Tracking Retail Investor Activity

Ekkehart Boehmer, Charles M. Jones, Xiaoyan Zhang and Xinran Zhang
Journal of Finance, 2020

Long Zhen

Contents

Introduction – Backgrounds

- Can retail equity investors predict future stock returns? Or do they make mistakes?

- Literature:
  - Mistake: Barber and Odean (2000,2008)
  - Predict: Barrot, Kaniel, and Sraer (2016)……

Introduction – Motivation

- Existing studies are based on proprietary datasets
- Many researchers use trade size as the proxy

- → We need an easily implementable method to isolate retail order flow

- A useful fact: retail orders, instead of institutional orders can receive price improvement.

Introduction – Research Problem

- How to identify retail orders?
  - Price improvement
- Whether retail orders have predictability for future returns?
  - Yes
- Why?
  - Persistence in retail order flow
  - Liquidity provision
  - Informed trading

Introduction – Contribution

- Propose a novel methodology for identifying and signing marketable trades using publicly available data.
- The marketable retail trades we identify can predict the cross-section of future stock returns.
- Half of its predictability is likely driven by order imbalance persistence and liquidity provision, while the other half is consistent with informed trading.

Research Design – Data

- TAQ(Trade and Quote Database) trade data that occur off-exchange
- Stock and accounting: NYSE Trade and Quote Database data, CRSP and Compustat
- news data from Thomson Reuters News Analytics (TRNA)
  - provides key information about each news item, such as the ticker, the time stamp of the news story, the news topics the story belongs to, and sentiment scores for each article.
- Time: 2010.1.3~2015.12.31
  - Why not full available data?
    - Strong upward trend of sub-penny trades before 2010
    - A new SEC program was adopted in 2016 which may affect the price improvement.

Research Design – Data

- Institutional Background
  - Retail orders are typically executed by wholesalers or via internalization, while institutional orders are sent to exchanges and dark pools.
  - Many orders that are internalized or executed by wholesalers are given a small price improvement (a small fraction of a cent), while exchange orders can’t have sub-penny prices, except midpoint trades.
  - Brokerage firms are required to provide regular summary statistics for non-directed orders (retail orders)
  - Limit orders are required to be priced at round pennies, so we only identify market retail orders.

Retail orders identification

- Let P_{it} be the transaction price in stock i at time t
- Let $Z_{it} \equiv 100 * \text{mod}(P_{it}, 0.01)$ be the fraction of a penny associated with that transaction price. Z_{it} can take any value in the unit interval [0,1)
  - $0 \leq Zit < 0.4$: retail sell transaction
  - $0.6 < Zit < 1$: retail buy transaction
  - $0.4 \leq Zit \leq 0.6$: not assigned
- → over 4.6 mil stock-day observations

	N	Mean	Std	Median	Q1	Q3
Round lots and odd lots						
Vol: 交易量	4,628,957	1,229,004	6,849,849	221,234	51,768	819,615
Trd: 交易笔数	4,628,957	5,917	13,909	1,505	312	5,502
Mrbvol: buy	4,628,957	42,481	280,474	5,165	1,200	20,681
Mrsvol: sell	4,628,957	42,430	264,704	5,635	1,369	21,828
Mrbtrd	4,628,957	110	410	22	5	79
Mrstrd	4,628,957	108	355	24	6	81
Mroibvol: order imbalance	4,628,957	-0.038	0.464	-0.027	-0.301	0.217
Mroibtrd	4,628,957	-0.032	0.437	-0.010	-0.276	0.205

$$mroibvol(i,t) = \frac{mrbvol(i,t) - mrsvol(i,t)}{mrbvol(i,t) + mrsvol(i,t)},$$

$$mroibtrd(i,t) = \frac{mrbtrd(i,t) - mrstrd(i,t)}{mrbtrd(i,t) + mrstrd(i,t)},$$

Research Design – Data

Cross validation with NASDAQ TRF data

- The NASDAQ sample covers all intraday transactions on its TRF for 117 stocks for the month of October 2010.
- “buy,” “sell,” or “cross.”

