Journal of Financial Economics 145 (2022) 616–641

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

The democratization of investment research and the informativeness of retail investor trading

Michael Farrell[a], T. Clifton Green[b,*], Russell Jame[c], Stanimir Markov[d]

[a] Lubar School of Business, University of Wisconsin-Milwaukee, WI, USA
[b] Goizueta Business School, Emory University, GA, USA
[c] Gatton College of Business and Economics, University of Kentucky, KY, USA
[d] Naveen Jindal School of Management, University of Texas at Dallas, TX, USA

ARTICLE INFO

Article history:
Received 2 July 2021
Revised 23 July 2021
Accepted 25 July 2021
Available online 24 September 2021

Keywords:
Investment research
Seeking alpha
Retail investors
Informed trading

ABSTRACT

We study the effects of social media on the informativeness of retail trading. Our identification strategy exploits the editorial delay between report submission and publication on Seeking Alpha, a popular crowdsourced investment research platform. We find the ability of retail order imbalances to predict the cross-section of stock returns and cash-flow news increases sharply in the intraday post-publication window relative to the pre-publication window. The findings are robust to controlling for report tone and stronger for reports authored by more capable contributors. The evidence suggests that recent technology-enabled innovations in how individuals share information help retail investors become better informed.

1. Introduction

Investing has always been social, and a large literature highlights the influence of peers on investment decisions (e.g., Shiller and Pound 1989, Duflo and Saez 2002, 2003, Ivković and Weisbenner 2007, Ouimet and Tate 2020). In recent years, improvements in technology have greatly expanded the scope for sharing information. Retail investors have embraced finance social media sites as users as well as creators of content, discussing news events, sharing investment research, and debating investment strategies (Grennan and Michaely, 2020). While innovations in social media offer the potential for improved access to investment information, theory suggests that peer interactions may also exacerbate behavioral biases (Han et al., 2021).

Empirical evidence on the effects of social media on retail investors is limited. Heimer (2016), Cookson et al. (2020), and Chawla et al. (2017) suggest that social media intensifies behavioral biases and spreads stale news. On the other hand, several recent studies find evidence that certain types of social media can provide investment value (Chen et al., 2014; Jame et al., 2016; Bartov et al., 2018; Crawford et al., 2018), yet it is unclear the extent to which social media informs retail traders. In this article, we study a popular investor social media site, Seeking Alpha, to identify when individual investors produce and share investment research, and we examine whether these activities increase the informativeness of their trading.

The Seeking Alpha platform, which curates crowdsourced investment research from non-professional analysts, offers several features that make it a natural setting to examine this question. First of all, Seeking Alpha (SA) provides broader access to in-depth investment analysis than most other social media platforms.[1] Consistent with this view, SA research reports and the comments they

* Corresponding author.
E-mail address: clifton.green@emory.edu (T.C. Green).

[1] For example, StockTwits limits the character length of posts, Estimize focuses primarily on short-term earnings forecasts, and the Motley Fool CAP system's stock picks lack detailed analysis. SumZero focuses on pro-

https://doi.org/10.1016/j.jfineco.2021.07.018
0304-405X/

engender have been shown to predict future stock returns and earnings surprises (Chen et al., 2014).[2]

Seeking Alpha research reports provide investment analysis rather than break news, and their publication process includes an editorial review to ensure quality.[3] This review-induced publication delay permits us to separate the impact of SA research from earlier news events that may also influence trading. Specifically, we use the intraday window immediately after SA report publication to measure the level of social-network-induced trading, and we use the intraday window prior to publication (but after potential information events that may have influenced the report) to capture the counterfactual level of trading that would have occurred in the absence of the SA report. The review-induced delay injects an element of randomness into the intraday timing of publication, and consistent with our identifying assumption, we find no evidence that media articles, brokerage research, or earnings announcements systematically precede or follow SA research publications over intraday windows.

We begin our analysis by documenting that Seeking Alpha's investor-authored research caters to retail investor information demand. We analyze roughly 180,000 research reports discussing 4900 stocks and find that after controlling for other firm characteristics, SA coverage is higher among firms with low institutional ownership and greater breadth of ownership, whereas the opposite is true for brokerage research coverage. The research coverage evidence confirms Seeking Alpha's emphasis on providing an investment analysis platform for retail investors.

Our analysis points toward a causal relation between Seeking Alpha research and retail investor trading. We analyze retail trading using ten half-hour intraday event windows around Seeking Alpha report publication using trade and quote data from NYSE TAQ and the method of (Boehmer et al., 2020) (BJZZ) to identify retail investor trades. Our regression approach includes individual report fixed effects, which benchmark the post-publication intraday period to the pre-publication intraday period. The results indicate that retail trading is markedly higher after the publication of Seeking Alpha research. For example, aggregate retail trading in the first half-hour after Seeking Alpha report publication is 7.68% higher than in the half-hour before publication. Moreover, measures of report sentiment that predict future returns, such as report tone and contributors' investment positions (Campbell et al., 2019; Chen et al., 2014), explain retail investor trade order imbalances in the post-publication period. In contrast, we find no evidence of an increase in retail trading or report-sentiment-driven order flows prior to report publication, which is inconsistent with retail investors reacting to unobserved information events. The evidence suggests that Seeking Alpha has a distinct influence on the intensity and direction of retail trading.

To assess the effect of Seeking Alpha research on the informativeness of retail investor trading, we compare the ability of aggregate retail order imbalances to predict the cross-section of five-day ahead stock returns in the period immediately before and after SA report publication.[4] The estimates indicate SA research publication leads to more informed retail trading. For example, the increase in future returns predicted by a one standard deviation increase in post-publication retail order imbalances is 0.26 percentage points larger than that predicted by pre-publication retail order imbalances. We do not find any increase in the informativeness of retail order imbalances over the five pre-publication half-hour windows, which suggests that the documented post-publication increase is not the continuation of pre-event trend.

Several additional tests suggest that the relation between retail order imbalances and future returns following SA research is at least partially attributable to informed trading, rather than price pressure or liquidity provision. First, we do not find evidence that the documented post-publication return predictability reverses over the subsequent quarter, alleviating concerns about price pressure. Second, similar to Boehmer et al., 2020, we decompose retail order imbalances into a persistent component which captures price pressure, a contrarian component which captures liquidity provision, and a residual component which captures informed trading. We find that the residual (informed) component remains a highly significant predictor of returns. Finally, retail order imbalances' ability to predict analyst earnings forecast revisions and traditional media sentiment over the subsequent five days strengthens in the five half-hours after SA research is published, further supporting the view that retail trades reveal fundamental information.

We observe thathe incremental information revealed by retail trades after Seeking Alpha research is largely orthogonal to the information revealed by SA research report tone and contributor investment position, consistent with retail investors actively gleaning valuable information rather than passively following opinions expressed by social media contributors. We hypothesize that higher quality research reports will lead to more informed trading, and we explore whether reports that are authored by more accomplished or capable contributors offer more opportunities for extracting valuable information. Consistent with our conjecture, we find that retail order imbalances predict future stock returns and cash flows news more convincingly after reports that receive more comments and

fessional investors employed by mutual funds, hedge funds, and private equity funds.

[2] In contemporaneous work, Gomez et al. (2020) show that Seeking Alpha coverage leads to lower bid-ask spreads around earnings announcements.

[3] One first-time SA contributor describes multiple rounds of revisions before acceptance, including requests to provide more sources, better flesh out investment theses, and offer additional financial statement analysis. The contributor concludes "There is so much that goes into an (SA) article that gives it substance and obviously a bit more difficult than I originally imagined." https://walletsquirrel.com/first-article-on-seeking-alpha/

[4] Hereafter, we refer to the relation between aggregate retail order imbalances and the cross-section of future stock returns as either trade informativeness or return predictability. Whether this translates to better individual retail investor trading performance is an empirical question which can be addressed only with investor-level data (e.g., Barrot et al. 2016).

those authored by contributors with strong academic backgrounds or a track record of impactful reports.[5]

Kogan et al. (2020), Mitts (2020), and Dyer and Kim (2021) find that a small percentage of Seeking Alpha research reports, identified ex ante as misleading or "fake", distort market prices. Drawing on these studies, we identify fake reports as those that are posted anonymously or have a low textual authenticity score and investigate whether they affect retail trading differently. We find that fake reports influence retail trading intensity and direction similarly or more than non-fake reports. In addition, retail order imbalances after the publication of fake reports predict the cross-section of one-week returns but not five-week returns, whereas retail order imbalances after non-fake reports predict the cross-section of five-week returns more strongly than one-week returns. These results are consistent with a small subset of SA research inducing uninformed retail trading that pushes prices from fundamentals over short horizons.

Our study contributes to the debate about the role of social media in capital markets. Since its arrival in the late 90s, regulators have repeatedly expressed concerns about social media impeding market efficiency and harming retail investors.[6] While a host of recent studies provide evidence that different types of social media contain investment value (Chen et al., 2014; Jame et al., 2016; Bartov et al., 2018), there is little evidence to suggest that it leads to more informative retail trading. To the contrary, existing evidence emphasizes that social media can exacerbate behavioral biases harmful to performance (Heimer, 2016; Cookson et al., 2020; Ammann and Schaub, 2020)[7]. Our results establish the role of crowdsourced investment research in informing retail investor decision-making, while at the same time validating concerns about misleading research content (Kogan et al., 2020; Mitts, 2020).

Our analysis also advances the literature that studies the informativeness of retail trading. Early studies conclude that individual investors are unsophisticated "noise" traders who tend to suffer from behavioral biases and may push prices away from fundamentals (e.g., Barber and Odean 2000, Kumar and Lee 2006, Frazzini and Lamont 2008, Hvidkjaer 2008, Barber et al., 2009). In contrast, more recent work finds evidence of informed trading by individuals and speculates that retail investors gain insights from geographic proximity to firms, relations with employees, or insights into consumer preferences (e.g., Kaniel et al. 2012, Kelley and Tetlock 2013, 2017, Boehmer et al., 2020). Our findings highlight a specific mechanism, technology-enabled improvements in how retail investors produce and share investment research, as a likely channel by which individual investors become better informed.

Another stream of literature examines the use of technology by regulators to level the informational playing field between institutional investors and retail investors.[8] Seeking Alpha is a technology-enabled market innovation whose ostensible purpose is to democratize the flow of investment analysis. Our findings illustrate how technological change enables new business models that can improve retail investors' access to investment research and level the informational playing field among investors.

2. Data and descriptive statistics

We discuss the Seeking Alpha sample in Section 2.1 and key variables in Section 2.2. We explore the determinants of Seeking Alpha research coverage in Section 2.3.

2.1. The Seeking Alpha sample

Seeking Alpha is one of the largest investment-related social media websites in the United States and epitomizes the democratization of investment research.[9] The website hosts curated investment research from a network of thousands of individual contributors. SA has roughly 40 million monthly visits and 15 million unique visitors.[10] Contributor testimonials indicate that some of the primary motivations for contributing research include direct compensation from SA, feedback on investment theses (via reader comments), and increased recognition and visibility which may lead to other professional opportunities.[11] Seeking Alpha research reports aim to provide analysis rather than break news, and each report is subject to an editorial review process that may involve multiple revisions. Chen et al. (2014) find that Seeking Alpha's crowdsourced investment research contains valuable investment information, with reports and user commentaries predicting future stock returns and earnings surprises.

We obtain all research reports published between 2006 and 2017 on the Seeking Alpha website.[12] For each report, we collect the following information: a report ID assigned by Seeking Alpha, report title, main text, date of

[5] Our primary analysis emphasizes intraday windows around the publication of SA reports for clean identification. We also conduct a daily analysis that includes reports released outside of market hours and find evidence consistent with SA reports having a larger effect on retail trade informativeness than brokerage research or traditional media articles.

[6] See, for example, the SEC's 2015 Investor Alert: *Social Media and Investing – Stock Rumors*, from the Office of Investor Education and Advocacy: https://www.sec.gov/oiea/investor-alerts-bulletins/ia_rumors.html

[7] A related set of studies conduct a micro-level analysis of information flows across peers and within a retail trader network to better understand trading behavior (Rantala, 2019; Ozsoylev et al., 2014; Kaustia and Knüpfer, 2012; and Ahern, 2017). Our focus is complementary in that we study the flow of information from a social finance media site, Seeking Alpha, to retail investors in aggregate.

[8] Examples include the launch of the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) in 1993 Asthana et al., 2004; and Gao and Huang, 2020), and the mandated use of eXtensible Business Reporting Language (XBRL) in corporate filings in 2009 (Blankespoor, Miller, and White, 2014; and Bhattacharya, Cho, and Kim, 2018)

[9] Seeking Alpha editor Douglas House commented in 2016 that "Seeking Alpha's raison d'être, of course, is to level the playing field for individual investors by leveraging the 'wisdom of crowds' via crowdsourcing." https://www.prnewswire.com/news-releases/slingshot-insights-partners-with-seeking-alpha-to-bring-transparency-to-expert-research-300308371.html

[10] https://seekingalpha.com/page/who_reads_sa

[11] See: https://seekingalpha.com/page/testimonials.

[12] Seeking Alpha was founded in 2004, but its stock coverage prior to 2006 is limited. For example, in 2005 Seeking Alpha covered less than 5% of all common stocks in CRSP.

Table 1
Summary Statistics for the Seeking Alpha (SA) Investment Research Report Sample
The table reports information on Seeking Alpha research reports by year. The sample is comprised of 183,969 single-ticker research reports (*SA Reports*) on 4910 unique firms, contributed by 8988 unique individuals (*SA Contributors*). The sample is limited to common stocks with available data in the CRSP-Compustat merged database and TAQ. *Firms Covered by SA* is the number of firms in the sample with at least one single-ticker report on Seeking Alpha. SA *Contributors* is the number of unique research report authors, *Reports per Firm* is the number of reports published for each firm with coverage, and *SA Coverage* is the average number of contributors for each firm with coverage. *Intraday SA Reports* is the number of reports that were published between 10:30 am and 3:30pm, and *Intraday Reports No Event* is the subset of intraday reports that are not confounded by other events: media articles, IBES research, or earnings announcements during the ten half-hour intervals surrounding the SA report publication.

Year	*Firms Covered by SA*	*SA Reports (Full Sample)*	*SA Contributors*	*Reports per Firm*	*SA Coverage*	*SA Reports (Intraday)*	*Intraday Reports (No Events)*
2006	724	2590	228	3.58	2.49	419	304
2007	1529	8560	610	5.60	3.31	1026	664
2008	1381	7650	851	5.54	3.31	1746	1116
2009	1296	8406	776	6.49	3.54	2424	1593
2010	1399	7721	782	5.52	3.25	2630	1635
2011	1539	11,389	1150	7.40	4.49	4499	3072
2012	1823	20,504	1616	11.25	6.56	7017	4620
2013	2503	20,659	2073	8.25	5.36	7439	5605
2014	2544	25,631	2216	10.08	5.97	8097	6349
2015	2756	27,101	2242	9.83	5.76	9276	7396
2016	2512	22,356	2183	8.90	5.03	8175	6154
2017	2305	21,402	2091	9.28	5.37	8534	6530
Average	1962	16,489	1508	8.01	4.57	5533	4067
Total	4910	183,969	8988	.	.	61,282	45,038

publication, author name, and the ticker (or tickers) assigned to each report. Following Chen et al. (2014), we limit the sample to reports that are associated with one ticker. We further limit the sample to common stocks (CRSP share codes 10 and 11) with available data in the CRSP-Compustat merged database and TAQ. Our final sample includes 183,969 single-ticker SA research reports covering 4910 firms.

Table 1 describes the increase in the breadth and depth of Seeking Alpha coverage over time. In 2006, there were 724 companies covered on SA, with 228 research contributors, and 2590 research reports. In 2017, coverage rose to 2305 companies, with 2091 contributors, and 21,402 reports.[13] In an average year in the sample, 1508 unique contributors publish 16,489 reports on 1962 different companies. Conditional on having Seeking Alpha coverage, the average firm has roughly 8.0 reports per year, written by 4.6 different contributors. Our analysis emphasizes the roughly one-third of research reports that are published during trading hours. On average, 5533 reports are published each year between 10:30 am and 3:30 pm, and of these, 4067 have no confounding information events (media articles, sell-side research, or earnings announcements released within the ten half-hour intervals around SA report publication).

2.2. Measuring retail investor trading and other key variables

Our approach for identifying retail trading relies on the methodology of Boehmer et al., 2020 (BJZZ).[14] Their approach exploits two key institutional features of retail trading. First, most equity trades by retail investors take place off-exchange, either filled from the broker's own inventory or sold by the broker to wholesalers (Battalio et al., 2016). TAQ classifies these types of trades with exchange code "*D*." Accordingly, we identify retail trades by limiting our analysis to trades executed on exchange code "*D*." Second, retail traders typical receive a small fraction of a cent price improvement over the National Best Bid or Offer (NBBO) for market orders (ranging from 0.01 to 0.4 cents), while institutional orders tend to be executed at whole or half-cent increments. Thus, we follow BJZZ and identify trades as retail purchases (sales) if the trade took place at a price just below (above) a round penny. The BJZZ approach is conservative in the sense that it has a low type 1 error (i.e., trades classified as retail are very likely to be retail). While this approach does omit some retail trading, including nonmarketable limit orders and retail traders that take place on registered exchanges, it "probably picks up a majority of overall retail trading activity" (BJZZ page 8).[15]

For each firm, we collect data on share price, shares outstanding, stock returns, and volume from CRSP. We obtain book value of equity, book value of debt, book value of assets, earnings before interest taxes depreciation and amortization (EBITDA), and total common shareholders from Compustat. We collect the number of shares held by institutions from the Thomson Reuters Institutional Holdings (S34) database. We obtain earnings announcement dates and sell-side analyst earnings forecast from the IBES unadjusted US detail history file and sell-side analyst recommendations from the IBES detail recommendation file.

