Does Market Selection Eliminate Poorly Performing Investors? Evidence from Chinese Brokerage Account Records[☆]

Neil D. Pearson,[a] Zhishu Yang,[b] and Qi Zhang[c]

June 2024

[☆] We gratefully thank Han Hou for excellent research assistance.

[a] University of Illinois at Urbana-Champaign and Canadian Derivatives Institute
[b] School of Economics and Management, Tsinghua University
[c] Antai College of Economics and Management, Shanghai Jiaotong University

Market selection

Do poorly performing investors survive and remain in financial markets? Or does the process of trading winnow them out?

The classic answer, which dates at least to Alchian (1950) and Friedman (1953), is that unskilled investors are selected out

The result is a market dominated by skilled investors with prices that are likely to reflect fundamental values (Cootner 1967)

Not surprisingly, there is a theoretical literature that explores conditions under which market selection does or does not reduce or eliminate the impact of unskilled or biased investors (De Long et al. (1991), Blume and Easley (1992), Sandroni (2000), Blume and Easley (2006), Yan (2008), Kogan et al. (2017), Dindo (2019), Sihvonen (2019), and Borovička (2020))

But there is limited empirical evidence regarding whether and how market selection functions

What we do and find

We study market selection using Chinese brokerage account records that cover the 15-year period running from 2006 through 2020.

The length of the sample period allows us to see the investors entering and exiting the market, and to study exit.

We find:

- Returns of the investors who exit *exceed* those of the investors who stay in the market.
- By the end of our sample period, more than 80% of the investors in the market have had negative cumulative returns since beginning trading
- These investors account for 64% of the total brokerage account balances of the sample investors
- The way in which market selection fails differs from the theoretical models

What we do and find

We also find

- The relation between exit and cumulative returns since beginning trading is non-monotonic, in the shape of an inverted V with a sharp peak at a cumulative return of zero
- This pattern is robust to controlling for other covariates that are plausibly related to exit
- The pattern is consistent with the Barberis (2012) model of gambling
- The inverted V allows our results to be reconciled with seemingly inconsistent results in the empirical literature on learning from trading based on Finnish data, without appealing to possible differences between Chinese and Finnish investors

Data

Brokerage account records come from a total of 313 branch offices located in 30 of the 31 regions within mainland China, where a "region" can be either a province (e.g., Fujian), a municipality (e.g., Shanghai), or an autonomous region (e.g., Xinjiang).

We use the records of individual investors' trading in A shares over the period running from January 1, 2006 through December 31, 2020

Individual investors dominate trading in the Chinese stock market:.

- For example, during 2017, individual investors held only 21% of market capitalization, but accounted for 82% of trading volume

We have a total of 1,954,613 investors, who executed 888,150,579 stock trades during the sample period

Entry and exit

We consider a new investor to enter the stock market on the day he or she places his or her first stock trade

An investor is considered to exit the stock market on a date if:

- He or she closes all of his or her stock positions on that date
- The date on which he or she does this is December 31, 2017 or earlier
- He or she does not trade again through the end of the sample period, which is December 31, 2020.

We require that positions be closed on or before December 31, 2017 because we require at least three years of inactivity before we consider an investor to have exited the market. Due to this requirement, we do not use data on investors who enter the market after 2017.

Definition of lifetime return

We need a measure of an investor's cumulative return since he or she began trading

At the end of each month or year, we define the cumulative gain as:

$$\textit{Gain} = \textit{Total proceeds from sales} + \textit{value of open positions} - \textit{Total cost of purchases}$$

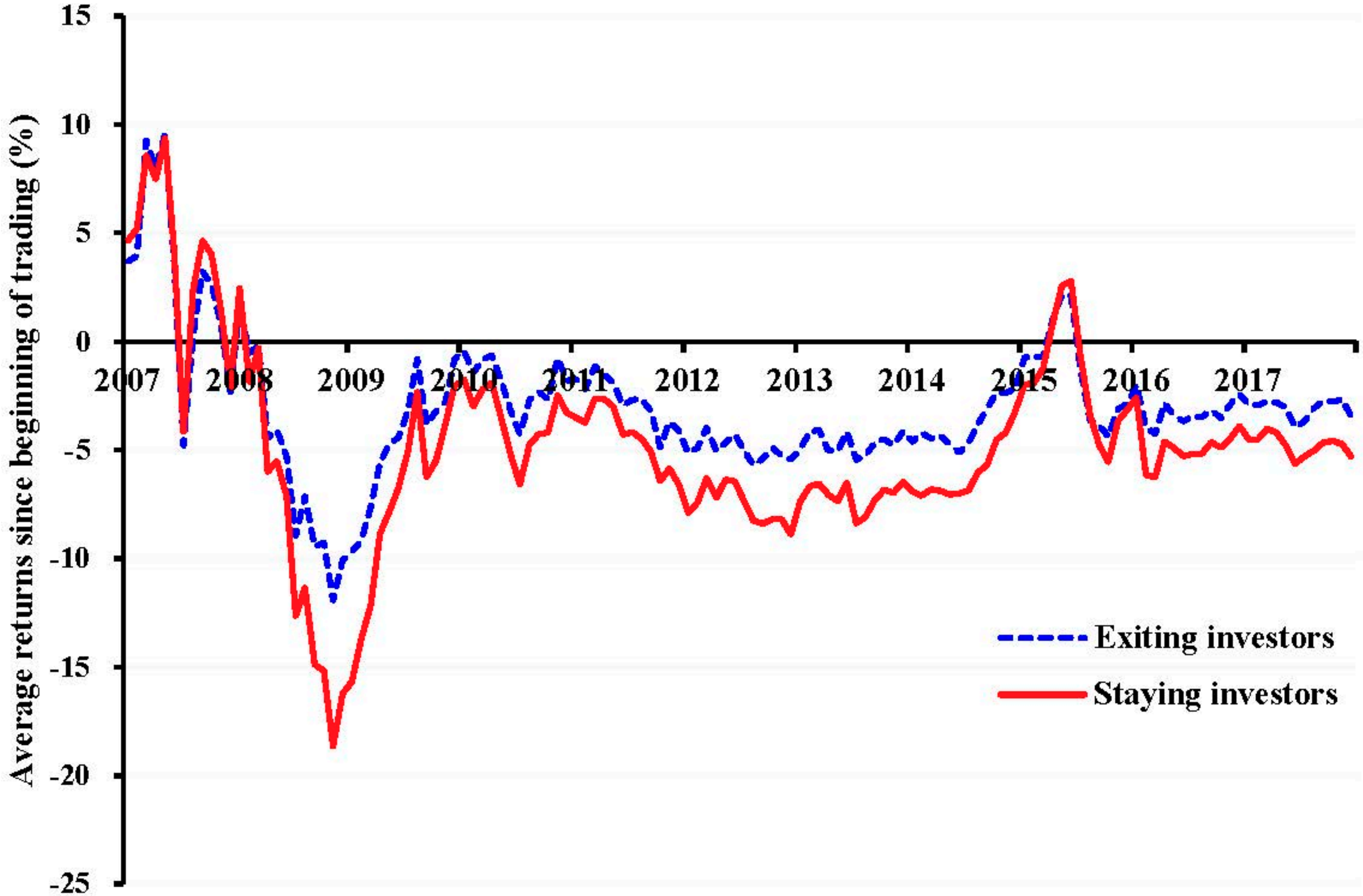
The return is then the gain divided by the total cost of purchases:

$$\textit{Return} = \textit{Gain}/(\textit{Total cost of purchases})$$

This is similar to a return measure that has been used in the private equity literature

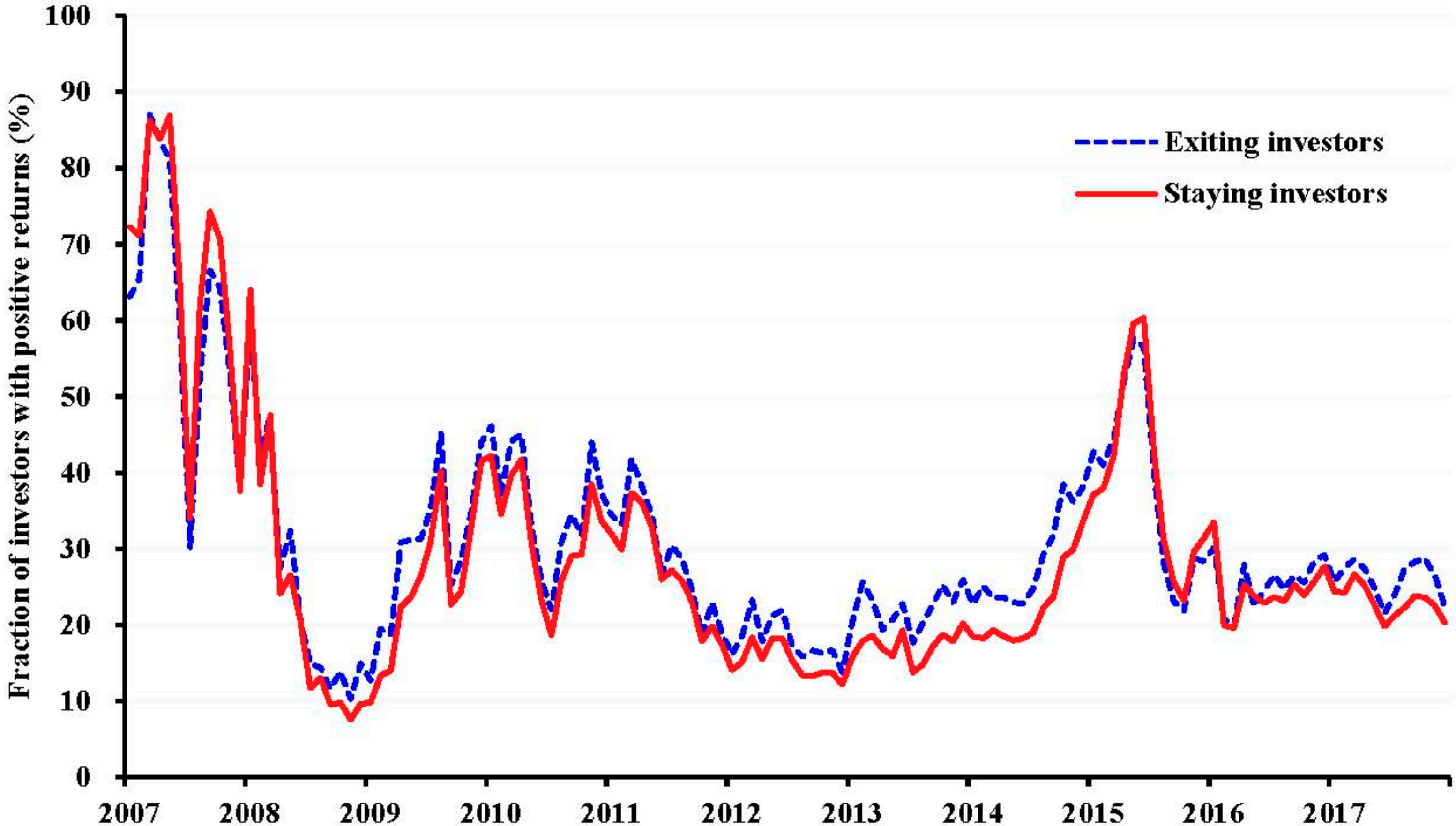
Using this measure of return, each month we compute the average returns of the exiting and staying investors (see the next slide)