- →The false rates turn out to be around 2%. Quite accurate.

Research Design

- A: what explains marketable retail investor order imbalances?
- B: Whether past marketable retail order imbalance measures can predict future stock returns using Fama-MacBeth regressions and long-short portfolios
- C: Compare alternative hypotheses for the predictive power
- D: Explore the nature of the information contained in marketable retail flow

Research Design

A: what explains marketable retail investor order imbalances?

- Fama-Macbeth regression:

$$mroib(i,w) = b0(w) + b1(w)'ret(i,w-1) + b2(w)'controls(i,w-1) + u1(i,w)$$

Week

Daily overlapping coefficients

Research Design

- **B**. Predicting Future Stock Returns with Marketable Retail Order Imbalance Measures

$$ret(i,w) = c0(w) + c1(w)mroib(i,w-1) + c2(w)'controls(i,w-1) + u2(i,w)$$

  - Overall predictive power
  - Subgroups
  - Longer horizons
  - Long-short portfolio

Research Design

- **C**. Alternative Hypotheses for Marketable Retail Order Imbalance Predictive Power for Future Returns
- Three hypotheses:
  - order flows are persistent, this could lead directly to the predictability
  - retail traders' contrarian trading provides liquidity to the market
  - retail investors may have valuable information about the firm

Research Design

- Two-stage decomposition:
- 1:

$$mroib(i,w-1) = d0(w-1) + d1(w-1)mroib(i,w-2) + d2(w-1)'ret(i,w-2) + u4(i,w-1).$$

$$\widehat{mroib}_{i,w-1}^{persistence} = \widehat{d1}(w-1)mroib(i,w-2),$$

$$\widehat{mroib}_{i,w-1}^{contrarian} = \widehat{d2}(w-1)'ret(i,w-2),$$

$$\widehat{mroib}_{i,w-1}^{other} = \widehat{u4}(i,w-1) + \widehat{d0}(w-1).$$

- 2:

$$ret(i,w) = e0(w) + e1(w)\widehat{mroib}_{i,w-1}^{persistence} + e2(w)\widehat{mroib}_{i,w-1}^{contrarian} + e3(w)\widehat{mroib}_{i,w-1}^{other} + e4(w)'controls(i,w-1) + u5(i,w).$$

- **D**. Public News and Marketable Retail Order Imbalance
  - compute a net sentiment score as the difference between the positive and negative sentiment score for each stock each day.

$$mroib(i,w-1) = h0(w-1) + h1(w)mroib(i,w-2) + h2(w-1)'ret(i,w-2) + h3(w-1)sent(i,w-1) + u8(i,w-1).$$

$$\widehat{mroib}_{i,w-1}^{persistence} = \widehat{h1}(w-1)mroib(i,w-2),$$

$$\widehat{mroib}_{i,w-1}^{contrarian} = \widehat{h2}(w-1)'ret(i,w-2),$$

$$\widehat{mroib}_{i,w-1}^{publicnews} = \widehat{h3}(w-1)sent(i,w-1),$$

$$\widehat{mroib}_{i,w-1}^{other} = \widehat{u8}(i,w-1) + \widehat{h0}(w-1).$$

$$ret(i,w) = j0(w) + j1(w)\widehat{mroib}_{i,w-1}^{persistence} + j2(w)\widehat{mroib}_{i,w-1}^{contrarian} + j3(w)\widehat{mroib}_{i,w-1}^{publicnews} + j4(w)\widehat{mroib}_{i,w-1}^{other} + j5(w)'controls(i,w-1) + u9(i,w).$$