[13] The fraction of firms with Seeking alpha coverage in the CRSP-Compustat-TAQ merged sample is 15.2% in 2006, rises to 72.2% in 2015, and is 64.6% in 2017.

[14] To be conservative, BJZZ focus on the sample period from 2010–2015 due to the gradual upward trend in sub-penny trading prior to 2010 and the potentially complicating effects of the tick size pilot program after 2015. Our findings are similar if we limit the sample to the 2010-2015 period. For example, in Fig. IA2 we present results by month over the sample period.

[15] BJZZ also note that, during a conference discussion of their work, Eric Kelley presented that the correlation between the BJZZ order imbalance measure and the imbalances calculated from Kelley and Tetlock (2013)'s proprietary retail data with observed trade directions is in the range of 0.345–0.507, with an average of 0.452.

We obtain data on traditional media coverage, measured using Dow Jones News Service articles from RavenPack, for the period from 2006 to 2017. Following Reed et al. (2020), we limit the RavenPack sample to articles with relevance and novelty scores of 100. For each article, we also collect the *Event Sentiment Score* (*ESS),* which ranges from 0 (very negative news) to 100 (very positive news) with a median value of 50 (neutral article).

Table IA1 in the Internet Appendix provides summary statistics on the characteristics of stocks covered in Seeking Alpha reports. We consider the following attributes: market capitalization (*Size*), book to market (*BM*), daily return volatility (*Volatility*), daily share turnover *(Turnover)*, past one-year return ($Return_{m\text{-}12,m\text{-}1}$), past one-year profitability (*Profitability*), the number of sell-side analysts covering the firm in the prior year (*IBES Coverage*), the number of unique media articles mentioning the firm the prior year (*Media Coverage*), the percentage of the firm's shares held by institutional investors in the prior year (*Institutional Ownership*), and the number of common shareholders in the prior year (*Breadth of Ownership*). See Appendix A for more detailed definitions. As a benchmark, we also report the average values for the value-weighted and equal-weighted market portfolios (*VW Market* and *EW Market*, respectively). We find that the average size of a firm covered by an SA report is roughly $61 billion, which is smaller than the corresponding size of the value-weighted average ($89 billion), but considerably larger than the equal-weighted market average ($4.6 billion). Relative to the *VW Market,* we also find that SA coverage tilts towards more volatile firms, more liquid firms, firms with stronger past returns, and firms with lower institutional ownership. However, the *VW Market* attribute almost always falls within the interquartile range of the SA attribute, suggesting that SA coverage is not dramatically different from the market portfolio. In the next section, we more carefully analyze the determinants of Seeking Alpha research coverage.

2.3. Determinants of Seeking Alpha research coverage

Seeking Alpha's business model is built on reaching a wide audience of do-it-yourself investors, and Seeking Alpha contributors are often individual investors. In contrast, prior survey evidence and empirical work suggests that brokerage analysts cater to institutional investors. For example, Brown et al. (2015) report that more than 80% of surveyed analysts view hedge funds and mutual fund clients as very important, while only 13% view retail clients as important. Consistent with this survey evidence, several papers find that sell-side research is strongly increasing in total institutional ownership (see, e.g., Bhushan 1989, Green et al. 2014).

We examine the determinants of Seeking Alpha coverage and sell-side coverage by estimating the following panel regression:

$$\begin{aligned} Coverage_{i,t} = \alpha &+ \beta_1 Inst.Ownership_{i,t-1} \\ &+ \beta_2 Breadth\ of\ Ownership_{i,t-1} \\ &+ \beta_3 Chars_{i,t-1} + Year_t + \varepsilon_{i,t.} \end{aligned} \tag{1}$$

where *Coverage* is the natural log of 1 plus the total number of unique Seeking Alpha contributors writing at least one report for stock *i* during the calendar year *t* (*SA Coverage*), or the natural log of 1 plus the total number of unique brokerage firms issuing at least one earnings forecast for the stock during the calendar year (*IBES Coverage*).

The two independent variables of primary interest are *Institutional Ownership,* defined as the percentage of the firm's shares held by institutional investors in year *t*-1, and *Breadth of Ownership*, defined as the number of common shareholders (both in logs). The vector of firm characteristics (*Chars*) includes: market capitalization (*Size*), book to market (*BM*), return volatility (*Volatility*), share turnover *(Turnover)*, past one-year return ($Return_{m\text{-}12,m\text{-}1}$), past one-year profitability (*Profitability*), and the number of unique media articles mentioning the firm during the prior year (*Media Coverage*). See Appendix A for detailed definitions. We log all continuous variables other than *Profitability* and *Return*, and we standardize all variables to have zero mean and unit variance. We include year fixed effects and cluster standard errors by firm.

Specification (1) of Table 2 examines the determinants of *SA Coverage* without controlling for *IBES Coverage.* In general, *SA Coverage* is higher for larger firms, firms with more frequent media coverage, and those with greater trading volume. In addition, *SA Coverage* is positively related to volatility, past one-year returns, and profitability. Consistent with our conjecture that Seeking Alpha research is a retail investor rather than an institutional investor phenomenon, we find a strong negative relation between *SA Coverage* and institutional ownership, and a strong positive relation between *SA Coverage* and total common shareholders. In particular, a one standard deviation increase in *Institutional Ownership* (*Breadth of Ownership*) is associated with a 25% decline (6% increase) in *SA Coverage*, and the findings are robust to controlling for *IBES Coverage*.

Specifications (3) and (4) present analogous results for brokerage analyst coverage (*IBES Coverage*). As expected and in sharp contrast to the *SA Coverage* patterns*, IBES Coverage* is strongly positively related to institutional ownership and strongly negatively related to breadth of ownership. Collectively, these results suggest that traditional sell-side research emphasizes institutional investors, whereas the Seeking Alpha platform caters to retail investors and provides a unique window into retail investors' information acquisition activities.

3. Identification strategy

The biggest obstacle to evaluating the impact of Seeking Alpha research on retail investor trading is estimating the counterfactual level of trading that would have occurred in the absence of a Seeking Alpha research report. SA research may be driven by an underlying information event, making it difficult to separate the effect of the event itself (news) from the subsequent analysis of the event (SA research). Our identification strategy exploits the time delay between potential unobserved information events and the publication of Seeking Alpha research reports.

The Seeking Alpha platform is designed to provide investment analysis rather than break news, and it naturally takes time for an SA contributor to read, process, and write reports based on any underlying news events. Moreover,

Table 2
Determinants of Seeking Alpha and IBES Coverage
The table presents the results from the estimation of Eq. (1):

$$
\begin{aligned}
Coverage_{i,t} &= \alpha + \beta_1 InstitutionalOwnership_{i,t-1} \\
&\quad + \beta_2 Breadth\ of\ Ownership_{i,t-1} \\
&\quad + \beta_3 Char_{i,t-1} + Year_t + \varepsilon_{i,t.}
\end{aligned}
$$

In Specifications (1) and (2), *Coverage* is the natural log of 1 plus the total number of unique Seeking Alpha contributors writing at least one report for the stock during the calendar year (*SA Coverage*). In Specifications (3) and (4), *Coverage* is the natural log of 1 plus the total number of unique brokerage firms issuing at least one earnings forecast for the stock during the calendar year (*IBES Coverage*). *Institutional Ownership* is the percentage of the firm's shares held by institutional investors at the end of the previous year, and *Breadth of Ownership* is the number of common shareholders at the end of the previous year. *Char* is a vector of firm characteristic controls. Detailed variable descriptions appear in Appendix A. All variables are standardized to have mean zero and unit variance. All specifications include year fixed effects. Standard errors are clustered by firm, with t-statistics reported in parentheses.

	SA Coverage		*IBES Coverage*	
	(1)	(2)	(3)	(3)
Institutional Ownership	-0.25	-0.27	0.13	0.14
	(-16.23)	(-17.26)	(13.03)	(13.96)
Breadth of Ownership	0.06	0.07	-0.07	-0.07
	(5.13)	(5.88)	(-9.52)	(-9.98)
Size	0.66	0.57	0.69	0.66
	(23.20)	(19.02)	(44.83)	(41.28)
Book-to-Market	0.02	0.01	0.00	0.00
	(1.40)	(0.85)	(0.11)	(0.02)
Volatility	0.21	0.21	0.05	0.04
	(14.18)	(13.96)	(5.34)	(4.28)
Turnover	0.13	0.09	0.28	0.27
	(10.15)	(6.88)	(23.50)	(23.03)
Return	0.02	0.02	0.04	0.03
	(3.57)	(3.47)	(4.04)	(3.98)
Profitability	0.03	0.03	-0.05	-0.05
	(2.72)	(3.59)	(-8.08)	(-8.43)
Media Coverage	0.06	0.05	0.04	0.03
	(4.49)	(4.03)	(3.96)	(3.68)
IBES Coverage		0.14		
		(8.76)		
SA Coverage				0.06
				(8.18)
Observations	35,261	35,261	35,261	35,261
Year fixed effects	Yes	Yes	Yes	Yes
R-squared	42.02%	42.49%	76.50%	76.67%

the editorial review process also introduces a lag between the creation of a report and its publication on Seeking Alpha. Discussions with SA representatives indicate that the report review process typically requires 24 hours from report submission to report publication, with turnaround times of less than three hours being extremely rare.

The publication delay implies that the period immediately prior to the publication release, during the report review process and after potentially unobserved information events, provides an opportunity to measure the counterfactual level of retail trading that would have occurred in the absence of the Seeking Alpha research report. Put another way, if SA research contributors and retail investors are reacting to earlier information events, we would expect the relation between SA research and retail trading estimated shortly before publication to be as strong as when it is estimated shortly after publication. On the other hand, if retail investors react to SA contributors' analysis of earlier events, we should observe a sharp change in retail investor behavior in the post-publication period relative to the pre-publication period.[16]

We analyze ten half-hour intervals around the publication of Seeking Alpha research, separated into 2.5 h pre- and post-event periods. The pre-event window consists of the five half-hour intervals [-5, -1], with the post-event covering half-hours [1, 5]. We also consider shorter [-1] and [1] pre- and post-event periods, which offers stronger identification by focusing immediately around the event but reduces statistical power and misses any effects that occur beyond the first half-hour after publication. Many Seeking Alpha users subscribe to alerts on the stocks they follow and therefore receive real-time notifications via text or email when reports are published. This feature of the platform makes it plausible to expect swift reactions to new research reports.

A simple comparison of post-event trading to pre-event trading is potentially confounded by intraday seasonality, which we address by including time-of-day fixed effects. We measure the intraday windows in calendar time to facilitate the inclusion of half-hour fixed effects,[17] and the fixed effects are allowed to vary each month in the sample to control for intraday seasonality that may vary over time. Event window [0] trades are excluded from our main tests but examined it in Figs. 1–3, where we plot each half-hour of the [-5, 5] window.

Trading outside regular market hours tends to be sparse, and trading during the first half-hour is often disproportionally driven by events occurring before the market opens. To ensure that we can reliably measure and compare pre- and post-event retail trading, we therefore analyze retail trades occurring between 10 am and 4 pm, and we require that reports be published between 10:30 am and 3:30 pm.[18] We note that for reports published before 12:30 pm (after 1:30 pm), the full pre- (post-) publication period will be less than 2.5 hours. This sample, which we refer to as *All Intraday* reports, contains 61,282 reports.

An important identifying assumption is that other confounding events that influence retail trading are just as likely to occur during the pre-publication window as in the post-publication window. This assumption could in principle be violated if Seeking Alpha's editorial team systematically seeks to release reports immediately before or after the arrival of important information events. While this seems unlikely, we empirically address this possibility by examining the distribution of earnings announcements, analyst reports, and media articles in the Seeking Alpha pre- and post-publication windows using a linear probability

[16] Seeking Alpha's research compensation arrangements reward exclusivity, yet it is possible that some contributors publish reports on their own personal blogs first, which could reduce the power of our tests. We explore this possibility in greater detail in Section IA.4 of the Internet Appendix, and we find no evidence to suggest that this practice is prevalent or consequential.

[17] For example, a report published at 2:15 pm has periods [-1], [0], and [1] which cover [1:30 pm–2:00 pm), [2:00 pm–2:30 pm), and [2:30 pm–3:00 pm).

[18] In robustness tests, we repeat the analysis after including reports published outside of market hours. The results are qualitatively similar (Section IA.6 and Table IA6 of the Internet Appendix).

model (more details available in Section IA.2 of the Internet Appendix and tabulated in Table IA2). We find no evidence of a significant relation between the intraday timing of SA research reports and the timing of earnings announcements, sell-side research reports, and media articles, which helps build confidence that any changes in retail trading immediately after SA research can be attributed to Seeking Alpha rather than the arrival of other information.

Even if the timing of Seeking Alpha reports is random, in some cases the [-5, 5] intraday publication window may nevertheless coincide with other major information events, which can add considerable noise to our analysis. Thus, in some specifications we also exclude SA research reports that have a confounding information event, defined as a media article, sell-side research report, or earnings announcement over the [-5, 5] window. The resulting *No Event* sample includes 45,038 SA research reports.[19] Fig. IA1 in the Internet Appendix confirms that the distribution of *All Intraday* reports and the subset of *No Event* reports is relatively uniform between 10:30 am and 3:30 pm. For example, in the full sample, the median number of reports in a 30 minute window is 5986, with a maximum of 7016 (11:30–11:59) and a minimum of 5451 (12:30–12:59).

4. The impact of Seeking Alpha research on the intensity and direction of retail trading

In this section, we analyze the effects of Seeking Alpha research on retail investor trading. Section 4.1 examines whether retail investors trade more actively after the publication of SA research, Section 4.2 explores whether the direction of retail trading is consistent with SA research sentiment, and Section 4.3 explores the potential effects of stale reports.

4.1. Seeking Alpha research and retail trading

We explore the effects of SA research on retail investor trading intensity using the following regression:

$$Retail\ Trd_{i,t} = \alpha + \beta_1 Post\_SA_{i,t} + Controls_{i,t} + Report_i + HalfHour_t \times Month + \varepsilon_{i,t.} \quad (2)$$

Retail_Trd is either *Retail Volume*, defined as log (1 + Retail Volume) in half-hour window *t* around the publication of report for firm *i*, or *Percent Retail Trading*, defined as total retail trading volume in half-hour window *t* for firm *i* scaled by total trading volume for firm *i* in the same window. Our primary variable of interest is *Post_SA*, which is an indicator equal to one if the trading is measured after the release of the report and zero if trading is measured prior to the release of the report. The sample is limited to the [-5, 5] event window around the report release, excluding event period [0]. Thus, *Post_SA* equals one over the [1, 5] window and zero over the [-1, -5] window. *Controls* includes the return and the absolute return over the previous half-hour ($Ret_{i,t-1}$, $AbsRet_{\ i,t-1}$) and the previous two to five half-hours ($Ret_{i,[t-5,t-2]}$, $AbsRet_{i,[t-5,t-2]}$). We include absolute returns as a proxy for any incrementally important new information that may capture retail investor attention, and we include signed returns to control for the possibility that retail investors may react differentially to good versus bad news. We also include indicators of extreme volume over the previous half-hour period ($High\ Vol_{i,t-1}$, $Low\ Vol_{i,t-1}$) and the previous two to five half-hour periods ($High\ Vol_{i,[t-5,t-2]}$, $Low\ Vol_{\ i,[t-5,t-2]}$) since extreme trading activity may signal shifts in demand for the stock that could influence trading and returns (Gervais et al., 2001; Mingelgrin, 2000). Report denotes fixed effects for each SA report. We also include half-hour fixed effects, and the loadings on the fixed effects are allowed to vary each month in the sample (i.e., *Half Hour* × *Month* fixed effects), which controls for intraday seasonality that may vary over time.[20] Standard errors are clustered by date.[21]

Specification (1) of Table 3 reports results for the full sample when the event window is [-5,5] and the dependent variable is the natural logarithm of retail volume. The coefficient on the post-publication dummy is statistically significant at 6.03%, implying a 6.22% ($e^{6.03\%}$-1) increase in retail volume in the post-publication five half hours relative to the pre-publication five half-hours. Specification (2) limits the sample to the 45,038 reports that do not coincide with earnings announcements, media coverage, or sell-side reports over the [-5, 5] window. We find that the coefficient on *Post_SA* increases to 8.77%. In Specification (3), we further limit the sample to observations in the [-1, 1] window and continue to find a large increase in retail volume.

The results from Section 2.3 suggest that the Seeking Alpha platform caters more towards retail investors than institutional investors. We examine retail investors' relative intensity of trading after SA research by setting *Retail_Trd* to *Percent Retail*. The *Post_SA* research coefficients are significant in each specification. For example, Specification (4) indicates that the percentage of retail trading increases by 0.17 percentage points (relative to a mean value of 7.75%) following SA reports, consistent with retail investors reacting to SA research more than institutional investors.

We next examine individual half-hour windows before and after the publication time by re-estimating Specification (2) after including retail trades in the half-hour of publication and replacing *Post_SA* with ten separate indicator variables for each half-hour period ranging from [-4] to [5], with period [-5] being the omitted group. Fig. 1 reports the results. We observe that the estimated coefficients in the pre-period [-4, -1] are economically small (less than

[19] Another potentially important confounding event is that some SA contributors may choose to post their research reports on their personal blogs prior to release on Seeking Alpha. We explore this possibility in greater detail in Section 4.3.

