A Tale of Two Zoos: Machine Learning Insights on Retail Investors

Pulak Ghosh, Huahao Lu, Hong Zhang, and Jian Zhang

Asian Bureau of Finance and Economic Research (ABFER) 11th Annual Conference

May 20, 2024

Discussion by Darwin Choi

Summary

- A tale of two zoos
  - Factor zoo: stock and firm characteristics that predict future stock returns
  - Bias zoo: investors' psychological heuristics and biases

Summary

- A tale of two zoos
  - Factor zoo: stock and firm characteristics that predict future stock returns
  - Bias zoo: investors' psychological heuristics and biases

Summary

- A tale of two zoos
  - Factor zoo: stock and firm characteristics that predict future stock returns
  - Bias zoo: investors' psychological heuristics and biases

- Retail investors earn poor returns by suffering from biases and/or choosing the wrong characteristics

Contributions

1. Machine learning
   - 15 million retail investor accounts in India
   - Feedforward Neural Network (FNN) and an enhanced Residual Neural Network (ResNN)
   - Compared with OLS, LASSO, Ridge, and Random Forest

Contributions

1. Machine learning
   - 15 million retail investor accounts in India
   - Feedforward Neural Network (FNN) and an enhanced Residual Neural Network (ResNN)
   - Compared with OLS, LASSO, Ridge, and Random Forest
2. Which zoo has more explanatory power?
   - Factors that affect investors' total returns: diversification, portfolio turnover, and momentum
   - Factors that affect investors' trading returns: portfolio turnover, the disposition effect, and diversification
   - Bias zoo is more important

Comment: The Composition of the Two Zoos

- Factor zoo: 23 charactertics, Bias zoo: 13 proxies

Comment: The Composition of the Two Zoos

- Factor zoo: 23 charactertics, Bias zoo: 13 proxies
- Why aren't there more animals?

Comment: The Composition of the Two Zoos

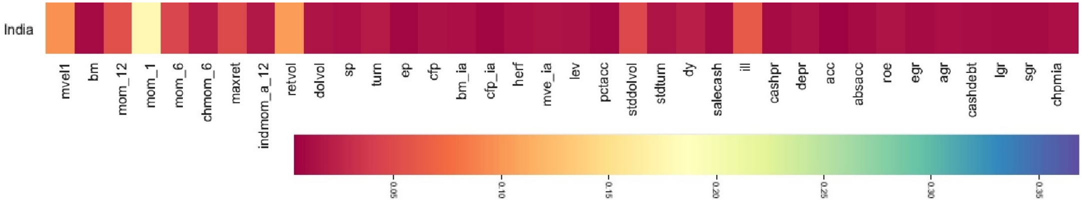
- Factor zoo: 23 charactertics, Bias zoo: 13 proxies
- Why aren't there more animals?
  - Kelly, Gu, and Xiu (RFS 2020): 94 characteristics, 8 macroeconomic predictors, and 74 industry dummies

Comment: The Composition of the Two Zoos

- Factor zoo: 23 charactertics, Bias zoo: 13 proxies
- Why aren't there more animals?
  - Kelly, Gu, and Xiu (RFS 2020): 94 characteristics, 8 macroeconomic predictors, and 74 industry dummies
  - Choi, Jiang, and Zhang (2024): 36 characteristics in 32 markets, including India

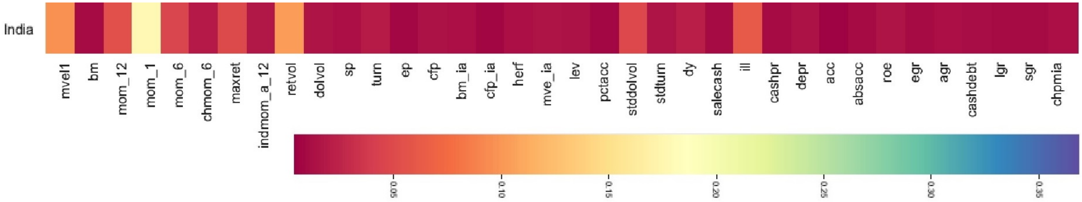
Comment: The Composition of the Two Zoos