Further discussion

- **A. Aggregate Marketable Retail Order Imbalance**
- **B. Subsample Analysis**
- **C. Market Conditions**
- **D. Odd Lots**
- **E. Order Sizes**
- **F. Wholesaler/Internalizer's Perspective: Profitability of Marketable Retail Order Flow**
- **G. Earnings Announcements and Marketable Retail Order Flow**
- **H. Controlling for Overall Order Imbalances**
- **I. When the Effective Spread is Less Than 1 Cent**

bid-ask midpoint returns, which do not have bid-ask bounce

Table II. Determinants of Marketable Retail Order Imbalances

Reg	I		II		III		IV	
Dep.var	Mroibvol		Mroibvol		Mroibtrd		Mroibtrd	
Return	Bid-ask return		CRSP return		Bid-ask return		CRSP return	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Intercept	-0.4013	-20.03	-0.4065	-20.19	-0.4326	-22.00	-0.4357	-22.01
Mroib(w-1)	0.2200	92.53	0.2201	92.57	0.2865	150.01	0.2866	150.06
Ret(w-1)	-0.9481	-40.60	-0.9620	-41.43	-0.9003	-35.92	-0.9156	-36.74
Ret(m-1)	-0.2778	-19.24	-0.2784	-19.30	-0.2258	-14.84	-0.2262	-14.87
Ret(m-7,m-2)	-0.0586	-11.49	-0.0584	-11.46	-0.0380	-6.50	-0.0378	-6.48
Lmto	0.0003	5.31	0.0003	5.19	0.0002	3.93	0.0002	3.83
Lvol	0.8100	8.37	0.8478	8.79	0.4366	4.24	0.4633	4.51
Size	0.0154	12.06	0.0157	12.31	0.0209	16.37	0.0211	16.48
Lbm	-0.0275	-17.66	-0.0274	-17.61	-0.0274	-18.09	-0.0273	-18.05
Adj.R2	6.00%		6.01%		9.49%		9.50%	

- Persistent
- Contrarian at weekly horizons (not at daily horizon)
- Buy more aggressively in larger firms, growth firms, and firms with higher turnover and higher volatility

Table III. Predicting Next-week Returns Using Marketable Retail Order Imbalances

Reg	I		II		III		IV	
Order imbalance	Mroibvol		Mroibvol		Mroibtrd		Mroibtrd	
Dep.var	Bid-ask return		CRSP return		Bid-ask return		CRSP return	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Intercept	0.0050	2.58	0.0056	2.85	0.0050	2.58	0.0056	2.85
Mroib(w-1)	0.0009	15.60	0.0010	16.29	0.0008	12.30	0.0008	13.20
Ret (w-1)	-0.0185	-5.83	-0.0220	-6.85	-0.0186	-5.88	-0.0222	-6.91
Ret (m-1)	0.0006	0.35	0.0006	0.34	0.0005	0.29	0.0005	0.29
Ret (m-7, m-2)	0.0008	1.16	0.0008	1.16	0.0008	1.12	0.0008	1.12
Lmto	0.0000	-3.37	0.0000	-3.76	0.0000	-3.36	0.0000	-3.75
Lvol	-0.0223	-1.41	-0.0205	-1.31	-0.0217	-1.37	-0.0198	-1.27
Size	-0.0001	-0.86	-0.0001	-0.92	-0.0001	-0.90	-0.0001	-0.96
Lbm	-0.0001	-0.39	0.0000	-0.07	-0.0001	-0.42	0.0000	-0.10
Adj.R2	3.85%		3.85%		3.84%		3.84%	
Interquartile	1.1888		1.1888		1.2292		1.2292	
Interquartile weekly return diff	0.1089%		0.1144%		0.0931%		0.0997%	

Table IV. Marketable Retail Return Predictability within Subgroups

Panel A. Market cap groups

Mroib measure Mkt cap	Mroibvol Coef.	t-stat	Interquartile	Weekly return diff	Mroibtrd Coef.	t-stat	Interquartile	Weekly return diff
Small	0.0013	13.90	1.662	0.219%	0.0012	11.58	1.736	0.207%
Medium	0.0007	9.18	1.323	0.087%	0.0004	5.63	1.346	0.059%
Big	0.0003	3.68	0.892	0.026%	0.0002	2.52	0.929	0.019%

