4-13-2022

Retail investor attention and the limit order book: Intraday analysis of attention-based trading

Artem Meshcheryakov
San Jose State University, artem.meshcheryakov@sjsu.edu

Drew B. Winters
Texas Tech University

International Review of Financial Analysis 81 (2022) 101627

Contents lists available at ScienceDirect

International Review of Financial Analysis

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

Review

Retail investor attention and the limit order book: Intraday analysis of attention-based trading

Artem Meshcheryakov[a,*], Drew B. Winters[b]

[a] San Jose State University, Lucas College and Graduate School of Business, BT 855, One Washington Square, San Jose, CA 95192-0066, United States
[b] Texas Tech University, Rawls College of Business, W302, Lubbock, TX 79409, United States

ARTICLE INFO

JEL:
G12
G14

Keywords:
Retail
Investors
Attention
Limit order book
Trading
Internet
Searches
Google
Market microstructure

ABSTRACT

We are the first to examine how intraday changes in retail investor attention, measured by hourly Google searches, affect trading activity and informativeness of trades. High levels of Google search activity are followed in the next hour by more intensive trading in all stocks. The increased trading activity is initiated by retail investors as evidenced by the reduced size of new orders. After googling a company, retail investors do not become informed in the traditional sense; rather, they act as noise traders, who mistake noise for information, as their orders are picked off by truly informed traders.

1. Introduction

Investor attention is a scarce resource. Prior studies document the importance of investor recognition and attention for asset pricing (see Arbel & Strebel, 1982; Merton, 1987) and for trading (see Barber & Odean, 2008; Fang & Peress, 2009). However, most of the attention-related studies use indirect and low-frequency (daily, weekly, monthly) proxies for investor attention. Da, Engelberg, and Gao (2011) and Joseph, Wintoki, & Zhang, 2011 introduce intensity of Google searches as an innovative and direct measure of aggregate retail investor attention.[1] Employing this new measure, we study if intraday (hourly) fluctuations of investor attention affect trading activity and informativeness of trades. Specifically, we investigate whether retail attention leads to more active trading. And if yes, what kind of trading does the retail attention translate into? Conventionally, retail investors are assumed to be uninformed providers of liquidity while institutional investors are informed demanders of liquidity. However, since Google searches serve an information purpose, do retail investors become informed and start demanding liquidity after "googling" the company? Alternately, Google searchers may believe that they have become informed but instead provide liquidity for the truly informed traders. In other words, retail investors may become noise traders (Black, 1986 and Bloomfield, O'Hara, and Saar (2009)).[2]

The objective of this study is to determine what type of investor trading activity, if any, follows high levels of intraday Google searches.[3] In particular, we want to know if intensive intraday Google searches lead

* Corresponding author.
E-mail addresses: Artem.meshcheryakov@sjsu.edu (A. Meshcheryakov), Drew.winters@ttu.edu (D.B. Winters).

[1] Researchers have used Google search data to predict a wide variety of issues of interest to society, such as: influenza epidemics (Ginsberg et al., 2009), unemployment rates (Askitas & Zimmermann, 2009; Choi & Varian, 2009), consumer confidence (Choi & Varian, 2012), business performance (Fantazzini & Toktamysova, 2015), changes in housing markets (Veldhuizen, Vogt, & Voogt, 2016; Wu & Brynjolfsson, 2009) and arsons (Meshcheryakov, 2018).

[2] Black (Black, 1986, p. 529) notes that people sometimes trade on noise as if it is information.

[3] We recognize the endogenous nature of Google searches: it is possible that an unobserved news event causes both an increase in Google search activity and trading activity. However, Google searches represent an aggregate measure of attention that will reflect investor reaction to a single or multiple news and other events, even in the absence of news. Additionally, it is much easier for a trader to observe Google searches, due to their almost instant availability, then keep track of all related news.

https://doi.org/10.1016/j.irfa.2020.101627
Received 29 April 2020; Received in revised form 12 September 2020; Accepted 30 October 2020
Available online 6 November 2020

to a significant increase in trading activity. Next, we attempt to determine whether retail or institutional investors are responsible for the change in trading activity. Further, we investigate if investors become informed after googling the company's stock ticker. For that, we analyze what type of investor trading activity follows high levels of Google searches: informed trading that demands liquidity, uninformed trading that provides liquidity, or noise trading by uninformed traders that believe that they are informed.

We analyze the effect of retail investor attention for NASDAQ traded stocks of different sizes: large-cap, mid-cap, and small-cap. Using high frequency NASDAQ TotalView-ITCH trading data, we build the limit order book for 100 stocks in each size group (300 stocks in total) for a seven-week period September 18, 2017 – November 3, 2017. To conduct our analysis, we examine the limit order book (LOB) for each stock for changes following periods of high Google search activity.[4] Specifically, we use hourly intensity of Google searches and analyze the LOB in the hour following an hour with high levels of Google search activity. We chose the following hour to determine if a direct link exists between Google searches and trading under the premise that the value of information decays with the passage of time.

Existing literature provides substantial guidance on the impact of investor attention on the LOB, which we use to develop a set of testable hypotheses. We assume that an increase in investor trading activity leads to the submission of new limit and market orders, their cancelations, updates, and executions, which will change the limit order book. We find indications of more active trading following hours with high levels of Google search activity. To determine the source of increased trading activity we follow Kumar & Lee, 2006, who argue that retail traders submit smaller orders. We find that average order size decreases for all stocks following hours with high levels of Google search activity. This allows us to draw an indirect conclusion that retail investors may be responsible for the increased trading activity.[5]

To investigate whether retail investors become informed after googling the company, we follow model developed in Glosten, 1994, that demonstrates that the bid-ask spread widens as a result of liquidity consumption by informed traders. Therefore, if retail investors become informed, we expect the bid-ask spread to widen. Kaniel, Saar, & Titman, 2008 argue that retail investors are not informed and simply provide liquidity. In this case, the bid-ask spread should narrow following hours with high levels of Google search activity (see also Kyle, 1985). Using multivariate analysis, we do not find any significant changes in the bid-ask spread in analyzed stocks. To test further if retail investors act as informed traders, we analyze changes in LOBs' depth. We find that sell-side depth tends to decrease and buy-side depth tends to increase following hours with high retail attention. Besides that, we find that order imbalance, defined as the difference in depth on the buy and sell sides of the limit order book measured in shares and scaled by the total LOB depth, is significantly larger in hours after high levels of Google search activity for all analyzed stocks. Our findings are consistent with the prediction of Barber and Odean (2008) that increased attention from retail investors may lead to temporary buying pressure for a stock. Therefore, we conclude that by exerting buying pressure, retail investors engage in directional trading, which is consistent with informed trading, but also noise trading.

We test for noise trading activity following Black's (Black, 1986, p. 531) description of noise trading as "trading on noise as if it were information". This position is consistent with Shleifer & Summers, 1990 argument that demand from retail investors is not fully justified by fundamental factors. Therefore, it is likely that after googling the company retail investors may believe that they are informed; however, in reality, they trade on noise. Linnainmaa, 2010 argues that limit orders of uninformed investors should be quickly picked off by informed traders, which should decrease limit order lifespan. We find that the average order life shortens significantly following hours of heightened retail attention among all analyzed stocks. Our previous finding that retail investors shift order imbalance by submitting buy orders is more consistent with informed trading. However, the reduced limit order life span implies that retail investors trade on noise and their orders are quickly picked off by truly informed traders. This suggests that after googling the company retail investors do not become informed but act as noise traders.

This paper makes important contributions to the existing literature on investor attention and market microstructure, by combining these two fields together. Previously, data limitations did not allow the study of intraday dynamics of attention-based trading. To our knowledge, our study is the first that uses a direct and high frequency (hourly) measure of investor attention and analyzes its effect on market microstructure. Our analysis uncovers the dynamics of retail attention-based trading and its impact on stock liquidity. Our results suggest that Google searches measure retail *attention* and signal activity but does not provide information in the traditional sense of an informed trader.

2. Literature review and hypotheses development

"The most valuable commodity I know of is information."[6]

Google searches are a timely measure of retail investor attention (Da et al., 2011; Joseph, Wintoki, & Zhang, 2011). Arbel and Strebel (1982) and Merton, 1987 document the importance of investor recognition and attention for asset pricing. According to Merton, attention grabbing stocks will experience a change in demand and prices. In more recent literature, Bank, Larch, & Peter, 2011 explore if Google searches influence liquidity and returns in German stocks. They find a negative association between online search intensity and trading illiquidity and positive, albeit temporary, association between search intensity and stock returns. Based on these findings, the authors conclude that Google searches proxy for attention from uninformed investors. Dimpfl and Jank (2011) demonstrate that investor attention measured by Google searches is positively related to the Dow Jones market index realized volatility. Aouadi, Arouri, and Teulon (2013) analyze the influence of investor attention, measured by weekly intensity of Google searches, on the French stock market. They find that online searches for companies' names are correlated with trading volume and affect liquidity and volatility of French stock market. Ding and Hou (2015) find that an increase in search volume is associated with a broader shareholder base and narrower bid-ask spreads.

Drake, Roulston, & Thornock, 2012 demonstrate that increased intensity of Google Searches before earnings announcements reduces the post-announcement price movements. In agreement with them, Fricke, Fung, & Sinan Goktan, 2014 find that Google searches tend to reduce post-earnings announcement drift up to 40 days after the announcement.

Prior literature demonstrates that the market responds to changes in investor attention measured by Google searches. However, all research papers analyze weekly or daily intensity of Google Searches to measure investor attention. Due to data limitations, the implicit assumption made by the literature is that if an investor searches for information on one day, then he/she will act on the acquired information in the next day or the next week. Since the value of information decays over time, we assume that after acquiring information via googling, an investor is likely to trade on it the same day. In other words, we hypothesize that a

[4] We employ the Google Search Volume Index (GSVI), which reflects relative intensity of online searches conducted via Google.

[5] Prior literature (see among others: Da et al., 2011, Joseph, Wintoki, & Zhang, 2011, Ben-Rephael et al., 2017) demonstrates that retail and not institutional investors use Google searches to collect information. We confirm this finding using intraday data.

[6] Wall Street. Directed by Oliver Stone (1987, Twentieth Century Fox, American Entertainment Partners L.P., Amercent Films).

surge of investor attention may be associated with an overall increase of their trading activity. Therefore, we expect to see an increase in the number of new limit orders, their cancelations, updates, and executions following spikes in Google searches.

H1. *Increased investor attention (measured by hourly Google search intensity) leads to an increased trading activity in the following hour.*

If increased attention leads to more active trading, then the logical next step is to determine who is responsible for the increased trading: retail investors or institutional investors. Among others Da et al. (2011), Joseph, Wintoki, & Zhang, 2011, Ben-Rephael, Da, and Israelsen (2017) demonstrate that Google searches are associated with attention of retail investors. Joseph, Wintoki, & Zhang, 2011 argue that institutional traders can use more sophisticated systems and technologies to acquire and analyze information. Since our paper is the first one that analyzes intraday Google searches, we would like to confirm conclusions made by previous researchers that Google searches even at an hourly frequency indeed proxy for retail investor attention. According to Kumar & Lee, 2006, retail investors tend to trade smaller quantities of shares compared to institutional investors.[7] Therefore, we expect to observe a reduction of the average order size in periods following periods of high retail attention. A reduction in order size will support, although indirectly, our assumption that increased trading activity is caused by retail investors.

H2. *Increased investor attention (measured by hourly Google searches) leads to a smaller average order size in the following hour.*

While Google searches serve an informational purpose, do investors become informed after googling the company? Conventionally, retail investors are assumed to be uninformed providers of liquidity while institutional investors are informed demanders of liquidity. Kaniel, Saar, & Titman, 2008 argue that retail investors act as liquidity providers that try to benefit from institutional demand for immediacy. Accordingly, increased attention of retail investors may be associated with narrower bid-ask spreads and increased LOB depth, as a result of increased liquidity on the market.