[20] For example, *Percent Retail* is much smaller in the last 30 minutes of the trading day, particularly in more recent periods, which is likely attributable to the growth in passive investment vehicles that typically rebalance at the end of the trading day.

[21] The serial correlation in firm residuals is close to zero ($\rho = 0.007$), which obviates the need for clustering by firm. In untabulated analysis, we confirm that clustering by both firm and date leads to similar standard errors.

Table 3
Seeking Alpha Research and the Intensity of Retail Investor Trading
This table presents results from the estimation of Eq. (2):

$$Retail\_Trd_{i,t} = \beta_1 Post\_SA_{i,t} + \beta_2 Controls_{i,t} + Report_i + HalfHour_t \times Month + \varepsilon_{i,t}$$

Retail Trd is either *Retail Volume*, defined as log (1 + Retail Volume) in half-hour window t around the publication of report i or *Percent Retail Trading*, defined as total retail trading volume in half-hour window t scaled by total aggregate trading volume in the same window. Trades are classified as retail using the approach of Boehmer et al., 2020. In specifications (1) and (4), the sample includes all intraday SA reports and the event period is [-5, 5]. In Specifications (2) and (5), the sample excludes reports accompanied by media articles, IBES research, or earnings announcements in the event period [-5, 5]. In Specifications (3) and (6), the event period is narrowed to [-1, 1]. *Post_SA* is equal to one for trading in the post-event period and zero for trading in pre-event period. *Controls* include lagged returns, absolute returns, and trading volume over the previous half-hour period and previous two to five half-hour periods. See Appendix A for detailed variable definitions. *Report* denotes SA research report fixed effects and *Half Hour* × *Month* denotes half-hour fixed effects interacted with a monthly fixed effect. All continuous variables are standardized. Standard errors are clustered by date, and *t*-statistics are reported below each estimate.

	Retail volume			Percent retail trading		
	(1)	(2)	(3)	(4)	(5)	(6)
Post_SA	6.03%	8.77%	7.68%	0.17%	0.14%	0.21%
	(6.40)	(8.15)	(8.17)	(5.73)	(3.86)	(4.19)
Abs $Ret_{i,t\text{-}1}$	9.74%	9.99%	7.41%	0.25%	0.28%	0.22%
	(29.08)	(24.86)	(10.79)	(13.89)	(11.74)	(4.86)
Abs $Ret_{i,[t\text{-}5,t\text{-}2]}$	3.38%	3.46%	0.49%	0.12%	0.15%	0.12%
	(10.17)	(8.29)	(0.70)	(6.89)	(6.04)	(2.68)
$Ret_{i,t\text{-}1}$	1.00%	1.03%	1.36%	0.01%	0.01%	0.03%
	(3.68)	(2.97)	(2.32)	(0.95)	(0.50)	(0.75)
$Ret_{i,[t\text{-}5,t\text{-}2]}$	0.64%	1.05%	0.65%	0.01%	0.00%	-0.04%
	(1.82)	(2.43)	(0.97)	(0.51)	(-0.06)	(-0.78)
High $volume_{i,t\text{-}1}$	13.61%	13.72%	11.22%	0.06%	0.07%	0.15%
	(16.63)	(12.84)	(6.36)	(1.59)	(1.44)	(1.73)
High $volume_{i,[t\text{-}5,\ t\text{-}2]}$	-1.54%	-2.16%	-17.89%	0.56%	0.59%	0.56%
	(-0.91)	(-0.94)	(-5.71)	(9.68)	(6.97)	(3.80)
Low $volume_{i,t\text{-}1}$	-6.99%	-6.01%	-1.19%	0.00%	-0.01%	0.03%
	(-4.49)	(-4.10)	(-0.60)	(0.04)	(-0.20)	(0.29)
Low $volume_{i,[t\text{-}5,\ t\text{-}2]}$	10.38%	12.79%	26.19%	-0.43%	-0.42%	-0.61%
	(4.17)	(4.71)	(6.70)	(-4.69)	(-3.85)	(-2.97)
Observations	485,710	354,755	90,076	485,710	354,755	90,076
SA Reports	Intraday	No Events	No Events	Intraday	No Events	No Events
Event Period	[-5, 5]	[-5, 5]	[-1, 1]	[-5, 5]	[-5, 5]	[-1, 1]
Report FE	Yes	Yes	Yes	Yes	Yes	Yes
Half Hour × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	81.9%	81.1%	90.5%	50.5%	49.9%	71.8%

2.66% in absolute value) and statistically insignificant. This is inconsistent with pre-trends explaining the increase in retail trading following the release of the report. In contrast, each of the post-event windows estimates are large (ranging from 8.55% to 10.06%) and all are highly statistically significant. We also observe a large increase in retail trading in period [0], the half-hour period that contains the publication, consistent with retail investors responding very quickly to SA research.

4.2. Seeking Alpha research sentiment and retail order imbalances

In this section, we examine whether investment research published on the Seeking Alpha platform influences the direction of retail trading by studying the relation between SA report sentiment and retail order imbalances. We classify SA research as having positive (negative) tone when the fraction of positive (negative) words in the SA report is above the sample median, using the word list of Loughran and McDonald (2011) as in Chen et al., (2014). We also measure sentiment using the SA contributor's disclosed investment position. We construct a long (short) indicator variable that takes the value of one if the contributor discloses a long (short) position (Campbell et al., 2019). We also consider a composite sentiment measure, constructed by aggregating the four indicator variables, (*Long* + *Pos. Tone*) – (*Short* + *Neg. Tone*).

We then estimate the following panel regression:

$$\begin{aligned} Retail\_OIB_{i,t} = {} & \alpha + \beta_1 Post\_SA \times SA\_Sentiment_{i,t} \\ & + \beta_2 Post\_SA_{i,t} + Controls_{i,t} + Report_i \\ & + HalfHour_t \times Month + \varepsilon_{i,t}. \end{aligned} \tag{3}$$

$Retail\_OIB_{i,t}$ is the retail order imbalance for firm *i* during half-hour *t*, defined as the difference between retail buy volume and retail sell volume, scaled by total retail trading volume (BJZZ). *Post_SA* is defined as in Eq. (2), and *Post_SA* × *SA_Sentiment* interacts the event-time indicators with the sentiment measures. The remaining controls and fixed effects are as in Eq. (2).

Specification (1) of Table 4 indicates that retail order imbalances are significantly related to three of the four tone measures in the predicted direction. For example, retail order imbalances decrease by -1.04 percentage points (pp) when a contributor discloses a short position (the co-

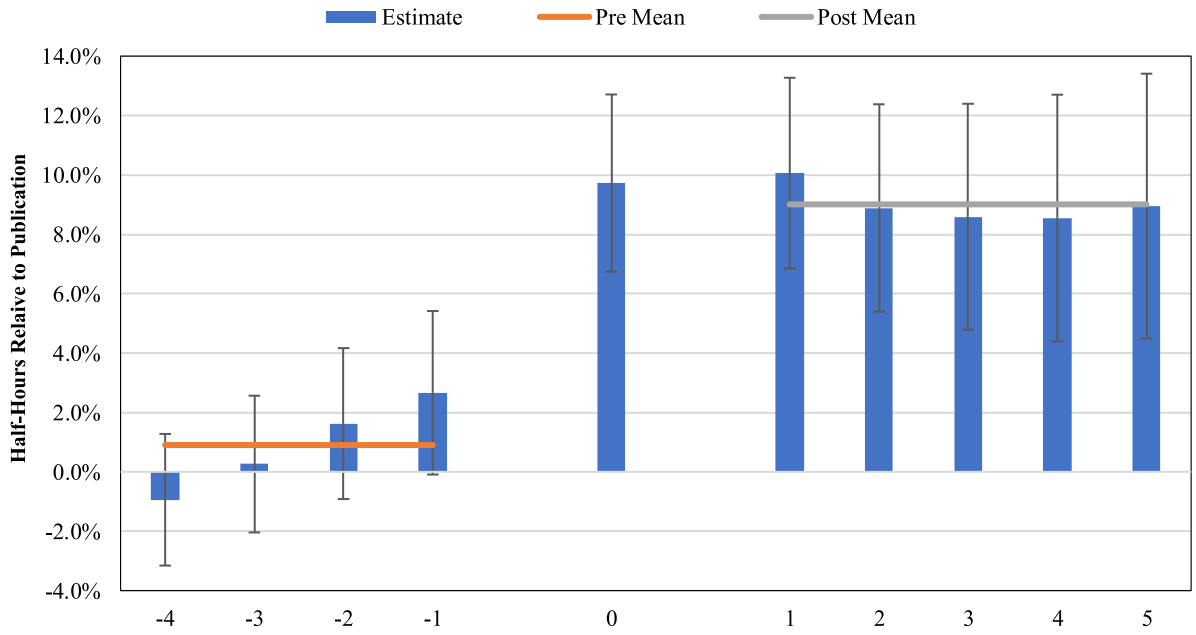

Fig. 1. Seeking Alpha Research and the Intensity of Retail Investor Trading: Event Time. This figure plots estimates from Specification (2) of Table 3 after including the half hour publication window and replacing the single post indicator variable with ten indicators marking off event windows [-4] through [5]. The coefficients for each indicator variables are reported as blue bars, and their 95% confidence intervals as error bars. The average of the pre-event and post-event coefficient estimates appear as orange and grey horizontal lines.

efficient on *Long* is positive but insignificant). Similarly, order imbalances change by 0.67pp (-1.31pp) when the fraction of positive (negative) words in the report exceeds the sample median. Specification (2) shows that a one-unit increase in *Composite Sentiment* is associated with a 0.79pp increase in *Retail_OIB*. Specifications (3) and (4) report very similar estimates for *Composite Sentiment* after excluding reports with confounding events and when shrinking the event window to [-1, 1].

Fig. 2 reports the estimates from Specification (2) after replacing *Post_SA* × *SA_Sentiment* with *SA_Sentiment* interacted with ten separate indicator variables for each half-hour period ranging from [-4] to [5], with period [-5] being the omitted group. All of the pre-event estimates are statistically insignificant. Similar to Fig. 1, we see a sharp significant increase in period [0], the calendar half-hour that contains the publication time. The increase remains stable through period [3] and significantly declines in periods [4] and [5]. The evidence suggests that retail investors react quickly to the investment analysis provided by SA contributors.

4.3. Stale Seeking Alpha research reports

The results in Tables 3 and 4 suggest that SA research reports induce significant amounts of retail trading that is directionally consistent with the sentiment of the report. One potentially important attenuating factor is that some contributors may post their research reports on other websites before posting on Seeking Alpha. In this case, attentive investors may be able to trade on these SA reports before they are posted to Seeking Alpha, and our approach would therefore underestimate the effects of SA research on retail trading.

To explore the potential impact of "stale" reports, for all contributors who have authored at least ten reports, we visit the contributor's author page to identify whether they provide a link to a website (more details are available in Section IA.4 in the Internet Appendix). We find that roughly 41% of contributors provide a link to a website. For the sample of contributors with webpages, we manually search the webpage to examine whether any of their SA research is available on their website. We find that only 8% of authors with websites post *any* of their research on their site.[22] Since we generally cannot determine the timing of the publication on personal websites, we classify an author's SA reports as stale if we are able to find any SA reports on their linked webpage. Using these criteria, only 5.2% of *No Event* reports are classified as stale (1847 reports by 63 contributors).

In Table IA3 in the Internet Appendix, we repeat Specification (2) of Table 3 and Specification (3) of Table 4 for stale and non-stale reports. The evidence is consistent with pre-posting attenuating our results, with the increase in *Retail* Volume following SA reports being roughly 50% larger for non-stale reports (6.54% versus 4.39%). Similarly, retail order imbalances are roughly 11% more correlated with report sentiment for non-stale reports (0.96pp versus 0.86pp). The evidence is consistent with stale Seeking Alpha reports inducing weaker retail trading responses. How-

[22] The reluctance of SA contributors to post research on their personal website may stem from SA's condition that authors will be compensated for articles that are exclusive to Seeking Alpha: https://seekingalpha.com/page/premium-partnership-faq.

Table 4
Seeking Alpha research Sentiment and the Direction of Retail Investor Trading
This table presents results from the estimation of Eq. (3):

$$Retail\ OIB_{i,t} = \beta_1 PostSA \times Sentiment_{i,t} + \beta_2 PostSA_{i,t} + \beta_3 Controls_{i,t} + Report_i + HalfHr_t \times Month + \varepsilon_{i,t.}$$

$Retail\_OIB_{i,t}$ is the retail order imbalance during half-hour t around the publication of report i, computed as the difference between retail buy volume and retail sell volume, scaled by total retail trading volume. In Specifications (1) and (2), the sample includes all intraday SA reports and the event period is [-5, 5]. In Specification (3), the sample excludes reports accompanied by media articles, IBES research, or earnings announcements in the event period [-5, 5]. In Specification (4), the event period is narrowed to [-1, 1]. *Post_SA* is equal to one in the post-event period and zero in the pre-event period. SA Sentiment includes a *Long* (*Short*) indicator equal to one if the contributor discloses a long (short) position, and a *Positive Tone* (*Negative Tone*) indicator equal to one if the fraction of positive (negative) words in the report exceeds the sample median. *Composite Sentiment* is defined as *Long* + *Positive Tone* – *Short* – *Negative Tone*. *Controls* include lagged returns, absolute returns, and trading volume for the previous half-hour period and previous two to five half-hour periods. See Appendix A for detailed variable definitions. All continuous variables are standardized. Standard errors are clustered by date, and t-statistics are reported below each estimate.

	(1)	(2)	(3)	(4)
Post × SA Long	0.36%			
	(1.29)			
Post × SA Short	-1.04%			
	(-2.19)			
Post × SA Negative Tone	-1.31%			
	(-4.70)			
Post × SA Positive Tone	0.67%			
	(2.49)			
Post × SA Composite Sentiment		0.79%	0.99%	0.90%
		(5.22)	(5.21)	(3.10)
Post × SA	0.83%	0.32%	0.43%	1.01%
	(2.97)	(1.95)	(2.11)	(3.61)
$Abs\ Ret_{i,t\text{-}1}$	0.26%	0.26%	0.28%	0.36%
	(3.42)	(3.46)	(2.92)	(1.94)
$Abs\ Ret_{i,[t\text{-}5,t\text{-}2]}$	0.23%	0.23%	0.34%	0.57%
	(2.72)	(2.75)	(3.18)	(2.82)
$Ret_{i,t\text{-}1}$	-1.45%	-1.45%	-1.50%	-1.79%
	(-21.71)	(-21.74)	(-17.21)	(-10.32)
$Ret_{i,[t\text{-}5,t\text{-}2]}$	-1.40%	-1.40%	-1.55%	-1.80%
	(-17.41)	(-17.45)	(-14.59)	(-8.92)
$High\ Volume_{i,t\text{-}1}$	-0.02%	-0.01%	-0.27%	0.19%
	(-0.10)	(-0.06)	(-0.94)	(0.36)
$High\ Volume_{i,[t\text{-}5,\ t\text{-}2]}$	0.63%	0.67%	0.49%	1.33%
	(1.92)	(2.03)	(1.03)	(1.45)
$Low\ Volume_{i,t\text{-}1}$	-0.17%	-0.16%	-0.26%	-0.44%
	(-0.63)	(-0.61)	(-0.85)	(-0.75)
$Low\ Volume_{i,[t\text{-}5,\ t\text{-}2]}$	-0.20%	-0.18%	-0.12%	-1.19%
	(-0.39)	(-0.35)	(-0.20)	(-1.05)
Observations	485,710	485,710	354,755	90,076
SA Reports	Intraday	Intraday	No Events	No Events
Event Period	[-5, 5]	[-5, 5]	[-5, 5]	[-1, 1]
Report FE	Yes	Yes	Yes	Yes
Half Hour × Month FE	Yes	Yes	Yes	Yes
R-squared	20.8%	20.8%	20.1%	54.5%

ever, stale reports are rare, and therefore their inclusion does not meaningfully impact the overall findings.

5. Seeking Alpha research and the informativeness of retail investor trading

There are at least two reasons to believe that SA may help retail investors trade in a more informed way. First, retail investors tend to trade in the direction of report and comment tone, and these variables have been shown to forecast stock returns (Chen et al., 2014). Second, retail investors may be skilled in gleaning additional valuable information from SA reports. This finding would be broadly consistent with growing evidence suggestive of retail investor skill (e.g. Kaniel et al. 2008, Kaniel et al. 2012, Kelley and Tetlock 2013, 2017; and Boehmer et al. 2020). On the other hand, SA reports could reinforce well-known biases of retail investors and potentially exacerbate mispricing (e.g., Heimer 2016, Cookson et al. 2020, Chawla et al. 2017).

