Singapore Management University

Institutional Knowledge at Singapore Management University

Dissertations and Theses Collection (Open Access) | Dissertations and Theses

4-2022

Information acquisition and market friction

Bo SANG
Singapore Management University, bo.sang.2017@pbs.smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/etd_coll

Part of the Finance Commons, and the Finance and Financial Management Commons

Citation

SANG, Bo. Information acquisition and market friction. (2022).
Available at: https://ink.library.smu.edu.sg/etd_coll/393

This PhD Dissertation is brought to you for free and open access by the Dissertations and Theses at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Dissertations and Theses Collection (Open Access) by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

INFORMATION ACQUISITION AND MARKET FRICTION

BO SANG

SINGAPORE MANAGEMENT UNIVERSITY

2022

Information Acquisition and Market Friction

Bo Sang

Submitted to Lee Kong Chian School of Business in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Finance

Dissertation Committee

Joe Zhang (Chair)
Associate Professor of Finance
Singapore Management University

Ekkehart Boehmer
Professor of Finance
Singapore Management University

Chi Shen Wei
Assistant Professor of Finance
Singapore Management University

Lin Tong
Assistant Professor of Finance
Fordham University

Singapore Management University

2022

Copyright (2022) Bo Sang

I hereby declare that this PhD dissertation is my original work
and it has been written by me in its entirety.
I have duly acknowledged all the sources of information
which have been used in this dissertation

This PhD dissertation has also not been submitted for any degree
in any university previously

Sang Bo

Bo Sang

28 Mar 2022

Information Acquisition and Market Friction

Bo Sang

Abstract

My dissertation consists of three papers related to information diversity, acquisition, and asymmetry. One part of the dissertation explores the implications of interactions among different market participants and subsequent price efficiency in the stock market. The empirical findings indicate the information diversity between individuals and institutional investors, as well as an important channel for retail investors to obtain useful information – through insider filings. The remaining part investigates the information asymmetry between issuers and naive investors in the cryptocurrency market.

In Chapter 2, I aggregate trading signals from hedge funds and retail investors, in order to examine their information diversity and the combined informational role in the stock market. I show that incorporating signals from both groups is necessary to identify firm-level information. Stocks that reflect consistent trading between two groups exhibit strong return predictability without reversal. When trading in the opposite direction to retail investors, hedge funds cannot yield any significant return, even in a longer horizon. I also document that consistent trading between two groups significantly predicts firm fundamentals, informational events, market reactions, and helps alleviate stock-level mispricing. Overall, the findings suggest combining signals that solely from hedge funds is incomplete, as there remain signals from retail investors who are informed in different aspects of stock fundamentals.

In Chapter 3, we examine the trading patterns of retail investors following insider trading and the corresponding price impact. Retail investors follow the

opportunistic purchases by insiders, but not their routine purchases. The abnormal retail downloads of the Form 4 filings from the EDGAR database also increase for opportunistic insider purchases. Neither investor attention nor common information such as earnings announcements or analysts forecast revisions explains the results. Moreover, for stocks with opportunistic insider purchases, those that retail investors bought yield higher cumulative abnormal returns than those that retail investors sold. The effect is mostly driven by the information component of the retail trades, rather than liquidity provision or temporary price pressure. Variance ratio tests also suggest price efficiency improvements for stocks bought by retail investors following opportunistic insider purchases. The evidence is mostly consistent with retail investors learning from opportunistic insider purchases, and their trading helping expedite price discovery.

In Chapter 4, we study the economics of financial scams by investigate the market for initial coin offerings (ICOs) using point-in-time data snapshots of 5,935 ICOs. Our evidence indicates that ICO issuers strategically screen for naïve investors by misrepresenting the characteristics of their offerings across listing websites. Misrepresented ICOs have higher scam risk, and misrepresentations are unlikely to reflect unintentional mistakes. Using on-chain analysis of Ethereum wallets, we find that less sophisticated investors are more likely to invest in misrepresented ICOs. We estimate that 40% of ICOs (U.S. $12 billion) in our sample are scams. Overall, our findings uncover how screening strategies are used in financial scams and reinforce the importance of conducting due diligence.

Table of Contents

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Joe Zhang for his valuable advice and continuous support during my PhD study. I truly enjoy discussing with him on research ideas, and his suggestions greatly help me develop my research skills and mindset. He is an integral part of me being able to make it to the end.

I would also like to thank Professor Ekkehart Boehmer and Professor Chi Shen Wei for their insightful comments and encouragement along this journey. Prof. Boehmer has given me continuous guidance in research with his immense knowledge. Prof. Wei greatly helps me sharpen my critical thinking. I would also like to thank my other committee member Professor Lin Tong for her support.