H3.1. *Increased investor attention (measured by hourly Google search intensity) leads to a decrease of the bid-ask spread in the following hour.*

H3.2. *Increased investor attention (measured by hourly Google search intensity) leads to an increase of depth on both sides of the Limit order book in the following hour.*

However, Drake, Roulston, & Thornock, 2012 find that "When investors search for more information in the days just prior to the announcement, preannouncement price and volume changes reflect more of the upcoming earnings news and there is less of a price and volume response when the news is announced." In other words, the authors argue that googling the company may help investors to become informed, then trade and incorporate acquired information in the preannouncement price, thus reducing the post-announcement price movement. According to the model of the electronic open limit order book of Glosten and Milgrom (1985) informed investors consume liquidity by picking stale orders from the limit order book, thus widening the bid-ask spread. Therefore, increased attention of retail investors may be associated with wider bid-ask spreads and reduced LOB depth, if investors become informed.

The other possibility is after googling the company, retail investors perceive themselves as informed and start behaving as such, by demanding liquidity, when in reality they trade on noise as in Fama (1965) and Black, 1986. Kyle, 1985 and Black, 1986 call such investors "noise traders". Shleifer & Summers, 1990 also argue that demand from retail investors is not fully justified by fundamental factors. According to Fama (1965), the unsophisticated investors that trade on noise, driving the stock price away from its intrinsic value, will be arbitraged away by sophisticated traders. Linnainmaa, 2010 argues that limit orders of such uninformed investors will be quickly picked off by informed traders. Therefore, if after googling a company retail investors act as noise traders, we expect limit orders' lifespans to decrease following hours of heightened retail attention.

H3.3. *Increased investor attention (measured by hourly Google search intensity) leads to a decrease of limit orders lifespan.*

Barber and Odean (2008) find that attention from retail investors leads to temporary buying pressure for a stock. The buying pressure results in an increased number of buy orders (market and/or limit). Incoming market buy orders are executed against limit sell orders from the book, pushing the inside ask price higher, and reducing the depth of the book on the sell side. Meanwhile, incoming limit buy orders, if not crossed immediately, enter the limit order book, increasing the depth of the book on the buy side and possibly improving the inside buy price. Therefore, retail buying pressure may cause an increase of order imbalance in the limit order book: more limit buy orders get submitted and more limit sell orders get executed.

H4. *Increased retail attention (measured by hourly Google search intensity) leads to a higher order imbalance in the following hour.*

Our article is close in spirit to Fink and Johann (2014) but it is also different in several important aspects. Both papers study the response of market microstructure variables to changes in investor attention as proxied by intensity of Google searches. For such studies, the frequency of measurement of investor attention is of outmost importance. Limited by the daily frequency of their Google data when conducting the lead/lag analysis, Fink and Johann (2014) make an implicit assumption that investors will trade next day on the information they gathered today. In our paper, we assume that investors will trade on information rather sooner than later (since information decays over time) and, thus, we analyze hourly Google searches and their effect on the next hour Limit Order Book structure. Other differences are related to the sample size and keywords used to collect Google search statistics. Fink and Johann (2014) analyze German securities and use companies' names as keywords. In our study, we analyze US stocks traded on NASDAQ and use stock ticker symbols to extract Google search statistics. As we argue below, using stock ticker symbols as opposed to company names may result in a smaller sample, but with more accurate search statistics. We argue that people searching for a company's ticker symbol are more likely to look for financial information about the stock than people searching for the company name.

3. Data and methodology

We examine the effect of intraday fluctuations of investors' attention on changes in their trading activity and their informativeness during a seven-week period starting Sept 17, 2017 through November 3, 2017. To proxy for investor attention, we use the intensity of Google searches. Due to its zero cost and near instant availability, the intensity of Google searches has recently become a popular tool to estimate aggregate investor attention. The search data are presented by the Google Search Volume Index (GSVI) and are provided by the Google Trends service. The GSVI represents the number of searches conducted for a specific key word or phrase during a specified time-period in a specified geographic region scaled by the total number of searches on all topics conducted

[7] O'Hara, Yao, & Ye, 2014 demonstrate that algorithmic and high-frequency traders (HFT) routinely split their orders into smaller pieces and increasingly use odd-lot orders (orders with a number of shares of less than 100) and, thus, the size of an order may not be used as a reliable indicator of retail trading. However, there is no reason to expect that the distribution of odd-lot orders from HFT depends on Google search activity. In other words, HFT submit their sliced orders when they need to trade without following Google search activity. Therefore, the size of a HFT order should not change as a result of more intensive Google search activity.

during the same period in the same region.[8] The scaling procedure ensures that the GSVI stays in the range between 0 and 100, with zero representing the lowest search activity and 100 representing the highest. In general, GSVI represents the relative popularity of a search topic among all submitted search queries in this geographic region. Thus, we assume that an increase/decrease of the GSVI can be interpreted as an increase/decrease of interest among internet users in a particular search topic and, thus, can be used as an aggregate measure of their attention.

To extract investor search queries from the enormous volume of daily searches and to measure investor attention, we need to define the key words that investors use when they "google" for financial information about companies. Two common alternatives used in the finance literature are: the company name and the ticker symbol. Following Da et al. (2011) and Drake, Roulston, & Thornock, 2012 in our analysis we employ ticker symbols as key words. An internet user that searches for a company name may be looking for a full spectrum of information about the company: its web site address, or branch location and working hours, or other information unrelated to finance. On the other hand, an investor that searches for the ticker symbol is more likely to be interested specifically in the financial or stock related information. Therefore, employing ticker symbols as search terms is a natural choice for the purpose of our study.

We conduct our analysis separately for large-, middle- and small-cap stocks. NASDAQ divides all traded companies by market capitalization into Mega-cap, Large-cap, Mid-Cap, Small-Cap, Micro-cap, and Nano-cap.[9,10] For our analysis we select the 100 largest companies from the NASDAQ large-cap group, the 100 largest companies from the Mid-cap group, and the 100 largest companies from the NASDAQ Small-cap group. Conventionally, we exclude stocks of exchanger traded funds, real-estate investment trusts and banks from our sample. Using the ticker symbols of our sample of 300 stocks, we download hourly GSVI from Google Trends web site for each day in our sample. We constrain Google searches only to searches that originated in the US. To reduce noise, we exclude ticker symbols that have generic meanings such as TOY, BABY, FIVE, CAT, TEAM, etc. As a result, we have a time series of hourly intensity of Google searches for each stock for seven weeks.

To test hour hypothesis, we build the limit order book for each stock for each trading day. Limit order books are built using microstructure data from NASDAQ Historical TotalView-ITCH 5.0 daily files. On NASDAQ data are transferred between traders and the exchange in the form of messages. Each file contains all transmitted messages with nanosecond precision on a specific day. There are 21 different types of messages divided into the following groups: system event, stock related, add order, modify order, trade and net order imbalance indicator. Messages carry information on new order submission, execution, updates, cancelation, etc. Messages are price and time prioritized. Due to the high volume of messages transmitted each day, NASDAQ started using nanoseconds instead of milliseconds for time prioritization in 2010. Thus, every message timestamp contains a nanoseconds portion. An average trading day contains about 326 million messages: this number includes all types of messages transmitted between traders and the exchange. For this study, we keep messages generated only during NASDAQ official trading hours: 9:30 am – 4:00 pm EST. We also filter out erroneous messages with negative or zero prices, or negative or zero number of shares. From the NASDAQ trading data, we compute values of the variables of interest after each change with nanosecond precision. Since the Google data have hourly frequency, we average the values of the limit order book variables for each hour and then we merge the datasets together for each ticker symbol.[11] All observations in our final data set have hourly frequency.

To test our hypotheses, we compute values of the following LOB variables. To estimate changes in the trading activity (H1), we calculate the total number of messages[12] transmitted by traders to the exchange each hour during the official trading day: we include add order messages (types A and F), and modify order messages (types E, C, X, D, U, P, Q) . To test if retail investors are responsible for changes in the trading activity (H2), we calculate the average size of new orders in shares, the average size of canceled orders in shares, and the average size of executed orders in shares. To examine if after googling a stock investors act as liquidity providers or become informed and start demanding liquidity (H3.1, H3.2), we calculate the relative bid-ask spreads and the limit order book depth on buy and sell sides as a number of shares and a number of orders. To test whether investors' trading behavior is more consistent with noise trading (H3.3), we calculate the average order lifespan as a difference between the time the order was placed and the time it was executed or canceled. To test whether a greater order imbalance results from an increase in investor attention (H4), we estimate the relative order imbalance from the previously computed total limit order book depth on both sides.[13]

Table 1 reports descriptive statistics for each variable relevant to our analysis separately for different stock size groups. Large-cap stocks demonstrate the highest average hourly trading activity with 24,737 messages, with a significantly smaller number of messages transmitted for mid-cap and small-cap stocks, at 5877 and 3743, respectively. However, the mean number of shares per order is the lowest for large-cap stocks 137 versus 162 and 168 for mid-cap and small-cap stocks respectively. Same pattern holds for the number of shares per canceled and executed orders. This can be attributed to the activity of high frequency traders, who tend to use orders of smaller size to check for hidden liquidity (hidden orders), and then cancel the unsuccessful ones. The mean relative bid-ask spread is the narrowest for large-cap stocks at 0.0007; it is 0.0032 for mid-cap stocks and 0.0048 for small-cap stocks. Different measures of depth on both sides of the LOB demonstrate that large-cap stocks have significantly higher depth than the other two groups. The mid-cap and small-cap stocks have similar depth regardless of the measurement units. The average order life, in seconds, is the shortest for large-cap stocks. On average, the execution of limit orders takes 76.10, 199.91 and 290.79 s for large-, medium- and, small-cap stocks, respectively. The order imbalance scaled by the total depth (relative order imbalance) is by far the smallest for the large-cap stocks at −0.08. Mid-cap and small-cap stocks demonstrate very close relative order imbalances of −0.18 and − 0.19 shares, respectively. The average hourly trading volume is the highest for large-cap stocks of 127,511 shares, with significantly lower numbers of 28,155 and 17,349 shares demonstrated by mid-cap and small-cap stocks respectively.

3.1. Univariate analysis

In this paper we investigate whether intraday fluctuations of investor attention lead to changes in their trading activity and informativeness of trades. We begin with two univariate analyses of the developed hypotheses.

3.1.1. Analysis of the impact of investor attention: high investor attention versus low investor attention

In the first univariate analysis, we hypothesize that high and low

[8] For more information see https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052

[9] For more details see http://www.nasdaq.com/screening/companies-by-industry.aspx?exchange=NASDAQ&sortname=marketcap&sorttype=1

[10] Since we use NASDAQ trading data to reconstruct the limit order books, we utilize the NASDAQ's classification of companies.

[11] For each hour we compute equally weighted average values of the variables of interest.

[12] On NASDAQ data between traders and the exchange are transferred in the form of messages. We assume that increase in the number of messages sent by traders to the exchange indicates increase in traders' activity.

[13] Please refer to the Appendix for variables definitions.

Table 1
Descriptive statistics.