- Factor zoo: 23 charactertics, Bias zoo: 13 proxies
- Why aren't there more animals?
  - Kelly, Gu, and Xiu (RFS 2020): 94 characteristics, 8 macroeconomic predictors, and 74 industry dummies
  - Choi, Jiang, and Zhang (2024): 36 characteristics in 32 markets, including India



Comment: The Composition of the Two Zoos

- Factor zoo: 23 charactertics, Bias zoo: 13 proxies
- Why aren't there more animals?
  - Kelly, Gu, and Xiu (RFS 2020): 94 characteristics, 8 macroeconomic predictors, and 74 industry dummies
  - Choi, Jiang, and Zhang (2024): 36 characteristics in 32 markets, including India



- Return volatility (retvol), standard deviation of dollar volume (stddolvol), and illiquidity (ill) seem to be important but missing

Comment: The Composition of the Two Zoos

- Bias zoo

Bias	Proxy
The disposition effect	Regression coefficient
Lottery preference	Ivol Iskew Stock price
Extrapolation	Excess return of holding stocks
Underdiversification	Number of stocks in an investor's portfolio
Local bias	Average distance between an investor's location and the headquarters of the stocks the investor bought
Turnover	The frequency of trading for the investors

Comment: Investor Sentiment

- Very reasonable list

Comment: Investor Sentiment

- Very reasonable list
- But it is missing one important feature: investor sentiment

Comment: Investor Sentiment

- Very reasonable list
- But it is missing one important feature: investor sentiment
- In the literature, investor sentiment is usually imperfectly measured

Comment: Investor Sentiment

- Very reasonable list
- But it is missing one important feature: investor sentiment
- In the literature, investor sentiment is usually imperfectly measured
  - Baker and Wurgler (JF 2006): price-based, market-based, not based on individual trades
  - Barber, Odean, and Zhu (RFS 2008): based on a small subset of household investors
  - Da, Engelberg, and Gao (JF 2011): based on Google search volume, not on trades

Comment: Investor Sentiment

- Very reasonable list
- But it is missing one important feature: investor sentiment
- In the literature, investor sentiment is usually imperfectly measured
    - Baker and Wurgler (JF 2006): price-based, market-based, not based on individual trades
    - Barber, Odean, and Zhu (RFS 2008): based on a small subset of household investors
    - Da, Engelberg, and Gao (JF 2011): based on Google search volume, not on trades
- The authors can try to extract some information from all individual investors' trades to proxy for investor sentiment

Comment: Next-month Return

- The paper uses machine learning to predict individual investors' next-month returns
  - The top 20% of predicted winners can generate a monthly return of 1.3% to 1.7%
  - The bottom 20% earns -2.0% to -2.8% monthly

Comment: Next-month Return

- The paper uses machine learning to predict individual investors' next-month returns
  - The top 20% of predicted winners can generate a monthly return of 1.3% to 1.7%
  - The bottom 20% earns -2.0% to -2.8% monthly
- How persistent are the top and bottom monthly returns?

Comment: Next-month Return

- The paper uses machine learning to predict individual investors' next-month returns
  - The top 20% of predicted winners can generate a monthly return of 1.3% to 1.7%
  - The bottom 20% earns -2.0% to -2.8% monthly
- How persistent are the top and bottom monthly returns?
- The paper makes some claims on investor welfare, but if their monthly returns are not persistent and they don't evaluate their portfolios monthly, we should examine longer-term returns

Comment: Next-month Return

- How often do individual investors turn over their portfolios? (not clear)

Comment: Next-month Return

- How often do individual investors turn over their portfolios? (not clear)

Panel B: Investor behavioral biases

Variable	N	Mean	SD	p10	p25	p50	p75	p90
Port. Value	1.523e+09	1.187	0.521	0.462	0.979	1.363	1.560	1.651
Diver	1.523e+09	9.669	17.013	1.000	2.000	5.000	11.000	22.000
Disp	1.523e+09	0.001	0.008	-0.004	-0.001	0.002	0.003	0.004
ivol	1.523e+09	-0.0570	1.018	-1.416	-0.944	-0.157	0.877	1.360
iskew	1.523e+09	-0.0150	1.031	-1.408	-0.924	-0.126	0.915	1.399
Distance	1.523e+09	877.144	498.822	258.547	528.240	858.926	1185.493	1484.389
Investor Tvr	1.523e+09	-0.128	0.909	-0.662	-0.627	-0.589	-0.534	1.498
Open Price	1.523e+09	0.317	0.773	-0.852	-0.181	0.426	0.910	1.258
High Price	1.523e+09	0.315	0.773	-0.854	-0.184	0.424	0.907	1.257
Low Price	1.523e+09	0.318	0.773	-0.849	-0.180	0.428	0.911	1.259
Close Price	1.523e+09	0.317	0.773	-0.851	-0.181	0.427	0.910	1.258
Extrapolat~n	1.523e+09	0.0380	0.174	-0.155	-0.0510	0.0430	0.135	0.228
Past Perform	1.523e+09	1.009	0.124	0.897	0.947	1.004	1.063	1.130