My sincere thanks also goes to my collaborators for their insights and great efforts in improving our research papers. I would like to thank my senior Li Guo for his remarkable suggestions and selfless help. I would also like to thank Professor Jun Tu, Kenny Phua and Gloria Yu for their great advice and patience. Furthermore, I want to thank the finance faculty here for their excellent guidance.

In addition, I am grateful to all my PhD friends for accompanying me, through ups and downs. I truly enjoy discussing fresh ideas and sharing PhD experience with them. They have made me a better person.

Finally, I would like to express my deepest thanks to my parents for their unconditional love. They have provided all I need, trained my logical thinking and open-minded outlook on life, and supported all my decisions. They are my mentors and also my best friends.

Chapter 1

Introduction

My dissertation mainly explores the implications of interactions among different market participants and subsequent market efficiency. Earlier studies portrait institutional investors as informed while individual investors as noise traders who make systematic mistakes on investment (Barber and Odean, 2000; Frazzini and Lamont, 2008; Barber, Odean, and Zhu, 2009). By contrast, recent evidence of retail trading, especially off-exchange retail order flow, proves its strong stock return predictability beyond liquidity provision (e.g., Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013; Boehmer, Jones, Zhang, and Zhang, 2021), suggesting an information role. Unaffected by the agency problems or liquidity constraints that institutional investors face, these informed individual investors have stronger incentives to trade on novel information, potentially benefiting from their geographic proximity to firms, relations with employees, or insights into customer preferences. Meanwhile, the informational playing field between institutional investors and retail investors is leveled recently given that modern information technologies such as social media further improve the informational role of individual investors (e.g., Farrell, Green, Jame, and Markov, 2018).

To the extent that informed retail investors and other sophisticated investors are likely to possess different types of information, research on their joint trading pattern is scant in the literature, where prior studies mostly focus on heterogeneity of trades that are from professional traders such as active mutual funds (Jiang and Sun, 2014) and hedge funds (Jiao, Massa, and Zhang, 2016). I am thus eager to investigate their interactions and the aggregated informational role in the stock market. In Chapter 2, I investigate the information diversity between retail investors and institutional investors. I select hedge funds to represent smart institutions, as they are widely portrayed as the most sophisticated institutional investors and weigh a significant portion in the overall stock market (e.g., Ackermann, McEnally, and Ravenscraft, 1999; Kosowski, Naik, and Teo, 2007; Aggarwal and Jorion, 2010; Cao, Liang, Lo, and Petrasek, 2018). As for individual trading signal, I follow Boehmer et al. (2021) algorithm to obtain marketable off-exchange retail order flow, which is shown to contain novel firm-level information. Preliminary results suggest that the information possessed by individuals and institutions are distinct in nature, in line with the rational expectations equilibrium proposed by Diamond and Verrecchia (1981). Trading from each group of investors contains both information and noise, while the noise portion is greatly reduced when combing the two trading signals together.

Furthermore, the analysis suggests that incorporating trading signals solely from institutions is incomplete. The reason is that these signals are likely to related to one type of firm fundamental, while there remain informed signals about other types of fundamentals potentially come from different investor groups (e.g., Grossman, 1976; Diamond and Verrecchia, 1981; Goldstein and Yang, 2015). Consistent with this intuition, I find trading signal from individuals provides indispensable information beyond hedge fund trading. In particular, when trading opposite to individuals, hedge funds cannot yield any significant returns. Also from retail investor side, negligence toward the signal from institutions would largely affect their stock performance. By contrast, consistent trading between these two groups exhibit strong return predictability without reversal, significantly forecasts firm fundamentals, and helps alleviate stock-level mispricing. Consistent with the story that retail investors and institutional investors are informed in different aspects of stock fundamentals, aggregating trading signals from both groups is necessary to identify firm-level information.

Given the informational role of retail trading, I find it is essential to figure out the sources of valuable information. How do skilled retail investors acquire their advantage? Do they learn from other informed traders? How quickly do they learn? As corporate insiders have privileged access to firm information ahead of others and they are required report the open market trades within 2 days after the trades, I am thus interested to investigate one channel that can potentially contribute to retail investor's information set – mimicking corporate insider trading. In Chapter 3, we examine the trading patterns of retail investors following insider trading and the corresponding price impact, which would help us understand how information is impounded into stock prices. Preliminary results show that retail investors tend to buy the stocks that insiders have purchased, and sell those that insider have sold, more than they usually sell. Moreover, retail investors do not blindly follow insider trades – they follow *opportunistic* (informed) insider purchases instead of *routine* (uninformed) insider purchases. To the extent that insider *opportunistic* trades reflect insider's private information, retail investors seem to be able to identify and learn from the actual informed trades and act quickly.