Interquartile range for Mroib

Panel B. Share price groups

Mroib measure Price groups	Mroibvol Coef.	t-stat	Interquartile	Weekly return diff	Mroibtrd Coef.	t-stat	Interquartile	Weekly return diff
Low	0.0014	13.34	1.432	0.205%	0.0012	10.34	1.586	0.185%
Medium	0.0007	10.00	1.289	0.089%	0.0005	7.56	1.309	0.070%
High	0.0002	3.23	0.961	0.020%	0.0002	2.19	0.961	0.015%

Panel C. Turnover groups

Mroib measure Turnover groups	Mroibvol Coef.	t-stat	Interquartile	Weekly return diff	Mroibtrd Coef.	t-stat	Interquartile	Weekly return diff
Low	0.0011	15.60	1.837	0.205%	0.0011	14.71	1.777	0.195%
Medium	0.0008	10.21	1.219	0.094%	0.0006	7.05	1.228	0.071%
High	0.0007	4.98	0.910	0.065%	0.0004	2.55	1.005	0.037%

- Significant within subgroups
- a clear cross-sectional pattern

Table V. Predicting Returns k-weeks Ahead

Panel A. Predict bid-ask average return k weeks ahead

	Mroibvol		Mroibtrd	
# of weeks ahead	Coef.	t-stat	Coef.	t-stat
1 week	0.00092	15.60	0.00076	12.30
2 weeks	0.00055	9.35	0.00048	7.89
4 weeks	0.00031	5.56	0.00026	4.66
6 weeks	0.00022	3.90	0.00015	2.60
8 weeks	0.00021	3.47	0.00011	1.75
10 weeks	0.00010	1.82	0.00002	0.35
12 weeks	0.00007	1.29	0.00009	1.52

Empirical Results

Table VI. Long-short Strategy Returns Based on Marketable Retail Order Imbalances

Panel A. Form portfolios on the previous week marketable retail order imbalance based on number of shares traded

Holding	Whole sample				Small		Medium		Big	
Period	Mean	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
1 week	0.092%	2.66	0.084%	2.43	0.403%	9.16	0.170%	6.24	0.067%	1.78
2 weeks	0.147%	2.45	0.135%	2.46	0.669%	9.01	0.292%	6.81	0.105%	1.70
4 weeks	0.223%	1.89	0.208%	2.00	1.124%	10.43	0.423%	6.36	0.143%	1.22
6 weeks	0.310%	1.72	0.277%	1.73	1.399%	13.02	0.558%	6.07	0.171%	1.05
8 weeks	0.448%	1.92	0.460%	2.26	1.709%	17.13	0.623%	4.18	0.342%	1.69
10 weeks	0.515%	1.99	0.484%	1.81	1.704%	11.17	0.578%	3.87	0.381%	1.53
12 weeks	0.588%	2.09	0.629%	1.89	1.857%	7.65	0.556%	3.20	0.477%	1.48

Panel A. First stage of projecting order imbalance on persistence and past return

Reg	I		II		III		IV	
Dep.var	Mroibvol(w-1)		Mroibvol(w-1)		Mroibtrd(w-1)		Mroibtrd(w-1)	
Return	Bid-ask return		CRSP return		Bid-ask return		CRSP return	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Intercept	-0.1413	-24.66	-0.1408	-24.61	-0.1054	-17.23	-0.1049	-17.19
Mroib(w-2)	0.2227	96.20	0.2228	96.20	0.2906	149.82	0.2907	149.85
Ret(w-2)	-0.9286	-38.93	-0.9422	-39.80	-0.8926	-34.92	-0.9076	-35.81
Ret(m-1)	-0.2029	-13.93	-0.2025	-13.90	-0.1591	-10.72	-0.1588	-10.70
Ret(m-7,m-2)	-0.0267	-4.98	-0.0268	-4.99	-0.0054	-0.86	-0.0055	-0.88
Adj.R2	5.62%		5.63%		8.99%		9.00%	