Comment: Next-month Return

- How often do individual investors turn over their portfolios? (not clear)

Panel B: Investor behavioral biases

Variable	N	Mean	SD	p10	p25	p50	p75	p90
Port. Value	1.523e+09	1.187	0.521	0.462	0.979	1.363	1.560	1.651
Diver	1.523e+09	9.669	17.013	1.000	2.000	5.000	11.000	22.000
Disp	1.523e+09	0.001	0.008	-0.004	-0.001	0.002	0.003	0.004
ivol	1.523e+09	-0.0570	1.018	-1.416	-0.944	-0.157	0.877	1.360
iskew	1.523e+09	-0.0150	1.031	-1.408	-0.924	-0.126	0.915	1.399
Distance	1.523e+09	877.144	498.822	258.547	528.240	858.926	1185.493	1484.389
Investor Tvr	1.523e+09	-0.128	0.909	-0.662	-0.627	-0.589	-0.534	1.498
Open Price	1.523e+09	0.317	0.773	-0.852	-0.181	0.426	0.910	1.258
High Price	1.523e+09	0.315	0.773	-0.854	-0.184	0.424	0.907	1.257
Low Price	1.523e+09	0.318	0.773	-0.849	-0.180	0.428	0.911	1.259
Close Price	1.523e+09	0.317	0.773	-0.851	-0.181	0.427	0.910	1.258
Extrapolat~n	1.523e+09	0.0380	0.174	-0.155	-0.0510	0.0430	0.135	0.228
Past Perform	1.523e+09	1.009	0.124	0.897	0.947	1.004	1.063	1.130

- Kaniel et al. (JFE 2023): 36-month returns

Comment: Next-month Return

- How often do individual investors turn over their portfolios? (not clear)

Panel B: Investor behavioral biases

Variable	N	Mean	SD	p10	p25	p50	p75	p90
Port. Value	1.523e+09	1.187	0.521	0.462	0.979	1.363	1.560	1.651
Diver	1.523e+09	9.669	17.013	1.000	2.000	5.000	11.000	22.000
Disp	1.523e+09	0.001	0.008	-0.004	-0.001	0.002	0.003	0.004
ivol	1.523e+09	-0.0570	1.018	-1.416	-0.944	-0.157	0.877	1.360
iskew	1.523e+09	-0.0150	1.031	-1.408	-0.924	-0.126	0.915	1.399
Distance	1.523e+09	877.144	498.822	258.547	528.240	858.926	1185.493	1484.389
Investor Tvr	1.523e+09	-0.128	0.909	-0.662	-0.627	-0.589	-0.534	1.498
Open Price	1.523e+09	0.317	0.773	-0.852	-0.181	0.426	0.910	1.258
High Price	1.523e+09	0.315	0.773	-0.854	-0.184	0.424	0.907	1.257
Low Price	1.523e+09	0.318	0.773	-0.849	-0.180	0.428	0.911	1.259
Close Price	1.523e+09	0.317	0.773	-0.851	-0.181	0.427	0.910	1.258
Extrapolat~n	1.523e+09	0.0380	0.174	-0.155	-0.0510	0.0430	0.135	0.228
Past Perform	1.523e+09	1.009	0.124	0.897	0.947	1.004	1.063	1.130

- Kaniel et al. (JFE 2023): 36-month returns
- The authors can also examine volatility and Sharpe ratio over the longer-term

Comment: Bias Zoo

- How do we proxy for biases?

Comment: Bias Zoo

- How do we proxy for biases?