Variable	Large-cap stocks				Mid-cap stocks				Small-cap stocks			
	Mean	P25	Median	P75	Mean	P25	Median	P75	Mean	P25	Median	P75
Number of messages (in thousands)	24.737	8.992	16.353	30.035	5.877	2.097	4.129	7.425	3.743	1.178	2.474	4.684
Number of shares per new order	137	93	108	136	162	100	119	162	168	108	130	175
Number of shares per canceled order	109	74	87	108	123	76	92	124	121	82	100	132
Number of shares per executed order	93	65	80	102	95	62	77	102	109	69	86	110
Relative B-A spread	0.0007	0.0003	0.0004	0.0007	0.0032	0.0008	0.0014	0.0026	0.0048	0.0013	0.0022	0.0037
Number of orders on buy side	1184	246	392	791	162	103	129	190	142	86	111	167
Number of shares on buy side	228,129	32,349	55,298	147,781	46,287	15,133	22,651	38,727	41,460	13,947	22,951	47,876
Number of orders on sell side	879	214	345	777	181	101	132	182	175	87	115	181
Number of shares on sell side	272,324	35,006	81,904	209,354	108,361	17,072	32,083	80,200	103,373	16,092	31,746	108,935
Order life (seconds)	76.10	42.29	61.51	90.08	199.91	71.88	113.97	209.64	290.79	99.31	175.73	357.17
Order life until cancelation (seconds)	73.51	40.73	59.37	87.29	198.60	70.74	112.16	206.95	290.31	98.00	173.46	353.78
Order life until execution (seconds)	109.71	42.03	69.34	118.96	245.45	67.55	116.73	221.91	313.02	90.15	169.40	328.97
Relative order imbalance	−0.08	−0.32	−0.07	0.16	−0.18	−0.41	−0.13	0.03	−0.19	−0.42	−0.18	0.03
Return	−0.0042	−0.0017	0.0000	0.0017	−0.0180	−0.0026	−0.0001	0.0023	−0.0246	−0.0031	−0.0001	0.0028
Trading volume (in thousands)	127.511	27.697	57.816	131.995	28.155	5.619	12.831	30.990	17.349	2.975	7.737	18.426
Google Search Volume Index	50.47	33.00	52.00	67.00	40.90	24.00	39.00	56.00	38.79	21.00	35.00	54.00

Note: In this table we report descriptive statistics for the variables we employ to test our hypotheses.

investor attention may influence the variables of interest differently. Thus, we test for the difference in means of each analyzed variable between the hours following the highest investor attention and the hours following the lowest investor attention for each stock. This allows us to study the lead/lag relationship between the intensity of Google searches and selected LOB characteristics.[14] To identify periods of high and low investor attention, we sort the time series of hourly Google Search Volume Indices for each stock and then we divide them into quintiles, so that quintile 1 (5) contains hours with the highest (lowest) intensity of Google searches for a given stock.

For example, to test our first hypothesis, if trading activity for stock XYZ changes following high investor attention, first, we calculate the number of messages transmitted each hour by traders to the exchange for this stock. Then, by sorting Google search data for XYZ over the entire analyzed period, we identify hours with the highest and the lowest investor attention (hours that belong to quintile 1 and quintile 5 respectively). Next, we compile two time series, so time series one contains the number of transmitted messages in the hours following hours of high investor attention (from quintile 1) and time series 2 contains the number of messages transmitted to the exchange during hours following hours of the lowest investor attention (from quintile 5) for XYZ ticker symbol. Lastly, we run the difference in means test between the two time series. A statistically significant positive (negative) difference indicates that for stock XYZ the average number of messages transmitted in the hour following high investor attention is significantly greater (less) than the average number of messages transmitted in the hour following low investor attention as measured by GSVI. We interpret this as an increase in trading activity for stock XYZ following periods of high investor attention.

We run this analysis for each variable for each stock.

3.1.2. Analysis of changes in variables of interest around periods of high investor attention

In the second univariate analysis, we study changes in the variables of interest surrounding hours of high investor attention. Thus, we test for the difference in means of each analyzed variable between its value in the hour following the highest investor attention and its value in the hour right before the highest investor attention for each stock. This analysis is similar to an event study at the intraday level: we treat an hour with the highest investor attention as an event, and we study the change of variables around that event.[15] We identify periods of high investor attention as in the first univariate analysis: we sort the time series of hourly Google Search Volume Indices for each stock and divide it into quintiles, so that quintile 1 (5) contains hours with the highest (lowest) intensity of Google searches for a given stock. On a per stock basis, we compile two datasets for variable of interest: in the first (second) dataset we include values of an analyzed variable computed in the hour following (right before) an hour with high investor attention. Then, we run the difference in means test between these two datasets for each variable and for each stock, by analogy with the first univariate analysis.

Fig. 1 demonstrates the distribution of hours with highest investor attention from quintile 1 for large-, mid-, and small-cap stocks. All three groups of stocks exhibit a similar distribution of hours with the highest investor attention.[16] We can infer that investors are more actively searching for ticker symbols between 10 am and 11 am.

Note: This figure demonstrates the distribution of hours with the highest investor attention measured by Google searches (the highest values of GSVI). The distribution of hours with the lowest attention is a mirrored version of this figure.

3.2. Multivariate analysis

The univariate analysis allows us to test one variable at a time for each stock individually in each group. With multivariate analysis, we attempt to account for other unobserved stock characteristics and effects using predictive panel regressions. In Model 1 we investigate whether lagged investor attention ($GSVI_{i,t-1}$) can explain changes in the variables of interest for each hypothesis ($LOBVAR_{i,t}$).

$$Model\ 1: LOBVAR_{i,t} = \gamma_i + \alpha_t + \beta GSVI_{i,t-1} + \epsilon_{i,t}$$

[14] All variables are calculated separately for each stock in our sample and all comparisons are completed on a per stock basis.

[15] Similar to the first univariate analysis, all variables are calculated separately for each stock in our sample and all comparisons are completed on a per stock basis.

[16] The distribution of low-attention hours from quintile 5 represents the mirrored image of Figure 2 and is available upon request.

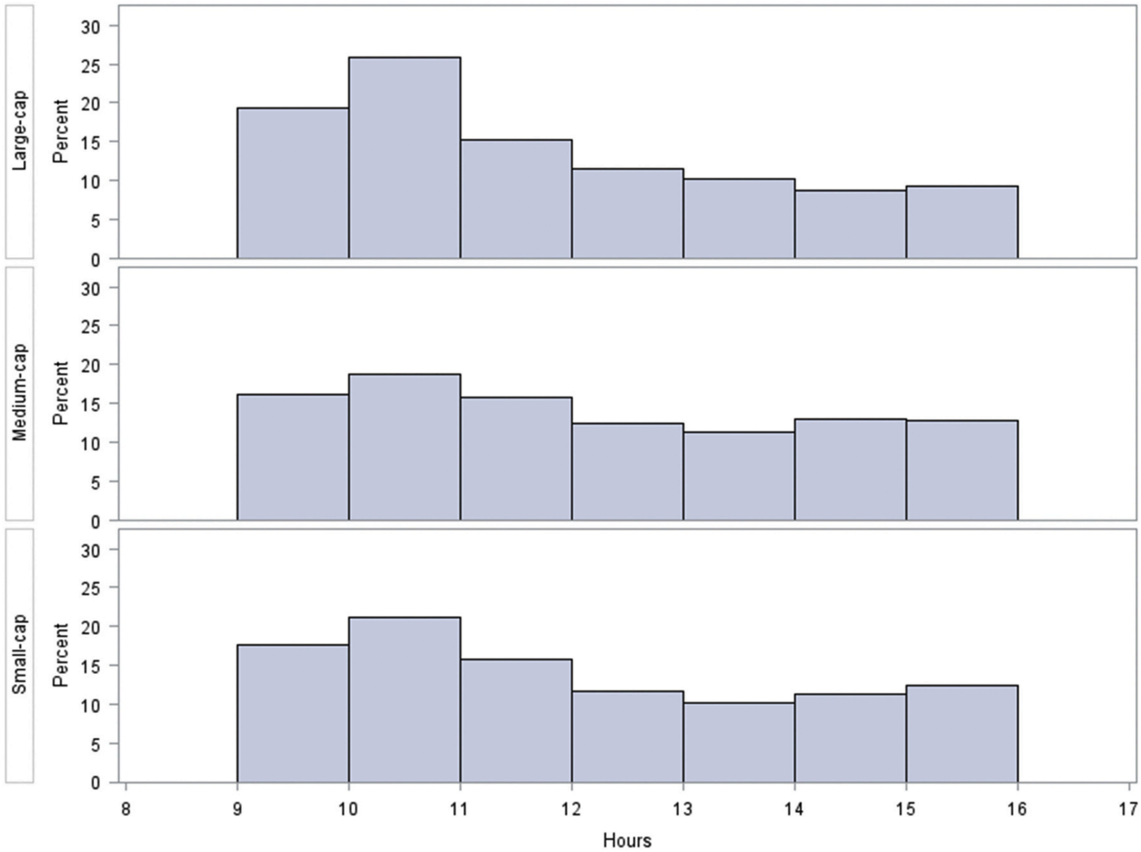

Fig. 1. Distribution of Hours with the highest investor attention.

In this model, i represents a stock, and t indexes the time; γ_i and α_t represent stock and hour fixed effects, respectively. The time fixed effect is included to address seasonality in Google search activity. We estimate Model 1 for each hypothesis separately for each stock size group, using the same set of variables as in our univariate analysis from Table 2 with some additional alternative specifications.

We recognize the simplicity of Model 1. Thus, in addition to Model 1, we estimate Model 2 with additional control variables. Since standard company's characteristics do not change significantly during the day, they would be ineffective as controls. Therefore, in Model 2 we include contemporaneous and lagged measures of trading volume (TV)[17] and contemporaneous and lagged measure of hourly rate of return.

$$Model\ 2: LOBVAR_{i,t} = \gamma_i + \alpha_t + \beta_1 GSVI_{i,t-1} + \beta_2 TV_{i,t} + \beta_3 TV_{i,t-1} + \beta_4 Ret_{i,t} + \beta_4 Ret_{i,t-1} + \epsilon_{i,t}$$

In this model, i represents a stock, and t indexes the time; γ_i and α_t represent stock and hour fixed effects, respectively. The time fixed effect is included to address seasonality in Google search activity. We estimate Model 2 for each hypothesis separately for each stock size group, using the same set of variables as in our univariate analysis.

4. Empirical results

4.1. Univariate analysis

4.1.1. Comparison of the effect of high and low investor attention

To investigate our hypotheses in a univariate setting, we conduct the difference in means test for each variable of interest for each stock between its mean value measured in hours following the hours of high investor attention and its mean value in hours following the hours of low investor attention. We aggregate results based on the significance and the sign of the difference in their respective means: we count the difference as significantly positive if the mean value of a variable following high Google search activity is significantly greater than its mean value following low Google search activity at the 10% level. The positive difference indicates that the respective LOB variable tends to increase after high Google search activity versus low Google search activity. We count the difference as significantly negative if the mean value of the LOB characteristic following high search activity is less than its mean value following low search activity at the 10% level. The negative difference indicates that the respective LOB variable tends to decrease after high Google search activity versus low Google search activity. Table 2 reports the percentage of stocks that demonstrate statistically significant difference in means of variables of interest grouped by the market cap category and the sign of the difference. The table is organized with respect to the hypotheses we test.

[17] To proxy for trade volume, we calculate the number of shares traded in each hour for each stock.

Table 2
Difference in means test for LOB characteristic calculated following an hour with high Google search intensity and an hour with low Google search intensity.