Figure 1A: Average returns of exiting and staying investors



Each month, we compute the average returns of the two groups of investors through the end of the previous month

Figure 1B: Fraction of investors with return > 0



Each month, we compute the percentage with positive returns through the end of the previous month for each group of investors

Contrast with theoretical literature

In most of the theoretical literature, market selection fails, or can fail for some parameter values, because biased investors have better returns and come to dominate the market

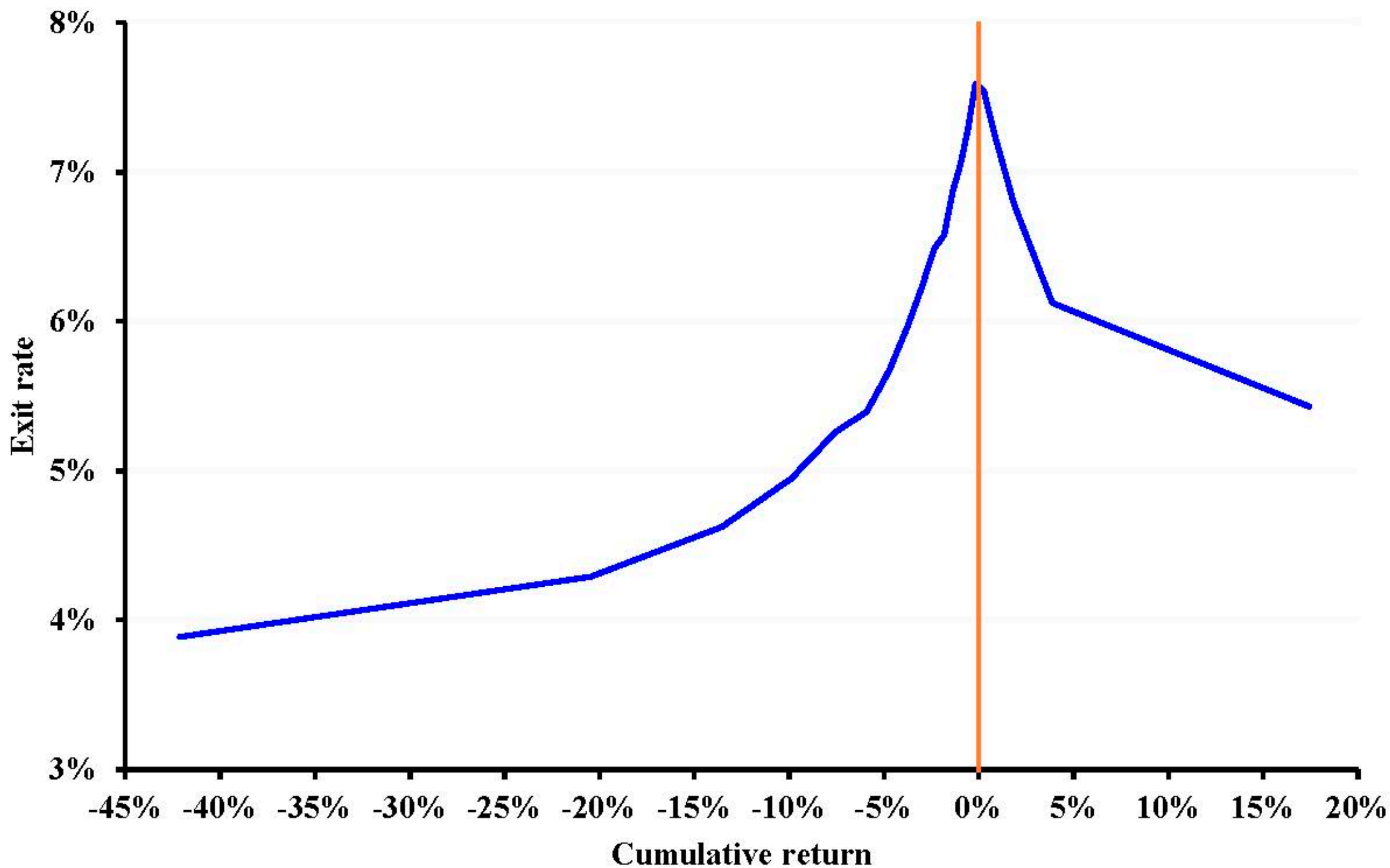
In contrast, we find that the better performing investors are more likely to exit the market

Why do the exiting investors have better returns?

The previous results showing that the investors who exit have better returns than the investors who stay is inconsistent with market selection

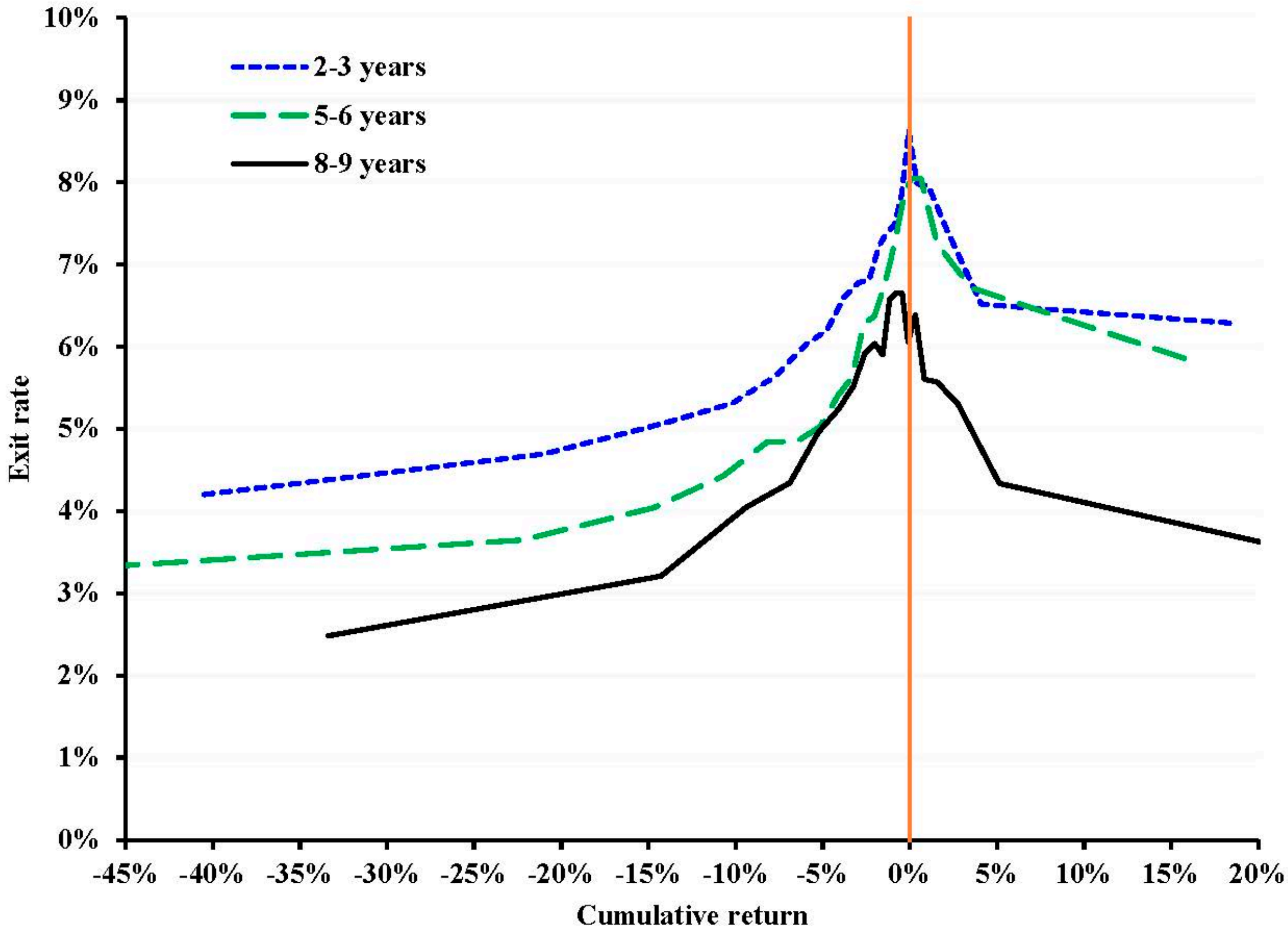
The explanation lies in a simple picture, on the next slide