5.1. Seeking Alpha research, retail investor trading, and future stock returns

Our primary measure of retail trade informativeness associates retail order imbalances and future stock returns. We focus on one-week ahead returns, as in BJZZ, but also consider longer-horizon returns. As in previous sections, we compare retail informativeness in the post-event window to the informativeness in the pre-event window. Specifically, we estimate the following intraday panel re-

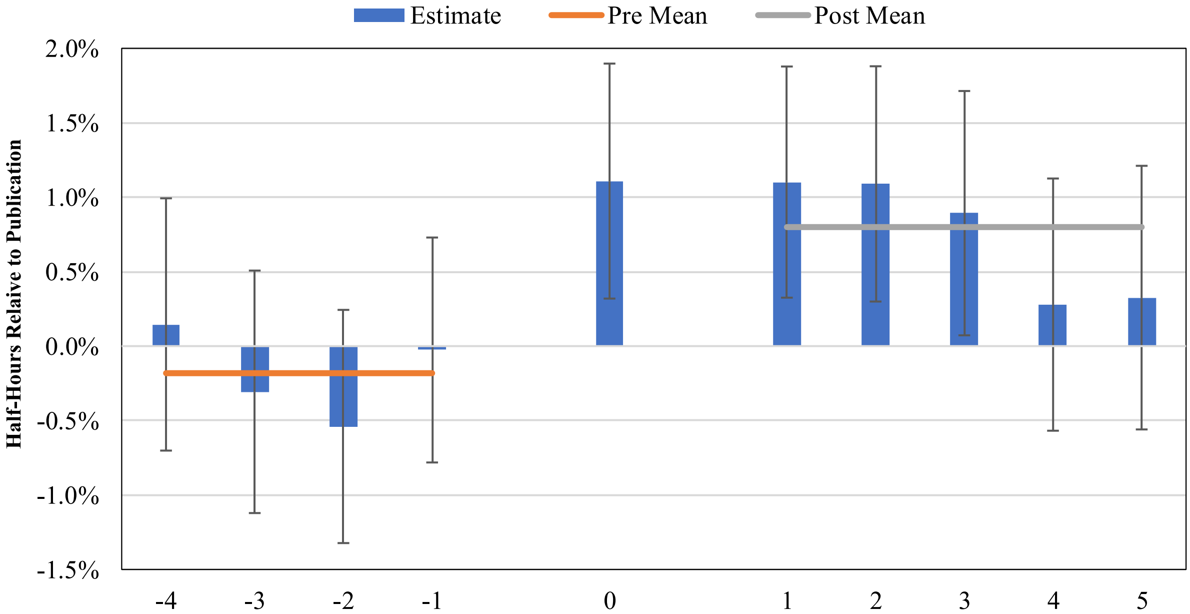

Fig. 2. Seeking Alpha Research Sentiment and the Direction of Retail Investor Trading: Event Time. This figure plots estimates from Specification (3) of Table 4 after replacing Post_SA × SA_Sentiment with SA_Sentiment interacted with ten indicators marking off event windows [-4] through [5]. We report the coefficients on these variables as blue bars, and their 95% confidence intervals as error bars. The average of the pre-event and post-event coefficient estimates appear as orange and grey horizontal lines.

gression.

$$Ret_{i,[t,\ t+5d]} = \beta_1 Retail\_OIB_{i,t} + \beta_2 Post\_SA_{i,t} \times Retail\_OIB_{i,t} + \beta_3 Inst\_OIB_{i,t} + \beta_4 Post\_SA_{i,t} \times Inst\_OIB_{i,t} + Controls_{i,t} + HalfHour_t \times Month + \varepsilon_{i,t.} \quad (4)$$

where $Ret_{i,[t,t+5d]}$ is the market-adjusted return, based on the bid-ask average price at the end of half-hour t until the close of trading after five full trading days.[23] *Retail_OIB* is defined as in Eq. (3); to facilitate interpretation, we standardize it to have mean zero and unit variance. *Post_SA* × *Retail_OIB* interacts *Retail_OIB* with the *Post_SA* indicator. Institutional order imbalance (i.e., non-retail order imbalance) variables are defined analogously. *Controls* and *Half Hour* × *Month* are defined as in Eq. (2). Standard errors are clustered by month. We no longer include *Report* fixed effects since the dependent variable (5-day ahead returns) exhibits very little variation for a given report.

Table 5 reports the results. In Specification 1, we find that *Retail_OIB* is negative and insignificant, suggesting that retail trading in the pre-event window is uninformed.[24] In contrast, we find the coefficient on *Retail_OIB* × *Post_SA* is positive and statistically significant. These point estimates suggest that a one-standard deviation increase in retail order imbalance is associated with a 0.054 percentage point decline in 5-day ahead returns prior to report release, but a 0.159pp increase (-0.054 + 0.213) after the report release (we confirm that the 0.159pp estimate is significantly greater than zero with $t = 3.41$). The coefficient on *Inst_OIB* × *Post_SA*, while positive, is smaller and statistically insignificant in each specification.[25] We find similar patterns in Specifications (2) and (3), where we exclude reports that coincide with confounding information events and limit the event window to [-1, 1].

Fig. 3 plots the estimates from Specification (2) of Eq. (4) after dropping *Retail_OIB* and *Post_SA* × $Retail\_OIB_{it}$ and adding *Retail_OIB* interacted with 11 separate SA indicator variables for each half-hour period ranging from -5 to +5.[26] We observe that four of the five the pre-event estimates are negative. In contrast, all the estimates over the [1, 5] window are positive, with the estimates for period [0] (the half-hour period of publication) and period [1] reliably greater than zero. The stark difference in trading between pre- and post-publication periods suggests that the informed trading in the half hour of publication is likely

[23] Thus, the returns reflect five-day returns plus the initial intraday return. Excluding the latter, i.e., assuming all trades occur at the closing price at the end of the publication day yields similar results.

[24] This finding raises the questions of whether retail trading is (i) only informative following SA reports and (ii) significantly less informative in the period immediately prior to the SA report. When we expand the sample to include retail trading on all firm-days (4.2 million firm-days) rather than just on days with SA research (61,282 firmdays), we find that (i) retail trading is informative in general (Table 11) and (ii) it is not significantly less informative in the days prior to the SA report (Table IA11).

[25] In the Internet Appendix, we find a weaker relationship between SA research and institutional trading intensity (Table IA4) and direction of trading (Table IA5). The difference in strength of findings is consistent with retail investors paying greater attention to Seeking Alpha, perhaps because institutional investors emphasize more exclusive information sources including private social networks (Crawford et al., 2018), private meetings with executives (Solomon and Soltes, 2015; Bradley, Jame, and Williams, 2021) or alternative data sets (Katona et al. (2019).

[26] In order to estimate period [0] effect, we also drop *Institutional_OIB* × *Post_SA*, which is undefined for period [0].

Table 5
Seeking Alpha Research and the Informativeness of Retail Investor Trading
This table presents results from the estimation of Eq. (4):

$$\begin{aligned} Ret_{i,[t,t+5d]} = {} & \beta_1 Retail\_OIB_{i,t} + \beta_2 Post\_SA_{i,t} \times \ Retail\_OIB_{i,t} \\ & + \beta_3 Inst\_OIB_{i,t} + \beta_4 Post\_SA_{i,t} \times Inst\_OIB_{i,t} \\ & + Controls_{i,t} + HalfHour_t \times \ Month + \varepsilon_{i,t.} \end{aligned}$$

In Specification (1), the sample includes all SA reports and the event period is [-5, 5]. In Specification (2), the sample excludes reports accompanied by media articles, IBES research, or earnings announcements in the event period [-5, 5]. In Specification (3), the event period is narrowed to [-1, 1]. $Ret_{i,[t,t+5d]}$ is the five-day market-adjusted return, computed from the bid-ask average price at the end of half-hour t and the closing price at the end of day 5, with the publication day being day 0. $Retail\_OIB_{i,t}$ is the retail order imbalance during half-hour t around the publication of report i, computed as the difference between retail buy volume and retail sell volume, scaled by total retail trading volume. *Post_SA* is equal to one in the post-event period and zero in the pre-event period. $InstOIB_{i,t}$ is non-retail buy volume less non-retail sell volume in window t, scaled by non-retail trading volume. All other control variables are defined in Appendix A. All continuous variables are standardized. Standard errors are clustered by month and t-statistics are reported below each estimate.

	(1)	(2)	(3)
Retail_OIB	-0.054	-0.105	-0.116
	(-1.36)	(-2.11)	(-1.22)
Retail_OIB × *Post_SA*	0.213	0.256	0.422
	(3.50)	(3.51)	(3.54)
Institutional_OIB	0.128	0.182	0.070
	(1.37)	(1.74)	(0.42)
Institutional_OIB × *Post_SA*	0.193	0.233	0.382
	(1.58)	(1.67)	(1.76)
$Abs\ Ret_{i,t\text{-}1}$	0.003	-0.024	0.002
	(0.05)	(-0.48)	(0.03)
$Abs\ Ret_{i,[t\text{-}5,t\text{-}2]}$	-0.060	-0.068	-0.094
	(-1.07)	(-1.01)	(-1.36)
$Ret_{i,t\text{-}1}$	0.024	0.028	0.045
	(1.06)	(1.06)	(1.06)
$Ret_{i,[t\text{-}5,t\text{-}2]}$	0.068	0.046	0.000
	(1.90)	(1.01)	(0.00)
$High\ volume_{i,t\text{-}1}$	0.030	0.028	-0.105
	(0.51)	(0.40)	(-1.11)
$High\ volume_{i,[t\text{-}5,\ t\text{-}2]}$	-0.126	-0.289	-0.251
	(-1.35)	(-2.28)	(-1.29)
$Low\ volume_{i,t\text{-}1}$	0.012	0.010	-0.118
	(0.30)	(0.22)	(-1.70)
$Low\ volume_{i,[t\text{-}5,\ t\text{-}2]}$	-0.092	-0.125	-0.043
	(-0.83)	(-1.07)	(-0.28)
Observations	484,244	353,557	89,774
SA reports	Intraday	No Events	No Events
Event period	[-5, 5]	[-5, 5]	[-1, 1]
Half hour × Month FE	Yes	Yes	Yes
R-squared	1.09%	1.16%	2.64%

a reflection of retail investors quickly responding to SA research.

We also consider the relation between *Retail_OIB* and stock returns over longer horizons. Specifically, we estimate Eq. (4) for weekly returns ranging from one-week ahead (the baseline analysis) through five-weeks ahead. We also report the cumulative returns for a five-week holding period and a 12-week holding period. The results, based on Specification (2) of Table 5, are presented in Table 6. The individual estimates for weeks 2 through 5 are all statistically insignificant. Similarly, the cumulative estimates for week 5 (0.291pp) and week 12 (0.224pp) are very similar to the week 1 estimate (0.256pp). The absence of a notable drift suggests that retail investors are trading primarily based on information that is impounded into prices within one week of the report publication, and the lack of a significant reversal is inconsistent with retail trading reflecting uninformed price pressure (additional price pressure tests in Sections 5.3 and 5.4, and Section IA.10.4 in the Internet Appendix reinforce this finding).

5.2. Seeking Alpha research, retail investor trading, and future stock returns – sensitivity tests

In this section, we examine the sensitivity of our baseline estimate of the effect of SA research (Specification 2 in Table 5) to alternative research design choices. The results are discussed in greater detail in Section IA.6 and Table IA6 of the Internet Appendix. We first explore the implication of stale reports. In Section 4.3, we report that stale research reports induce lower trading responses, which raises concerns about the measurement of counterfactual informed retail trading in the pre-SA publication period. We find that excluding stale reports increases our baseline estimate slightly from 0.256pp to 0.259pp.[27]

[27] Conducting the informativeness test solely on stale reports yields an estimate of -0.049pp (t = 0.17), consistent with information leakage diminishing the effect of SA research publication on retail trade informativeness (untabulated for brevity).

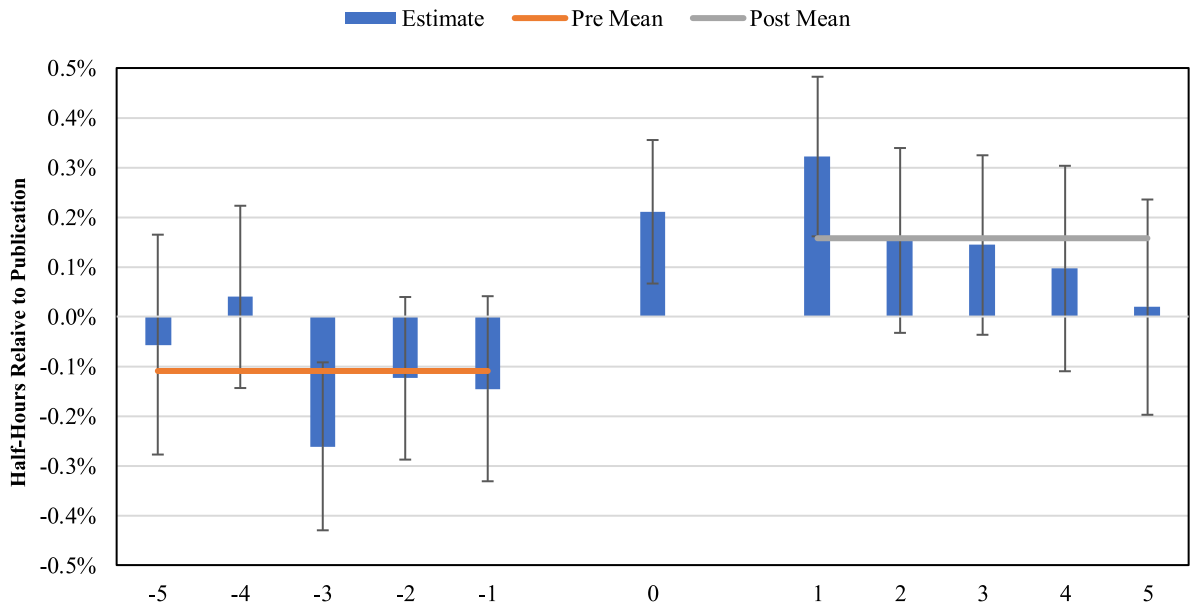

Fig. 3. Seeking Alpha Research and the Informativeness of Retail Investor Order Imbalances: Event Time. This figure plots estimates from Specification (2) of Table 5 after replacing *Retail_OIB* and *Post_SA* × *Retail_OIB* with *Retail_OIB* interacted with 11 separate indicator variables for each half-hour period ranging from [-5] to [5]. We report the coefficients on these variables as blue bars, and their 95% confidence intervals as error bars. The average of the pre-event and post-event coefficient estimates appear as orange and grey horizontal lines (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
SA Research and the Informativeness of Retail Investor Trading: Longer Horizon Evidence
This table repeats Specification (2) of Table 5 after replacing the dependent variable (one-week ahead returns) with returns over longer horizons. For reference, Specification 1 repeats the one-week analysis reported in Table 5. Specifications 2 through 5 replace one-week ahead returns with returns over weeks 2 through 5, respectively. The week 2 return is computed based on the buy-and-hold return over the six to ten trading days after the report releases, and other weekly returns are defined analogously. Specifications 6 and 7 report the cumulative buy-and-hold return over the 5-week and 12-week holding periods.

	Week 1 (1)	Week 2 (2)	Week 3 (3)	Week 4 (4)	Week 5 (5)	Weeks (1-5) (6)	Wks (1-12) (7)
Retail_OIB	-0.105	0.033	0.060	-0.014	-0.062	-0.088	-0.027
	(-2.11)	(0.78)	(1.30)	(-0.28)	(-1.37)	(-0.93)	(-0.14)
Retail_OIB × *Post_SA*	0.256	-0.010	-0.062	0.007	0.099	0.291	0.224
	(3.51)	(-0.15)	(-1.40)	(0.10)	(1.67)	(2.30)	(1.00)
Inst_OIB	0.182	0.098	-0.024	-0.076	0.104	0.283	0.596
	(1.74)	(1.15)	(-0.31)	(-0.75)	(1.09)	(1.42)	(1.52)
Inst_OIB × *Post_SA*	0.233	-0.063	-0.054	0.223	0.115	0.452	0.400
	(1.67)	(-0.57)	(-0.47)	(1.75)	(0.95)	(1.82)	(0.71)
$Abs\ Ret_{i,t-1}$	-0.024	-0.012	0.008	-0.064	-0.014	-0.103	-1.046
	(-0.48)	(-0.31)	(0.12)	(-1.42)	(-0.36)	(-0.72)	(-3.14)
$Abs\ Ret_{i,[t-5,t-2]}$	-0.068	-0.038	-0.068	-0.056	-0.056	-0.282	-1.436
	(-1.01)	(-0.76)	(-1.36)	(-1.04)	(-1.24)	(-1.69)	(-3.24)
$Ret_{i,t-1}$	0.028	-0.030	-0.011	0.021	0.006	0.014	0.172
	(1.06)	(-1.70)	(-0.51)	(1.05)	(0.25)	(0.28)	(2.04)
$Ret_{i,[t-5,t-2]}$	0.046	0.002	-0.035	0.053	-0.038	0.027	0.351
	(1.01)	(0.06)	(-0.95)	(1.55)	(-1.02)	(0.34)	(1.80)
$High\ volume_{i,t-1}$	0.028	-0.019	-0.064	-0.090	-0.024	-0.171	0.226
	(0.40)	(-0.42)	(-0.93)	(-1.83)	(-0.55)	(-1.21)	(0.75)
$High\ volume_{i,[t-5,\ t-2]}$	-0.289	-0.302	-0.096	-0.047	0.034	-0.694	-1.060
	(-2.28)	(-2.74)	(-0.85)	(-0.42)	(0.32)	(-3.21)	(-1.90)
$Low\ volume_{i,t-1}$	0.010	-0.080	-0.065	0.013	0.027	-0.093	-0.151
	(0.22)	(-1.83)	(-1.67)	(0.30)	(0.69)	(-0.98)	(-0.90)
$Low\ volume_{i,[t-5,\ t-2]}$	-0.125	-0.048	-0.034	0.236	0.097	0.129	0.101
	(-1.07)	(-0.38)	(-0.28)	(2.08)	(0.78)	(0.49)	(0.22)
Observations	353,557	353,557	353,557	353,557	353,557	353,557	353,557
Sample	No Events	No Events	No Events	No Events	No Events	No Events	No Events
Window	[-5,5]	[-5,5]	[-5,5]	[-5,5]	[-5,5]	[-5,5]	[-5,5]
Half hour × Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	1.16%	1.00%	1.11%	0.81%	0.76%	1.91%	1.99%

We also more carefully address earnings dates. Although our analysis excludes days with earnings announcements during the event windows, it is possible that the confounding effect of earnings news on the informativeness of retail trading may extend beyond day 0. We therefore further exclude reports published on days +1 and -1. The resulting estimates are 0.265pp and 0.235pp, respectively, and both are highly significant.[28]

We next expand the sample of intraday research reports (those published between 10:30 am and 3:30 pm) to include reports issued between 10 am and 3:30 pm, and we further expand the set to include all (including overnight) reports. For reports issued between 10 and 10:30, the pre-period is 9:30–10 am. For all reports issued after hours, the pre-period is the five half-hour periods at the end of the previous trading day (i.e., from 1:30 to 4:00 pm) and the post-period is the five-half hour at the beginning of the next trading day (i.e., from 9:30 am to 12:00 pm). The results after including articles issued between 10 am and 10:30 am are virtually identical. Including overnight reports results in somewhat smaller estimates (0.175pp) but they remain statistically significant (*t*-stat of 2.71).