Variable	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
	Significant Positive difference	Significant Negative difference	Significant Positive difference	Significant Negative difference	Significant Positive difference	Significant Negative difference
Panel A: Hypothesis 1						
Number of messages	53.61%	3.09%	18.18%	0.00%	27.54%	5.80%
Panel B: Hypothesis 2						
Number of shares per new order	3.09%	67.01%	2.60%	41.56%	7.25%	46.38%
Number of shares per canceled order	3.09%	55.67%	1.30%	37.66%	1.45%	43.48%
Number of shares per executed order	3.09%	43.30%	1.30%	24.68%	2.90%	14.49%
Panel C: Hypothesis 3.1						
Relative Bid-Ask spread	3.09%	91.75%	3.90%	68.83%	2.90%	57.97%
Panel D: Hypothesis 3.2						
Number of orders on buy side	71.13%	2.06%	53.25%	3.90%	53.62%	1.45%
Number of shares on buy side	41.24%	6.19%	28.57%	2.60%	31.88%	1.45%
Number of orders on sell side	55.67%	5.16%	50.65%	3.90%	52.17%	2.90%
Number of shares on sell side	37.11%	14.43%	24.68%	9.09%	24.64%	1.45%
Panel E: Hypothesis 3.3						
Order life (total)	2.06%	46.39%	11.69%	15.58%	15.94%	8.70%
Panel F: Hypothesis 4						
Relative Order imbalance	31.96%	24.74%	36.36%	15.84%	30.44%	17.39%

Note: In this table we report results of the difference in means test for each variable. The reported number indicates the percentage of stocks with statistically significant positive and negative differences. For each stock we ranked its hourly GSVI observations into quintiles: quintile 1 contains the highest attention hours (hours with highest GSVI), quintile 5 contains the lowest attention hours (hours with lowest GSVI). Then, for each stock and for each LOB variable we create two datasets: the first dataset contains value of respective variable calculated in the hour that follows the hour with highest investor attention from quintile 1. The second dataset contains values of the same variable calculated in the hour that follows the hour with the lowest investor attention from quintile 5. In the next step, we conduct the difference in means test between these two datasets for each LOB characteristic for each stock. We aggregate results based on the statistical significance and the sign of the difference in their respective means: we count the difference as significantly positive if the mean value of LOB variable following high Google search activity is significantly greater than its mean value following low Google search activity. Significantly positive difference indicates that the value of the respective LOB characteristic tends to increase following high Google search activity versus low search activity periods. We count the difference as significantly negative if the mean value of the LOB variable following high search activity is significantly less than its mean value following low search activity. Significantly negative difference indicates that the value of the respective LOB characteristic tends to decrease following high Google search activity versus low search activity periods. The analysis is conducted at the stock level, then the results are aggregated.

With H1 we test if trading activity changes in response to investor attention. To measure trading intensity, we compare the average number of messages transmitted during the hour following high investor attention to the number of messages transmitted during the hour following low investor attention. In Panel A of Table 2 we observe that 53.61% of large-cap stocks have a significantly positive difference in means. It indicates that the number of messages transmitted in the hours following high investor attention is significantly greater than the number of messages transmitted in the hours following low investor attention. In other words, 53.61% of large-cap stocks experience significant increase in trading activity following intense Google searches. Only 3.09% of large-cap stocks exhibit a significant decrease in trading activity. The results are similar, albeit less pronounced for mid-cap and small-cap stocks. The 18.18% of mid-cap stocks and 27.54% of small-cap stocks demonstrate significantly positive difference in means, indicating an increase in trading activity, with 0.00% and 5.80% of mid-cap and small-cap stocks demonstrating a fall in trading activity, respectively. These results suggest that retail investors tend to trade more actively in stocks of large, well-known companies following high levels of Google search activity than in stocks of mid- and small-cap firms. Overall, we conclude that googling the company's ticker leads to an increase of trading activity for large stocks.

With H2, we investigate whether retail or institutional investors are responsible for the increased trading activity. Findings in the prior literature indicate that retail investors are more likely to utilize Google to collect information. We would like to confirm this finding for intraday data. Kumar & Lee, 2006 suggest that retail investors tend to trade smaller quantities of shares compared to institutional investors. Following Kumar & Lee, 2006, we hypothesize that if retail investors are responsible for the increased trading activity, the average order size will drop following the hours with high investor attention.

To examine order size, we calculate an average order size measured in a number of shares. Results in Panel B of Table 2 reveal that 67.01%, 41.56%, and 46.38% of large-cap, mid-cap, and small-cap stocks, respectively, demonstrate a significant negative difference in the mean number of shares per new order. This finding indicates, that the size of a new order measured in shares significantly drops in a large percentage of stocks in each size group following hours of high investor attention. The effect is the strongest for large-cap stocks: 67.01% of them experience a decline in order size. This finding suggests that the increased trading activity following high attention hours is consistent with retail trading, since the size of new submitted orders significantly decreases across the board.

For robustness, we also examine the size of canceled orders and executed orders. The results are qualitatively similar to the results on the size of new orders, but weaker. However, canceled and executed orders may have been placed before the high search activity hours and therefore should be less responsive.

With the next set of hypotheses (3.1 and 3.2) we test whether retail investors act as uninformed providers of liquidity following hours of high investor attention. If retail investors act as uninformed providers of liquidity, we expect the bid-ask spread to narrow and the limit order book depth to increase on both sides. Alternatively, retail investors may become informed and start demanding liquidity following increase in their google search activity. If investors indeed become informed, then we expect the bid-ask spread to widen and limit order book depth to reduce on both buy and sell sides.

The results reported in Panel C of Table 2 demonstrate that the relative bid-ask spread has a statistically significant negative difference for 91.75%, 68.83% and 57.97% of large-cap, mid-cap, and small-cap stocks respectively. The negative difference indicates that the relative quoted bid-ask spread is lower in the hour following high retail attention than in the hour following low retail attention measured by Google searches. In other words, after investors search more for stocks' ticker symbols online, the relative bid-ask spread narrows in the following hour. For robustness, we also analyze the change of the quoted bid-ask spread and find a similar pattern. The quoted spread is significantly lower in the hours that follow high retail attention than in hours that follow low retail attention for 91.75% of large-cap stocks, 66.23% of mid-cap stocks, and 59.42% of small-cap stocks. These results are consistent with the results on the relative quoted spread. Only about 3% of stocks in each size group demonstrate increased relative bid-ask spread.

Panel D of Table 2 reports percentage of stocks in each group that demonstrate statistically significant change in depth in hours following high investor attention versus low investor attention. Results exhibit significantly positive difference across all depth measures and across all size groups. A positive difference indicates that depth measured in a number of shares or number of orders tends to increase following hours of high Google searches compared to hours with low Google searches. A relatively small number of stocks demonstrate a significant negative difference in means of the depth variables.

Overall, our findings are consistent with Kaniel, Saar, & Titman, 2008, who argue that retail investors act as uninformed liquidity providers.

In H3.3 we test if order lifespan decreases following increased retail attention. We summarize the results of the univariate test in Panel E of Table 2. The difference in order's life is significantly negative for 46.39%, 15.58%, and 8.70% of large-cap, mid-cap, and small-cap stocks, respectively. In other words, the order life becomes shorter following hours of high retail attention versus hours of low retail attention hours. However, this relationship holds mainly for large stocks. The results for mid-cap and small-cap stocks are inconclusive due to a comparable percentage of stocks with significantly positive difference.

With H4, we test Barber and Odean's (2008) prediction that retail investor attention may exert a temporary buying pressure on a stock. The results presented in Panel F of Table 2 exhibit that during hours that follow hours of increased retail attention the order imbalance tends to be higher for 31.96%, 36.36%, and 30.44% of large-, mid-, and small-cap stocks, respectively, than during hours that follow periods of low retail attention. It means that an increased retail attention leads to some buying pressure and our results support the H4. The order imbalance analysis serves as an additional confirmation that retail investors participate in the market more as liquidity providers.

The results of the first set of univariate tests can be summarized as follows. Increased investor attention leads to higher trading activity, which is expressed as an increased number of transmitted messages. At the same time, the order size decreases, implying that retail investors are responsible for the increased trading activity. Moreover, the increased investor activity leads to a reduction in the bid-ask spread. This narrowing of the bid-ask spread, coupled with the increased depth on both buy and sell sides of the limit order book, are consistent with retail investors acting as uninformed providers of liquidity. Order imbalance increases, consistent with an increased buying pressure in these stocks. Lastly, the order life shortens in hours that follow increased retail attention.

Our univariate evidence suggests that high Google search activity is followed closely by trading activity that provides liquidity to the market. Providing liquidity is consistent with uninformed trading as informed traders demand liquidity. Our results suggest that retail trading following spikes in Google search activity is not informed trading as defined in academic research.

4.1.2. Analysis of periods of high investor attention in the event-study setting

In our second univariate analysis, we conduct the difference in means test for each variable of interest for its values measured in hours surrounding hours of high investor attention: in the hour after and in the hour right before an hour of high investor attention. We aggregate results based on the significance and the sign of the difference in their respective means. We count the difference as significantly positive if the mean value of a variable following high Google search activity is significantly greater (at the 10% level) than its mean value right before high Google search activity. We count the difference as significantly negative (at the 10% level) if the mean value of the LOB characteristic following high search activity is less than its mean value right before high search activity. Table 3 reports the percentage of stocks with significantly different means grouped by the market cap category and the sign of the difference. The table is organized by analogy with Table 2.

In Panel A of Table 3 we observe that 70.10% of large-cap stocks have a significantly positive difference in means. It indicates that the number of messages transmitted in the hours following high investor attention is significantly greater than the number of messages transmitted in the hours right before high investor attention. In other words, 70.10% of large-cap stocks experience significantly increase in trading activity following periods of intense Google searches. Only 2.06% of large-cap stocks exhibit a negative difference in means. The results are similar, albeit less pronounced for mid-cap and small-cap stocks. 18.18% of mid-cap stocks and 27.54% of small-cap stocks demonstrate significantly positive difference in means, indicating an increase in trading activity, with 2.60% and 1.45% of mid-cap and small-cap stocks demonstrating a fall in trading activity, respectively. These results suggest that retail investors tend to trade more actively in stocks of large, well-known companies following high levels of Google search activity than in stocks of mid- and small cap firms.

With H2, we investigate whether retail or institutional investors are responsible for the increased trading activity. To examine order size, we calculate an average order size measured in a number of shares. Results in Panel B of Table 3 reveal that 43.30%, 44.16%, and 57.97% of large-cap, mid-cap, and small-cap stocks, respectively, demonstrate a significant negative difference in means. This finding indicates, that the size of a new order measured in shares significantly drops in the hour following high investor attention compared to an hour right before in a large percentage of stocks in each size group. The results are qualitatively similar when we test change in size of canceled and executed orders.

With Hypotheses 3.1 and a 3.2 we test whether retail investors act as uninformed providers of liquidity following hours of high investor attention. If retail investors act as uninformed providers of liquidity, we expect the bid-ask spread to narrow and the limit order book depth to increase on both sides.

The results reported in Panel C of Table 3 demonstrate that the relative bid-ask spread has a statistically significant negative difference for 98.97%, 98.70% and 98.55% of large-cap, mid-cap, and small-cap stocks respectively. The negative difference indicates that the relative quoted bid-ask spread is lower in the hour following high retail attention than in the hour right before high retail attention measured by Google searches. In other words, after investors search more for stocks' ticker symbols online, the relative bid-ask spread narrows in the following hour. For robustness, we also analyze the change of the quoted bid-ask spread and find a similar pattern.

The difference in means test of the limit order book depth variables, reported in Panel D of Table 3 demonstrate a significant positive difference across all depth measures and across all size groups. A positive difference indicates that depth measured in a number of shares or number of orders tends to increase following hours of high Google

Table 3
Difference in means test for LOB characteristic calculated following an hour with high Google search intensity and a prior hour.

