>RES</th>
<th>RES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oib.var</td>
<td>OibRT1</td>
<td>OibRT2</td>
<td>OibRT3</td>
<td>OibRT4</td>
<td>OibRT5</td>
<td>OibINST</td>
</tr>
<tr>
<td>Oib(-1)</td>
<td>-0.0099***</td>
<td>-0.0054***</td>
<td>-0.0057***</td>
<td>-0.0019***</td>
<td>0.0019***</td>
<td>0.0010**</td>
</tr>
<tr>
<td>Adj.R2</td>
<td>12.75%</td>
<td>12.57%</td>
<td>12.41%</td>
<td>12.31%</td>
<td>12.34%</td>
<td>12.51%</td>
</tr>
<tr>
<td>Interquartile spread</td>
<td>-0.22%</td>
<td>-0.10%</td>
<td>-0.10%</td>
<td>-0.05%</td>
<td>0.09%</td>
<td>0.07%</td>
</tr>
</tbody>
</table>
Using account level data from January 2016 to June 2019, we examine millions of retail investors with different account sizes.

- Retail investors with smaller account sizes
  - buy and sell in the wrong directions
  - public information doesn’t help them
  - behavior biases hurt them
  - follow daily momentum, demand immediate liquidity, but reduce future transaction costs.

- Retail investors with the largest account sizes
  - buy and sell in the correct directions
  - public information help them
  - they take advantage of other peoples’ behavior biases
  - follow daily contrarian, provide immediate liquidity, but their potentially informed trades raise future transaction costs for the long run.
Chinese Retail Investors: A 20-Year Debate

- Investors:
  - Investing vs. Gambling?

- Capital Market:
  - Financing platforms vs. Casinos?

- Market Efficiency
  - Information
  - Liquidity
What Next?

- Retail investors continue to be an important part of the market
  - Investor education
  - Restrict riskier trading for the smallest accounts
  - Channel some of the investments into wealth products and pension plans

- The rise of institutions
  - Rapid growth of fund industry
    - New question: how to choose funds?
  - Information environments
  - Foreign institutions
Thanks!