Finally, we estimate and plot informativeness estimates by month in Fig. IA2. We observe a sharp increase in the effect of SA research on retail trade informativeness from July 2008 to December 2008, followed by a subsequent steady increase. Our results are robust to excluding the financial crisis period from July to December 2008 (the coefficient is 0.230pp). We also split the sample into three equal calendar periods and find the estimates are statistically significant at the 10% level or higher for each period.

5.3. Decomposing retail trading into price pressure, liquidity provision, and informed trading

Boehmer et al. (2020) document that retail investor order imbalances are persistent, which raises concerns that buying or selling pressure could explain our findings. In particular, Seeking Alpha research may amplify noise trading among retail investors, generating price pressure that results in short-term return predictability. Moreover, it is possible that that the relation between retail order imbalances and future returns may be attributable to liquidity provision rather than informed trading. For example, Table 4 documents that retail investor order imbalances are contrarian (e.g. the coefficients on Ret_{t-1} and $Ret_{[t-5,t-2]}$ are significantly negative), and short-term contrarian trading is a common proxy for liquidity provision (e.g., Nagel 2012, Jame 2018).

We explore the potential roles of price pressure and liquidity provision using an approach similar to BJZZ. In particular, we decompose retail order imbalances into three components: *OIB Persistence* (a proxy for price pressure), *OIB Contrarian* (a proxy for liquidity provision), and *OIB Other* (a proxy for informed trading). The three components are estimated using the following panel regression: $Retail\ OIB_{it} = \alpha + \beta_1 Retail\ Oib_{i,[t-5,t-1]} + \beta_2 Ret_{i,[t-5,t-1]} + \varepsilon_{it},$ where $OIB\ Persistence = \hat{\beta}_1\ Retail\ Oib_{[t-5,t-1]}$; $OIB\ Contrarian = \hat{\beta}_2\ Ret_{i,[t-5,t-1]}$; and $OIB\ Other = \hat{\varepsilon}_{it}$. We then re-estimate Specification 2 of Table 5 after replacing total retail order imbalance (*Retail OIB)* with *OIB Persistence, OIB Contrarian,* or *OIB Informed*.

The decomposition results are reported in Table 7. We observe that the coefficient on *OIB Contrarian* × *Post SA* is insignificant, which is inconsistent with liquidity provision contributing to the incremental informativeness of retail order imbalances after SA research. The coefficient on *OIB Persistence* × *Post SA*, which captures the price pressure component, is marginally significant ($p < 0.10$). Importantly, however, we find that the informed component of retail order flow is a highly significant predictor of the cross-section of future returns (the coefficient on *OIB Informed* × *Post SA* is significant with $p < 0.01$), which is consistent with SA research contributing to more informative retail trading.

5.4. Seeking Alpha research, retail investor trading, and future cash flows news

If SA research results in more informative retail trading, then retail order imbalances following SA research should also more strongly predict the cross-section of future cash flow news. To test this prediction, we estimate the following panel regression:

$$\begin{aligned} CFNews_{i,[t,t+5d]} = \alpha &+ \beta_1 Retail\_OIB_{i,t} + \beta_2 Post\_SA_{i,t} \\ &\times Retail\_OIB_{i,t} + \beta_3 Inst\_OIB_{i,t} \\ &+ \beta_4 Post\_SA_{i,t} \times Inst\_OIB_{i,t} + Controls_{i,t} \\ &+ HalfHour_t \times Month + \varepsilon_{i,t}. \end{aligned} \tag{5}$$

The dependent variable $CFNews_{i,[t,t+5d]}$ is a proxy for innovations in expected firms' cash flows over days t+1 through t+5, as measured by either the sentiment of media articles or the direction of analysts' earnings forecast revisions.

We contend that two media articles with the same negative sentiment convey more information about cash flows than a single article with the same negative sentiment. We therefore construct a measure of aggregate *Media Article Tone* by summing the *Event Sentiment Scores* of all articles published in the window [t+1, t+5], after subtracting 50 from each of them to ensure that summing articles with negative sentiment is meaningful. We define analyst forecast revisions (*Revisions*) as the total number of upward forecast revisions less the total number of downward forecast revisions over the [t+1, t+5] window. We exclude observations where there are no media articles (or forecast revisions) over days t+1 through t+5. In addition to the independent variables from Eq. (4), we also include lags of *Media Sentiment* or *Revisions* to control for potential persistence in public news. Following Kelley and Tetlock (2013), we construct lagged *Media Sentiment* and *Revisions* over day [0], days [-5, -1], and days [-26, -6].

Specifications (1)–(3) of Table 8 report the results for *Media Article Tone.* The estimate in Specification (1) indicates that a one-standard deviation increase in retail order

[28] We also separately examine trade informativeness for reports issued immediately prior to or after an earnings announcement. We find that our results are stronger for the former and similar for the latter. (see Section IA.7 of the Internet Appendix for additional details).

Table 7
Seeking Alpha Research and the Informativeness of Retail Order Imbalances: Decomposition Analysis
This table presents coefficients from the estimation of Specification (2) of Table 5 when retail trading is replaced with one of its three components: *Persistence* (a proxy for price pressure), *Contrarian* (a proxy for liquidity provision), or *Other* (a proxy for informed trading). These components are estimated as the fitted values from the panel regression:

$$Retail\_OIB_{i,t} = \alpha + \beta_1 Retail\_OIB_{i,[t-5,t-1]} + \beta_2 Ret_{i,[t-5,t-1]} + \varepsilon_{i,t},$$

where $\widehat{\mathrm{OIB}}^{\mathrm{Persistence}}_{\mathrm{i,t}} = \hat{\beta}_1 \mathrm{OIB}_{\mathrm{i,[t-5,t-1]}}$; $\widehat{\mathrm{OIB}}^{\mathrm{Contrarian}}_{\mathrm{i,t}} = \hat{\beta}_2 \mathrm{Ret}_{\mathrm{i,[t-5,t-1]}}$; and $\widehat{\mathrm{OIB}}^{\mathrm{Other}}_{\mathrm{i,t}} = \hat{\varepsilon}_{\mathrm{i,t}}$, respectively. All continuous variables are standardized. Standard errors are clustered by month, and t-statistics are reported below each estimate.

	Persistence (1)	**Contrarian** (2)	**Other (Informed)** (3)
Retail_OIB	-0.080	-0.090	-0.100
	(-1.08)	(-0.64)	(-2.11)
Retail_OIB × SA	0.193	-0.080	0.243
	(1.93)	(-0.39)	(3.35)
Institutional_OIB	0.150	0.146	0.178
	(1.41)	(1.37)	(1.70)
Institutional_OIB × SA	0.311	0.321	0.243
	(2.21)	(2.29)	(1.74)
Abs $Ret_{i,t\text{-}1}$	-0.024	-0.024	-0.024
	(-0.48)	(-0.48)	(-0.47)
Abs $Ret_{i,[t\text{-}5,t\text{-}2]}$	-0.068	-0.067	-0.067
	(-1.01)	(-0.99)	(-1.00)
$Ret_{i,t\text{-}1}$	0.028	0.008	0.029
	(1.06)	(0.53)	(1.07)
$Ret_{i,[t\text{-}5,t\text{-}2]}$	0.045		0.046
	(0.98)		(1.00)
High $Volume_{i,t\text{-}1}$	0.026	0.026	0.026
	(0.37)	(0.37)	(0.37)
High $Volume_{i,[t\text{-}5,\ t\text{-}2]}$	-0.287	-0.287	-0.288
	(-2.24)	(-2.24)	(-2.24)
Low $Volume_{i,t\text{-}1}$	0.011	0.011	0.011
	(0.24)	(0.24)	(0.25)
Low $Volume_{i,[t\text{-}5,\ t\text{-}2]}$	-0.123	-0.123	-0.124
	(-1.05)	(-1.05)	(-1.06)
Observations	353,557	353,557	353,557
Sample	Intraday-Clean	Intraday-Clean	Intraday-Clean
Half Hour × Month FE	Yes	Yes	Yes
R-squared	1.15%	1.15%	1.16%

imbalances over the pre-event window [-5, -1] is associated with a 1.42 increase in five-day ahead media tone, but this estimate increases to 2.63 (1.42 + 1.21) over the [1, 5] post-event window (roughly 1/10 of the cross-sectional standard deviation of 28). We find similar results after excluding confounding events in Specification (2); shrinking the event window to [-1, 1] in Specification (3), however, reduces the coefficient magnitude by roughly 50% and the estimate is no longer statistically significant.

Specifications (4)–(6) of Table 8 document analogous results for forecast revisions. Using the [-5, 5] event window, we find a positive and significant coefficient (0.20) on *Retail_OIB × Post_SA* for the sample of reports without confounding events. The point estimate is similar (0.31) and significant when we shrink the event window to [-1, 1]. Collectively, the evidence is consistent with retail trading being a stronger predictor of cash flows following the release of SA research.[29]

5.5. *Exploring the mechanism underlying Seeking Alphas's effect on retail investor informativeness*

The evidence from the prior sections suggests that Seeking Alpha reports contribute to retail investor trading being more informed. It is possible that the increased informativeness could simply be a consequence of retail investors trading in the direction of report sentiment which has been shown to forecast stock returns (Chen et al., 2014). Alternatively, retail investors may be skilled in gleaning additional valuable information from SA reports. In particular, SA users often subscribe to real-time alerts for stocks that they already follow, which likely provides important context for interpreting the analysis in SA research reports. This section offers evidence on the relative importance of these two channels.

[29] In Fig. IA3 in the Internet Appendix, we calculate the cash flow effects for each half-hour interval analogous to Fig. 3. Between the two measures of cash flow news, only one of the pre-event half-hour coefficient estimates is statistically significant, whereas 8 of the 10 post-event coefficients are significant.

Table 8
Seeking Alpha research and the informativeness of retail investor trading: predicting cash flow news
This table reports results from the estimation of Eq. (5):

$$CFNews_{i,[t,t+5d]} = \alpha + \beta_1 Retail\ OIB_{i,t} + \beta_2 PostSA_{i,t} \times\ Retail\ OIB_{i,t} +\ \beta_3 Inst\ OIB_{i,t} + \beta_4 Inst\ OIB_{i,t}$$
$$+ Controls_{i,t} + HalfHour_t \times\ Month + \varepsilon_{i,t.}$$

CFNews is *Media Article Tone*, defined as the sum of the Adjusted Event Sentiment Score (ESS) across all stock-focused media articles in the five-day period following the day of SA report publication, or *Forecast Revisions*, defined as the number of upward forecast revisions less the number of downward forecast revisions over the same period. SA reports that are not followed by any media articles or any forecast revisions within five days are dropped. *Controls* include all Table 5 control variable, as well as *Media Article Tone* and *Forecast Revisions* calculated over daily intervals [0], [-5, -1], and [-26, -6]. In Specifications (1) and (4), the sample includes all SA reports and the event period is [-5, 5]; in Specifications (2) and (5), the sample is limited to SA reports unaccompanied by media articles, IBES research, or earnings announcements in the event period; and in Specifications (3) and (6), the sample is further limited to event period [-1, 1]. All other variables are as defined in Table 5. Standard errors are clustered by month and t-statistics are reported below each estimate.

	Media Article Tone			Forecast Revisions		
	(1)	(2)	(3)	(4)	(5)	(6)
Retail_OIB	1.42	0.91	0.68	0.09	0.01	-0.04
	(3.89)	(2.34)	(1.13)	(1.29)	(0.09)	(-0.33)
Retail_OIB × *Post_SA*	1.21	1.04	0.56	0.11	0.20	0.31
	(2.97)	(2.27)	(0.77)	(1.20)	(1.99)	(2.23)
Institutional_OIB	0.25	0.63	-0.54	0.17	0.32	0.46
	(0.38)	(1.01)	(-0.57)	(1.12)	(1.75)	(1.83)
Institutional_OIB × *Post_SA*	0.12	0.19	0.67	0.03	0.18	-0.08
	(0.16)	(0.23)	(0.58)	(0.20)	(0.92)	(-0.22)
$Abs\ Ret_{i,t\text{-}1}$	-1.83	-1.94	-1.84	-0.13	-0.12	-0.11
	(-7.82)	(-8.87)	(-6.41)	(-4.54)	(-3.94)	(-2.47)
$Abs\ Ret_{i,[t\text{-}5,t\text{-}2]}$	-2.39	-2.71	-2.43	-0.21	-0.24	-0.25
	(-9.37)	(-10.44)	(-7.67)	(-5.97)	(-5.98)	(-5.44)
$Ret_{i,t\text{-}1}$	0.29	0.14	-0.04	0.09	0.07	0.05
	(2.90)	(1.29)	(-0.18)	(5.48)	(3.44)	(1.24)
$Ret_{i,[t\text{-}5,t\text{-}2]}$	0.13	-0.04	0.12	0.14	0.10	0.14
	(0.66)	(-0.21)	(0.47)	(5.42)	(2.51)	(2.80)
$High\ volume_{i,t\text{-}1}$	1.61	3.41	4.04	0.18	0.25	0.20
	(3.74)	(6.03)	(4.41)	(2.57)	(2.62)	(1.58)
$High\ volume_{i,[t\text{-}5,\ t\text{-}2]}$	-2.98	1.14	-0.25	-0.17	-0.01	0.01
	(-3.25)	(1.05)	(-0.20)	(-1.26)	(-0.07)	(0.06)
$Low\ volume_{i,t\text{-}1}$	-2.68	-2.61	-2.69	-0.19	-0.22	-0.27
	(-7.21)	(-6.94)	(-4.31)	(-2.67)	(-3.30)	(-2.17)
$Low\ volume_{i,[t\text{-}5,\ t\text{-}2]}$	-5.05	-4.88	-4.78	-0.28	-0.35	-0.20
	(-5.26)	(-5.33)	(-4.27)	(-1.53)	(-1.72)	(-0.93)
$Media\ tone\ /\ For\ Rev_{[0]}$	1.04	1.86	1.76	1.01	1.75	1.68
	(2.29)	(3.19)	(3.08)	(22.93)	(10.32)	(10.39)
$Media\ tone\ /\ For\ Rev_{[-5,-1]}$	2.37	1.43	1.41	0.65	0.59	0.59
	(5.94)	(3.86)	(3.72)	(11.63)	(10.99)	(11.08)
$Media\ tone\ /\ For\ Rev_{[-26,-6]}$	8.03	7.28	7.23	0.89	0.83	0.82
	(13.40)	(12.51)	(12.58)	(9.12)	(8.62)	(8.80)
Observations	396,314	276,097	70,195	238,905	157,680	40,058
SA Reports	Intraday	No Events	No Events	Intraday	No Events	No Events
Event Period	[-5, 5]	[-5, 5]	[-1, 1]	[-5, 5]	[-5, 5]	[-1, 1]
Half-Hour × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	7.87%	7.40%	9.13%	13.46%	13.40%	15.73%

5.5.1. Controlling for report tone

We begin by exploring whether retail investors' tendency to trade in the direction of report tone and position disclosures (as evidenced in Table 4) significantly contributes to retail order imbalances' increased informativeness following SA research. Specification (1) of Table 9 repeats Specification (2) of Table 5 after including *Composite Sentiment* (as defined in Table 4) as a control. We find that *Composite Sentiment* predicts future returns (Chen et al., 2014) but leaves the estimates on *Retail_OIB* × *Post_SA* virtually unchanged.[30] Specifications (2) and (3) present analogous results after replacing 5-day ahead returns with 5-day ahead media sentiment and forecast revisions, respec-

[30] This finding is perhaps surprising given the evidence that retail order imbalances are in the direction of report sentiment (Table 4), and the evidence that report sentiment tends to predict returns. However, while SA sentiment is a significant predictor of returns, the economic relation over our sample period is relatively weak. For example, the inclusion of Composite Sentiment only increases the R-squared by 0.07% (from 1.16% to 1.23%).

Table 9
Retail Investor Trading Informativeness Tests: Controlling for Seeking Alpha Research Sentiment

In this table, we re-estimate Specification (2) in Table 5 and Specifications (2) and (4) in Table 8 after including a measure of SA Research Sentiment (*Composite Sentiment*): computed as *Long* + *Positive Tone* – *Short* – *Negative Ton*e, where *Long* (*Short*) is an indicator equal to one if the contributor discloses a long (short) positions and *Positive* (*Negative*) *Tone* is an indicator equal to one if the fraction of positive (negative) words in the report exceeds the sample median. Standard errors are clustered by month and *t*-statistics are reported below each estimate.