Panel B: Investor behavioral biases

Variable	N	Mean	SD	p10	p25	p50	p75	p90
Port. Value	1.523e+09	1.187	0.521	0.462	0.979	1.363	1.560	1.651
Diver	1.523e+09	9.669	17.013	1.000	2.000	5.000	11.000	22.000
Disp	1.523e+09	0.001	0.008	-0.004	-0.001	0.002	0.003	0.004
ivol	1.523e+09	-0.0570	1.018	-1.416	-0.944	-0.157	0.877	1.360
iskew	1.523e+09	-0.0150	1.031	-1.408	-0.924	-0.126	0.915	1.399
Distance	1.523e+09	877.144	498.822	258.547	528.240	858.926	1185.493	1484.389
Investor Tvr	1.523e+09	-0.128	0.909	-0.662	-0.627	-0.589	-0.534	1.498
Open Price	1.523e+09	0.317	0.773	-0.852	-0.181	0.426	0.910	1.258
High Price	1.523e+09	0.315	0.773	-0.854	-0.184	0.424	0.907	1.257
Low Price	1.523e+09	0.318	0.773	-0.849	-0.180	0.428	0.911	1.259
Close Price	1.523e+09	0.317	0.773	-0.851	-0.181	0.427	0.910	1.258
Extrapolat~n	1.523e+09	0.0380	0.174	-0.155	-0.0510	0.0430	0.135	0.228
Past Perform	1.523e+09	1.009	0.124	0.897	0.947	1.004	1.063	1.130

Comment: Bias Zoo

- How do we proxy for biases?

Panel B: Investor behavioral biases

Variable	N	Mean	SD	p10	p25	p50	p75	p90
Port. Value	1.523e+09	1.187	0.521	0.462	0.979	1.363	1.560	1.651
Diver	1.523e+09	9.669	17.013	1.000	2.000	5.000	11.000	22.000
Disp	1.523e+09	0.001	0.008	-0.004	-0.001	0.002	0.003	0.004
ivol	1.523e+09	-0.0570	1.018	-1.416	-0.944	-0.157	0.877	1.360
iskew	1.523e+09	-0.0150	1.031	-1.408	-0.924	-0.126	0.915	1.399
Distance	1.523e+09	877.144	498.822	258.547	528.240	858.926	1185.493	1484.389
Investor Tvr	1.523e+09	-0.128	0.909	-0.662	-0.627	-0.589	-0.534	1.498
Open Price	1.523e+09	0.317	0.773	-0.852	-0.181	0.426	0.910	1.258
High Price	1.523e+09	0.315	0.773	-0.854	-0.184	0.424	0.907	1.257
Low Price	1.523e+09	0.318	0.773	-0.849	-0.180	0.428	0.911	1.259
Close Price	1.523e+09	0.317	0.773	-0.851	-0.181	0.427	0.910	1.258
Extrapolat~n	1.523e+09	0.0380	0.174	-0.155	-0.0510	0.0430	0.135	0.228
Past Perform	1.523e+09	1.009	0.124	0.897	0.947	1.004	1.063	1.130

- Why four prices?

Comment: Bias Zoo

- How do we proxy for biases?

Panel B: Investor behavioral biases

Variable	N	Mean	SD	p10	p25	p50	p75	p90
Port. Value	1.523e+09	1.187	0.521	0.462	0.979	1.363	1.560	1.651
Diver	1.523e+09	9.669	17.013	1.000	2.000	5.000	11.000	22.000
Disp	1.523e+09	0.001	0.008	-0.004	-0.001	0.002	0.003	0.004
ivol	1.523e+09	-0.0570	1.018	-1.416	-0.944	-0.157	0.877	1.360
iskew	1.523e+09	-0.0150	1.031	-1.408	-0.924	-0.126	0.915	1.399
Distance	1.523e+09	877.144	498.822	258.547	528.240	858.926	1185.493	1484.389
Investor Tvr	1.523e+09	-0.128	0.909	-0.662	-0.627	-0.589	-0.534	1.498
Open Price	1.523e+09	0.317	0.773	-0.852	-0.181	0.426	0.910	1.258
High Price	1.523e+09	0.315	0.773	-0.854	-0.184	0.424	0.907	1.257
Low Price	1.523e+09	0.318	0.773	-0.849	-0.180	0.428	0.911	1.259
Close Price	1.523e+09	0.317	0.773	-0.851	-0.181	0.427	0.910	1.258
Extrapolat~n	1.523e+09	0.0380	0.174	-0.155	-0.0510	0.0430	0.135	0.228
Past Perform	1.523e+09	1.009	0.124	0.897	0.947	1.004	1.063	1.130

- Why four prices?
- Disposition effect: how about V-shaped? (Ben-David and Hirshleifer, RFS 2012)

Comment: Bias Zoo

- How do we proxy for biases?