With the advancement in technology and the abundance information, savvy retail investors can monitor and research about insider trades on their own. Indeed, we find an increase in the abnormal retail downloads of the insider trading filing Form 4 in the EDGAR database. Higher abnormal retail downloads during the event window are also associated with more retail buys in the following days. This evidence is most significant for stocks with *opportunistic* insider purchases.Our results are not driven by elevated investor attention or common information such as earnings news and analyst revisions. Furthermore, for stocks with opportunistic insider purchases, those that retail investors bought yield higher cumulative abnormal returns than those that retail investors sold, with an improved price efficiency. Such effect is mostly driven by the information component of the retail trades and stronger for stocks with higher arbitrage costs.

Overall, the results suggest that retail investors learn from opportunistic insider purchases, and their trading helps expedite price discovery.

In addition to stock market analysis, I have a keen interest in investigating the cryptocurrency market. Significant information asymmetry between firm issuers and investors in this nascent market has drawn much attention. In Chapter 3, we investigate initial coin offerings (ICO) scams and find that ICO issuers strategically screen for naïve investors. In our analysis, we collect 13 months of point-in-time snapshots of self-reported ICO data from five leading ICO listing websites. In our sample of 5,935 ICOs, 34% of tokens have discrepancies at their first appearances. A discrepancy implies that the issuer has misrepresented the offering because at least one of the reported material facts must be untrue.

To better understand why misrepresentations are so prevalent, we model the behavior of a malicious ICO issuer who faces a pool of naïve and astute investors. Naïve investors are unable to conduct proper due diligence and are likely to fall for an ICO scam, while astute investors carefully evaluate the offering and eventually refrain from funding it. We hypothesize that malicious ICO issuers use misrepresentations, along with other suspicious actions, to screen out astute investors and target naïve investors. Our main finding is that ICOs with misrepresentations are significantly more likely to be scams. Further, less sophisticated investors are more likely to invest in misrepresented ICOs. Additional tests rule out the alternative interpretation that misrepresentations are unintentional mistakes made by careless issuers. We estimate that 40% of ICOs (U.S. $12 billion) in our sample are scams. Our findings uncover how screening strategies are used in financial scams and reinforce the importance of conducting due diligence.

Chapter 2

Aggregating Diverse Information from Institutions and Individuals

2.1 Introduction

In a competitive security market, there exist multiple informational sources and traders tend to specialize in different types of information. Theoretical papers that study market efficiency highlight the importance to investigate the interaction between different types of informed investors (e.g., Grossman and Stiglitz, 1980; Diamond and Verrecchia, 1981; Goldstein and Yang, 2015). Prior empirical studies mostly focus on heterogeneity of trades that are from professional traders such as active mutual funds (Jiang and Sun, 2014) and hedge funds (Jiao, Massa, and Zhang, 2016). However, to the extent that individual investors and institutions are likely to differ in their information sources, research interacting trades from these two groups is scant in the literature.

One potential reason for such lack of analysis could be that earlier studies assume only institutional investors are informed, while individual investors are noise traders who make systematic mistakes on investment.[1] As a result, researchers tend to ignore the latter group and concentrate on interaction within informed institutions. However, recent evidence of retail trading, especially off-exchange retail order flow, proves its strong stock return predictability beyond liquidity provision (e.g., Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013; Boehmer, Jones, Zhang, and Zhang, 2021).[2] The smart retail orders

[1]E.g., Barber and Odean (2000); Frazzini and Lamont (2008); Barber, Odean, and Zhu (2009).

[2]The discrepancy in the informational role of individual trading could be due to different sample periods or data sources. Indeed, Kelley and Tetlock (2013) provide evidence that the

identified here represent a significant share of the overall retail trading[3] and potentially contain valuable private information. For example, both Kelley and Tetlock (2013) and Boehmer et al. (2021) point out that retail market orders correctly predict firm-level news, including earnings surprises. Unaffected by the agency problems or liquidity constraints that institutional investors face, these informed individual investors have stronger incentives to trade on novel information, potentially benefiting from their geographic proximity to firms, relations with employees, or insights into customer preferences (e.g., Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013). Meanwhile, the informational playing field between institutional investors and retail investors is leveled recently given that modern information technologies such as social media further improve the informational role of individual investors (e.g., Farrell, Green, Jame, and Markov, 2018). Therefore, an analysis on interaction of trading signals solely within institutional investors tend to overlook the distinct source of information from informed retail trading.

In this paper, I address the literature gap by interacting trading signals from two investor groups, i.e., institutional investors and individual investors, who are likely to be informed in different aspects of the stock value. By observing subsequent return patterns and fundamental predictions, this paper aims to investigate the information diversity between two parties and their aggregated informational role in the stock market. I select hedge funds to represent smart institutions, as they are widely portrayed as the most sophisticated institutional investors and weigh a significant portion in the overall stock market (e.g., Ackermann, McEnally, and Ravenscraft, 1999; Kosowski, Naik, and Teo, 2007; Aggarwal and Jorion, 2010; Cao, Liang, Lo, and Petrasek, 2018). I use

trading skill of retail clientele vary across brokers, as well as the aggregate skill of retail traders have changed over time.