	Returns (1)	Media Article Tone (2)	Forecast Revisions (3)
Composite Sentiment	0.289	0.74	0.27
	(6.20)	(2.05)	(3.96)
Retail_OIB	-0.114	0.90	0.00
	(-2.28)	(2.31)	(-0.02)
Retail_OIB × *Post_SA*	0.254	1.02	0.20
	(3.48)	(2.22)	(1.98)
Institutional_OIB	0.188	0.63	0.33
	(1.79)	(1.01)	(1.78)
Institutional_OIB × *Post_SA*	0.195	0.17	0.16
	(1.41)	(0.21)	(0.86)
$Abs\ Ret_{i,t\text{-}1}$	-0.010	-1.91	-0.11
	(-0.16)	(-8.79)	(-3.59)
$Abs\ Ret_{i,[t\text{-}5,t\text{-}2]}$	-0.048	-2.68	-0.22
	(-0.71)	(-10.32)	(-5.65)
$Ret_{i,t\text{-}1}$	0.020	0.13	0.06
	(0.74)	(1.17)	(3.20)
$Ret_{i,[t\text{-}5,t\text{-}2]}$	0.029	-0.06	0.09
	(0.63)	(-0.34)	(2.35)
$High\ Volume_{i,t\text{-}1}$	0.034	3.42	0.25
	(0.48)	(6.05)	(2.60)
$High\ Volume_{i,[t\text{-}5,\ t\text{-}2]}$	-0.287	1.14	-0.01
	(-2.25)	(1.04)	(-0.06)
$Low\ Volume_{i,t\text{-}1}$	0.014	-2.60	-0.22
	(0.31)	(-6.94)	(-3.26)
$Low\ Volume_{i,[t\text{-}5,\ t\text{-}2]}$	-0.121	-4.86	-0.35
	(-1.03)	(-5.31)	(-1.70)
$Media\ Tone\ /\ Forecast\ Rev_{[0]}$		1.39	1.74
		(3.77)	(10.26)
$Media\ Tone\ /\ Forecast\ Rev_{[-5,-1]}$		7.24	0.58
		(12.40)	(10.75)
$Media\ Tone\ /\ Forecast\ Rev_{[-26,-6]}$		1.83	0.81
		(3.14)	(8.44)
Observations	353,557	276,097	157,680
SA Reports	No Events	No Events	No Events
Event Period	[-5, 5]	[-5, 5]	[-5, 5]
Half-Hour × Month FE	Yes	Yes	Yes
R-squared	1.23%	7.42%	13.51%

tively. Controlling for tone does not meaningfully attenuate the relation between retail order imbalances and future cash flow news. The findings suggest that retail investors' incremental informativeness extends beyond a cursory assessment of report tone.

5.5.2. The role of Seeking Alpha research report quality

We conjecture that if retail investors are able to glean value relevant information, then higher quality research reports will lead to more informed trading. Our first measure of report quality is based on contributor's academic accomplishments, as self-reported in her bio. Chaudhuri et al. (2020) find that funds managed by PhDs outperform otherwise similar funds, and Chevalier and Ellison (1999) find that managers with MBAs or degrees from universities with higher average SAT scores outperform other fund managers. Accordingly, we define the indicator variable *Academic Quality* equal to one if the contributor bio mentions that the contributor has a PhD, an MBA, or graduated from a school in the top 50 of SAT scores based on the 75^{th} percentile, as reported in the 2015 vintage of stateuniversity.com.

Recent work finds that SA contributor skill is highly persistent (Farrell et al., 2020), with contributors that have issued more impactful research, as measured by the price impact of prior reports, being more likely to publish impactful reports in the future. We consider two measures of contributor skill. Our first measure is *Signed Returns*, computed as the two-day market-adjusted reaction multiplied by the *sign* of the report, where *sign* = 1 for positive reports and *sign* = -1 for negative reports. Our approach to signing reports follows Farrell et al. (2020), and is based on investment position disclosure and report tone (more details are available in the Appendix). Signing reports is noisy and excludes roughly 25% reports that are classified as neutral. Therefore, we also consider *Unsigned*

Returns, which equals one if the average absolute two-day market-adjusted reaction to a contributor's last five reports exceeds the yearly median and zero otherwise. The correlation between *Signed Returns* and *Unsigned Returns* is low ($\rho = 0.05$), suggesting that both may contain independently useful information.

We expect that higher quality reports will garner more attention and discussion. Our fourth measure of research quality is *Comments*, which is an indicator variable equal to one if the number of comments elicited by the report within 24 hours of the report release exceeds the yearly median. We also compute a composite quality measure (*Composite Quality),* defined as the sum of the four report quality measures.

We augment Eqs. (4) and (5) by interacting *Retail_OIB* and *Post_SA* × *Retail_OIB* with the different quality measures *(Academic Quality, Signed Return, Unsigned Return, Comments,* or *Report Quality*). We report results for the composite quality measure in Table 10 and for the individual components in Table IA8. The findings are consistent with retail investors becoming more informed following higher quality reports. In particular, we find that *Post_SA* × *Retail_OIB* × *Report Quality* is statistically significant and economically large. The negative (insignificant) coefficients on *Post_SA* × *Retail_OIB* are consistent with uninformed trading after reports with a composite quality score of 0 (roughly 23% of reports). However, for each one unit increase in composite quality, retail trade informativeness increases in the post-event window by an additional 0.347pp. Specifications (2) and (3) indicate that retail investors ability to forecast media tone and forecast revisions are also significantly stronger following higher quality reports.[31]

5.6. Seeking Alpha and the informativeness of retail trading:daily approach

For identification purposes, our approach thus far has focused exclusively on Seeking Alpha reports that are published within the trading day, which eliminates 2/3 of the sample (61,282 SA reports are published between 10:30 am and 3:30 pm out of a total of 183,969 SA reports). In this section, we consider an alternative daily empirical approach, which compares the informativeness of retail trading on *days* with SA research to the informativeness of their trading on all other days (hereafter *daily* approach).[32]

Measuring retail order flows over a longer horizon increases the possibility of confounding information events and makes it difficult to cleanly isolate the effects of SA research. Yet the daily approach has several benefits. It allows us to extend the sample to include all SA reports, including overnight reports, and benchmark the influence of SA research relative to media articles, brokerage research, and earnings announcements. It also facilitates comparison with Boehmer et al., 2020, who introduce the measure of retail trading and study retail trade informativeness at the daily level.

We examine the informativeness of retail order imbalances at the daily level by estimating the following panel regression:

$$\begin{aligned} Y_{i,[t+1,t+5]} = {} & \alpha + \beta_1 Retail\_OIB_{i,t} + \beta_2 Retail\_OIB_{i,t} \\ & \times Event_{i,t} + \beta_3 Retail\_OIB_{it} \times Log(Size)_{i,t} \\ & + \beta_4 Inst\_OIB_{i,t} + \beta_5 Inst\_OIB_{i,t} \times Event_{i,t} \\ & + \beta_6 Inst\_OIB_{i,t} \times Log(Size)_{i,t} + \beta_7 Event_{i,t} \\ & + \beta_8 Char_{i,t} + Day_t + \varepsilon_{i,t,} \end{aligned} \tag{6}$$

where $Y_{i,[t+1,t+5]}$ is stock *i*'s return from the close of day *t* to the close of day *t*+5 *(Stock Returns),* the sum of the Adjusted Event Sentiment Score (ESS) across all media article over the same period (*Media Article Tone*), or the number of upward forecast revisions less the number of downward forecast revisions over the same period (*Forecast Revisions*). *Retail_OIB* is the daily retail buy volume less daily retail sell volume, scaled by daily retail trading volume, and *Institutional_OIB* is the total non-retail buy volume less total non-retail sell volume, scaled by total non-retail trading volume.

$Event_{i,t}$ is a vector of event indicators: $SA_{i,t}$, $IBES_{i,t}$, $Media_{i,t}$, and $Earnings_{i,t}$. We define $SA_{i,t}$ as one if an SA research report is published between 1:30 pm on day *t*-1 and 4 pm on day *t*, and zero otherwise, in effect assuming that reports published between 1:30 and 4 pm influence retail trading on the day of publication and the day after. This assumption is motivated by the intraday evidence that retail trading remains elevated for at least five periods after report release (Fig. 1).[33] We define all other events: *IBES, Media,* and *Earnings* analogously.

Char is a vector of firm characteristics taken from BJZZ, and it includes past returns estimated over the prior week ($Ret_{i,w-1}$), the prior month ($Ret_{i,m-1}$), and the prior two-to-seven months ($Ret_{i,[m-7,m-2]}$), market capitalization (*Size*), monthly turnover (*Turnover*), volatility of daily returns (*Volatility*), and book-to-market (*BM*). We also add indicators for whether trading volume in the stock in the previous day was in the top or bottom 10% relative to the stock's trading volume in the previous fifty trading days (*High Volume* and *Low Volume*). We further include *Retail_OIB* × *Size* and *Inst_OIB* × *Size* to control for the relation between the informativeness of retail and institutional trading and firm size (BJZZ). With the exception of returns, *High Volume,* and *Low Volume,* all control variables are measured at the end of the previous year and are in natural logs, and all continuous variables are standardized to have mean zero and unit variance.

Specification (1) of Table 11 presents the results. Consistent with BJZZ, we find that retail order imbalance is a strong positive predictor of future returns on non-event days. More importantly, order flow is a considerably

[31] The coefficient on *Post_SA* × *Retail_OIB* × *Quality* using the four individual quality measures are all positive and at least marginally significant ($p < 0.10$) predictors of one-week ahead returns. The individual quality measures are less robust predictors of media (none of the four estimates are significant in isolation) and forecast revisions (two of the four predictors are significant at the 10% level).

[32] In the Internet Appendix, we conduct tests of retail trading intensity and direction that parallel the intra-day tests in Tables 3 and 4, and we find similar results (see Tables IA9 and IA10).

[33] The results are robust to extending the window (e.g., 9:30 am on day *t*-1 through 4 pm on day *t*) or shrinking the window (e.g., 4 pm on day *t*-1 through 4 pm on day *t*).

Table 10
Retail Investor Trading Informativeness Tests: Conditioning on Seeking Alpha Report Quality
In this table, we augment specification (2) in Table 5 and specifications (2) and (4) in Table 8 to include the following terms: *Retail_OIB* × *Report Quality* and *Retail_OIB* × *Post_SA* × *Report Quality*. *Report Quality* is the sum of *Academic Quality*, *Signed Return*, *Unsigned Return*, and *Comments* (all defined in Appendix A). All continuous variables are standardized. Standard errors are clustered by month, and t-statistics are reported below each estimate.

	Returns (1)	Media Article Tone (2)	Forecast Revisions (3)
Retail_OIB	0.063	1.21	0.08
	(0.72)	(2.10)	(0.47)
Retail_OIB × *Report Quality*	-0.113	-0.24	-0.05
	(-1.68)	(-0.59)	(-0.48)
Retail_OIB × *Post_SA*	-0.263	-0.45	-0.37
	(-2.30)	(-0.65)	(-1.80)
Retail_OIB × *Post_SA* × *Rep. Quality*	0.347	1.01	0.37
	(3.89)	(1.94)	(3.07)
Report Quality	-0.032	1.95	0.08
	(-0.62)	(6.69)	(1.29)
Institutional_OIB	0.186	0.62	0.30
	(1.78)	(0.99)	(1.65)
Institutional_OIB × *Post_SA*	0.216	0.21	0.16
	(1.56)	(0.26)	(0.82)
Abs $Ret_{i,t-1}$	-0.022	-2.08	-0.13
	(-0.44)	(-9.42)	(-4.06)
Abs $Ret_{i,[t-5,t-2]}$	-0.065	-2.89	-0.25
	(-0.99)	(-10.96)	(-6.22)
$Ret_{i,t-1}$	0.028	0.14	0.06
	(1.06)	(1.30)	(3.47)
$Ret_{i,[t-5,t-2]}$	0.046	-0.04	0.09
	(1.00)	(-0.20)	(2.34)
High $Volume_{i,t-1}$	0.026	3.47	0.26
	(0.37)	(6.06)	(2.73)
High $Volume_{i,[t-5,\ t-2]}$	-0.287	1.17	0.08
	(-2.24)	(1.07)	(0.45)
Low $Volume_{i,t-1}$	0.012	-2.60	-0.20
	(0.26)	(-6.89)	(-3.11)
Low $Volume_{i,[t-5,\ t-2]}$	-0.127	-4.73	-0.31
	(-1.09)	(-5.22)	(-1.51)
Media Tone / Forecast $Revisions_{[0]}$		1.99	1.81
		(3.48)	(10.21)
Media Tone / Forecast $Revisions_{[-5,-1]}$		1.41	0.57
		(3.79)	(10.98)
Media Tone / Forecast $Revisions_{[-26,-6]}$		7.20	0.82
		(12.32)	(8.56)
Observations	353,557	276,097	157,680
SA Reports	No Events	No Events	No Events
Event Period	[-5, 5]	[-5, 5]	[-5, 5]
Half Hour × Month FE	Yes	Yes	Yes
R-squared	1.17%	7.53%	13.44%

stronger predictor of future returns on days with Seeking Alpha research. Specifically, a one standard deviation increases in daily retail order imbalance (roughly 0.40) is associated with a 0.036 percentage point increase in 5-days returns on days without SA research, and a 0.125pp (0.036 + 0.089) increase on days with SA research. We find modest evidence that retail investor trade informativeness increases after other information events, yet the magnitudes are considerably weaker. For example, the incremental informativeness following media articles (0.022pp) or sell-side research (0.023pp) is roughly one quarter of the estimated effect following SA research. This finding supports the view that Seeking Alpha serves a unique role in broadening access to investment research and helping retail investors make more informed trading decisions.[34]

Specifications (2) and (3) of Table 11 present analogous tests after replacing returns with *Media Tone* and *Revisions*. As in Table 8, we also add controls for *Media Tone* or *Revisions* over day [0], days [-5, -1], and days [-26, -6]. Consistent with the intraday evidence, we find that the ability

[34] In contemporaneous work, Akbas and Subasi (2019) find evidence that the informativeness of retail trading increases following corporate news events, but they do not benchmark their findings to other information sources such as Seeking Alpha or brokerage research.

Table 11
Retail investor trading informativeness tests: daily analysis
This table presents the results from the estimation of Eq. (6):

$$Y_{i,[t+1,t+5]} = \alpha + \beta_1 Retail\_OIB_{i,t} + \beta_2 Retail\_OIB_{i,t} \times Event_{i,t} + \beta_3 Retail\_OIB_{i,t} \times Log(Size)_{i,t} + \beta_4 Inst\_OIB_{i,t} + \beta_5 Inst\_OIB_{i,t} \times Event_{i,t} + \beta_6 Inst\_OIB_{i,t} \times Log(Size)_{i,t} + \beta_7 Event_{i,t} + \beta_8 Char_{i,t} + Day_t + \varepsilon_{i,t},$$

where $Y_{i,[t+1,t+5]}$ is stock i's return from the close of day t to the close of day t+5 (*Stock Returns*), the sum of the Adjusted Event Sentiment Score (ESS) across all media article over the same period (*Media Article Tone*), or the number of upward forecast revisions less the number of downward forecast revisions over the same period (*Forecast Revisions*). $Retail\_OIB_{it}$ is the total retail buy volume for stock i on day t minus the respective sell volume, scaled by total retail trading volume. $Event_{i,t}$ is a vector of indicator variables: $SA_{i,t}$, $IBES_{i,t}$ $Media_{i,t}$, and $Earnings_{i,t}$. $SA_{i,t}$ is equal to one if an SA research report on stock i is published between 1:30 pm on day t-1 and 4 pm on day t. $IBES_{i,t}$, $Media_{i,t}$, and $Earnings_{i,t}$ indicate the release of an IBES report, media article, and earnings, and are defined analogously. *Inst_OIB* is non-retail buy volume less non-retail sell volume scaled by non-retail trading volume. *Char* is a vector of the following firm characteristics: past returns estimated over the prior week ($Ret_{i,w-1}$), prior month ($Ret_{i,m-1}$), and prior two to seven months ($Ret_{i,[m-7,m-2]}$), market capitalization (*Size*), monthly turnover (*Turnover*), volatility of daily returns (*Volatility*), book-to-market (*BM*), and indicators for whether trading volume in the stock is the top or bottom 10% relative to the stock's trading volume in the previous fifty trading days (*High Volume* and *Low Volume*). When the dependent variable is *Media Article Tone* (*Forecast Revisions*), *Char* also includes *Media Tone* (*Revisions*) calculated over daily intervals [0], [-5, -1], and [-26, -6]. In Specifications (4)–(6), we include report *Quality* and its interaction with *Retail_OIB* × *SA*. Detailed variable definitions appear in Appendix A. All continuous variables are standardized. Standard errors are clustered by month, and t-statistics are reported below each estimate.