Variable	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
	Significantly Positive difference	Significantly Negative difference	Significantly Positive difference	Significantly Negative difference	Significantly Positive difference	Significantly Negative difference
Panel A: Hypothesis 1						
Number of messages	70. 10%	2.06%	18.18%	2.60%	27.54%	1.45%
Panel B: Hypothesis 2						
Number of shares per new order	17.53%	43.30%	7.79%	44.16%	2.90%	57.97%
Number of shares per canceled order	17.53%	42.27%	9.09%	53.25%	1.45%	66.67%
Number of shares per executed order	8.25%	47.42%	5.20%	35.07%	1.45%	30.44%
Panel C: Hypotheses 3.1						
Relative Bid-Ask spread	0.00%	98.97%	0.00%	98.70%	0.00%	98.55%
Panel D: Hypotheses 3.2						
Number of orders on buy side	100.00%	0.00%	94.81%	0.00%	92.75%	0.00%
Number of shares on buy side	86.60%	1.03%	85.71%	0.00%	85.51%	0.00%
Number of orders on sell side	94.85%	0.00%	97.40%	0.00%	94.20%	0.00%
Number of shares on sell side	68.04%	2.06%	77.92%	1.30%	82.61%	2.90%
Panel E: Hypothesis 3.3						
Order life (total)	67.01%	1.03%	55.84%	0.00%	60.87%	0.00%
Panel F: Hypothesis 4						
Relative order imbalance	29.90%	6.19%	19.48%	11.69%	15.94%	4.35%

Note: In this table we report results of the difference in means test for each variable. The reported number indicates the percentage of stocks with statistically significant positive and negative differences. For each stock we ranked its hourly GSVI observations into quintiles: quintile 1 contains the highest attention hours (hours with highest GSVI), quintile 5 contains the lowest attention hours (hours with lowest GSVI). Then, for each stock and for each LOB variable we create two datasets: the first dataset contains value of respective variable calculated in the hour that follows the hour with highest investor attention from quintile 1. The second dataset contains values of LOB characteristics calculated in the hour just before the hour with the highest investor attention (from quintile 1). In the next step, we conduct the difference in means test between these two datasets for each LOB characteristic for each stock. We aggregate results based on the statistical significance and the sign of the difference in their respective means: we count the difference as significantly positive if the mean value of LOB variable following high Google search activity is significantly greater than its mean value in the prior hour. Significantly positive difference indicates that the value of the respective LOB characteristic tends to increase following high Google search activity versus prior hour. We count the difference as significantly negative if the mean value of the LOB variable following high search activity is significantly less than its mean value in the prior hour. Significantly negative difference indicates that the value of the respective LOB characteristic tends to decrease following high Google search activity versus prior hour. The analysis is conducted at the stock level, then the results are aggregated.

searches. These findings are completely in line with results of our first univariate analysis.

Linnainmaa, 2010 argues that orders of uninformed retail investors will be quickly picked off by informed traders. The results summarized in Panel E of Table 3 demonstrate that the difference in order's life is significantly positive for 67.01%, 55.84%, and 60.87% of large-cap, mid-cap, and small-cap stocks, respectively. In other words, the order life tends to get longer after hours following high retail attention. This finding is inconsistent with our expectations and results of the univariate analysis 1.

With Hypothesis 4, we test Barber and Odean's (2008) prediction that retail investor attention may exert a temporary buying pressure on a stock. The results presented in Panel F of Table 3 show that during hours that follow hours of increased retail attention the order imbalance tends to be higher for 29.90%, 19.48%, and 15.94% of large-, mid-, and small-cap stocks, respectively, than during hours right before high retail attention. It means that an increased retail attention leads to some buying pressure for large-cap stocks and is less pronounced in mid- and small- cap stocks. These findings provide strong support of the hypothesis 4 but mainly for large-cap stocks.

Overall, results of the second univariate analysis are generally in line with findings produced by the first univariate analysis. The one exception is order lifespan. Our univariate analyses provide snapshots of one dimension of the limit order book at a time. It is done one stock at a time with no controls for the timing of the search activity. Accordingly, in the next section we conduct cross-sectional multivariate analysis.

4.2. Multivariate analysis of response of LOB variables to changes in investor attention

We estimate Model 1 and Model 2 for each hypothesis using the panel data compiled individually for each stock size group. To test our hypotheses, we employ the same set of variables as in our univariate analysis. The estimation results are summarized in Table 4. Table 4 is organized by analogy with Tables 2 and 3 and contains same panels. We report results from 78 panel regressions: two regressions (Model 1 and Model 2) for each LOB variable (we analyze 13 variables) and for our three stock size groups – large-cap, mid-cap, and small-cap stocks. In each regression the variable of interest is the lagged Google Search Volume Index ($GSVI_{t-1}$). Each regression is estimated with time and stock fixed effects to address possible seasonality and unobserved stock characteristics, however their coefficient are not reported for brevity. Standard errors are clustered by stock.

We begin with tests of Hypothesis 1, to determine if an increase in intensity of Google searches leads to an increase of trading activity, measured by the number of messages transmitted to the exchange every hour. The results are reported in Table 4 Panel A. The Model 1 estimated coefficients for lagged Google search activity are positive and highly statistically significant for all three groups of stocks: 0.0187, 0.0048, and 0.0118 for large-, mid-, and small-cap stocks, respectively. Model 2 produces coefficients of 0.0136, 0.0033 and 0.0063 for large- mid- and small-cap stocks respectively. All coefficients are positive and highly statistically significant. This finding suggests that an increased volume of Google searches tends to be followed by a surge in the number of

Table 4
Multivariate analysis of effect of investor attention on selected LOB variables (full sample).

Dependent variable	Independent Variables	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
		Model 1 (*N* = 23,668)	Model 2 (*N* = 16,975)	Model 1 (*N* = 18,788)	Model 2 (*N* = 13,475)	Model 1 (N = 16,836)	Model 2 (*N* = 12,075)
Panel A: Hypothesis 1							
Trading activity (Number of messages transferred to the exchange)	*Intercept*	24.0585***	−36.3779***	7.1197***	−3.8148***	3.3761***	0.2230***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.4358)
	$GSVI_{t-1}$	0.0187***	0.0136***	0.0048***	0.0033**	0.0118***	0.0063***
		(<0.0001)	(0.0024)	(0.0005)	(0.0171)	(<0.0001)	(<0.0001)
	TV_t		11.0042***		1.6109***		0.8777***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		0.8547***		0.5968***		0.3308***
			(0.0010)		(<0.0001)		(<0.0001)
	Ret_t		31.5595		18.0525***		−0.3153
			(0.2387)		(0.0003)		(0.9220)
	Ret_{t-1}		10.1922***		0.3794		0.3177**
			(0.0002)		(0.1177)		(0.0272)
	R-Squared	0.7794	0.8321	0.7423	0.8083	0.7604	0.8109
Panel B: Hypothesis 2							
Number of shares per new order	*Intercept*	149.7574***	0.1264***	1376.711***	1.4648***	205.7082***	0.1648***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0018	0.000002	−0.0649*	−0.0001	−0.2296***	−0.0002***
		(0.9143)	(0.7003)	(0.0619)	(0.2813)	(0.0087)	(0.0021)
	TV_t		0.0065***		0.0087***		0.0124***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−0.0021*		−0.0063***		0.0031*
			(0.0704)		(<0.0001)		(0.0994)
	Ret_t		−0.0277		−0.0009		0.0304
			(0.8196)		(0.9953)		(0.8966)
	Ret_{t-1}		0.0291**		−0.0019		−0.0129
			(0.0203)		(0.7953)		(0.2164)
	R-Squared	0.8254	0.8494	0.7920	0.8092	0.2758	0.3049
Number of shares per canceled order	*Intercept*	114.9748***	0.0885***	1117.96***	1.2169***	141.4973***	0.0902***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.00204	0.00001	−0.0412	−0.0001**	−0.1172*	−0.0001***
		(0.8870)	(0.6570)	(0.1472)	(0.0626)	(0.0753)	(0.0098)
	TV_t		0.0056***		0.0074***		0.0136***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−0.0006		−0.0039***		0.0031**
			(0.5500)		(<0.0001)		(0.0177)
	Ret_t		−0.0647		0.0277		0.0892
			(0.5256)		(0.8152)		(0.5737)
	Ret_{t-1}		0.0212**		−0.0053		−0.0194***
			(0.0440)		(0.3539)		(0.0062)
	R-Squared	0.8149	0.8448	0.7931	0.8323	0.2180	0.3154
Number of shares per executed order	*Intercept*	99.4824***	0.0196***	638.5994***	0.5812***	116.0131***	−0.0270***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.0159)	(0.4746)
	$GSVI_{t-1}$	−0.0048	0.00001	−0.02422	0.0001*	−0.1320	−0.0003*
		(0.4944)	(0.5775)	(0.2737)	(0.0944)	(0.4073)	(0.0554)
	TV_t	0.8885	0.0194***		0.0235***		0.0576***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−0.0050***		−0.0034***		−0.0129***
			(<0.0001)		(<0.0001)		(0.0002)
	Ret_t		0.1356***		0.1624*		0.0527
			(0.0034)		(0.0879)		(0.9011)
	Ret_{t-1}		−0.0067		0.0041		0.0198
			(0.1638)		(0.3705)		(0.2952)
	R-Squared		0.9085	0.6788	0.7192	0.1334	0.1421
Panel C: Hypotheses 3.1							
Relative bid-ask spread	*Intercept*	0.00001	0.0003***	0.00051	0.0026***	0.00187***	0.0021***
		(0.9278)	(<0.0001)	(0.3948)	(<0.0001)	(0.0097)	(<0.0001)
	$GSVI_{t-1}$	0.00001	0.0000001	0.00001	0.0000001	0.00001	0.0000001
		(0.8848)	(0.9442)	(0.1439)	(0.9781)	(0.9121)	(0.9369)
	TV_t		0.000001***		0.000001***		−0.0001***
			(0.0027)		(0.0013)		(0.0002)
	TV_{t-1}		0.0000001		0.0000001		0.0000001***
			(0.8605)		(0.3598)		(0.0081)
	Ret_t		0.0004		−0.0028**		0.0033**
			(0.6369)		(0.0284)		(0.0269)
	Ret_{t-1}		0.0000001		0.0000001		0.0000001
			(0.9294)		(0.4864)		(0.7444)
	R-Squared	0.3271	0.3543	0.5379	0.8113	0.6167	0.8315
Panel D: Hypotheses 3.2							
Number of orders on buy side	*Intercept*	508.4975***	403.7904***	368.3914***	363.1627***	124.9440***	120.9076***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	0.2583***	−0.1020	0.0447***	0.0317***	0.0416***	0.0243**

(*continued on next page*)

Table 4 (*continued*)