Figure 2: Relation between cumulative return and exit



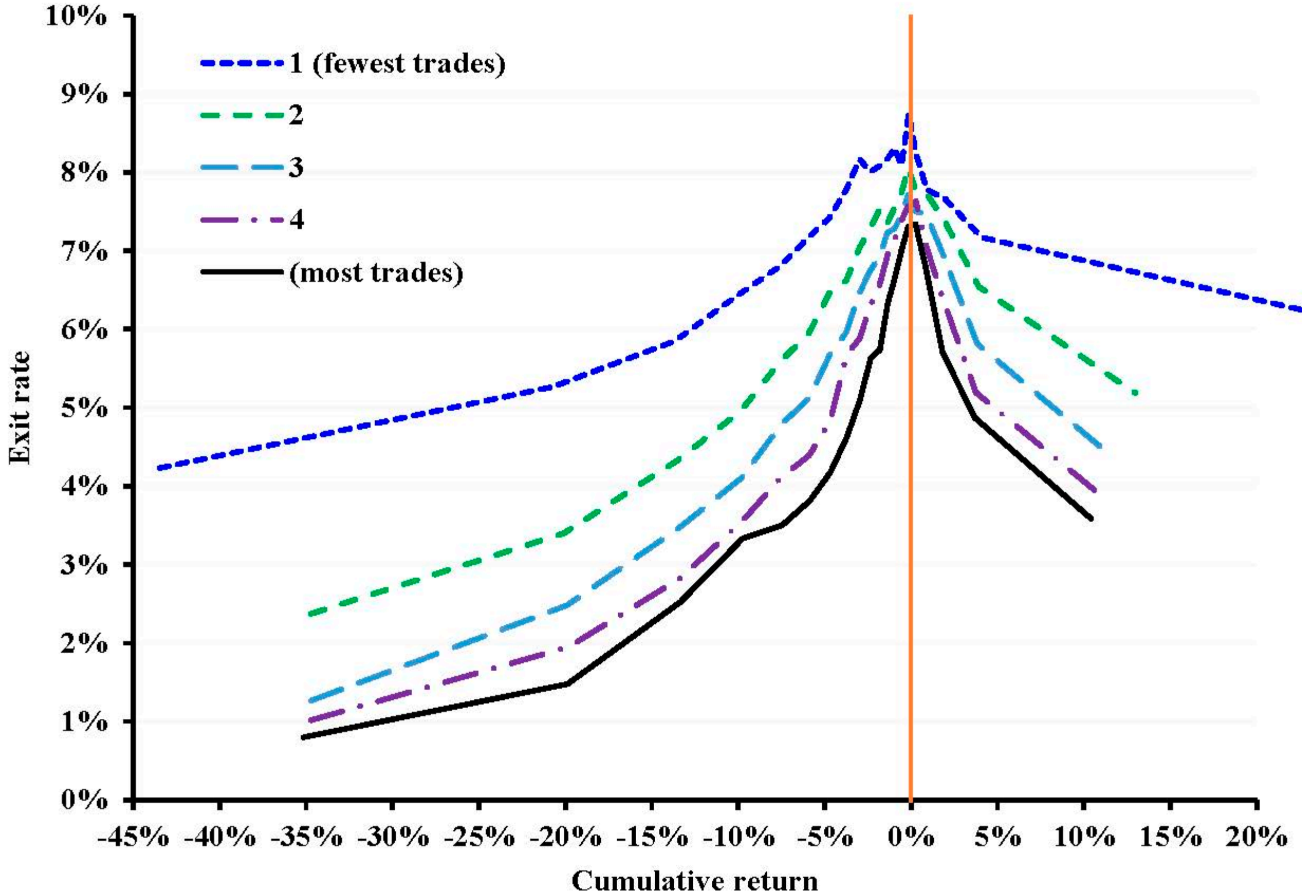
Bin scatterplot illustrating how the exit rate varies with the cumulative return since the beginning of trading

Figure 3: Cumulative return and exit by experience



Same pattern is found in investors with different levels of experience

Figure 4: Cumulative return and exit by trading frequency



Average numbers of trades per year in the five trading frequency groups are 2.8, 10.5, 23.3, 50.0, and 206.3.

Table 3: Linear and logistic regressions explaining exit

Panel A. Results for cumulative return indicator variables

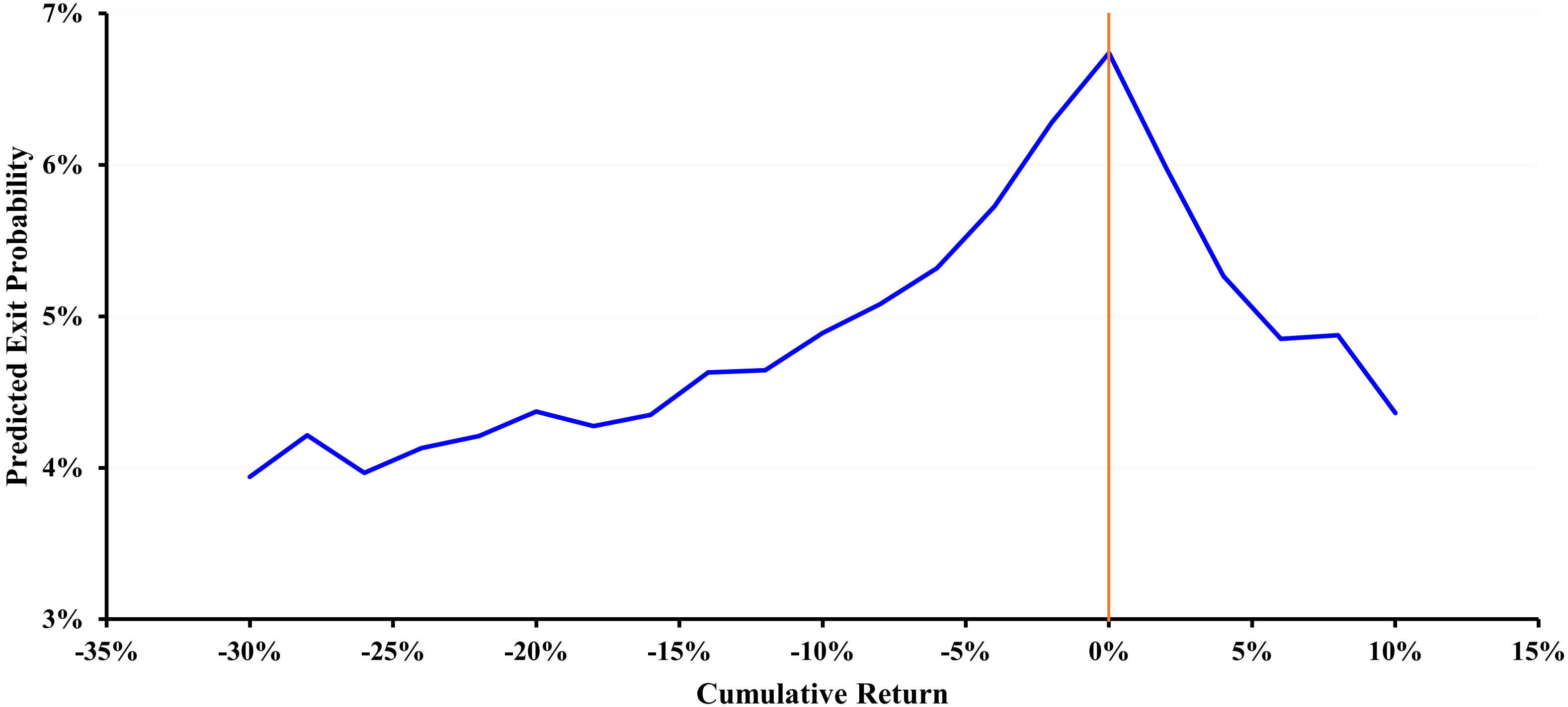
	Linear regression		Logistic regression		
Indicator variable or constant	Coefficient estimate	Standard error	Coefficient estimate	Standard error	Predicted Probability
Constant	0.04562	0.00218	−3.12923	0.08799	0.03941
[−29%, −27%)	0.00261	0.00072	0.07008	0.02341	0.04215
[−27%, −25%)	0.00011	0.00095	0.00656	0.02403	0.03966
[−25%, −23%)	0.00188	0.00120	0.04865	0.02727	0.04129
[−23%, −21%)	0.00262	0.00101	0.06851	0.02241	0.04209
[−21%, −19%)	0.00424	0.00090	0.10781	0.01950	0.04370
[−19%, −17%)	0.00329	0.00049	0.08453	0.01618	0.04274
[−17%, −15%)	0.00410	0.00143	0.10309	0.04363	0.04350
[−15%, −13%)	0.00690	0.00048	0.16831	0.02487	0.04630
[−13%, −11%)	0.00707	0.00083	0.17162	0.03605	0.04645
[−11%, −9%)	0.00951	0.00103	0.22531	0.03737	0.04888
[−9%, −7%)	0.01146	0.00099	0.26572	0.04586	0.05080
[−7%, −5%)	0.01387	0.00103	0.31435	0.04883	0.05319
[−5%, −3%)	0.01801	0.00077	0.39244	0.04803	0.05727
[−3%, −1%)	0.02376	0.00069	0.49004	0.04780	0.06277
[−1%, 1%)	0.02891	0.00112	0.56596	0.05452	0.06738
[1%, 3%)	0.02092	0.00185	0.43779	0.06889	0.05976
[3%, 5%)	0.01293	0.00179	0.30369	0.06240	0.05266
[5%, 7%)	0.00812	0.00320	0.21734	0.08578	0.04851
[7%, 9%)	0.00838	0.00293	0.22220	0.07960	0.04874
[9%, max%]	0.00217	0.00261	0.10570	0.06938	0.04361