[3] Kelley and Tetlock (2013) analyze over $2.6 trillion retail executed trades, which is roughly one-third of all self-directed retail trading in the United States. The sample in Boehmer et al. (2021) covers marketable retail order flow from TAQ, which contains all off-exchange trading.

short interest and change in hedge fund holdings as two alternative proxies for hedge fund trading in this paper. The reason I choose to use short interest as one proxy is that the vast majority of short sellers are hedge funds[4] and the literature document that low (high) level of short interest contain positive (negative) information (e.g., Asquith, Pathak, and Ritter, 2005; Karpoff and Lou, 2010; Dechow, Hutton, Meulbroek, and Sloan, 2001; Desai, Ramesh, Thiagarajan, and Balachandran, 2002; Diether, Lee, and Werner, 2009; Boehmer, Huszar, and Jordan, 2010). As for individual trading signal, I follow Boehmer et al. (2021) algorithm to obtain marketable off-exchange retail order flow, which is shown to contain novel firm-level information.[5]

Aggregating informational signals from different groups of investors has been studied in theoretical papers such as Grossman (1976); Diamond and Verrecchia (1981); Admati and Pfleiderer (1987). There are in general two assumptions regarding the relationship between the nature of information from two groups, either to be complements or to be substitutes. For example, Goldstein and Yang (2015) analyze a model in which two types of traders are informed in different aspects of firm value and acquiring these two types of information can be complementary. Paul (1993), on the other hand, documents a substitution effect that arises from competition among two types of investors trading on the same type of information. In this paper, I test information diversity between institutional and individual investors and propose two hypotheses that their informational signals could be complements or substitutes. For the two hypotheses, both groups of investors are assumed to be informed to some extent. In addition, I propose a third hypothesis that institutions are smart traders while individual investors are noise traders who do not possess valuable information.

[4]Jiao, Massa, and Zhang (2016) also use short interest to represent hedge fund trading. According to Goldman Sachs's report "Hedge fund trend monitor", hedge funds accounted for 85% of total short interest positions, or $361 billion, as of December 31, 2009.

[5]Short interest are short positions submitted to exchanges and are less likely to overlap with off-exchange retail order flow. Still, I exclude retail short selling from the retail order flow for the sake of conservation.

I investigate these alternative hypotheses that explain information diversity between institutions and individual investors. The *Complementarity Hypothesis* posits individual investors as smart traders who may have access to novel information sources that institutional investors either do not have or are constrained to trade on, due to concerns such as conflict of interest or binding regulation. Hence, trading signal from retail investors provides indispensable value beyond the information contained in hedge fund trading and it may be necessary to consider the bundle of both signals when analyzing informed trading. On the other hand, the *Substitutability Hypothesis* assumes smart retail investors and hedge funds obtain the same type of information and both make investments on such information. Informational signals from two parties in this case serve as substitutes. Under the former (latter) hypothesis, combining trading signals from individuals and institutions yields (does not yield) superior stock performance than checking each signal unconditionally. As for the *Noise trader Hypothesis*, individual investors are assumed to make investment mistakes and smart institutions make profits serving as counterparty to individuals.

To disentangle these hypotheses, I conduct empirical analysis for stocks with nonmissing data in short interest, hedge fund holdings, and retail order flow over the period 2010–2018.[6] First, I focus on the subsets of stocks for which hedge funds and retail investors show consistent trading direction. "Consistent" trading direction is defined as scenarios when individuals intensively buy stocks that hedge funds regard as undervalued, or when they intensively sell stocks that hedge funds treat as overvalued. As comparison, benchmark stocks are selected based on trading signal from only one investor group, either individuals or hedge funds. If information sources from individuals and hedge funds are complements, stocks with *consistent* trading direction are likely to exhibit a stronger informational signal compared to benchmark stocks. Alternatively, the

[6]The sample period does not include the years before 2010 because the subpenny trade practice did not stabilize until 2009.

informational signal from stocks with *consistent* trading direction is no larger than that of the benchmark stocks, if information sources from individuals and hedge funds are substitutes.

The first set of the empirical results suggests that the information possessed by individuals and institutions are distinct in nature, which corroborates the *Complementarity Hypothesis*. Specifically, I examine ex post abnormal returns, which reflect valuable information when it is subsequently revealed, for stocks with *consistent* trading direction. These stocks exhibit strong return predictability, regardless of whether I use short interest or hedge fund holding changes to proxy for hedge fund trading signal. Constructing monthly-rebalanced portfolios with double sorts on both individual and hedge fund trading signals, I find that stocks with intensive retail buying (selling) and with short interest in the bottom (top) 25% of the sample, i.e., those with *consistent* trading direction, earn a significant Carhart alpha as 89 bps (-63 bps) on a monthly basis. Similarly, the monthly Carhart alpha for stocks with intensive retail buying (selling) and an increase (decrease) of hedge fund holdings is significant and as high as 47 bps (-30 bps). Interestingly, benchmark stocks that are constructed from either retail side or institution side yield small and insignificant return, for both the long leg and the short leg.[7] Furthermore, the alphas earned from *consistent* trading are not driven by temporary price pressure. By extending the investment horizon to one year, I find these stocks continue to earn significant abnormal returns without return reversal.