Dependent variable	Independent Variables	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
		Model 1 (*N* = 23,668)	Model 2 (*N* = 16,975)	Model 1 (*N* = 18,788)	Model 2 (*N* = 13,475)	Model 1 (N = 16,836)	Model 2 (*N* = 12,075)
		(0.0042)	(0.2176)	(<0.0001)	(0.0051)	(0.0013)	(0.0271)
	TV_t		2.4804		1.5984***		0.8546**
			(0.6170)		(<0.0001)		(0.0334)
	TV_{t-1}		17.4195***		2.8533***		2.4686***
			(0.0003)		(<0.0001)		(<0.0001)
	Ret_t		3709.2380***		341.8298***		610.1945***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_{t-1}		−146.2970***		2.7774		2.0137
			(0.0041)		(0.1626)		(0.3520)
	R-Squared	0.9860	0.9923	0.9591	0.9617	0.9158	0.9186
Number of shares on buy side	*Intercept*	108.1446***	69.7801***	327.4255***	339.9020***	23.8598***	17.8476***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0240	−0.0632	0.0335***	0.0251*	0.0447***	0.0352***
		(0.4240)	(0.1778)	(<0.0001)	(0.0139)	(<0.0001)	(0.0011)
	TV_t		2.6219		1.0744***		0.5414**
			(0.2221)		(0.0003)		(0.0462)
	TV_{t-1}		5.0125**		1.3872***		1.8177***
			(0.0156)		(<0.0001)		(<0.0001)
	Ret_t		1294.3490***		102.6073***		187.3535***
			(<0.0001)		(0.0053)		(<0.0001)
	Ret_{t-1}		11.0021		1.4716		1.0081
			(0.6183)		(0.4112)		(0.4907)
	R-Squared	0.9692	0.9703	0.9379	0.9407	0.8618	0.8655
Number of orders on sell side	*Intercept*	348.7269***	416.7776***	833.4336***	844.8495***	163.1487***	164.5801***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	0.0006	−0.1848	−0.0386***	−0.0586***	−0.0618***	−0.0938***
		(0.9948)	(0.1083)	(0.0004)	(<0.0001)	(<0.0001)	(<0.0001)
	TV_t		−4.4623		0.8574***		−0.5339*
			(0.5179)		(0.0231)		(0.0649)
	TV_{t-1}		−12.5128*		1.7602***		0.2919
			(0.0603)		(<0.0001)		(0.3038)
	Ret_t		−3003.6800***		−335.1540***		−422.7360***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_{t-1}		−124.1870*		−1.6930		−0.5779
			(0.0801)		(0.4535)		(0.7106)
	R-Squared	0.9628	0.9660	0.9763	0.9785	0.9874	0.9906
Number of shares on sell side	*Intercept*	110.8331***	145.1910***	1597.7730***	1620.4660***	47.4970***	57.9232***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0456	−0.0637	−0.0468	−0.0838**	−0.1242***	−0.1403***
		(0.1963)	(0.1294)	(0.1217)	(0.0218)	(<0.0001)	(<0.0001)
	TV_t		−2.1154		2.2875**		−2.6707***
			(0.4011)		(0.0328)		(<0.0001)
	TV_{t-1}		−6.0809**		2.5585**		−1.1463**
			(0.0124)		(0.0126)		(0.0393)
	Ret_t		−1130.0200***		−238.1750*		−112.2410*
			(<0.0001)		(0.0707)		(0.1004)
	Ret_{t-1}		−34.0911		−9.8046		−4.2488
			(0.1882)		(0.1263)		(0.1638)
	R-Squared	0.9843	0.9851	0.9523	0.9270	0.9762	0.9770
Panel E: Hypothesis 3.3							
Order life	*Intercept*	114.3668***	175.3102***	586.2028***	958.7084***	556.8672***	810.4372***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0950***	−0.1034***	−0.4186***	−0.1891*	−0.4559***	−0.3696***
		(<0.0001)	(<0.0001)	(<0.0001)	(0.0631)	(<0.0001)	(0.0089)
	TV_t		5.6075***		−6.5059**		−16.3728***
			(<0.0001)		(0.0294)		(<0.0001)
	TV_{t-1}		−21.4152***		−62.1706***		−68.2507***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_t		157.0466		579.1305		−610.5970
			(0.1991)		(0.1150)		(0.1542)
	Ret_{t-1}		−22.2245*		41.2892**		32.6586*
			(0.0783)		(0.0210)		(0.0880)
	R-Squared	0.4833	0.5009	0.4891	0.5581	0.4564	0.5009
Order life until execution	*Intercept*	64.2018***	−19.3559	1404.7520***	2814.2380***	535.5054***	1244.9250***
		(0.0001)	(0.4249)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0727	−0.0132	−0.5156***	−0.2609	−0.5501**	−0.0822
		(0.1405)	(0.8425)	(0.0076)	(0.2799)	(0.0129)	(0.7780)
	TV_t		91.5956***		−155.1000***		−135.8580***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−93.4922***		−52.5867***		−104.3350***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_t		586.5539		621.5162		−1029.2200
			(0.1379)		(0.4760)		(0.2447)

(*continued on next page*)

Table 4 (*continued*)

Dependent variable	Independent Variables	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
		Model 1 ($N =$ 23,668)	Model 2 ($N =$ 16,975)	Model 1 ($N =$ 18,788)	Model 2 ($N =$ 13,475)	Model 1 (N = 16,836)	Model 2 ($N =$ 12,075)
	Ret_{t-1}		14.8765		29.3373		59.5242
			(0.7154)		(0.4893)		(0.1320)
	R-Squared	0.2657	0.2880	0.3593	0.4626	0.2430	0.3046
Order life until cancelation	*Intercept*	123.2848***	210.1027***	546.0242***	895.2537***	552.2184***	794.1466***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0963***	−0.1129***	−0.4138***	−0.1866*	−0.3838***	−0.2751*
		(<0.0001)	(<0.0001)	(<0.0001)	(0.0761)	(0.0009)	(0.0514)
	TV_t		−6.1059***		−6.5906**		−15.5122***
			(<0.0001)		(0.0329)		(<0.0001)
	TV_{t-1}		−13.1449***		−59.7501***		−64.1247***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_t		166.4268		532.7937		−499.3350
			(0.1601)		(0.1609)		(0.2440)
	Ret_{t-1}		−20.4119*		39.6236**		26.8329
			(0.0950)		(0.0321)		(0.1611)
	R-Squared	0.4883	0.5073	0.4763	0.5396	0.4601	0.5076
Panel F: Hypothesis 4							
Relative Order imbalance	*Intercept*	0.0315**	−0.0645***	−0.7106***	−0.7552***	−0.2248***	−0.2950***
		(0.0411)	(0.0012)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	0.0003***	0.0003***	0.0004***	0.0003***	0.0006***	0.0005***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	TV_t		0.0058*		0.0058***		0.0122***
			(0.0768)		(0.0084)		(<0.0001)
	TV_{t-1}		0.0148***		0.0064***		0.0157***
			(<0.0001)		(0.0024)		(<0.0001)
	Ret_t		2.5689***		1.5097***		1.7573***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_{t-1}		−0.0344		0.0006		0.0141
			(0.3033)		(0.9641)		(0.2332)
	R-Squared	0.8656	−0.0645	0.7739	−0.7552	0.7653	−0.2950

Note: In this table we report results of estimation of Model 1 and Model 2 using full sample. Each panel regression is estimated for each LOB variable individually for large-cap, mid-cap, and small-cap stocks. Each regression is estimated with time and stock fixed effects to address possible seasonality issues and unobserved stock characteristics. Standard errors are clustered by stock. We report respective *p*-values below each estimated coefficient. *, **, *** denote 10%, 5% and 1% significance levels.

messages transmitted by traders to the NASDAQ exchange in the following hour. These results support Hypothesis 1.

For Hypothesis 2, we analyze order size to investigate whether retail or institutional investors are responsible for the increased trading. Following Kumar & Lee, 2006, a decrease in order size would be consistent with retail investors. The output reported in Table 4 Panel B suggests a negative relationship between the size of new orders and the lagged value of Google searches in all three size groups. The estimated coefficients of Model 1 are −0.0018, −0.0649 and − 0.2296 for large-cap, mid-cap and small-cap stocks respectively. The negative sign indicates that new order size tends to decrease following a high-volume hour of Google searches. The estimated coefficients on the lagged Google searches are statistically significant for mid-cap and small-cap stocks in full sample. Inclusion of additional controls in Model 2 does not produce results materially different from Model 1. In both models only small-cap stocks group demonstrates highly statistically significant relationship. The insignificant coefficient for large-cap stocks may be explained by the fact that large stocks are the most actively traded stocks and the retail orders represent only a fraction of a total number of newly submitted orders. Thus, the influence of retail investors tends to be lower in larger stocks.

We conduct two additional tests of Hypothesis 2. We analyze if average order size of canceled and executed orders changes in response to change in investor attention. The results are reported in the Panel B of Table 4. All estimated coefficients of Model 1 have negative sign, which represents the tendency for the order size to decrease. Estimation results of Model 2 also produce negative coefficients that are statistically significant for mid- and small-cap stocks. The size of orders that are canceled or executed depends on orders that have already been posted in the book. Therefore, it is possible that those orders have been submitted in prior hours by a different set of investors.

Thus, we confirm that size of new orders tends to decrease following high investor attention, which provides indirect evidence that retail investors are the likely cause of the increased trading activity but mainly for small stocks. These findings are in agreement with prior literature and with our univariate analysis.

After establishing that retail investors are responsible for the increase in trading activity, we test if retail investors become informed (H3.1) and demand liquidity (H3.2) or if they stay uniformed and provide liquidity. If retail investors stay uninformed, we expect the bid-ask spread to decrease and the depth of the limit order book to increase. Alternatively, if retail investors become informed, then we expect the bid-ask spread to widen and the limit order book depth to become thinner. The results of the bid-ask spread regressions are reported in Panel C of Table 4. The estimated coefficients are all positive but very small in magnitude and statistically insignificant for all three groups of stocks. These results suggest that, after controlling for time and stock fixed effects, the bid-ask spreads are not significantly affected by Google search activity. Hence, we conclude that the relative bid-ask spread does not significantly change in response to the increased retail attention. Our findings here are counter to our univariate results but demonstrate the importance of the fixed effects in our regression analysis.

In Panel D of Table 4 we report the output of the limit order book depth variable regressions. We find the depth measured in a number of orders significantly increases on the buy sides of the book, with Model 1 estimated coefficients of 0.2583, 0.0447, and 0.0416 for large-, mid-, and small-cap stocks, respectively. All coefficients are highly statistically significant at 1%. We also find that sell-side depth measured in the number of orders decreases. The estimated coefficients for Model 1 are 0.0006, −0.0386, and − 0.0618 for large-, mid-, and small-cap stocks, respectively. These results suggest that after controlling for time and stock fixed effects the number of buy-side orders increases following more intensive Google searches and the number of sell-side orders decreases. The effect is stronger and highly significant in mid-cap and small-cap stocks, and this is specifically where we expect retail investors to be able to make a difference. Results of estimation of Model 2 fully support our finding that number of buy-side orders tend to increase following increase in investor attention, and number of sell-side orders tend to decrease. Similarly, the estimated coefficients are highly statistically significant for mid- and small-cap stocks.

For robustness, we estimate alternative specifications of Model 1 and Model 2, with the book depth measured with a number of shares. The test results reported in Panel D of Table 4 are consistent with our previous findings. As before, the estimated coefficients are statistically significant for mid-cap and small-cap stocks.

Overall, our findings are consistent with informed trading. We demonstrate that increased Google search activity tends to add bid depth and reduce ask depth with more of the influence in mid-cap and small-cap stocks. This can be interpreted as retail investors' attempting to buy stocks after googling a company: by submitting limit buy orders, they increase depth on the buy side of the book, and by submitting market buy orders they consume orders on the sell side of the book. Retail investors seem to act as informed traders in stocks where they have an opportunity to be informed: mid-cap and small-cap stocks. These results are counter to our univariate results on the ask side and support the importance of time and stock fixed effects in our cross-sectional analysis.

Even though retail investors may be acting like informed traders, in reality they could be noise traders – investors that trade on noise as if it were information. We hypothesize that if retail investors trade on noise and per se post uninformed orders, their orders will be picked off by informed traders as in Linnainmaa, 2010. To test the H3.3, we regress the average order life, in seconds in a given hour, on the lagged intensity of Google searches. The results of the tests are reported in Panel E of Table 4. We find that the average order life tends to decrease across all stocks following increase in retail attention. The estimated coefficients for Model 1 are −0.0950, −0.4185, and − 0.4559 for large-cap, mid-cap and small-cap respectively and they are statistically significant at 1%. Our findings still hold even after inclusion of additional control variable in Model 2. The estimated coefficients are −0.1034 -0.1891 and − 0.3696 for large-cap, mid-cap and small-cap respectively and all of them are statistically significant.

To better understand the drivers behind the reduced order life, we separately examine order life until execution and order life until cancelation. We start with the order life until execution. The coefficient for lagged Google search activity is negative but statistically insignificant for large-cap stocks while negative and statistically significant for mid-cap and small-cap stocks. It means that an increase in retail attention leads to faster execution of limit orders in the next hour, but mainly for mid-cap and small-cap orders. This finding, combined with the reduced order size and the increased book depth, confirms that retail orders receive a faster execution. This is in agreement with Linnainmaa, 2010 argument that uninformed orders of retail investors will be quickly picked off by informed investors. It is worth noting, that Model 2 estimation results also produce negative albeit insignificant coefficients on lagged Google search volume.

Next, we examine the time it takes for orders to be canceled. The Model 1 estimated coefficient for lagged Google search activity is highly statistically significant and negative for all three groups of stocks for all orders. It means that an increased Google search intensity tends to be followed by a faster order cancelation in the next hour. In Our findings stay immune to the inclusion of additional control variable in Model 2. All estimated coefficients are negative albeit less significant.