Table 3: Linear and logistic regressions explaining exit

Panel B. Results for control variables

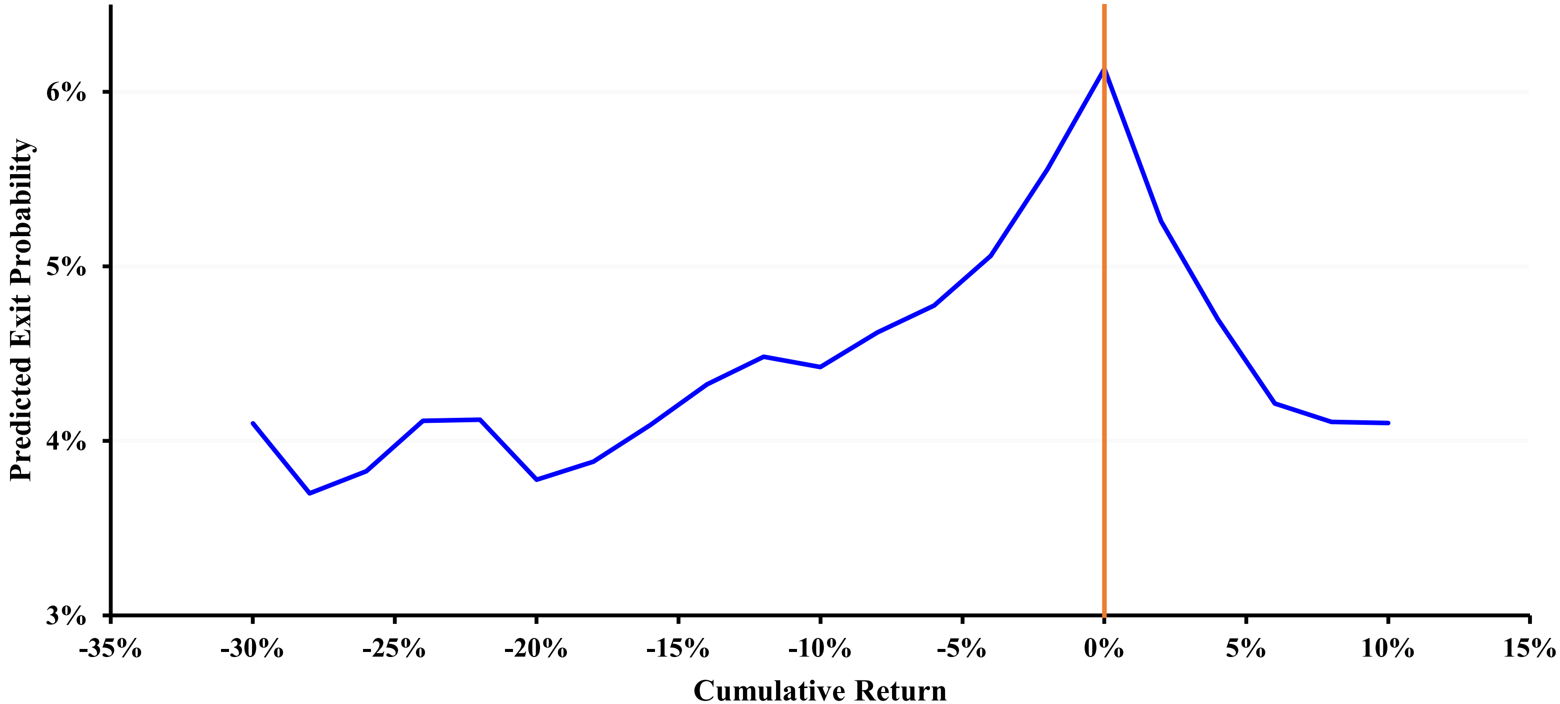
	Linear regression		Logistic regression		
Control variable	Coefficient estimate	Standard error	Coefficient estimate	Standard error	Marginal effect
Male	0.00356	0.00049	0.06189	0.00730	0.00328
Age	–0.00005	0.00004	–0.00102	0.00084	–0.00005
Wealthy	–0.00851	0.00160	–0.14833	0.01788	–0.00761
CumulTrades	–0.00030	0.00005	–0.00872	0.00064	–0.00046
StkNum	0.00002	0.00000	0.00055	0.00005	0.00003
Adj. R^2	0.0123		0.0249		
No. of obs.	4,794,784		4,794,784		

Predicted exit rates computed from the logistic regression



Predicted exit rates computed by setting the indicator variable for each return range to one and the indicator variables for the other return ranges to zero

Predicted exit rates of wealthy investors



Predicted exit rates of wealthy investors based on a logistic regression model. A wealthy investor is one whose average brokerage account balance is in the top quartile of all investors' account balances

The inverted V explains a puzzling feature of the data

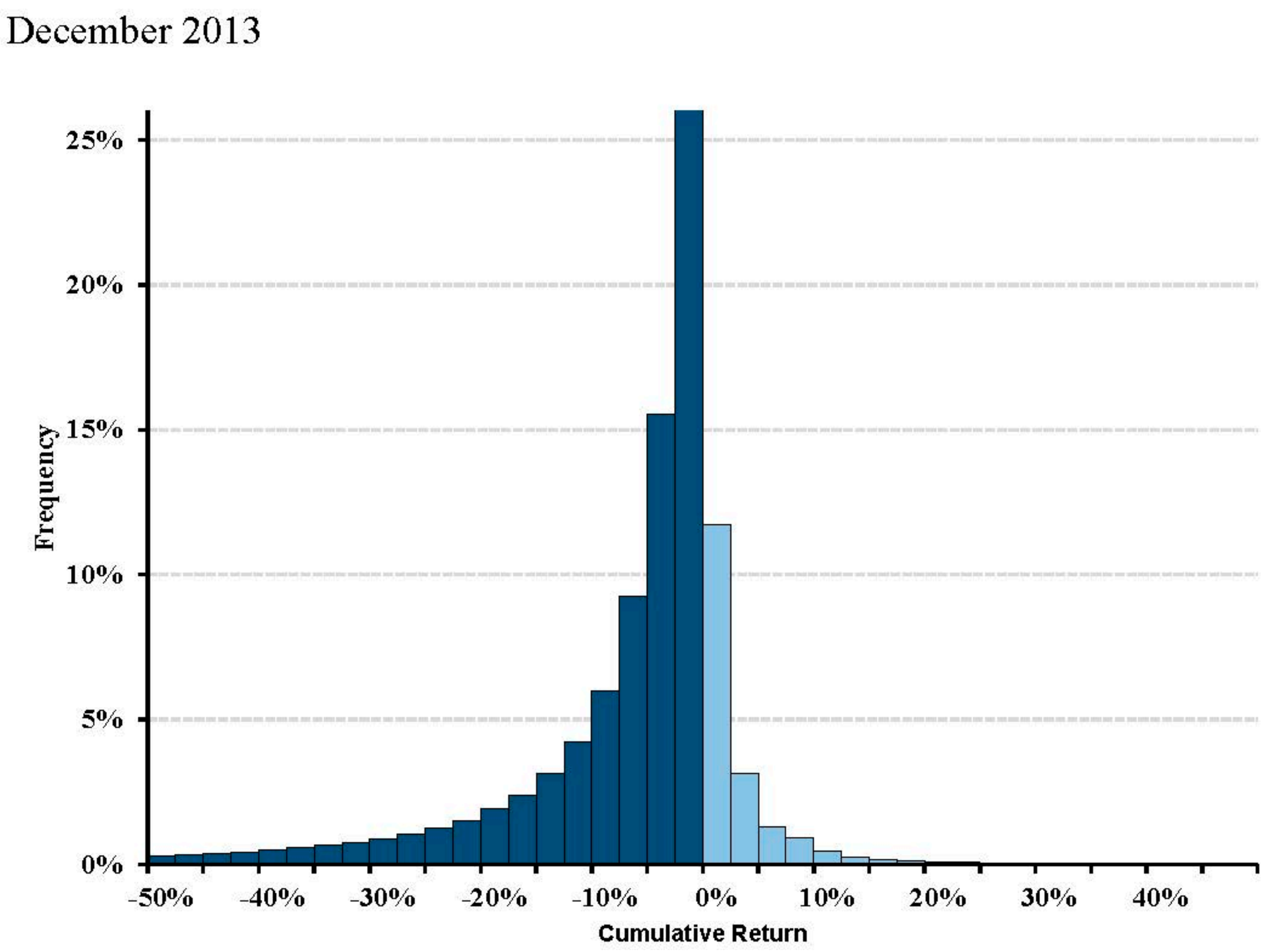
Figure 1 showed that average returns of exiting investors exceed those of staying investors during most months

But in a few months, average returns of exiting investors were less than those of staying investors—why?

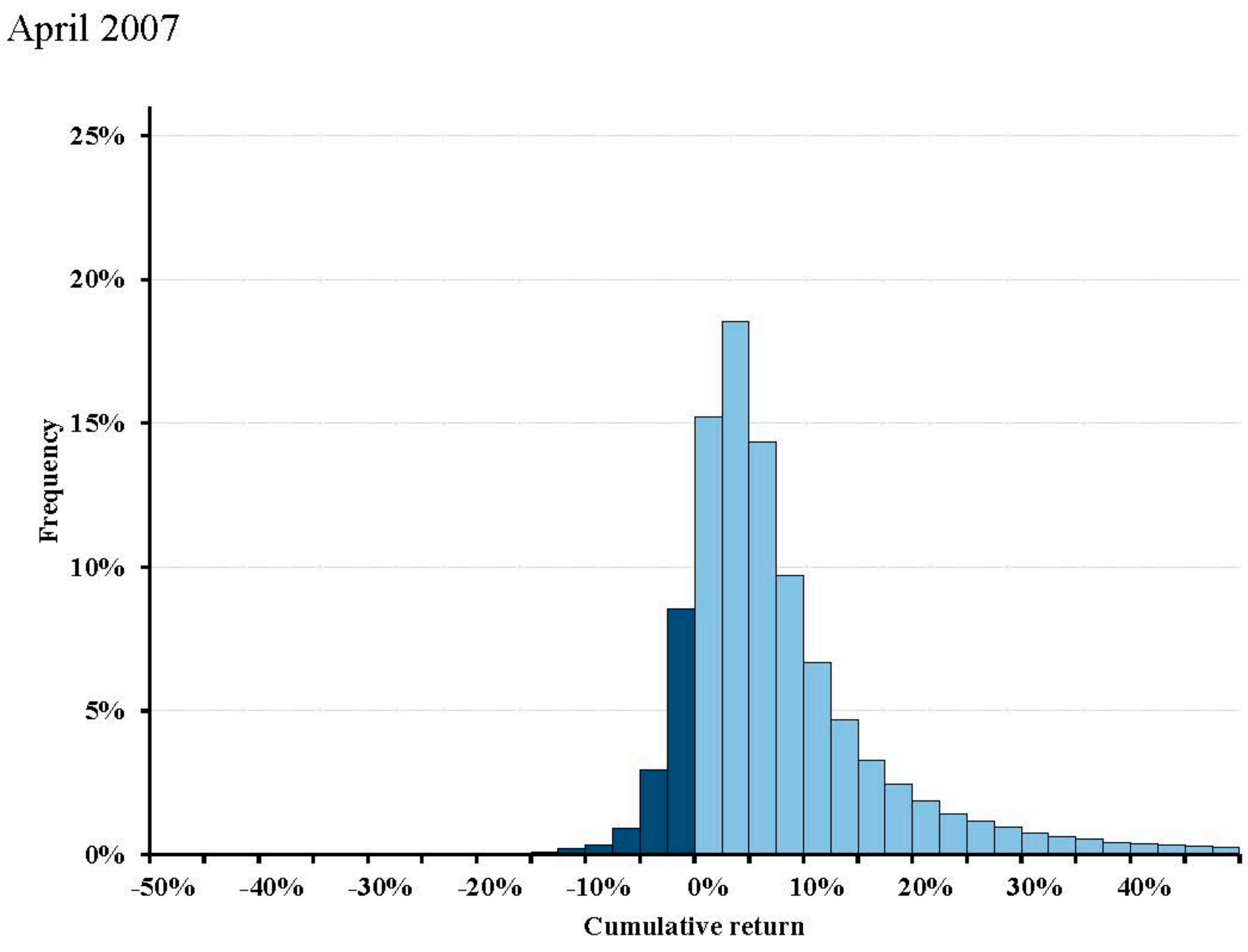
In most months, most investors had negative lifetime returns that fall on the positively-sloped part of the inverted V to the left of zero