To further distinguish the main hypotheses, I next investigate stocks for

[7]For benchmark stocks with single signal from retail trading, I only include those with short interest (change of hedge fund holdings) that falls within 0.1% distance from its median, in order to reduce the influence from the short interest (change of hedge fund holdings) as much as possible. In the same spirit, I construct benchmark stocks with signal from hedge fund trading by only include those whose retail order flow is is within [-0.1,0.1].

which hedge funds and retail investors show opposite trading directions. Specifically, I define "opposite" trading directions as scenarios when individuals intensively buy stocks that hedge funds regard as overvalued, or when they intensively sell stocks that hedge funds treat as undervalued. The portfolio results show that the abnormal returns of stocks with *opposite* trading directions are small and statistically insignificant. This finding rules out the *Noise trader Hypothesis* that portraits hedge funds as smart traders and retail investors as noise traders, as hedge funds do not earn any abnormal return when trading against individuals. The main results in *consistent* and *opposite* trading are also in line with the rational expectations equilibrium proposed by Diamond and Verrecchia (1981). They document that there exist noises that represent factors other than information which cause prices to vary in the security market, and these noises are imperfectly observed. Consistent with this intuition, trading from each group of investors contains both information and noise, while the noise portion is greatly reduced when combing the two trading signals together.

Overall, the portfolio results of *consistent* and *opposite* trading altogether corroborate the *Complementarity Hypothesis.* But it is possible that information that serve as complements solely come from hedge funds. To explain, Jiao, Massa, and Zhang (2016) treat changes in hedge fund holdings and short interest as two sides of hedge fund trading and suggest the two sides complement each other in terms of information. Therefore, my previous results may still hold in a scenario when trading signal from individual investors does not contribute to stocks' information content. However, results that I show in Fama-MacBeth regressions rule out this possibility, as stocks with *consistent* trading between hedge funds and individual investors exhibit strong return predictability even controlling for changes in hedge fund holdings and short interest simultaneously.[8]

[8] In particular, returns from the long leg of *consistent* trading is much larger compared to those earned from the hedge fund long-side demand. Returns from the short leg of *consistent* trading and returns from the hedge fund short-side demand are comparable in magnitude.

The above results highlight the importance to observe signals from different investor groups when identifying informed trading. Even for smart institutional investors such as hedge funds, combining information signals that come from the institution is not enough. The reason is that these signals are likely to related to one type of firm fundamental, while there remain informed signals about other types of fundamentals potentially come from different investor groups (e.g., Grossman, 1976; Diamond and Verrecchia, 1981; Goldstein and Yang, 2015). Consistent with this intuition, I find trading signal from individuals provides indispensable information beyond hedge fund trading. In particular, when trading opposite to individuals, hedge funds cannot yield any significant returns. Also from retail investor side, negligence toward the signal from institutions would largely affect their stock performance. It is therefore important to incorporate signals from both investor groups to identify valuable firm-level information.

The stock-level information asymmetry greatly affects subsequent return premiums earned by informed investors (e.g., O'Hara, 2003; Easley and O'hara, 2004). If *consistent* trading from individual investors and hedge funds reflects valuable information that is yet absorbed by the market, the corresponding return premium is expected to be higher for stocks with larger information asymmetry. I therefore divide stocks into sub-groups based on their ex ante information asymmetry, for which I use firm size, age, and turnover as proxies (e.g., Llorente, Michaely, Saar, and Wang, 2002; Zhang, 2006), and examine their return predictability. Consistent with this intuition, I find that for stocks with *consistent* trading, the return premiums are larger for firms with smaller size, younger age, and lower turnover. Also, the return premiums are significant in the remaining sub-samples, suggesting that the main findings are not likely to be driven by these firm characteristics.[9]

[9] I further examine firm characteristics of portfolios constructed using *consistent* or *opposite* trading, as well as those with single trading signal. The results suggest that aforementioned return premiums are not driven by specific firm characteristics, such as low stock price, investor disagreement, or lottery features.