Overall, our regression results demonstrate that the average order life decreases following high levels of Google search activity. This fact is consistent with retail investors acting as noise traders that provide liquidity to the truly informed investors. Our multivariate findings are in agreement with results of the univariate analysis.

Lastly, we hypothesize that increased retail attention may result in temporary buying pressure that translates into a more significant order imbalance according to Barber and Odean (2008). We compute relative order imbalance as the depth of the limit order book (in shares) on the buy side minus the depth of the book on the sell side and then we scale it by the total depth on both sides of the LOB. The regression results are reported in Panel F of Table 4. We find that indeed, the order imbalance increases following periods of high retail attention. The estimated coefficients on the lagged Google searches are 0.0003, 0.0004, and 0.0006 for large-cap, mid-cap, and small-cap stocks, respectively. The estimated coefficients are statistically significant at 1% for all size groups of stocks. It is also worth noting that the estimated coefficients are increasing monotonically from large-cap to small-cap stocks, affecting the small stocks the most. This confirms our previous conclusion, that retail investors are more active in smaller stocks. Model 2 estimation results are fully in line with Model 1 findings: all estimated coefficients are positive and statistically significant at 1%. For robustness we estimate Model 1 and Model 2 using order imbalance measured in shares as a dependent variable. The results confirm our findings.

The multivariate tests of the hypotheses allow us to confirm that retail investors indeed trade more actively after googling the company, and retail investors perceive themselves informed and start consuming liquidity on the sell side of the book and providing liquidity on the buy side of the book, thus acting as noise traders. We also confirm that their uninformed orders are quickly picked off by truly informed investors, thus reducing their life span.

4.3. Robustness checks

In this study we analyze the seven-week period that covers September 18, 2017 – November 3, 2017. Our sample period coincides with "the earnings season" which is a time period during which many large companies release their quarterly financial results.[18] Thus, the days of earnings announcements may naturally be characterized by heightened investor attention, and, potentially affect our findings. In addition to earnings announcements, a company may release other material information to shareholders, which may also attract investor attention. To address this concern, separately for each company, we identify and drop all dates from our sample period when a company issued form 8-K. By SEC definition "Form 8-K is the 'current report' companies must file with the SEC to announce major events that shareholders should know about"[19] including the releases of financial results. In addition, Kiymaz and Berument (2003) demonstrate the effect of the day of the week on stock market volatility and volume. Specifically, they find that in the USA Fridays are characterized by the highest stock market volatility. Therefore, to address these concerns, we re-estimate each regression model for a reduced sample, from which we exclude Fridays and form 8 K release dates. The Model 1 and Model 2 estimation results are summarized in Table 5. They demonstrate that our findings are not affected by the exclusion of Fridays and dates of form 8-K releases, and in some cases produce even more statistically significant coefficients.

5. Conclusion

In our study we employ intensity of Google searches, a new, direct, and high-frequency measure of investor attention, to analyze how its intraday (hourly) fluctuations affect trading. We hypothesize, that after googling a company an investor will trade in the next hour before the value of the gathered information decays. Since the purpose of a Google search is to gather information, we also examine if, after googling a company, investors act as uninformed providers of liquidity, or if they

[18] We would like to thank anonymous referees of the "International Review of Financial Analysis" for this valuable suggestion.

[19] https://www.sec.gov/fast-answers/answersform8khtm.html

Table 5
Multivariate analysis of effect of investor attention on selected LOB variables (reduced sample).

Dependent variable	Independent Variables	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
		Model 1 (N = 18,915)	Model 2 (N = 13,580)	Model 1 (*N* = 15,015)	Model 2 (*N* = 10,780)	Model 1 (N = 13,455)	Model 2 (*N* = 9660)
Panel A: Hypothesis 1							
Trading activity (Number of messages transferred to the exchange)	*Intercept*	26.0200***	−33.5928***	8.2815***	−3.2168***	3.9335***	0.6725**
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.0399)
	$GSVI_{t-1}$	19.6751***	0.0097*	5.1012***	−0.0024**	12.4415***	0.0070***
		(0.0002)	(0.0627)	(0.0017)	(0.0454)	(<0.0001)	(<0.0001)
	TV_t		10.8193***		1.6397***		0.9202***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		0.6734**		0.6052***		0.3573***
			(0.0246)		(<0.0001)		(<0.0001)
	Ret_t		−113.8050***		17.1374***		−6.3364
			(0.0006)		(0.0037)		(0.1099)
	Ret_{t-1}		10.1970***		0.5733*		0.5059***
			(0.0015)		(0.0503)		(0.0032)
	R-Squared	0.7913	0.8384	0.7496	0.8093	0.7610	0.8081
Panel B: Hypothesis 2							
Number of shares per new order	*Intercept*	0.1469***	0.1192***	1.3750***	1.4628***	0.1899***	0.1483***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.00002	−0.00001*	−0.00006	−0.0001	−0.00025***	−0.0003***
		(0.3758)	(0.1004)	(0.1265)	(0.3286)	(0.0066)	(0.0019)
	TV_t		0.0074***		0.0084***		0.0132***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−0.0023*		−0.0059***		0.0054**
			(0.0829)		(<0.0001)		(0.0264)
	Ret_t		−0.1362		−0.1185		−0.0191
			(0.3556)		(0.5097)		(0.9508)
	Ret_{t-1}		0.0368**		−0.0001		−0.0253*
			(0.0105)		(0.9893)		(0.0589)
	R-Squared	0.8331	0.8561	0.7913	0.8104	0.3102	0.2830
Number of shares per canceled order	*Intercept*	0.1124	0.0828***	1.1166	1.2160***	0.1293	0.0737***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.00001	0.00001	−0.00003	−0.00001*	−0.0001**	−0.0002**
		(0.4137)	(0.1393)	(0.3381)	(0.0571)	(0.0458)	(0.0122)
	TV_t		0.0064***		0.0071***		0.0158***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−0.0009		−0.0037***		0.0045***
			(0.4441)		(0.0009)		(0.0055)
	Ret_t		−0.1470		−0.0802		0.0601
			(0.2394)		(0.5706)		(0.7719)
	Ret_{t-1}		0.0273		−0.0034		−0.0285***
			(0.0250)		(0.6284)		(0.0015)
	R-Squared	0.8207	0.8488	0.7915	0.8322	0.2676	0.2918
Number of shares per executed order	*Intercept*	0.0995***	0.0152***	0.6273***	0.5505***	0.0932***	−0.0454
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.0240)	(0.1122)
	$GSVI_{t-1}$	−3.44E-08	0.00001	0.00003	0.000001	−0.00025*	−0.0003***
		(0.9966)	(0.2779)	(0.2134)	(0.9022)	(0.0792)	(0.0022)
	TV_t		0.0203***		0.0248***		0.0525***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−0.0052***		−0.0038***		−0.0032
			(<0.0001)		(<0.0001)		(0.2336)
	Ret_t		0.2044***		0.1246		0.0779
			(0.0003)		(0.2775)		(0.8221)
	Ret_{t-1}		−0.0054		0.0049		−0.0139
			(0.3303)		(0.3882)		(0.3565)
	R-Squared	0.8924	0.9108	0.6767	0.7046	0.1771	0.2177
Panel C: Hypotheses 3.1							
Relative bid-ask spread	*Intercept*	−0.0000008	0.0004***	0.0006	0.0029***	0.0020**	0.0022***
		(0.9588)	(<0.0001)	(0.3402)	(<0.0001)	(0.0135)	(<0.0001)
	$GSVI_{t-1}$	0.000001	0.0000001	0.00008	0.0000001	0.000006	0.0000001
		(0.1148)	(0.8171)	(0.1518)	(0.7325)	(0.8557)	(0.9001)
	TV_t		0.00002***		0.0000***		−0.0001***
			(0.0002)		(0.0013)		(<0.0001)
	TV_{t-1}		0.00003		0.0000**		0.0000***
			(0.9956)		(0.0129)		(0.0020)
	Ret_t		0.0001		−0.0067***		0.0047**
			(0.9493)		(<0.0001)		(0.0102)
	Ret_{t-1}		0.000002		−0.0001		0.0001
			(0.9347)		(0.1893)		(0.4694)
	R-Squared	0.2921	0.3135	0.5298	0.8284	0.6169	0.8310
Panel D: Hypotheses 3.2							
Number of orders on buy side	*Intercept*	466.1132***	423.5563***	365.7642***	359.4616***	121.7169***	115.4441***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	0.3301***	0.0289	0.0340***	0.0148	0.0708***	0.0562***

(*continued on next page*)

Table 5 (*continued*)

Dependent variable	Independent Variables	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
		Model 1 (N = 18,915)	Model 2 (N = 13,580)	Model 1 (*N* = 15,015)	Model 2 (*N* = 10,780)	Model 1 (N = 13,455)	Model 2 (*N* = 9660)
		(0.0013)	(0.7611)	(0.0021)	(0.2606)	(<0.0001)	(0.0017)
	TV_t		−3.0381		1.6302***		0.9501**
			(0.5907)		(<0.0001)		(0.0353)
	TV_{t-1}		9.5011*		3.0132***		2.5556***
			(0.0812)		(<0.0001)		(<0.0001)
	Ret_t		4682.8490***		324.1946***		651.7960***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_{t-1}		−134.4660**		3.8388		1.2458
			(0.0215)		(0.1039)		(0.6076)
	R-Squared	0.9867	0.9926	0.9594	0.9619	0.9191	0.9223
Number of shares on buy side	*Intercept*	97.7384***	65.3540***	328.7171***	340.1321***	18.3574***	12.9769***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0459	−0.0810*	0.0328***	0.0202*	0.0535***	0.0407***
		(0.1953)	(0.0574)	(0.0013)	(0.1023)	(<0.0001)	(0.0011)
	TV_t		2.2559		1.1845***		0.6425**
			(0.3735)		(0.0011)		(0.0414)
	TV_{t-1}		3.9371***		1.5160***		1.7351***
			(0.1073)		(<0.0001)		(<0.0001)
	Ret_t		1725.7500***		87.0303*		207.9746***
			(<0.0001)		(0.0514)		(<0.0001)
	Ret_{t-1}		7.6406		2.4495		1.1719
			(0.7709)		(0.2686)		(0.4889)
	R-Squared	0.9684	0.969	0.9333	0.9368	0.8619	0.8658
Number of orders on sell side	*Intercept*	359.9816***	372.7148***	826.2742***	837.5596***	165.8843***	169.4921***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.0927	−0.2233*	−0.0476***	−0.0703***	−0.0605***	−0.0873***
		(0.4081)	(0.0906)	(0.0002)	(<0.0001)	(<0.0001)	(<0.0001)
	TV_t		4.1184		0.9239**		−0.6975**
			(0.6000)		(0.0358)		(0.0271)
	TV_{t-1}		−9.7955		1.8015***		0.2040
			(0.1959)		(<0.0001)		(0.5059)
	Ret_t		−1903.4800**		−308.5040***		−504.6820***
			(0.0223)		(<0.0001)		(<0.0001)
	Ret_{t-1}		−124.9550		−2.5587		−1.0075
			(0.1242)		(0.3398)		(0.5525)
	R-Squared	0.9637	0.967	0.9764	0.9788	0.9885	0.9915
Number of shares on sell side	*Intercept*	106.1418***	137.2734***	1613.8450***	1627.1810***	52.4164***	65.4006***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.01332	−0.0172	−0.03145	−0.0693**	−0.11655***	−0.1333***
		(0.7423)	(0.7232)	(0.4200)	(0.0391)	(<0.0001)	(<0.0001)
	TV_t		−0.7384		3.3197**		−3.3410***
			(0.7980)		(0.0160)		(<0.0001)
	TV_{t-1}		−6.8256**		3.8629***		−1.3698**
			(0.0142)		(0.0035)		(0.0337)
	Ret_t		−1233.9700***		−241.5990		−185.0560**
			(<0.0001)		(0.1536)		(0.0245)
	Ret_{t-1}		−42.3706		−13.2742		−4.8275
			(0.1559)		(0.1137)		(0.1760)
	R-Squared	0.9844	0.9852	0.9130	0.9161	0.9758	0.9764
Panel E: Hypothesis 3.3							
Order life	*Intercept*	98.6628***	166.1795***	581.4845***	949.9733***	558.7153***	778.1040***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.1153***	−0.1250***	−0.4061***	−0.1548*	−0.5465***	−0.4789***
		(<0.0001)	(<0.0001)	(<0.0001)	(0.0902)	(<0.0001)	(0.0029)
	TV_t		4.2562***		−4.3072		−11.4345***
			(0.0023)		(0.2152)		(0.0048)
	TV_{t-1}		−21.2300***		−66.1715***		−66.8196***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_t		−192.2230		534.3834		−1733.9100***
			(0.1942)		(0.2109)		(0.0006)
	Ret_{t-1}		−13.8556		60.2657		15.9295
			(0.3376)		(0.0044)		(0.4642)
	R-Squared	0.4894	0.5090	0.4865	0.5596	0.4674	0.5052
Order life until execution	*Intercept*	39.5950**	−61.2831**	1394.2770***	2827.6100***	515.2834***	1138.6000***
		(0.0293)	(0.0262)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.1272**	−0.0975	−0.4207*	−0.1896	−0.5120**	−0.1993
		(0.0279)	(0.2099)	(0.0618)	(0.4953)	(0.0322)	(0.5305)
	TV_t		89.8482***		−152.1590***		−125.0810***
			(<0.0001)		(<0.0001)		(<0.0001)
	TV_{t-1}		−89.0104***		−53.6547***		−86.2876***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_t		−554.5170		1040.0320		−2246.4100**
			(0.2584)		(0.3006)		(0.0237)