Panel B: Investor behavioral biases

Variable	N	Mean	SD	p10	p25	p50	p75	p90
Port. Value	1.523e+09	1.187	0.521	0.462	0.979	1.363	1.560	1.651
Diver	1.523e+09	9.669	17.013	1.000	2.000	5.000	11.000	22.000
Disp	1.523e+09	0.001	0.008	-0.004	-0.001	0.002	0.003	0.004
ivol	1.523e+09	-0.0570	1.018	-1.416	-0.944	-0.157	0.877	1.360
iskew	1.523e+09	-0.0150	1.031	-1.408	-0.924	-0.126	0.915	1.399
Distance	1.523e+09	877.144	498.822	258.547	528.240	858.926	1185.493	1484.389
Investor Tvr	1.523e+09	-0.128	0.909	-0.662	-0.627	-0.589	-0.534	1.498
Open Price	1.523e+09	0.317	0.773	-0.852	-0.181	0.426	0.910	1.258
High Price	1.523e+09	0.315	0.773	-0.854	-0.184	0.424	0.907	1.257
Low Price	1.523e+09	0.318	0.773	-0.849	-0.180	0.428	0.911	1.259
Close Price	1.523e+09	0.317	0.773	-0.851	-0.181	0.427	0.910	1.258
Extrapolat~n	1.523e+09	0.0380	0.174	-0.155	-0.0510	0.0430	0.135	0.228
Past Perform	1.523e+09	1.009	0.124	0.897	0.947	1.004	1.063	1.130

- Why four prices?
- Disposition effect: how about V-shaped? (Ben-David and Hirshleifer, RFS 2012)
- (Under)diversification: the current proxy is the number of stocks. How about covariance between the stocks? One idea is to look at portfolio volaility vs. average stock volatility

Comment: Machine Learning

- Factors that affect investors' total returns: diversification, portfolio turnover, and momentum

Comment: Machine Learning

- Factors that affect investors' total returns: diversification, portfolio turnover, and momentum
    - I assume investors are trading opposite to momentum?
    - Can we show that in the NN models?

Comment: Machine Learning

- Factors that affect investors' total returns: diversification, portfolio turnover, and momentum
  - I assume investors are trading opposite to momentum?
  - Can we show that in the NN models?
- The paper examines total returns and trading returns separately

Comment: Machine Learning

- Factors that affect investors' total returns: diversification, portfolio turnover, and momentum
  - I assume investors are trading opposite to momentum?
  - Can we show that in the NN models?
- The paper examines total returns and trading returns separately
  - A more direct way is to put trading volume on the left-hand-side
  - Goyenko et al. (2024) "Trading Volume Alpha"

Comment: Machine Learning

- Factors that affect investors' total returns: diversification, portfolio turnover, and momentum
  - I assume investors are trading opposite to momentum?
  - Can we show that in the NN models?
- The paper examines total returns and trading returns separately
  - A more direct way is to put trading volume on the left-hand-side
  - Goyenko et al. (2024) "Trading Volume Alpha"
- ResNN

Comment: Machine Learning

- Factors that affect investors' total returns: diversification, portfolio turnover, and momentum
  - I assume investors are trading opposite to momentum?
  - Can we show that in the NN models?
- The paper examines total returns and trading returns separately
  - A more direct way is to put trading volume on the left-hand-side
  - Goyenko et al. (2024) "Trading Volume Alpha"
- ResNN
  - "Each layer of neural in the residual learning framework only needs to figure out the additional information–compared to the inputs–that helps to improve maximization"
  - "Roughly speaking, this feature of ResNN also allows machine learning to be benchmarked against some economically important inputs"
  - In the empirical result, ResNN predicts the low group better. Why?

- Very interesting and ambitious paper!

- Very interesting and ambitious paper!

- Important contribution: first paper to use machine learning to study retail investors
- I encourage the authors to think more about investor sentiment and welfare
- Good luck!