Although the previous results show strong return predictability of the joint signal by retail trading and hedge fund trading, there is no direct evidence of its informational content. Therefore, I next investigate whether stocks with *consistent* trading from two groups predict firm fundamentals. The first set of proxies for firm fundamentals is related to the future earnings. I use both standardized unexpected earnings (SUE) and the 3-day cumulative abnormal returns (CAR) around the upcoming earnings announcements to measure stocks' cash flow realizations. I find that the long (short) leg of the *consistent* trading significantly predicts both a CAR of 5.3% (-6.6%) and an SUE of 1.0% (-1.3%). Furthermore, I check whether *consistent* trading can predict recommendation revisions by analysts, changes in returns on assets (ROA), as well as public news in the media. Fama-MacBeth regression results suggest that both long and short legs of the *consistent* trading has significant predicting power for firm fundamentals, informational events as well as market reactions.

Prior studies on the informational content of trades from both hedge funds and individual investors suggest that these two investor groups possess private information (e.g., Massoud, Nandy, Saunders, and Song, 2011; Kelley and Tetlock, 2013; Qian and Zhong, 2018; Boehmer, Jones, Zhang, and Zhang, 2021). However, the above results have yet provided evidence of whether the *consistent* trading by two groups contains any private information. It is likely that the fundamental predictability is solely due to these investors' superior ability to interpret public news information. To further examine their informational content, I follow Boehmer et al. (2020) and decompose both the long and the short leg of *consistent* trading into trades driven by public news and trades driven by private information, where I use earnings announcements, analyst actions and media news to proxy for public news.[10] The regression results suggest that both

[10]In the panel regressions, I include four event dummies that are related to earnings announcement, analyst recommendation change, analyst earnings forecast change, and public news from Thomson Reuters News.

legs of *consistent* trading contain information that goes beyond public news for predicting firm fundamentals.[11]

Furthermore, I examine the relation between *consistent* trading and stock anomalies, in order to find out whether both legs of such trading is in the same direction with stock mispricing. I construct two composite mispricing scores (MGMT and PERF) by following Stambaugh, Yu, and Yuan (2012) and also use the mispricing measure (MISP) constructed by Stambaugh, Yu, and Yuan (2015) to proxy for stock-level mispricing.[12] By sorting stocks into quintile portfolios based on their three-month historical average of each mispricing proxy, I find that the percentage of the number of stocks that in the long (short) side of *consistent* trading significantly increases (decreases) from the overvalued portfolio to the undervalued portfolio. Such finding indicates that when hedge funds and retail investors trade in the same direction, they jointly tilt their trades to be in line with well-known anomalies. As a result, *consistent* trading by these two group of traders is indeed smart trading.

After observing that *consistent* trading by hedge funds and individual investors is in the same direction with anomaly prediction, I next test whether hedge funds and individuals jointly serves as arbitragers that improve stock market efficiency. Indeed, I find that *consistent* trading greatly alleviate stock-level mispricing. By examining the return spread based on aforementioned mispricing proxies for stocks in both long and short side of *consistent* trading, I show that large and significant Carhart alphas in the full sample have greatly attenuated and become statistically insignificant for stocks in both sides of *consistent* trading. The evidence here corroborates the arbitrager hypothesis that by trading in the same direction, hedge funds and individual investors jointly correct

[11] The results are robust when I apply the same decomposition method to check return predictability.

[12] I classify net stock issuance, accrual, asset growth, and investment to assets into MGMT cluster, and include financial distress, medium-term momentum, gross profitability, and return on assets in PERF cluster.

stock-level mispricing and improve stock efficiency.

One of the main contributions of this paper is to further the understanding of the information diversity between institutions and individual investors and their aggregated informational role in the stock market. To the best of my knowledge, this paper is the first to combine trading signals from institutions and individuals to study their information content. The paper contributes to a bunch of theoretical studies that highlight the importance to investigate the interaction between different types of informed investors to understand the corresponding market efficiency (e.g., Grossman and Stiglitz, 1980; Diamond and Verrecchia, 1981). While related empirical studies mostly focus on heterogeneity of trades that are from professional traders such as active mutual funds (Jiang and Sun, 2014) and hedge funds (Jiao, Massa, and Zhang, 2016), this paper shows the importance to take trading signal from both individual investors and institutional investors when identifying informed demand, as these two groups are likely to be informed about different aspects of the firm's value (Goldstein and Yang, 2015). I show that negligence toward the signal from either group would largely affect subsequent stock performance.

This paper is also closely related to the literature on the relation between the nature of information from different groups. Studies such as Goldstein and Yang (2015) and Paul (1993) suggest that information signals from two groups can be either complements or substitutes. Results in my paper suggest that the information possessed by individuals and institutions are distinct in nature, which corroborates the hypothesis that they serve as complements. Furthermore, my findings are in line with the rational expectations equilibrium proposed by Diamond and Verrecchia (1981), where trading solely by institutions or individuals contains both information and noise, while the noise portion is greatly reduced when combing the two trading signals together.