But, at some points during the 2006-2007 and 2014-2015 stock market booms, a majority of investors had positive lifetime returns that fall on the negatively sloped part of the inverted V to the right of zero

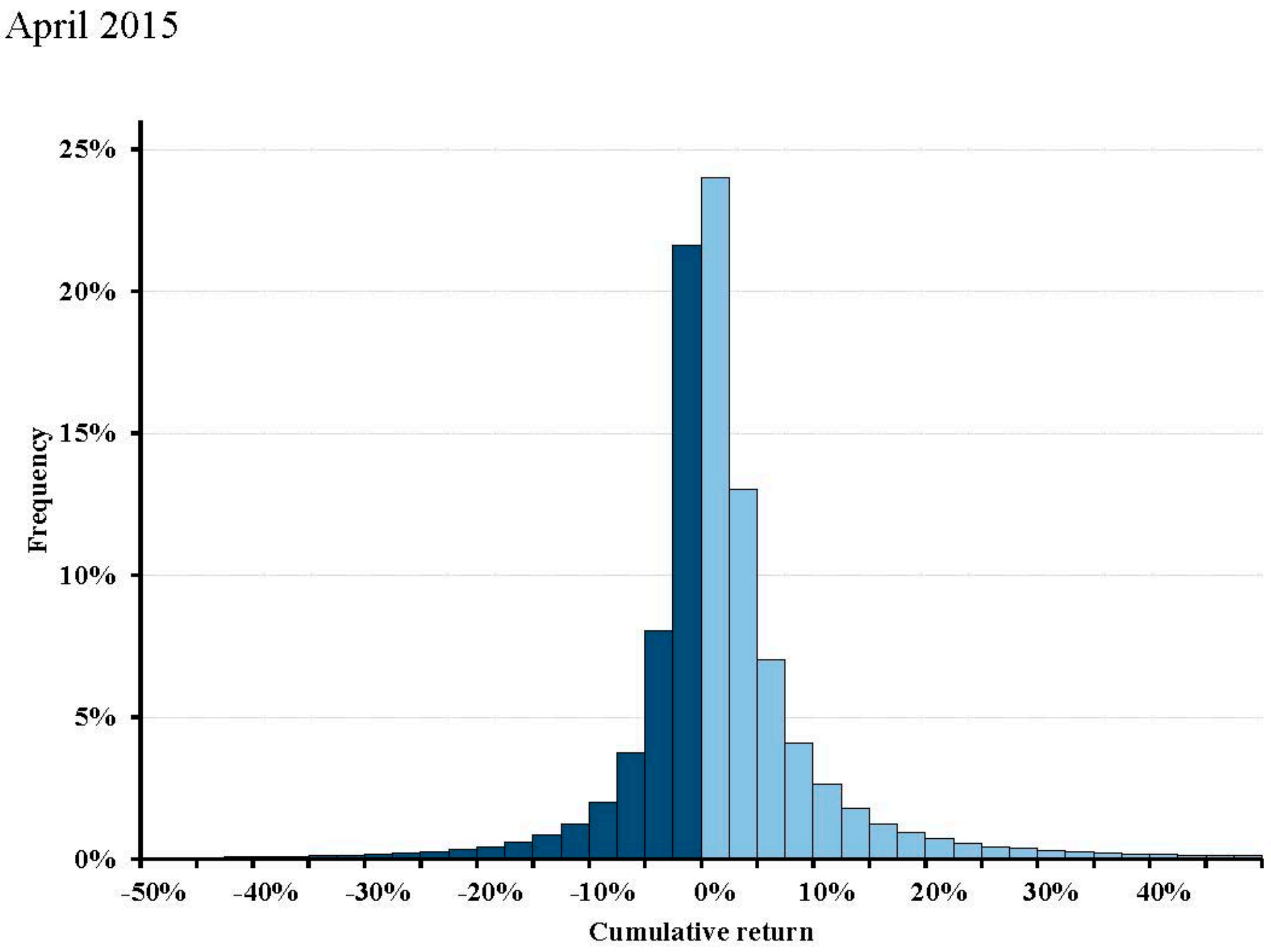
Distribution of lifetime returns of investors in the market at end of December 2013



Distribution of lifetime returns of investors in the market at end of April 2007



Distribution of lifetime returns of investors in the market at end of April 2015



The inverted V allows our results to be reconciled with empirical evidence on learning through trading

Seru, Shumway, and Stoffman (2010) and Linnainmaa (2011) report empirical results showing that Finnish investors display a negative relation between returns and exit

This seems inconsistent with our finding that the average returns of the Chinese retail investors who exit the market exceed those of the investors who stay

The inverted V allows our results to be reconciled with those in Seru, Shumway, and Stoffman (2010) and Linnainmaa (2011), without appealing to possible differences between Finnish and Chinese investors

The Finnish stock market index performed very well during the sample periods in Seru, Shumway, and Stoffman (2010) and Linnainmaa (2011), so that the Finnish investors in their sample were on the negatively-sloped region of the inverted V to the right of zero

Why is the exit rate highest at a cumulative return of zero?

The sharp peak at zero in the inverted V strongly suggests a reference point (of zero), which in turn suggests prospect theory

But our setting of a sequence of trades is different from the typical analysis of a single gamble in prospect theory

However, there is a prospect theory model of a sequence of trades, the Barberis (2012) model of casino gambling, that derives predictions for agents' decisions to continue or quit gambling (exit the casino) as a function of their cumulative gains since beginning gambling

If we analogize a lifetime of stock trading to an evening in a casino, then the Barberis (2012) model speaks to a lifetime of trading and predicts the patterns we find

Barberis (2012) model of gambling

Agents face a sequence of mean zero gambles, interpreted as an evening in a casino

After the outcome of each gamble, agents decide whether to take the next bet, or stop betting (exit the casino)

Agents have cumulative prospect theory preferences, which include probability weighting

Some stopping rules, notably continuing if cumulative gains are positive but stopping if they reach zero, create positively-skewed payoffs

Probability weighting causes agents to desire such payoffs and begin betting, even if the odds are fair or even somewhat unfair

Agents begin betting with the plan to exit if they start to experience losses

Barberis (2012) model of gambling

Three types of agents:

- Naïve agents with time-inconsistent preferences, who do not follow their planned strategies
- Sophisticated agents who are able to commit to their planned strategies
- Sophisticated agents who understand the time-inconsistency but are unable to commit to their plans.

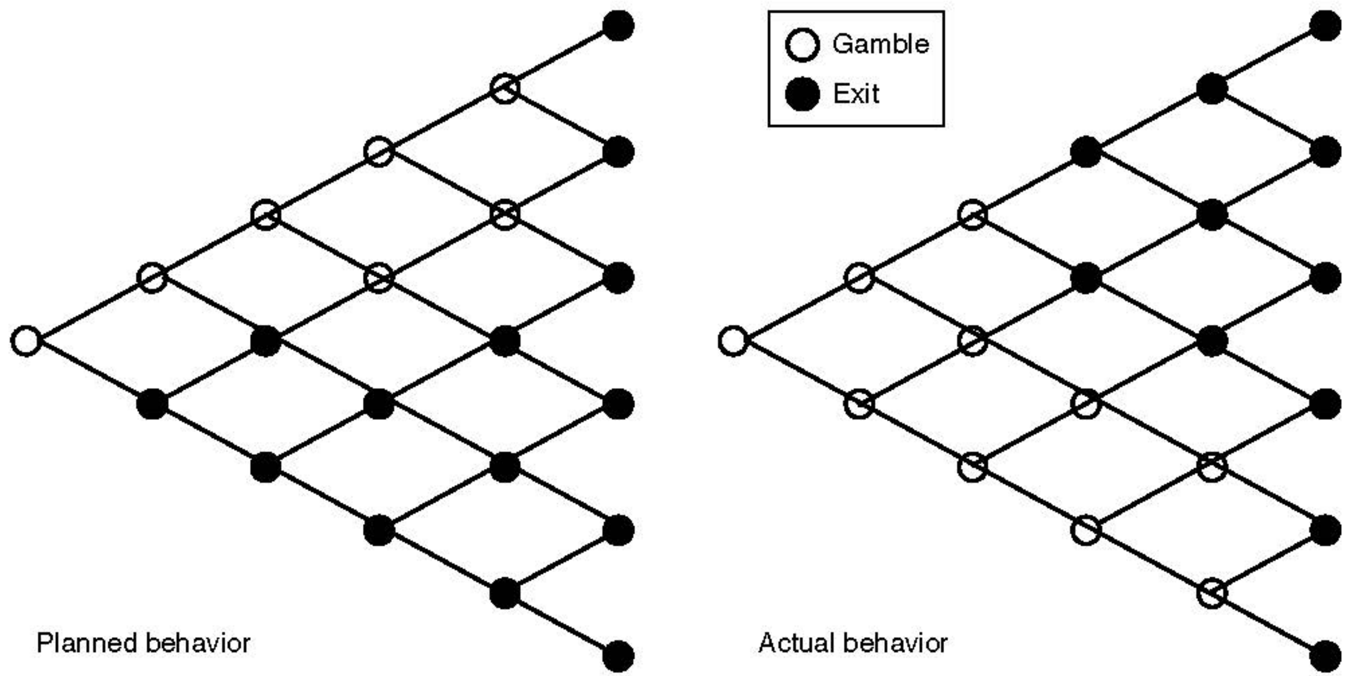
Naïve agents end up following gain-exit strategies in which they continue betting in the region of losses, but quit if they enter the region of gains

Sophisticated agents who can commit follow their planned strategies and continue betting in the region of gains, but quit if they enter the region of losses

Sophisticated agents who cannot commit do not begin betting

Barberis (2012) Figure 4

Figure 4 **A Naive Prospect Theory Agent's Planned and Actual Gambling Behavior**



If a node has a black color, then the agent does not gamble at that node. At the remaining nodes, he or she gambles.