	Stock Returns (1)	Media Article Tone (2)	Forecast Revisions (3)	Stock Returns (4)	Media Article Tone (5)	Forecast Revisions (6)
Retail_OIB	0.036	0.28	0.01	0.037	0.28	0.01
	(7.68)	(9.38)	(2.38)	(7.69)	(9.41)	(2.39)
Retail_OIB × *SA*	0.089	1.29	0.11	-0.022	0.43	-0.01
	(3.30)	(6.91)	(2.46)	(-0.45)	(1.49)	(-0.13)
Retail_OIB × *SA* × *Quality*				0.079	0.58	0.08
				(1.96)	(2.94)	(2.01)
Retail_OIB × *Media*	0.022	0.03	-0.02	0.022	0.03	-0.02
	(2.51)	(0.62)	(-1.56)	(2.51)	(0.61)	(-1.57)
Retail_OIB × *IBES*	0.023	0.48	0.01	0.023	0.47	0.01
	(1.69)	(7.01)	(0.83)	(1.68)	(6.97)	(0.81)
Retail_OIB × *Earnings*	0.066	0.07	-0.07	0.066	0.06	-0.07
	(1.45)	(0.39)	(-2.49)	(1.43)	(0.34)	(-2.53)
Retail_OIB × *Size*	-0.028	0.21	0.00	-0.028	0.21	0.00
	(-5.46)	(7.24)	(0.26)	(-5.45)	(7.24)	(0.26)
Institutional_OIB	-0.051	0.06	0.02	-0.051	0.06	0.02
	(-7.32)	(1.97)	(2.12)	(-7.32)	(1.97)	(2.12)
Institutional_OIB × *SA*	0.035	0.26	0.08	0.035	0.26	0.08
	(1.24)	(1.45)	(2.43)	(1.23)	(1.45)	(2.43)
Institutional_OIB × *Media*	0.019	-0.07	-0.03	0.019	-0.07	-0.03
	(1.77)	(-1.60)	(-2.45)	(1.77)	(-1.61)	(-2.45)
Institutional_OIB × *IBES*	0.008	0.16	-0.02	0.008	0.16	-0.02
	(0.52)	(1.84)	(-1.18)	(0.52)	(1.82)	(-1.18)
Inst_OIB × *Earnings*	0.022	-0.28	0.04	0.022	-0.28	0.04
	(0.53)	(-1.25)	(1.41)	(0.53)	(-1.25)	(1.42)
Institutional_OIB × *Size*	0.006	0.10	0.00	0.006	0.10	0.00
	(1.12)	(3.45)	(0.48)	(1.12)	(3.44)	(0.47)
$Ret_{i,w-1}$	-0.092	-0.66	0.17	-0.092	-0.66	0.17
	(-3.78)	(-10.34)	(20.01)	(-3.78)	(-10.34)	(20.01)
$Ret_{i,m-1}$	-0.040	-0.88	0.27	-0.040	-0.88	0.27
	(-1.25)	(-11.05)	(21.59)	(-1.25)	(-11.04)	(21.59)
$Ret_{i,[m-7,\ m-2]}$	-0.003	0.34	0.34	-0.003	0.34	0.34
	(-0.08)	(4.02)	(20.13)	(-0.08)	(4.04)	(20.13)
$Turnover_{i,m-1}$	-0.048	-1.53	-0.04	-0.048	-1.54	-0.04
	(-1.90)	(-17.34)	(-2.78)	(-1.91)	(-17.43)	(-2.80)
$Volatility_{i,m-1}$	0.062	-0.04	-0.03	0.062	-0.05	-0.03
	(1.53)	(-0.46)	(-1.68)	(1.52)	(-0.54)	(-1.68)
Log (Size)	-0.002	2.94	0.07	-0.002	2.94	0.07
	(-0.07)	(14.32)	(2.81)	(-0.07)	(14.31)	(2.81)
Log (BM)	0.016	0.52	-0.10	0.016	0.52	-0.10
	(0.59)	(6.64)	(-8.64)	(0.59)	(6.65)	(-8.64)
$High\ Volume_{i,d-1}$	0.198	0.92	0.01	0.197	0.91	0.01
	(6.80)	(7.94)	(0.52)	(6.80)	(7.84)	(0.51)
$Low\ Volume_{i,d-1}$	-0.125	-0.34	-0.07	-0.125	-0.33	-0.07
	(-5.31)	(-3.25)	(-4.11)	(-5.31)	(-3.23)	(-4.10)
SA	0.005	3.67	0.08	-0.010	1.76	0.08
	(0.13)	(12.99)	(2.47)	(-0.16)	(4.77)	(1.57)
Media	0.017	-1.13	-0.01	0.008	-1.13	-0.01
	(1.60)	(-11.88)	(-0.93)	(0.20)	(-11.91)	(-0.93)

(*continued on next page*)

Table 11 (*continued*)

	Stock Returns (1)	Media Article Tone (2)	Forecast Revisions (3)	Stock Returns (4)	Media Article Tone (5)	Forecast Revisions (6)
IBES	-0.025	0.51	-0.01	0.017	0.51	-0.01
	(-0.94)	(4.89)	(-0.66)	(1.60)	(4.87)	(-0.65)
Earnings	-0.072	-9.81	-0.13	-0.025	-9.83	-0.13
	(-1.30)	(-15.99)	(-3.01)	(-0.94)	(-16.02)	(-3.01)
Quality				-0.072	1.24	0.00
				(-1.30)	(6.54)	(0.08)
Media / Revisions$_{[0]}$		0.11	0.48		0.11	0.48
		(25.21)	(95.89)		(25.15)	(95.96)
Media / Revisions$_{[-5,\ -1]}$		0.05	0.12		0.05	0.12
		(14.61)	(30.72)		(14.61)	(30.73)
Media / Revisions$_{[-26,-6]}$		0.04	0.07		0.04	0.07
		(25.51)	(19.55)		(25.45)	(19.55)
Day Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,216,191	2,928,492	1,220,545	4,216,191	2,928,492	1,220,545
SA Sample	Full Sample	Full Sample	Full Sample	Full Sample	Full Sample	Full Sample
R-squared	15.57%	5.49%	10.58%	15.57%	5.49%	10.58%

of retail order imbalances to forecast *Media Tone* or *Revisions* is significantly larger after the release of SA research reports. Specifications (4)–(6) of Table 11 repeat Specifications (1)–(3) after including *Report Quality*, as defined in Section 5.5.2, and interacting *Report Quality* with *Retail_OIB* × *SA*. We continue to find that the incremental informativeness of retail trading following SA research reports is concentrated in higher quality reports. Overall, the evidence from the daily analysis is highly consistent with the intraday tests.[35]

6. Fake research reports

Investors' increasing reliance on social media for investment information creates incentives to disseminate inaccurate or misleading analysis for the purpose of price manipulation. Seeking Alpha takes steps to prevent fake research, including mandating that contributors disclose investment positions publicly and requiring that pseudonymous contributors disclose their identity to SA.[36] However, recent evidence suggests that a subset of Seeking Alpha research reports are indeed inauthentic and the market is misled by them, as evidenced by an initial reaction followed by a reversal (Kogan et al., 2020; Mitts, 2020; Dyer and Kim, 2021). Drawing on these studies, we identify potentially fake reports and examine their effects on retail trading intensity, direction, and informativeness. If many retail investors treat fake research reports as authentic, we expect to observe elevated trading as well order imbalances that are directionally consistent with report tone and predictive of short-term but not long-term returns.

We identify potentially fake research in two ways. First, we classify research reports by anonymous contributors as potentially fake (Dyer and Kim, 2021). Specifically, a contributor is anonymous if her SA bio includes none of the following: (i) a complete human name (i.e., first and last name); (ii) a human face; (iii) a tag to a company name; (iv) a link to a LinkedIn account; or (v) a link to a Twitter account[37]. We find that 17% of all reports are authored by anonymous contributors, and we classify the remaining 83% as non-fake.

Second, we classify as potentially fake all reports that exhibit linguistic characteristics indicative of deception, as captured by a low authenticity score (Pennebaker, 2011). As in Kogan et al. (2020), we rely on the Linguistic Inquiry Word Count textual algorithm (LIWC2015) from Pennebaker et al. (2015), which is designed to detect deception.[38] The evidence in Kogan et al. (2020) suggests that the relation between authenticity score and the probability of being inauthentic is non-linear. We therefore classify reports whose scores are in the bottom quintile relative to all other reports published in the same month as potentially fake (low authenticity), and the remaining reports as high authenticity.

We partition the sample into potentially fake and non-fake reports and repeat the analysis on retail trading volume (Specification 2 of Table 3), retail order imbalances (Specification 3 of Table 4), retail informativeness over a one-week holding period (Specification 2 of Table 5), and retail informativeness over holdings periods of five weeks and 12 weeks (Specifications 6 and 7 of Table 6). The results are reported in in Table 12. To test for a differential effect between fake and non-fake reports, we conduct the analyses on the full sample and include a fake report indicator. Panel A shows that anonymous reports induce a significant increase in retail trading. Fake report tone also influences the direction of retail trades, and these effects

[35] Similar to Section 5.3, we also decompose daily *Retail OIB* into *OIB Persistence, OIB Contrarian*, and *OIB Other* (a proxy for informed trading). The results of this decomposition, reported in Table IA12 (Section IA.10.4 provides more details), are consistent with the intraday decomposition results in Table 7.

[36] Contributors are also not allowed to write under multiple pseudonyms or change from one pseudonym to another. https://seekingalpha.com/page/policy_anonymous_contributors

[37] For examples of contributors with different amounts of attributable information, consider the following: Five pieces of biographic information (https://seekingalpha.com/author/donovan-jones#regular_articles); Two pieces of biographic information (https://seekingalpha.com/author/paul-lebo-cfa#regular_articles); and Zero pieces of information included in the bio: (https://seekingalpha.com/author/bull-s-run#regular_articles).

[38] See http://liwc.wpengine.com/

Table 12
Fake Seeking Alpha Reports

This table repeats the retail trading volume tests (Specification 2 of Table 3); retail order imbalance tests (Specification 3 of Table 4); one-week retail trading informativeness test (Specification 2 of Table 5), and the five-week and 12-week retail trade informativeness tests (Specifications 6 and 7 of Table 6) after partitioning the sample into reports that are more or less likely to be fake. In Panel A, our proxy for fake reports is author anonymity. We define a contributor as anonymous if her SA bio includes none of the following: (i) a complete human name (i.e., first and last name); (ii) a human face; (iii) a tag to a company name; (iv) a link to a LinkedIn account; or (v) a link to a Twitter account. All other contributors are classified as non-anonymous. In Panel B, our proxy for fake reports is the authenticity score of Pennebaker 2011. We classify a report as low (high) authenticity if its authenticity score is in the bottom 20 (top 80%) of the distribution of scores of reports issued during the same calendar month. We also test whether the coefficients for the fake reports are significantly different from non-fake reports by estimating the same regression model on the full sample and interacting the main variable of interest with a fake report indicator. Standard errors are clustered by time and t-statistics are reported below each estimate.

	Obs.	Retail Volume	Percent Retail	Retail OIB	Return 1-week	Return 5-weeks	Return 12-weeks
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Anonymous Contributors							
Anonymous	58,729	7.38	0.11	1.70	0.378	0.021	-0.056
		(3.12)	(1.08)	(3.67)	(2.48)	(0.06)	(-0.09)
Non-Anonymous	293,386	9.06	0.15	0.81	0.229	0.431	0.305
		(8.02)	(3.82)	(3.91)	(2.69)	(2.83)	(1.32)
Anonymous Interaction Term		-1.57	-0.05	0.88	0.155	-0.461	-0.342
		(-0.64)	(-0.49)	(1.74)	(0.86)	(-1.24)	(-0.49)
Panel B: Authenticity Score							
Low Authenticity	73,965	18.19	0.18	1.60	0.405	0.117	-0.578
		(9.64)	(1.96)	(3.78)	(2.91)	(0.39)	(-1.14)
High Authenticity	279,423	6.24	0.14	0.83	0.200	0.388	0.427
		(5.17)	(3.50)	(3.96)	(2.60)	(2.40)	(1.74)
Low Authenticity Interaction Term		12.05	0.04	0.78	0.232	-0.283	-1.059
		(5.73)	(0.46)	(1.65)	(1.60)	(-0.79)	(-1.91)

are similar to the effects of non-anonymous reports, as evidenced by the statistically insignificant anonymous interaction term.

The incremental informativeness of retail trading induced by anonymous reports is economically and statistically significant when the return window is one week (0.378pp) but inconsequential when the holding period is extended to five weeks (0.021pp) or 12 weeks (-0.056pp), consistent with the one-week market reaction later reversing itself. In Table IA.13 of the Internet Appendix, we formally test whether retail trading after anonymous reports is associated with significant return reversals in weeks 2 through 5. We fail to reject the null hypothesis that the informativeness of retail trading after anonymous reports is significantly different from zero; however, we do find that retail trading following anonymous reports is associated with significantly lower returns relative to non-anonymous reports (i.e., the anonymous interaction term). The evidence is consistent with non-anonymous reports being associated with more informative retail trading whereas anonymous reports cause price pressure.

Panel B of Table 12 examines reports with low and high authenticity scores, and the findings are generally similar, with one exception. The effect of low authenticity reports on retail trading is more than twice the effect of high authenticity reports, suggesting that fake research influences retail trading more than real research. In Table IA14 of the Internet Appendix, we present evidence that much of the differential effect is due to fake reports targeting firms with more opaque information environments and using language designed to project confidence and sophistication. We also find evidence that the fake news trading effect is concentrated in the first few articles by the contributor, where assessing article truthfulness may be more difficult.

7. Conclusion

We examine whether social media enhances the informativeness of retail investor trading. Our empirical strategy exploits the editorial delay between Seeking Alpha report submission and publication to identify the effect of social media on retail trading from the effects of earlier events. We use the intraday window immediately after SA report publication to estimate the level of social-network-induced retail trading and the intraday window prior to publication to estimate the counterfactual level of trading.

We find that the level of retail investor trading increases significantly during the intraday post-publication window relative to the pre-publication window, consistent with Seeking Alpha encouraging retail trading. More importantly, we document a substantial increase in the ability of retail order imbalances to predict both future stocks returns and cash flow news, consistent with SA research informing retail trading. The incremental information revealed by post-research retail trading is largely orthogonal to the information revealed by report tone and contributor investment position, consistent with retail investors actively gleaning valuable information from SA research rather than trading quickly on report sentiment. Post-publication trades are especially informative after reports authored by more capable contributors and those that attract more comments, supporting the view of retail investors as capable of sophisticated information processing. These findings suggest that social media can play a positive role in informing retail investors.

We also present evidence that speaks to potentially negative aspects of social media. A small subset of SA research reports, those authored by anonymous contributors or exhibiting linguistic attributes associated with deception, induce retail trading and order imbalances that predict returns measured over one week but not over longer windows. At a minimum, these findings suggest that anonymous reports and those with certain linguistic features deserve extra scrutiny by investors and SA management.

We acknowledge that the documented effects of Seeking Alpha research on retail traders may not generalize to other social media sites, particularly those organized much differently from Seeking Alpha. For example, it is doubtful that retail trade informativeness would similarly increase following tweets on StockTwits, which limits the character length of posts, or posts on SumZero, which limits access to professional analysts. We leave it to future research to identify features of social media that contribute to more informative retail trading. A1–A3

Acknowledgments

We thank Dan Bradley, Suzanne Chang (discussant), Huaizi Chen (discussant), Alex Chinco (discussant), Frank Heflin (discussant), Byoung-Hyoung Hwang, Eric Kelley (discussant), Petra Vokata (discussant), Ming Shou (discussant), Musa Subasi (discussant), and seminar participants at the 2019 SFS Cavalcade, the 2019 Western Finance Association Meeting, the 2019 UConn Finance Conference, the 2019 Northern Finance Association, the 2019 Midwest Finance Association, the 2019 FARS Midyear Meeting, the 2019 Smokey Mountain Finance Conference, the 2019 Hawaii Accounting Research Conference, the Second Conference on Intelligent Information Retrieval in Accounting and Finance, Auburn University, Boston College, Middle Tennessee State University, Tulane University, and the University of Kentucky for helpful comments, and we thank Yanbin Wu for excellent research assistance. We thank Seeking Alpha for providing the data and George Moriarty (Executive Editor, VP Content) for details of the platform. Part of the work was completed while Farrell was a Darden Post-Doctoral Fellow at the University of Virginia.

Funding

Jame acknowledges a research grant from University of Kentucky's Institute for Free Enterprise

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jfineco.2021.07.018.

Appendix A. Variable definitions

A1. Outcome variables

- *SA Coverage* (Table 2) – the number of Seeking Alpha contributors writing a report for a firm during the calendar year. (Source: Seeking Alpha).
- *Retail Volume* (Table 3) – the natural log of 1 + Retail Share Trading Volume. Retail Trading is estimated using the approach outlined in Boehmer et al. (2020). (Source: TAQ).
- *Percent Retail Trading* (Table 3) – retail share volume scaled by total share volume. (Source: TAQ).
- *Retail_OIB* (Table 4) – retail buy volume less retail sell volume, scaled by total retail trading volume. (Source: TAQ).
- *Ret* (Tables 5–12) – the equally-weighted market adjusted return over the subsequent five trading days.
  - In intraday tests, returns are based on the bid-ask average price at the end of half-hour t until the close of trading after five full trading days.
  - In daily tests, returns are based on the bid-ask average price at the end of the trading day until the close of trading after five full trading days.
- *Media Tone* (Tables 8–11) – the sum of the *Adjusted Event Sentiment Score* across all articles written about the firm on days *t+1* through *t+5*, where the *Adjusted Event Sentiment Score* is the *Event Sentiment (ESS) Score* of each article, after centering the variable at 0 by subtracting 50 from the ESS score reported by RavenPack. (Source: RavenPack).
- *Forecast Revisions* (Tables 8–11) – The number of upward annual and quarterly forecast revisions less the number of downward annual and quarterly forecast revision for a firm over days $t+1$ through $t+5$. (Source: IBES).