Lastly, this study contributes to the literature related to the informativeness of both retail investors and hedge funds. Earlier papers document that retail investors are generally uninformed and make systematic mistakes on investment (e.g., Barber and Odean, 2000; Frazzini and Lamont, 2008; Barber, Odean, and Zhu, 2009), while recent studies show that retail trading, especially off-exchange retail order flow, exhibits strong stock return predictability and potentially contain valuable information (e.g., Kaniel, Saar, and Titman, 2008; Kelley and Tetlock, 2013; Boehmer, Jones, Zhang, and Zhang, 2021). This paper provides further evidence of retail trading's informational role by showing that it provides additional valuable information beyond hedge fund trading. More importantly, I add to the hedge fund trading literature (e.g., Ackermann, McEnally, and Ravenscraft, 1999; Kosowski, Naik, and Teo, 2007; Aggarwal and Jorion, 2010; Cao, Liang, Lo, and Petrasek, 2018), by pointing out that when trading in the opposite direction to individuals, hedge funds cannot yield any significant returns subsequently. Such finding again indicates the importance of incorporating signal from informed retail trading when analyzing institutions' information role.

The remaining part of this paper is organized as follows. Section 2.2 describes the data and reports the summary statistics. Section 2.3 presents the return predictability of combined trading signal from hedge funds and individual investors. Section 2.4 links the combined trading signal from two investor groups with firm fundamentals, stock anomalies and price efficiency. Section 2.5 concludes.

2.2 Data and summary statistics

2.2.1 Data sources and sample construction

There are three main data sources used to construct the sample, which are retail order imbalance data, short selling data, as well as hedge fund holding data. I use short interest and change in hedge fund holdings as two alternative proxies for hedge fund trading in this paper. The reason I choose to use short interest as one proxy is that the vast majority of short sellers are hedge funds (Jiao, Massa, and Zhang, 2016) and the literature document that low (high) level of short interest contain positive (negative) information (e.g., Asquith, Pathak, and Ritter, 2005; Karpoff and Lou, 2010; Dechow, Hutton, Meulbroek, and Sloan, 2001; Desai, Ramesh, Thiagarajan, and Balachandran, 2002; Diether, Lee, and Werner, 2009; Boehmer, Huszar, and Jordan, 2010). For example, Goldman Sachs's report "Hedge fund trend monitor" documents that hedge funds accounted for 85% of total short interest positions, or $361 billion, as of December 31, 2009. First, to get the short interest ratio (SR), I obtain the monthly short interest data from Compustat, and scale it by the number of shares outstanding extracted from Center for Research in Security Prices (CRSP). The short interest data, i.e., the number of shares held short for a specified stock, is collected on the 15th business day of each month.[13]

Daily retail order flow originates from TAQ trade dataset between January 3, 2010 and December 31, 2018. The sample period does not include the years before 2010 because the subpenny trade practice did not stabilize until 2009. I follow Boehmer et al. (2021) using subpenny price improvement measure to isolate the marketable orders of retail investors from institutional orders. Specifically, I identify a transaction as a retail buy if the subpenny price ranges between

[13]After September 2007, the data has been reported both at the middle and at the end of each month. For the main analysis, I use the short interest data reported in the middle of each month during the sample period (2010-2018). The results are not materially affected if I instead use the short interest reported at the end of the month.

60 and 100 basis points, and a retail sale if the subpenny price ranges between 0 and 40 basis points. I thus obtain retail buy volume and retail sell volume for each stock on each trading day. Retail order imbalance is computed as the difference between the number of shares bought and sold by off-exchange retail investors, divided by the total number of shares traded by them. The daily retail order imbalance is then aggregated on a monthly basis.

[Insert Figure 2.1 here]

Although short interest are short positions submitted to exchanges and are less likely to overlap with off-exchange retail order flow, I still remove shares that are shorted by retail short sellers when calculating retail order imbalance, for the sake of conservation. I first obtain short sale trade data from the Financial Industry Regulatory Authority (FINRA).[14] Next, I use the subpenny price improvement algorithm (Boehmer et al. (2021)) to identify retail short sale transactions from the short sale transactions reported on FINRA, and scale the retail short selling volume by total trading volume and also aggregate it on a monthly basis, in order to get standardized retail short selling (RSS) for each stock in each month. Lastly, I exclude the retail short selling (RSS) from general retail order imbalance calculated previously to get a cleaner version of retail order imbalance (OIB). The time series of the monthly average retail order imbalance (OIB) and short interest (SR) are plotted in Figure 2.1. Trading pattern from hedge funds and retail investors show opposite signals. Except for some episodes of draw downs, both of the values have continuously grown over time. While an increasing order imbalance indicates a tilt towards buying, a greater short interest level suggests that hedge funds short sell stocks more intensively.

[14] To increase the transparency of short selling, FINRA makes short sale transactions public available since August 2009. Our retail short selling sample covers from January 2010 to December 2016.

Therefore, it seems that hedge funds and individual investors do not trade in a consistent way in the aggregate level.