Table 4A&B: Annual exit rates for double-sort

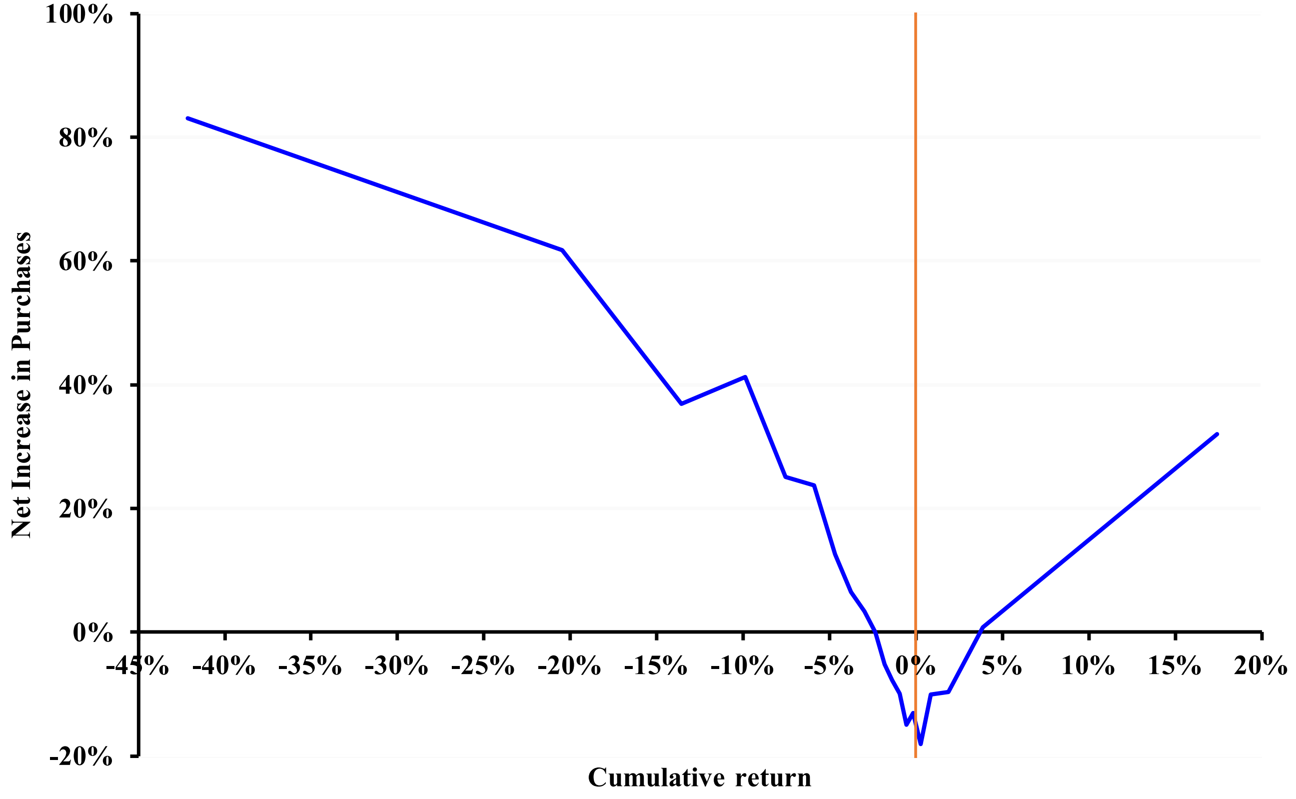
Panel A. Annual exit rates

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)	4.04%	6.67%	7.32%	7.92%	8.73%
[–35%, –25%)	2.71%	3.81%	6.07%	6.84%	7.69%
[–25%, –15%)	2.81%	2.92%	4.31%	6.75%	7.30%
[–15%, –5%)	2.61%	2.74%	3.05%	5.19%	7.41%
[–5%, 5%)	2.45%	2.76%	2.79%	3.82%	6.73%
[5%, 15%)	2.27%	2.21%	2.35%	3.26%	4.88%
[15%, max%]	2.41%	2.61%	1.87%	2.70%	4.13%

Panel B. Standard errors

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)	0.06%	0.16%	0.19%	0.22%	0.23%
[–35%, –25%)	0.11%	0.09%	0.14%	0.17%	0.19%
[–25%, –15%)	0.20%	0.10%	0.05%	0.08%	0.11%
[–15%, –5%)	0.35%	0.20%	0.07%	0.03%	0.05%
[–5%, 5%)	0.49%	0.33%	0.15%	0.04%	0.02%
[5%, 15%)	0.75%	0.55%	0.30%	0.17%	0.05%
[15%, max%]	0.84%	0.65%	0.36%	0.26%	0.07%

Relation between cumulative return and change in purchases

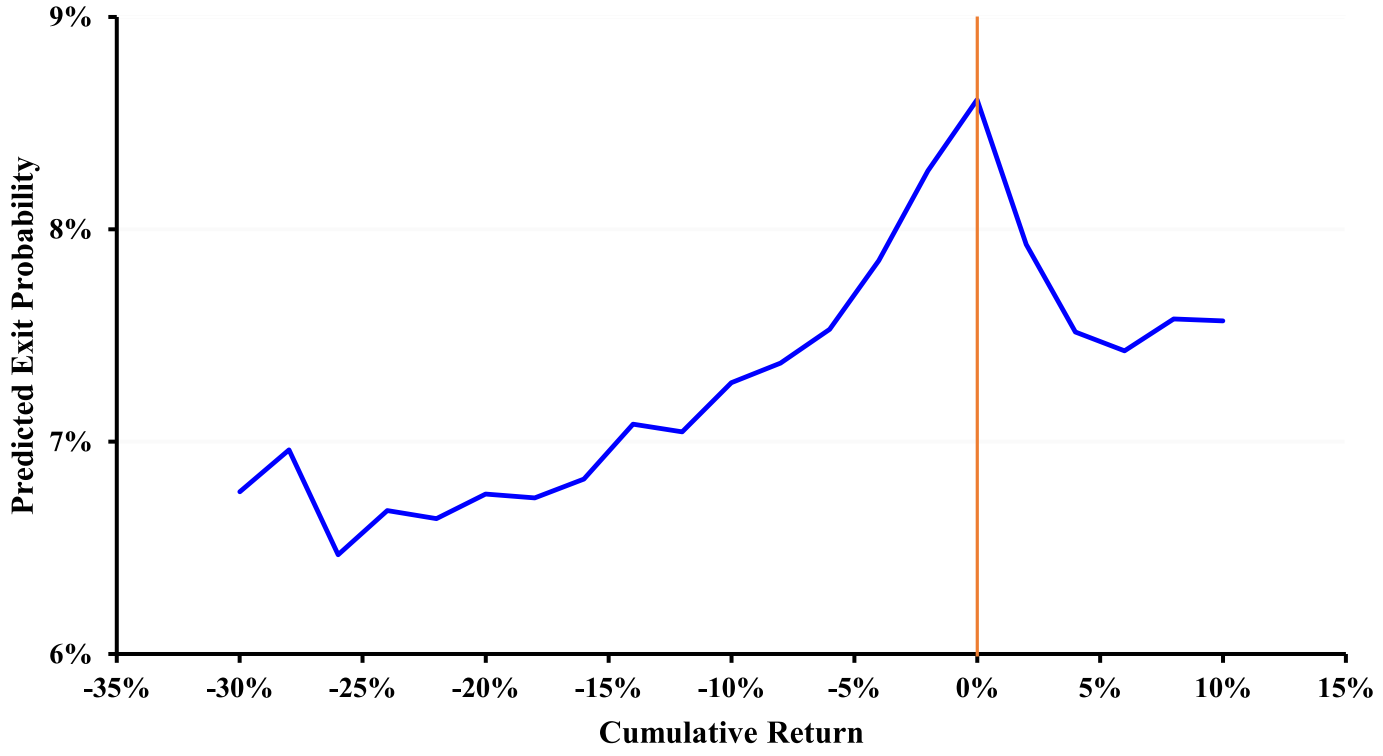


For each group, we compute the difference between the stock purchases in year t and year $t-1$, and then scale by the purchases in year $t-1$

Relation between cumulative return and numbers of investors who increase purchases

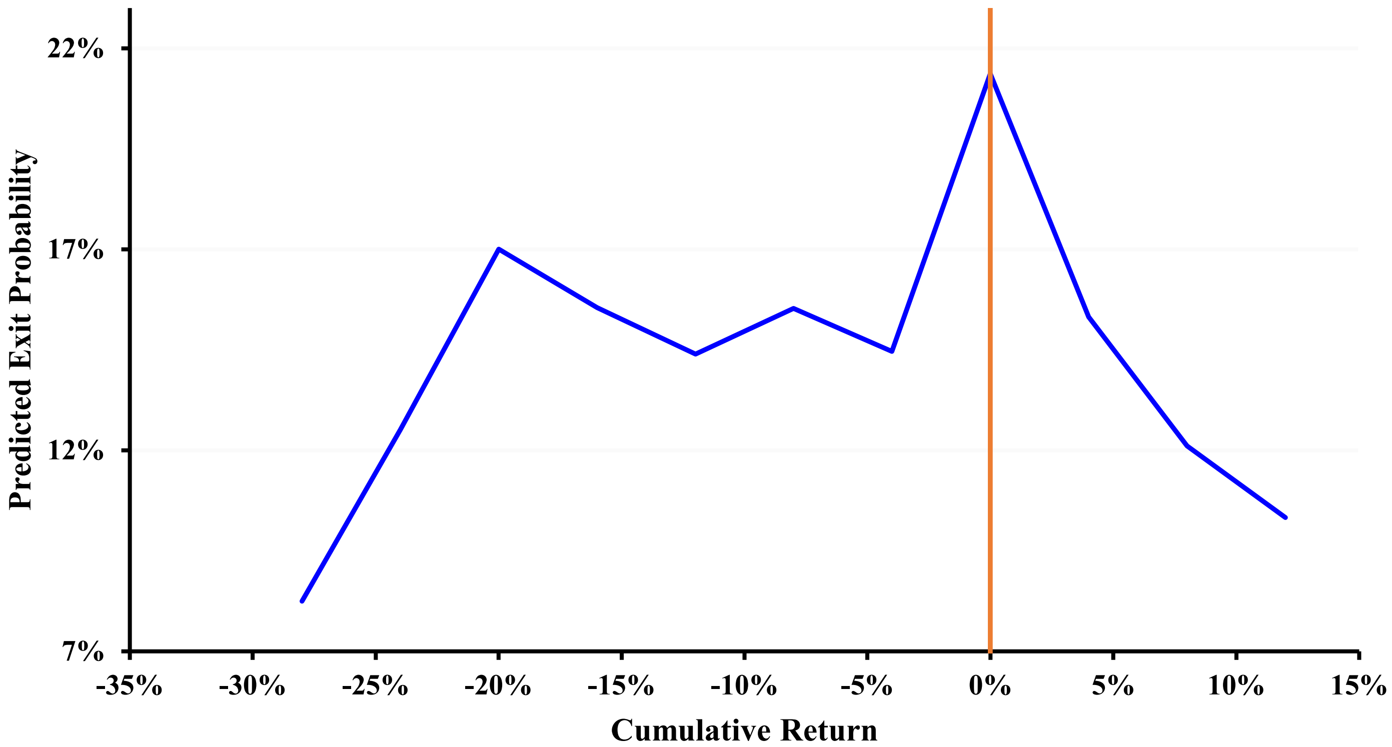
For each group, we compute the difference between the no. of investors who increase the purchases in year *t* minus the number who do not increase, and then scale by the number of investors