(*continued on next page*)

Table 5 (*continued*)

Dependent variable	Independent Variables	Large-cap stocks		Mid-cap stocks		Small-cap stocks	
		Model 1 (N = 18,915)	Model 2 (N = 13,580)	Model 1 (*N* = 15,015)	Model 2 (*N* = 10,780)	Model 1 (N = 13,455)	Model 2 (*N* = 9660)
	Ret_{t-1}		17.3488		81.8781*		23.5072
			(0.7172)		(0.1001)		(0.5852)
	R-Squared	0.2697	0.2904	0.3609	0.4711	0.2483	0.2943
Order life until cancelation	*Intercept*	107.6328***	202.0578***	539.3088***	885.5769***	568.8943***	780.4578***
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	−0.1125***	−0.1260***	−0.4087***	−0.1600	−0.4822***	−0.3747**
		(<0.0001)	(<0.0001)	(0.0001)	(0.1904)	(0.0002)	(0.0203)
	TV_t		−6.9000***		−5.0458		−10.5454***
			(<0.0001)		(0.1604)		(0.0096)
	TV_{t-1}		−13.6794***		−63.4400***		−65.0227***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_t		−25.4324		455.4374		−1596.6500***
			(0.8602)		(0.3025)		(0.0016)
	Ret_{t-1}		−12.5986		57.7727***		16.3165
			(0.3715)		(0.0083)		(0.4559)
	R-Squared	0.4934	0.5147	0.474	0.5417	0.4716	0.5111
Panel F: Hypothesis 4							
Relative order imbalance	*Intercept*	0.0287*	−0.0499**	−0.7155***	−0.7567***	−0.2490***	−0.3233***
		(0.0825)	(0.0257)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
	$GSVI_{t-1}$	0.0003***	0.0003***	0.0003***	0.0003***	0.0006***	0.0005***
		(<0.0001)	(<0.0001)	(<0.0001)	(0.0020)	(<0.0001)	(<0.0001)
	TV_t		0.0048		0.0048*		0.0151***
			(0.2028)		(0.0625)		(<0.0001)
	TV_{t-1}		0.0118***		0.0066***		0.0153***
			(0.0011)		(0.0071)		(<0.0001)
	Ret_t		2.7394***		1.4733***		1.8603***
			(<0.0001)		(<0.0001)		(<0.0001)
	Ret_{t-1}		−0.0160		0.0092		0.0155
			(0.6812)		(0.5578)		(0.2602)
	R-Squared	0.8662	0.8671	0.7677	0.7656	0.7629	0.7681

Note: In this table we report results of estimation of Model 1 and Model 2 for a reduced sample – the sample that excludes Fridays and form 8 K release dates. Each panel regression is estimated for each LOB variable individually for large-cap, mid-cap, and small-cap stocks. Each regression is estimated with time and stock fixed effects to address possible seasonality issues and unobserved stock characteristics. Standard errors are clustered by stock. We report respective p-values below each estimated coefficient. *, **, *** denote 10%, 5% and 1% significance levels.

become informed and start demanding liquidity. The third alternative we explore, is if after conducting a Google search, investors become noise traders trading on noise as if it were information.

We collect hourly Google search statistics for 300 NASDAQ stocks: 100 large-cap, 100 mid-cap, and 100 small-cap. To test our hypotheses, we estimate values of limit order book variables on hourly basis.

Our results clearly indicate that trading activity intensifies in the hour following the hour with an increased investor attention. Second, we confirm, that retail investors are responsible for increased trading activity, as indicated by a decrease in the size of new orders. Next, we analyze retail investors' trading behavior. If Google search activity leads to liquidity trading, we would expect to find a narrower bid-ask spread and increased depth on both sides of the limit order book across all three size groups of stocks. We have conflicting results on the spread. However, we demonstrate an increase in depth on the buy side and a decrease in depth on the sell side of the book, with an overall increase in order imbalance, which is not consistent with liquidity trading. Our results are more consistent with informed trading that consumes liquidity. We also find support for the Barber and Odean (2008) prediction that retail attention will lead to a temporary buying pressure.

If Google search activity is not enhancing the liquidity of the limit order book, then the retail investors are either informed traders or noise traders. The limit order book imbalance is stronger in the smaller stocks, which is where we expect the retail investors could have an information advantage. However, Linnainmaa, 2010 argues that limit orders of uninformed (individual) investors should be quickly picked off by informed traders. Our regression results suggest shorter average order lives across all three sets of stocks. These results, in conjunction with an order imbalance suggest that high levels of Google search activity lead to noise training as defined by Black, 1986.

Google searches are how society becomes informed. However, our results suggest that retail investors do not become informed through Google searches in the traditional finance sense of an informed investor, but they rather perceive themselves as informed and act as noise traders, since their orders are being picked off by truly informed traders.

Declaration of Competing Interest

None.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Definitions of Limit Order Book variables

All measures are computed on a per stock basis.

Quoted bid-ask spread is calculated as a difference between Best Ask and Best Bid prices.

Quoted BA spread = Best Ask − Best Bid

Mid-point price is calculated as an arithmetic average of the Best Ask and Best Bid prices.

$$\text{Midpoint Price} = \frac{\text{Best Ask} + \text{Best Bid}}{2}$$

Relative quoted bid-ask spread is calculated by dividing the quoted bid-ask spread by the midpoint price.

$$\text{Relative BA spread} = \frac{\text{Quoted BA spread}}{\text{Midpoint Price}}$$

Bid depth is measured two ways: 1) as a total number of orders on the bid side of the book and 2) as total number of shares on the bid side of the book at all price levels.

Ask depth is measured two ways: 1) as a total number of orders on the ask side of the book and 2) as total number of shares on the ask side of the book at all price levels.

Order imbalance is measured in shares as a difference in depth on the buy and sell sides of the limit order book scaled by the total number of shares on both sides.

$$\text{Relative Order imbalance} = \frac{\text{Bid Depth}_{\text{shares}} - \text{Ask Depth}_{\text{shares}}}{\text{Total Depth}_{\text{shares}}}$$

Order size is measured in a number of shares per newly submitted order.

Order life is measured in seconds as a difference between the time the order is removed from the book for any reason (e.g execution, cancelation, or change) and its creation time.

Order life until execution is measured in seconds as a difference between the time the order is fully executed and its creation time.

Order life until cancelation is measured in seconds as a difference between the time the order is canceled and removed from the book and its creation time.

Return is calculated as an hourly rate of change in mid-point price.

Trading volume is measured as a natural logarithm of a number of shares in thousands of a specific stock traded on NASDAQ per hour.

Trading activity is measured as a number of messages transmitted to the exchange by traders in thousands. This includes following types of messages: new orders, order executions, order cancelations, and order changes (updates of existing orders).

References

Aouadi, A., Arouri, M., & Teulon, F. (2013). Investor attention and stock market activity: Evidence from France. *Economic Modeling, 35*, 674–681.

Arbel, A., & Strebel, P. (1982). The neglected and small firm effects. *The Financial Review, 17*(4), 201–218.

Askitas, N., & Zimmermann, K. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly, 2*, 107–120.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Market and Protfolio Management, 25*, 239–264.

Barber, B., & Odean, T. (2008). All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies, 21*(2), 785–818.

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. *Review of Financial Studies, 30*(9), 3009–3047.

Black, F. (1986). Noise. *The Journal of Finance, 41*(3), 528–543.

Bloomfield, R., O'Hara, M., & Saar, G. (2009). How Noise Trading Affects Markets: An Experimental Analysis. *The Review of Financial Studies, 22*(6), 2275–2302.

Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc (Workin paper)*, 1–5.

Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *Economic Record*, 2–9.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance, 66*(5), 1461–1499.

Dimpfl, T., & Jank, S. (2011). Can internet search queries help to predict stock market volatility? *European Financial Management, 22*(2), 171–192.

Drake, M. S., Roulston, D. T., & Thornock, J. R. (2012). Investor information demand: Evidence from Google searches around earnings announcements. *Journal of Accounting research, 50*(4), 1001–1040.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business, 38*(1), 34–105.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance, 64*(5), 2023–2052.

Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics, 170*(Part A), 97–135.

Fink, C., & Johann, T. (2014). May I Have Your Attention, Please: The Market Microstructure of Investor Attention. September 17 https://ssrn.com/abstract=2139313. (Accessed September 2017).

Fricke, E., Fung, S., & Sinan Goktan, M. (2014). Google search, Information uncertainty, and Post-Earningas Announcement Drift. *Journal of Accoutning and Finance, 14*(2), 11–27.

Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature., 457*, 1012–1014.

Glosten, L. R. (1994). Is the Electronic Open Limit Order Book Inevitable? *The Journal of Finance, 49*(4), 1127–1161.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics, 14*(1), 71–100.

Joseph, K., Wintoki, M. B., & Zhang, Z.. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting, 27*(4), 1116–1127.

Kaniel R.; Saar G.; Titman S.; Individual investor trading and stock returns. 63 (1). The Journal of Finance. (2008) Wiley Online Library 273–310.

Kumar, A., & Lee, C. M. C. (2006). Retail investor sentiment and return comovements. *The Journal of Finance, 61*(5), 2451–2486.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica, 53*(6), 1315–1335.

Linnainmaa, J. T. (2010). Do limit orders alter inferences about investor performance and behavior? *The Journal of Finance, 65*(4), 1473–1506.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance, 42*(3), 483–510.

Meshcheryakov, A. V. (2018). Using online search queries in real estate research with an empirical example of arson forecast. *Journal of Real Estate Literature, 26*(2), 331–361.

O'Hara, M., Yao, C., & Ye, M. (2014). What's not there: Odd Lots and Market data. *The Journal of Finance LXIX, 5*, 2199–2236.

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic perspectives, 4*(2), 19–33.

Veldhuizen, S. V., Vogt, B., & Voogt, B. (2016). Internet searches and transactions on the Dutch housing market. *Applied Economics Letters, 23*(18), 1321–1324.

Wu, L., & Brynjolfsson, E. (2009). The future of prediction: how Google searches foreshadow housing prices and quantities. *ICIS 2009 Proceedings*, 147.