A.2. Intraday control variables (Tables 3–10 and 12)

- *Post_SA* – an indicator equal to one if the trading is measured after the release of an SA research report and zero if trading is measured prior to the release. For example, in intraday tests where the event-window spans the ten half-hours centered around the release of SA research [-5, 5], *Post_SA* equals one over the [1, 5] window and zero over the [-1, -5] window. (Source: Seeking Alpha).
- $Ret_{i,t\text{-}1}$ – the intraday return over the prior 30 minutes (-1), computed using bid-ask midpoints. (Source: TAQ).
- $Ret_{i,[t\text{-}5,t\text{-}2]}$ – the intraday return over period (-5) through (-2), where each period is 30 minutes long. Returns are computed using bid-ask midpoints. (Source: TAQ).
- *Abs* $Ret_{i,t\text{-}1}$ – the absolute value of $Ret_{i,t\text{-}1}$. (Source: TAQ).
- *Abs* $Ret_{i,[t\text{-}5,t\text{-}2]}$ – the absolute value of $Ret_{i,[t\text{-}5,t\text{-}2]}$. (Source: TAQ).
- *High* $Volume_{i,t\text{-}1}$ – an indicator equal to one if the trading volume for firm *i* in the prior 30 minutes is larger than any of the trading volumes for the same firm during the same half hour interval over the previous 9 trading days. (Source TAQ).
- *High* $Volume_{i,[t\text{-}5,t\text{-}2]}$ – an indicator equal to one if the trading volume for firm *i* over period (-5 through (-2), where each period in 30 minutes long, is larger than any of the trading volumes for the same interval over the previous 9 trading days. (Source: TAQ).
- *Low* $Volume_{i,t\text{-}1}$ – an indicator equal to one if the trading volume for firm *i* in the prior 30 min is smaller than any of the trading volumes for the same firm during the same half hour interval over the previous 9 trading days. (Source: TAQ).
- *Low* $Volume_{i,[t\text{-}5,t\text{-}2]}$ – an indicator equal to one if the trading volume for firm *i* over period (-5 through (-2), where each period is 30 minutes long, is smaller than any of the trading volumes for the same interval over the previous 9 trading days. (Source: TAQ).
- *Negative (Positive) Tone* – An indicator equal to one when the average fraction of negative (positive) words in the *Seeking Alpha* report exceeds the sample median. (Source: Seeking Alpha). We identify negative and positive words using Loughran and McDonald's (2011) list.
- *Short (Long)*– An indicator equal to one if the contributor discloses a short (long) investment position in the researched company. (Source: Seeking Alpha).
- *Composite Sentiment* – Calculated as (*Long* + *Pos. Tone*) - (*Short* + *Neg. Tone*). (Source: Seeking Alpha).
- *Institutional_OIB* – the non-retail share buy volume less the non-retail share sell volume, scaled by the non-retail share volume. Non-retail trading is signed used the Lee and Ready (1991) algorithm. When Daily Trade and Quote (DTAQ) data is available (2015–2017), the Lee and Ready (1991) algorithm as classified by WRDS. For the Monthly Trade and Quote *(MTAQ)* data sample (2007–2014), the *Interpolated Lee and Ready Algorithm* of Holden and Jacobsen (2014) is used. (Source: TAQ).
- *Media* $Tone_{[0]}$ – the sum of the *Adjusted Event Sentiment Score* across all articles written about the firm on day *t*. where the *Adjusted Event Sentiment Score* is the *Event Sentiment (ESS) Score* of each article, after centering the variable at 0 by subtracting 50 from the ESS score reported by RavenPack. (Source: RavenPack).
  - *Media* $Tone_{[-5,\ -1]}$ – the sum of the *Adjusted Event Sentiment Score* across all articles written about the firm on days *t*-1 through *t*-5.
  - *Media* $Tone_{[-26,\ -6]}$ – the sum of the *Adjusted Event Sentiment Score* across all articles written about the firm on days *t*-6 through *t*-26.
- *Forecast* $Revisions_{[0]}$ – The number of upward annual and quarterly forecast revisions less the number of downward annual and quarterly forecast revision for a firm on day *t*.
  - *Forecast* $Revisions_{[-5,\ -1]}$ – The number of upward annual and quarterly forecast revisions less the number of downward annual and quarterly forecast revision for a firm computed over days *t*-1 through *t*-5.
  - *Forecast* $Revisions_{[-26,\ -6]}$ – The number of upward annual and quarterly forecast revisions less the number of downward annual and quarterly forecast revision for a firm computed over days *t*-6 through *t*-26.
- *Academic Quality* – An indicator equal to one if the contributor's self-reported bio mentions that the contribu-

tor has any of the following: (a) a PhD, (b) an MBA, or (c) a degree from a school in the top 50 of SAT scores based on the 75th percentile, as reported by the 2015 vintage of stateuniversity.com. (Source: https://www.stateuniversity.com/rank/sat_75pctl_rank.html).

- *Unsigned Returns* – An indicator equal to one when the average market reaction to a contributor's last five reports exceeds the yearly median. Market reaction is measured as two-day absolute market-adjusted return. (Source: Seeking Alpha/CRSP).
- *Signed Returns* – An indicator equal to one if the average signed return to a contributor's last five (non-neutral) reports exceeds the yearly median. Signed returns are based on two-day market-adjusted reactions multiplied by the sign of the article, where sign is 1 (-1) for positive (negative) reports. Reports are signed using a two-step procedure. First, we classify reports with long (short) position disclosures as positive (negative). For remaining reports, we compute the tone of the report as the percentage of negative words in the report (Loughran and McDonald, 2011). We assign reports in the bottom (top) tercile of percent negative relative to the distribution of report tone on the previous day as positive (negative).
- *Comments* – An indicator equal to one when the number of comments on an SA report exceeds the yearly median. We exclude comments made more than 24 hours after report publication. (Source: Seeking Alpha).
- *Composite Report Quality* – a measure of aggregate informativeness defined as: *Academic Quality* + *Signed Return* + *Unsigned Return* + *Comments.* (Source: Seeking Alpha).
- *Anonymous* – an indicator equal to one if a contributor's bio includes *none* of the following: (i) a complete human name (i.e., first and last name); (ii) a human face; (iii) a tag to a company name; (iv) a link to a LinkedIn account; or (v) a link to a Twitter account. (Source: Seeking Alpha).
- *Non-Anonymous* – an indicator equal to one if the contributor is not *Anonymous*. (Source: Seeking Alpha).
- *Low Authenticity* – an indicator equal to one if the report is in the bottom quintile of authenticity relative to all other reports issued in the same month, where authenticity is measures using the Linguistic Inquiry Word Court (LIWC) model from Pennebaker et al. (2015). (Source: http://liwc.wpengine.com).
- *High Authenticity* – an indicator equal to one if the report is not classified as *Low Authenticity*.

A3. Daily and annual control variables (Tables 2 and 11)

- *Size* – the market capitalization computed as share price times total shares outstanding at the end of the year. (Source: CRSP).
- *Book-to-Market (BM)* – the book-to-market ratio computed as the book value of equity scaled by the market capitalization, both measured at the end of the calendar year. Negative values are deleted and positive values are winsorized at the 1st and the 99th percentile. (Source: CRSP/Compustat).
- *Volatility* – the standard deviation of daily returns during the calendar year (Source: CRSP).
- *Turnover* – the average daily turnover (i.e., share volume scaled by shares outstanding) during the calendar year.
- *Profitability* – EBITDA scaled by book value of assets, and winsorized at the 1st and the 99th percentiles. (Source: Compustat).
- $Return_{i,[m\text{-}12,\ m\text{-}1]}$ – the buy-and-hold gross return over the prior 12 months. (Source: CRSP).
    - $Ret_{i,w\text{-}1}$ – the buy-and-hold gross return over the prior one week.
    - $Ret_{i,m\text{-}1}$ – the buy-and-hold gross return over the prior one month.
    - $Ret_{i,[m\text{-}7,\ m\text{-}2]}$ – the buy-and-hold gross return over the prior two to seven months.
- *High* $Volume_{i,t\text{-}1}$ – an indicator equal to one if the firm's trading volume is in the top 10% relative to the firm's trading volume in the previous fifty days. (Source: CRSP).
- *Low* $Volume_{i,t\text{-}1}$ – an indicator equal to one if the firm's trading volume is in the bottom 10% relative to the firm's trading volume in the previous fifty days. (Source: CRSP).
- *Institutional Ownership* – the percentage of the firm's shares held by institutions at year end. (Source: Thomson Reuters Institutional Holdings S34).
- *Breadth of Ownership* – the total number of common shareholders. (Source: Compustat).
- *IBES Coverage* – the number of unique brokerage houses issuing earnings forecast for a firm during the calendar year. (Source: IBES).
- *Media Coverage* – the total number of media articles about a firm during the calendar year. The sample is limited to articles with a RavenPack relevance and novelty scores of 100. (Source: RavenPack).
- *SA* – an indicator equal to one if at least one SA research report is published between 1:30 pm on day *t*-1 and 4 pm on day *t.* (Source: Seeking Alpha).
- *Media* – an indicator equal to one if at least one media article was released between 1:30 pm on day *t*-1 and 4 pm on day *t.* (Source: Ravenpack).
- *IBES* – an indicator variable equal to one if an IBES earnings forecast or IBES investment recommendation is released between 1:30 pm on day *t*-1 and 4 pm on day *t.* (Source: IBES).
- *Earnings* – an indicator variable equal to one if earnings are announced between 1:30 pm on day *t*-1 and 4 pm on day *t.* (Source: IBES).

References

Akbas, F., Subasi, M., 2019. Corporate News Releases and the Profitability of Retail Trades. University of Illinois at Chicago Unpublished working paper.

Ahern, K.R., 2017. Information networks: evidence from illegal insider trading tips. J. Financ. Econ. 125, 26–47.

Ammann, M., Schaub, N., 2020. Do individual investors trade on investment-related internet postings? Manag. Sci. 67 (9), 5679–5702.

Asthana, S., Balsam, S., Sankaraguruswamy, S., 2004. Differential response of small versus large investors to 10-K filings on EDGAR. Account. Rev. 79, 571–589.

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

Barber, B., Odean, T., Zhu, N., 2009. Do retail trades move markets? Rev. Financ. Stud. 22, 151–186.
Barrot, J.N., Kaniel, R., Sraer, D., 2016. Are retail traders compensated for providing liquidity? J. Financ. Econ. 120, 146–168.
Bartov, E., Faurel, L., Mohanram, P., 2018. Can Twitter help predict firm-level earnings and stock returns? Account. Rev. 93, 25–57.
Battalio, R., Corwin, S., Jennings, R., 2016. Can brokers have it all? On the relation between make-take fees and limit order execution quality. J. Financ. 71, 2238–2913.
Bhattacharya, N., Cho, Y., Kim, J., 2018. Levelling the playing field between large and small institutions: evidence from the SEC's XBRL mandate. Account. Rev. 93, 51–71.
Bhushan, R., 1989. Firm characteristics and analyst following. J. Account. Econ. 11, 255–274.
Blankespoor, E., Miller, B., White, H., 2014. Initial evidence on the market impact of the XBRL mandate. Rev. Account. Stud. 19, 1468–1503.
Boehmer, E., Jones, C., Zhang, X., Zhang, X., 2020. Tracking retail investor activity. J. Financ. 76, 2249–2305.
Bradley, D., Jame, R., Williams, J, 2021. Non-Deal roadshows, informed trading, and analyst conflicts of interest. J. Financ Forthcoming.
Brown, L., Call, A., Clement, M., Sharp, N., 2015. Inside the "black box" of sell-side financial analysts. J. Account. Res. 53, 1–47.
Campbell, J., DeAngelis, M., Moon, J., 2019. Skin in the game: personal stock holdings and investors' response to stock analysis on social media. Rev. Account. Stud. 24, 731–779.
Chaudhuri, R., Ivković, Z., Pollet, J., Trzcinka, C., 2020. A tangled tale of training and talent: PhDs in institutional asset management. Manag. Sci. 66 (12), 5623–5647.
Chawla, N., Da, Z., Xu, J., Ye, M., 2017. Information Diffusion on Social Media: Does it Affect Trading, Return, and Liquidity?. University of Notre Dame Unpublished working paper.
Chen, H., De, P., Hu, J., Hwang, B.H., 2014. Wisdom of the crowds: the value of stock opinions transmitted through social media. Rev. Financ. Stud. 27, 1367–1403.
Chevalier, J., Ellison, G., 1999. Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance. J. Financ. 54, 875–899.
Cookson, J.A., Engelberg, J., Mullins, W., 2020. Echo Chambers. University of Colorado at Boulder Unpublished working paper.
Crawford, S., Gray, W., Johnson, B., Price, R., 2018. What motivates buy-side analysts to share recommendation online? Manag. Sci. 64, 2473–2972.
Duflo, E., Saez, E., 2002. Participation and investment decisions in a retirement plan: the influence of colleagues' choices. J. Public Econ. 85, 121–148.
Duflo, E., Saez, E., 2003. The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. Q. J. Econ. 118, 815–842.
Dyer, T., Kim, E., 2021. Anonymous equity research. J. Account. Res. 59, 575–611.
Farrell, M., Jame, R., Qiu, T., 2020. The Cross-Section of non-Professional Analyst Skill. University of Virginia Unpublished working paper.
Frazzini, A., Lamont, O., 2008. Dumb money: mutual fund flows and the cross-section of stock returns. J. Financ. Econ. 88, 299–322.
Gao, M., Huang, J., 2020. Informing the market: the effect of modern information technologies on information production. Rev. Financ. Stud. 33 (4), 1367–1411.
Gervais, S., Kaniel, R., Mingelgrin, D., 2001. The high-volume return premium. J. Financ. 56, 877–919.
Gomez, E., Heflin, F., Moon, J., Warren, J., 2020. Crowdsourced Financial Analysis and Information Asymmetry at Earnings Announcements. University of Georgia Unpublished working paper.
Green, T.C., Jame, R., Markov, S., Subasi, M., 2014. Broker-hosted investor conferences. J. Account. Econ. 58, 142–166.
Grennan, J., Michaely, R., 2020. FinTechs and the market for financial analysis. J. Financ. Quant. Anal. 56 (6), 1877–1907.
Han, B., Hirshleifer, D., Walden, J., 2021. Social transmission bias and investor behavior. J. Financ. Quant. Anal. Forthcoming.
Heimer, R., 2016. Peer pressure: social interaction and the disposition effect. Rev. Financ. Stud. 29, 3177–3209.
Holden, C.W., Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions. J. Financ. 69, 1747–1785.
Hvidkjaer, S., 2008. Small trades and the cross-section of stock returns. Rev. Financ. Stud. 21, 1123–1151.
Ivković, Z., Weisbenner, S., 2007. Information diffusion effects in individual investors' common stock purchases: covet thy neighbors' investment choices. Rev. Financ. Stud. 20, 1327–1357.
Jame, R., Johnston, R., Markov, S., Wolfe, M., 2016. The value of crowdsourced earnings forecasts. J. Account. Res. 54, 1077–1110.
Jame, R., 2018. Liquidity provision and the cross-section of hedge fund returns. Manag. Sci. 64, 3288–3312.
Kaniel, R., Saar, G., Titman, S., 2008. Individual investor sentiment and stock returns. J. Financ. 63, 273–310.
Kaniel, R., Liu, S., Saar, G., Titman, S., 2012. Individual investor trading and return patterns around earnings announcements. J. Financ. 67, 639–680.
Katona, Z., Painter, M., Patatoukas, P., Zeng, J., 2019. Alternative Data and Price Discovery: Evidence from Outer Space. University of California, Berkeley Unpublished working paper.
Kaustia, M., Knüpfer, S., 2012. Peer performance and stock market entry. J. Financ. Econ. 104, 321–338.
Kelley, E., Tetlock, P., 2013. How wise are crowds? Insights from retail orders and stocks returns. J. Financ. 68, 1229–1265.
Kelley, E., Tetlock, P., 2017. Retail short-selling and stock prices. Rev. Financ. Stud. 30, 801–834.
Kogan, S., Moskowitz, T., Niessner, M., 2020. Fake News in Financial Markets. Massachusetts Institute of Technology Unpublished working paper.
Kumar, A., Lee, C., 2006. Retail investor sentiment and return comovements. J. Financ. 61, 2451–2486.
Lee, Charles M., Ready, Mark J., 1991. Inferring trade direction from intraday data. J. Financ. 46, 733–746.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Financ. 66, 35–65.
Mingelgrin, D., 2000. The information content of trading activity. Doctoral dissertation, University of Pennsylvania.
Mitts, J., 2020. Short and distort. J. Leg. Stud. 49 (2), 287–334.
Nagel, S., 2012. Evaporating liquidity. Rev. Financ. Stud. 25, 2005–2039.
Ouimet, P., Tate, G., 2020. Learning from Coworkers: peer effects on individual investment decisions. J. Financ. 75, 133–172.
Ozsoylev, H., Walden, J., Yavuz, M., Bildik, R., 2014. Investor networks in the stock market. Rev. Financ. Stud. 27, 1323–1366.
Pennebaker, J., 2011. The Secret Life of Pronouns: What Our Words Say About Us. Chapter 6. Bloomsbury Press, New York.
Pennebaker, J., Booth, R., Boyd, R., Francis, M., 2015. Linguistic Inquiry and Word Count. LIWC operator's manual.
Rantala, V., 2019. How do investment ideas spread through social interaction? Evidence from a Ponzi scheme. J. Financ. 74, 2349–2389.
Reed, A., Samadi, M., Sokobin, J., 2020. Shorting in broad daylight: short sales and venue choice. J. Financ. Quant. Anal. 55 (7), 2246–2269.
Shiller, R.J., Pound, J., 1989. Survey evidence on diffusion of interest and information among investors. J. Econ. Behav. Organ. 12, 47–66.
Solomon, D., Soltes, E., 2015. What are we meeting for? The consequences of private meetings with investors. J. Law Econ. 58, 325–355.