Alternative definition of exit: Stop trading for at least three years, but possibly continue to hold stocks



The figure plots predicted exit rates from a logistic regression that explains exit

Predicted exit rates for "legal persons"



The figure plots predicted exit rates from a logistic regression that explains exit of legal persons

Conclusion

Market selection fails among Chinese individual investors

It is explained by the inverted V-shaped relation between exit frequency and cumulative returns

The inverted V in turn is predicted by the Barberis (2012) model of gambling

We think our results using data from one securities firm likely generalize to all Chinese retail investors, as there is not much differentiation in the customers of different Chinese securities firms

Do they generalize more broadly?

They might—they are consistent with cumulative prospect theory, which has been proposed as a theory that is broadly and perhaps universally applicable

Extra slides

Extra slides not to be used in short presentation

Additional figures

Additional figures not to be used in short presentation

Contrast with theoretical literature

In most of the theoretical literature, market selection fails, or can fail for some parameter values, because biased investors have better returns and come to dominate the market

In contrast, we find that the better performing investors are more likely to exit the market

Additional results

Results of additional linear and logistic regression models

Tables 7A&B: Linear probability model explaining exit

Panel A. Estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.85%	1.34%	2.16%	2.78%
[–35%, –25%)	–0.84%	0.00%	0.46%	1.01%	1.66%
[–25%, –15%)	–0.87%	–0.58%	0.38%	1.15%	1.33%
[–15%, –5%)	–0.90%	–0.85%	–0.42%	1.08%	1.70%
[–5%, 5%)	–1.24%	–0.90%	–0.93%	0.27%	2.17%
[5%, 15%)	–1.52%	–1.57%	–1.56%	–0.71%	0.53%
[15%, max%]	–1.49%	–1.33%	–2.18%	–1.50%	–0.14%

Panel B. Standard errors of estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.54%	0.82%	0.56%	0.64%
[–35%, –25%)	0.06%	0.10%	0.28%	0.33%	0.25%
[–25%, –15%)	0.33%	0.18%	0.08%	0.13%	0.09%
[–15%, –5%)	0.13%	0.20%	0.14%	0.10%	0.06%
[–5%, 5%)	0.54%	0.57%	0.27%	0.20%	0.11%
[5%, 15%)	0.70%	0.49%	0.49%	0.35%	0.27%
[15%, max%]	1.11%	0.45%	0.52%	0.25%	0.26%

Table 7C: Linear prob. model of exit (control variables)

Panel C. Estimates of coefficients on the control variables

Control variable	Coefficient estimate	Standard error
Male	0.00365	0.00051
Age	–0.00005	0.00004
Wealthy	–0.00853	0.00163
CumulTrades	–0.00029	0.00005
StkNum	0.00003	0.00001
Adj. R^2	0.0122	
No. of obs.	4,794,784	

Table 8A&B: Logistic regression model explaining exit

Panel A. Estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.23	0.30	0.41	0.48
[–35%, –25%)	–0.31	–0.03	0.16	0.25	0.34
[–25%, –15%)	–0.30	–0.24	0.07	0.28	0.29
[–15%, –5%)	–0.33	–0.32	–0.19	0.24	0.36
[–5%, 5%)	–0.44	–0.32	–0.33	0.03	0.44
[5%, 15%)	–0.53	–0.57	–0.54	–0.22	0.13
[15%, max%]	–0.48	–0.42	–0.79	–0.45	–0.03

Panel B. Standard errors of estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.06	0.09	0.04	0.04
[–35%, –25%)	0.06	0.02	0.03	0.03	0.03
[–25%, –15%)	0.08	0.03	0.02	0.01	0.04
[–15%, –5%)	0.07	0.04	0.02	0.04	0.04
[–5%, 5%)	0.23	0.17	0.06	0.05	0.05
[5%, 15%)	0.33	0.22	0.16	0.08	0.08
[15%, max%]	0.40	0.16	0.23	0.07	0.08

Table 8C&D: Incremental effects of the indicator variables

Panel C. Incremental effects of the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)	4.17%	5.20%	5.55%	6.17%	6.59%
[–35%, –25%)	3.08%	4.05%	4.87%	5.29%	5.74%
[–25%, –15%)	3.13%	3.32%	4.47%	5.42%	5.51%
[–15%, –5%)	3.03%	3.08%	3.47%	5.22%	5.85%
[–5%, 5%)	2.74%	3.07%	3.03%	4.28%	6.30%
[5%, 15%)	2.50%	2.40%	2.48%	3.38%	4.70%
[15%, max%]	2.61%	2.77%	1.94%	2.69%	4.06%

Table 8D: Logistic regression model (control variables)

Panel D. Estimates of coefficients on the control variables

Control variable	Coefficient estimate	Standard error	Marginal effect
Male	0.0636	0.0070	0.0032
Age	–0.0011	0.0008	0.0001
Wealthy	–0.1490	0.0184	–0.0072
CumulTrades	–0.0081	0.0006	–0.0004
StkNum	0.0007	0.0001	0.0000
Adj. R^2	0.0249		
No. of obs.	4,794,784		

Proxy for disposition effect

Is the phenomenon we observe subsumed by the usual disposition effect?

To measure the extent to which investors are subject to the disposition effect, we follow Odean (1998) and calculate the difference between Proportion of Gains Realized (PGR) and Proportion of Losses Realized (PLR) in investors' portfolios

Specifically, PGR and PLR are

$$PGR = \frac{\#\ of\ Realized\ Gains}{\#\ of\ Realized\ Gains + \#\ of\ Paper\ Gains}$$

$$PLR = \frac{\#\ of\ Realized\ Losses}{\#\ of\ Realized\ Losses + \#\ of\ Paper\ Losses}$$

Then, for each investor,

$$DOD = PGR\ - PLR$$

Table 9A&B: Linear probability model of exit w/*DOD*

Panel A. Estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.47%	0.82%	1.96%	2.75%
[–35%, –25%)	–0.21%	0.76%	0.80%	1.53%	2.00%
[–25%, –15%)	–0.43%	0.01%	1.15%	1.79%	1.83%
[–15%, –5%)	–0.34%	–0.20%	0.27%	1.80%	2.42%
[–5%, 5%)	–0.50%	–0.22%	–0.25%	1.00%	2.82%
[5%, 15%)	–0.99%	–0.52%	–1.08%	–0.29%	1.17%
[15%, max%]	–1.92%	–1.02%	–1.43%	–1.12%	0.33%

Panel B. Standard errors of estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.30%	0.51%	0.49%	0.34%
[–35%, –25%)	0.12%	0.14%	0.36%	0.34%	0.27%
[–25%, –15%)	0.30%	0.14%	0.11%	0.25%	0.14%
[–15%, –5%)	0.41%	0.21%	0.11%	0.13%	0.12%
[–5%, 5%)	0.72%	0.37%	0.23%	0.14%	0.13%
[5%, 15%)	0.71%	0.74%	0.52%	0.27%	0.18%
[15%, max%]	0.83%	0.53%	0.71%	0.22%	0.26%

Table 9C: Linear prob. Model (control variables), w/*DOD*

Panel C. Estimates of coefficients on the control variables

Control variable	Coefficient estimate	Standard error
Male	0.00270	0.00047
Age	–0.00006	0.00004
Wealthy	–0.00855	0.00160
CumulTrades	–0.00028	0.00005
StkNum	0.00002	0.00001
DOD	–0.02311	0.00184
Adj. R^2	0.0121	
No. of obs.	4,481,792	

Table 10A&B: Logistic regression model of exit w/*DOD*

Panel A. Estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.31	0.39	0.57	0.69
[–35%, –25%)	–0.20	0.24	0.38	0.51	0.58
[–25%, –15%)	–0.26	–0.07	0.36	0.56	0.56
[–15%, –5%)	–0.19	–0.13	0.06	0.53	0.65
[–5%, 5%)	–0.14	–0.06	–0.08	0.31	0.71
[5%, 15%)	–0.26	–0.15	–0.35	–0.03	0.40
[15%, max%]	–0.77	–0.26	–0.46	–0.28	0.19

Panel B. Standard errors of estimates of coefficients on the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)		0.05	0.08	0.05	0.03
[–35%, –25%)	0.05	0.03	0.06	0.04	0.04
[–25%, –15%)	0.14	0.04	0.02	0.03	0.04
[–15%, –5%)	0.19	0.07	0.03	0.03	0.03
[–5%, 5%)	0.29	0.14	0.08	0.05	0.04
[5%, 15%)	0.30	0.32	0.20	0.08	0.07
[15%, max%]	0.63	0.23	0.34	0.10	0.09

Table 10C&D: Incr. effects of indicator variables, w/*DOD*

Panel C. Incremental effects of the indicator variables

Cum. return through $t-2$	Cumulative return through end of year $t-1$ [min%, –35%)	[–35%, –25%)	[–25%, –15%)	[–15%, –5%)	[–5%, max%]
[min%, –35%)	3.15%	4.26%	4.56%	5.45%	6.09%
[–35%, –25%)	2.60%	3.96%	4.53%	5.15%	5.50%
[–25%, –15%)	2.45%	2.96%	4.47%	5.36%	5.38%
[–15%, –5%)	2.61%	2.79%	3.32%	5.21%	5.87%
[–5%, 5%)	2.76%	2.97%	2.93%	4.26%	6.23%
[5%, 15%)	2.44%	2.71%	2.24%	3.06%	4.62%
[15%, max%]	1.48%	2.44%	2.02%	2.40%	3.80%

Table 10D: Logistic model w/*DOD* (control variables)

Panel D. Estimates of coefficients on the control variables

Control variable	Coefficient estimate	Standard error	Marginal effect
Male	0.04558	0.00687	0.00230
Age	–0.00137	0.00104	–0.00007
Wealthy	–0.14652	0.01780	–0.00716
CumulTrades	–0.00766	0.00062	–0.00039
StkNum	0.00056	0.00007	0.00003
DOD	–0.41876	0.02812	–0.02110
Adj. R^2	0.0252		
No. of obs.	4,481,792		

Additional results

The next few slides display additional results showing that there is also a disposition effect at the level of individual trades

Transaction cycles and cycle returns

The next of analyses focus on transaction cycles and cycle returns

Investors sometimes build up a position by multiple purchases and liquidate the position through multiple sales.