Table VII. Predictability Decomposition

Panel B. Second-stage decomposition of order imbalance's predictive power

Reg	I		II		III		IV	
Order Imbalance	Mroibvol(w)		Mroibvol(w)		Mroibtrd(w)		Mroibtrd(w)	
Dep.var	Bid-ask return		CRSP return		Bid-ask return		CRSP return	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Intercept	0.0046	2.25	0.0052	2.54	0.0046	2.23	0.0052	2.52
Mroib(w-1,persistence)	0.0027	8.75	0.0029	9.41	0.0018	7.80	0.0019	8.56
Mroib(w-1,contrarian)	-0.0044	-0.42	-0.1310	-1.46	-0.0073	-0.73	0.0328	1.62
Mroib(w-1,other)	0.0008	14.47	0.0009	15.48	0.0006	10.51	0.0007	11.64
Ret(w-1)	-0.0176	-5.41	-0.0206	-6.27	-0.0177	-5.45	-0.0207	-6.30
Ret(m-1)	-0.0060	-0.67	0.0002	0.03	0.0017	0.56	0.0093	1.13
Ret(m-7,m-2)	-0.0009	-0.65	-0.0127	-1.12	0.0017	0.95	-0.0008	-0.34
Lmto	0.0000	-3.49	0.0000	-3.80	0.0000	-3.48	0.0000	-3.78
Lvol	-0.0230	-1.48	-0.0231	-1.50	-0.0224	-1.44	-0.0225	-1.46
Size	-0.0001	-0.61	-0.0001	-0.67	-0.0001	-0.65	-0.0001	-0.72
Lbm	-0.0001	-0.46	0.0000	-0.14	-0.0001	-0.56	-0.0001	-0.23
Adj.R2	4.26%		4.27%		4.25%		4.26%	
	Int'quartile range	return diff	Int'quartile range	return diff	Int'quartile range	return diff	Int'quartile range	return diff
Mroib(w-1,persistence)	0.2591	0.0688%	0.2593	0.0739%	0.3498	0.0620%	0.3500	0.0679%
Mroib(w-1,contrarian)	0.0627	-0.0277%	0.0631	-0.8265%	0.0614	-0.0445%	0.0619	0.2031%
Mroib(w-1,other)	1.1141	0.0915%	1.1141	0.0977%	1.1326	0.0718%	1.1327	0.0792%

Table IX. Relation between Public News and Marketable Retail Order Flow

Panel A. Predicting returns using public news and marketable retail order flow

Reg	I		II		III		IV	
Dep.var	Bid-ask return		CRSP return		Bid-ask return		CRSP return	
Order Imbalance					Mroibvol		Mroibvol	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Intercept	0.0057	2.48	0.0066	2.83	0.0061	2.66	0.0070	3.02
Sent(w-1)	0.0008	3.31	0.0009	3.64	0.0008	3.33	0.0009	3.66
Mroib(w-1)					0.0009	10.61	0.0010	11.53
Ret(w-1)	-0.0088	-2.65	-0.0105	-3.10	-0.0090	-2.70	-0.0107	-3.16
Ret(m-1)	0.0008	0.38	0.0009	0.44	0.0013	0.64	0.0015	0.72
Ret(m-7,m-2)	0.0001	0.15	0.0001	0.11	0.0002	0.21	0.0001	0.18
Lmto	0.0000	-1.03	0.0000	-1.29	0.0000	-1.14	0.0000	-1.41
Lvol	-0.0435	-2.15	-0.0465	-2.34	-0.0444	-2.20	-0.0477	-2.40
Size	-0.0001	-0.90	-0.0001	-1.04	-0.0001	-0.99	-0.0002	-1.14
Lbm	0.0001	0.29	0.0001	0.61	0.0001	0.44	0.0002	0.79
Adj.R2	5.01%		5.01%		5.06%		5.08%	

Panel B. contemporaneous relation between sentiment and order imbalance

Topic	Type	Description	N	g1	t-stat
RESF	Equities	results forecast	102,515	0.0054	3.90
AAA	money/debt	debt rating news	23,405	0.0131	3.66
DIP	general news	diplomacy	6,057	0.0362	3.34
DIV	Equities	Dividend	24,282	0.0093	2.77
IGD	money/debt	investment grade debt	6,760	0.0261	2.76
DRV	cross market	derivatives	18,061	0.0238	2.58
DBT	money/debt	debt markets	73,600	0.0060	2.55
MTG	money/debt	mortgage-backed debt	7,764	0.0264	2.52
RES	equities	corporate results	176,699	0.0031	2.36
JUDIC	general news	Judicial	28,280	0.0096	2.31

- **Table X. Predictability Decomposition using Public News Releases**

Panel A. First stage of projecting order imbalance on persistence, past return, and public news.

Reg	I		II		III		IV	
Dep.var	Mroibvol(w-1)		Mroibvol(w-1)		Mroibtrd(w-1)		Mroibtrd(w-1)	
Return	Bid-ask return		CRSP return		Bid-ask return		CRSP return	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Intercept	-0.1510	-24.08	-0.1505	-24.03	-0.1132	-16.89	-0.1127	-16.85
Mroib(w-2)	0.2208	112.64	0.2209	112.67	0.2827	125.44	0.2829	125.48
Ret(w-2)	-0.8918	-36.22	-0.9051	-37.00	-0.8940	-34.04	-0.9098	-35.01
Ret(m-1)	-0.2169	-13.68	-0.2156	-13.60	-0.1702	-10.56	-0.1687	-10.47
Ret(m-7,m-2)	-0.0264	-4.54	-0.0264	-4.55	-0.0080	-1.17	-0.0081	-1.19
Sent(w-1)	0.0249	11.60	0.0249	11.60	0.0305	13.81	0.0305	13.81
Adj.R2	5.49%		5.50%		8.58%		8.59%	

Panel B. Second-stage decomposition of order imbalance's predictive power

Reg	I		II		III		IV	
Order Imbalance	Mroibvol(w)		Mroibvol(w)		Mroibtrd(w)		Mroibtrd(w)	
Dep.var	Bid-ask return		CRSP return		Bid-ask return		CRSP return	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Intercept	0.0055	2.40	0.0060	2.63	0.0055	2.39	0.0060	2.62
Mroib(w-1,persistence)	0.0027	8.50	0.0029	9.16	0.0018	7.28	0.0020	8.02
Mroib(w-1,contrarian)	0.0088	0.55	0.6947	1.03	-0.0182	-1.00	0.0419	1.18
Mroib(w-1,public news)	0.1150	1.17	-0.0134	-0.35	-0.0386	-0.84	0.0020	0.04
Mroib(w-1,other)	0.0008	13.98	0.0009	15.16	0.0006	10.15	0.0007	11.35
Ret(w-1)	-0.0217	-6.25	-0.0250	-7.12	-0.0218	-6.28	-0.0251	-7.16
Ret(m-1)	0.0059	0.54	0.6585	1.03	0.0106	1.05	0.0119	1.41
Ret(m-7,m-2)	0.0014	1.01	0.0511	1.02	0.0059	0.99	-0.0004	-0.22
Lmto	0.0000	-2.40	0.0000	-2.75	0.0000	-2.35	0.0000	-2.71
Lvol	-0.0273	-1.63	-0.0252	-1.52	-0.0266	-1.59	-0.0244	-1.47
Size	-0.0001	-0.68	-0.0001	-0.73	-0.0001	-0.73	-0.0001	-0.79
Lbm	0.0001	0.31	0.0001	0.66	0.0001	0.25	0.0001	0.59
Adj.R2	4.22%		4.23%		4.21%		4.22%	
	Interquartile	return diff	Interquartile	Return diff	Interquartile	return diff	Interquartile	return diff
Mroib(w-1,persistence +contrarian+public news)	0.2760	0.0707%	0.2763	0.0761%	0.3609	0.0614%	0.3611	0.0678%
Mroib(w-1,other)	1.1202	0.0932%	1.1203	0.1006%	1.1654	0.0745%	1.1654	0.0829%

Conclusion

- We use this methodology to characterize the trading behavior and information content of marketable retail orders.
- We find that the marketable retail order flow can predict the cross section of future stock returns