We address this by introducing the notion of a *transaction cycle*

- Begins when an investor opens a position in a stock
- Ends when the investor closes all of positions in the stock

Cycle return:

- Weighted sum of the sale prices, weighted by the quantities sold in the various sell orders
- Divided by the weighted sum of the purchase prices, where the weights are the quantities purchased in the various buy orders
- Minus one

Cox hazard rate model of an investor's decision to sell

Cox model specifies that $\lambda_{i,k,t}(\tau)$, the hazard function of selling shares to close an open position by investor i in stock k on date t, τ trading days after the end of the investor's last transaction cycle, takes the form

$$\lambda_{i,k,t}(\tau) = \lambda(\tau) \times e^{x_{i,k,t}\beta}$$

where $\lambda(\tau)$ is the baseline hazard rate and $x_{i,k,t}$ is a vector of covariates that proportionally shift the baseline hazard

Two key covariates are:

- $UnrealizedRet_{i,k,t}$: return on an open position by investor i in stock k on date t
- $I(UnrealizedRet_{i,k,t} > 0)$: indicator variable that equals one if $UnrealizedRet_{i,k,t}$ is positive

We also estimate a logistic regression model using the same covariates

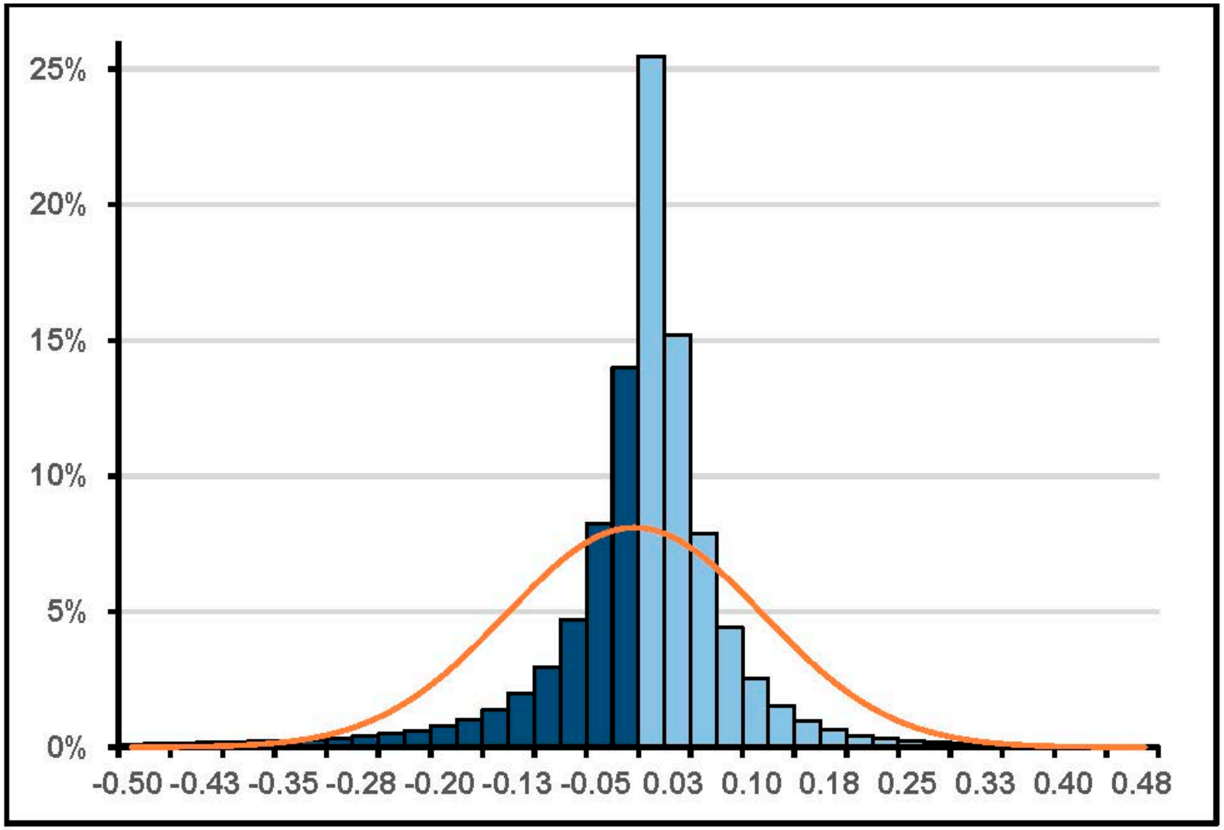
Table A1: Results of empirical models of sales

	Cox hazard rate model		Logistic Regression Model	
	(1)	(2)	(3)	(4)
Explanatory variable	Coefficient	*p*-value	Coefficient	*p*-value
UnrealizedRet	1.356	<0.0001	0.4186	<0.0001
I(*UnRealizedRet* > 0)	0.367	<0.0001	0.9226	<0.0001
HoldingRet	0.5232	<0.0001	0.6362	<0.0001
I(*HoldingRet* > 0)	0.0407	<0.0001	0.2011	<0.0001
WithHoldingRet	–0.0741	<0.0001	–0.0796	<0.0001
MktRet1Day	0.1129	<0.0001	1.9015	<0.0001
MktRet4Day	0.0451	<0.0001	0.1403	<0.0001
MktRet3Week	0.0388	<0.0001	–0.0849	<0.0001
Turnover1Day	0.0013	<0.0001	0.0033	<0.0001
Turnover4Day	0.0004	0.0003	0.0032	<0.0001
Turnover3Week	0.0003	0.0076	0.0029	<0.0001
No. of observations	3,503,718		3,503,718	
Stock fixed effects	Yes		Yes	
Date fixed effects	Yes		Yes	
Duration fxd. effects			Yes	

Table A2: Distributions of cycle durations and returns

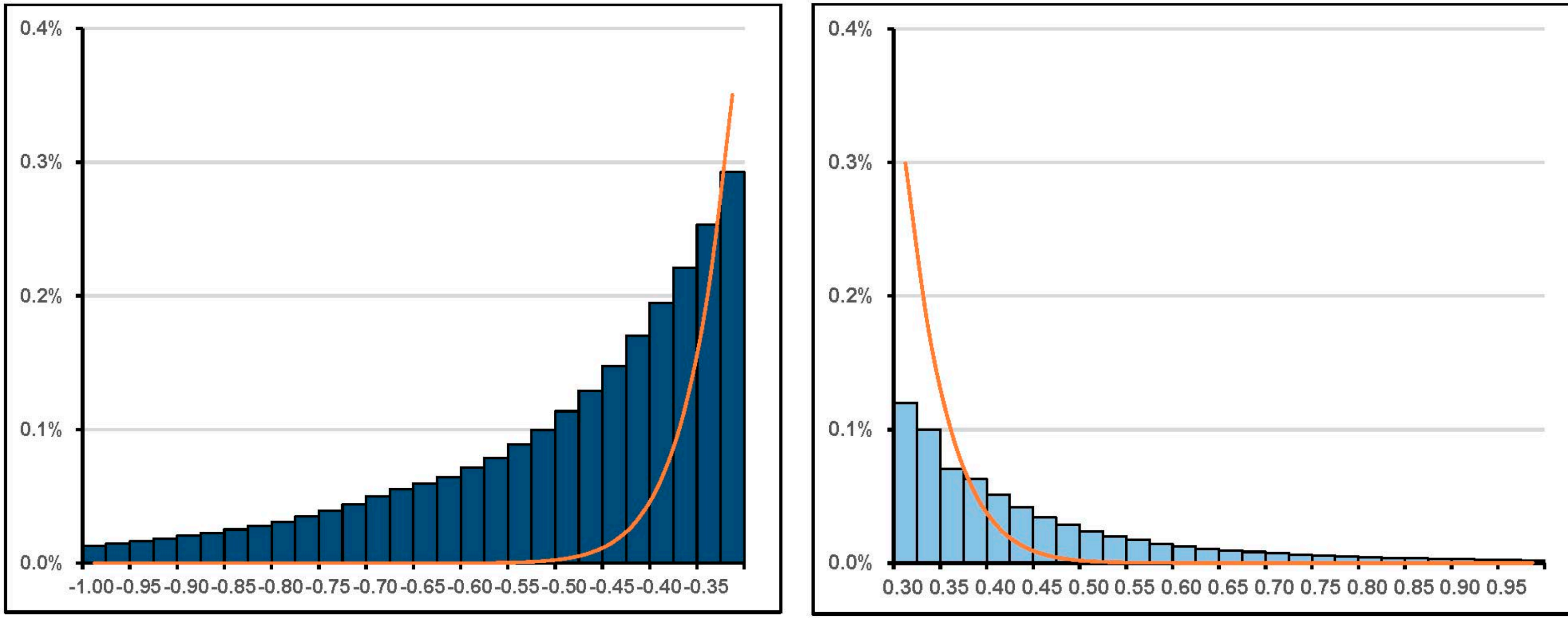
Variable	Mean	Skewness	P1	P5	P10	P25	Median	P75	P90	P95	P99
Log return	-0.004	-2.594	-0.498	-0.179	-0.095	-0.026	0.010	0.040	0.085	0.125	0.261
Duration (days)	27.373	9.792	2	2	2	2	4	12	42	96	497

Figure A1: Distribution of cycle returns



41.3% of the cycle (log) returns are between 0 and 5%
Only 22.6% are between 0 and –5%.

Figures A1 Left and right-hand tails of the distribution



Left-hand tail is fatter than the right-hand tail