I also obtain hedge fund holding data from the Thomson Reuters Institutional Holdings (13F) database. As 13F database does not distinct hedge funds from other institutions, I identify hedge funds by manually checking their SEC ADV forms, following Griffin and Xu (2009) and Jiao, Massa, and Zhang (2016). Specifically, I first exclude institutions that have are classified as "bank trust departments" and "insurance companies" from 13F institutions. Next, I download SEC ADV forms for the remaining institutions and use several filters to identify hedge funds. I refine the sample that only includes institutions with more than 50% of its investments listed as "other pooled investment vehicles," which are private investment companies, private equity, and hedge funds, or more than 50% of its clients listed as "high net worth individuals". I then exclude institutions that do not charge performance-based fees, and also check the remaining institutions one by one through their company websites or third-party aggregator websites to only include the ones whose primary business is hedge fund-related activity. Lastly, I obtain 13F quarterly holding data for institutions that are identified as hedge funds.

To construct the sample, I merge stocks that have non-missing retail order imbalance data, short interest ratio, and hedge fund holdings with CRSP sample, after selecting the common stocks (share code equals 10 or 11) listed on the NYSE, AMEX or Nasdaq. The sample period covers from January 2010 to December 2018. The accounting variables are obtained from Compustat. The analyst forecast and recommendation data is acquired from the Institutional Brokers' Estimate System (IBES), and the data on general institutional holdings is gathered from Thomson Reuters (13F). Moreover, I use the web-scraping technique to collect weekly Google's Search Volume Index (SVI) for each stock ticker in the sample, and construct abnormal retail investor attention in the same way as in Da, Engelberg, and Gao (2011).[15] The weekly SVI is collected for 5,524 distinct firm tickers between April 2009 and December 2018, and are merged to the main sample. Following Da, Engelberg, and Gao (2011), I use abnormal search volume index (ASVI) as proxy for retail investor abnormal attention. ASVI is calculated as the log of weekly SVI minus the log of median SVI during the previous eight weeks, and is aggregated on a monthly basis.

2.2.2 Key variables and descriptive summary

The main independent variables are constructed as follows. At the end of each month, I rank all sample stocks based on short interest (SR) or retail order imbalance (OIB) or change in hedge fund holdings (Δ HF holding) and sort them into quartile portfolios. The portfolios are held for one month. The indicator of the long leg of *consistent* trading between individuals and hedge funds (CON_LONG) equals one if the stock has retail order imbalance in the highest quartile and has short interest in the lowest quartile, and equals zero otherwise. The indicator of the short leg of *consistent* trading between individuals and hedge funds (CON_SHORT) equals one if the stock has retail order imbalance in the lowest quartile and has short interest in the highest quartile, and equals zero otherwise. Alternatively, I also use changes in hedge fund holdings instead of short interest as proxy for hedge fund trading. In this scenario, "CON_LONG" equals one if the stock has both retail order imbalance and changes in hedge fund holdings in the highest quartile, and equals zero otherwise. "CON_SHORT" equals one if the stock has both retail order imbalance and changes in hedge fund holdings in the lowest quartile, and equals zero otherwise.

There are two scenarios for *opposite* trading between individuals and hedge funds. The indicator of the first scenario (OPP_1) equals one if the stock has

[15] Similar to Da, Engelberg, and Gao (2011), I manually remove stocks with tickers that have generic meanings such as "ALL", "B", and "GPA", as these generic-meaning tickers would cause ambiguity and create noise.

both retail order imbalance and short interest in the lowest quartile, and equals zero otherwise. The indicator of the second scenario (OPP_2) equals one if the stock has both retail order imbalance and short interest in the highest quartile, and equals zero otherwise. Following Jiao, Massa, and Zhang (2016), I define the indicator of the long leg of hedge fund trading (HF_LONG) to be one if the stock has changes in hedge fund holdings in the highest quartile and has changes in short interest in the lowest quartile, and to be zero otherwise. I also define the indicator of the short leg of hedge fund trading (HF_SHORT) to be one if the stock has changes in hedge fund holdings in the lowest quartile and has changes in short interest in the highest quartile, and to be zero otherwise.

I use standard control variables in the main regressions. Firm size (LSIZE) is taken as the natural logarithm of market capitalization. Book-to-market ratio (BM) is the most recent fiscal year-end book value divided by the market capitalization. Momentum (MOM) is the past cumulative returns from month -12 to month -2. Short-term reversal (RET1) is the prior month's return. Retail investor attention (ASVI) is the log of weekly SVI minus the log of median SVI during the previous eight weeks, and is aggregated on a monthly basis. Institutional Ownership (IO) is the number of shares held by institutions (13F filings) in the most recent quarter, normalized by the number of shares outstanding. Turnover (TURN) is the monthly trading volume divided by the number of shares outstanding, averaged over the past 12 months. Retail short selling (RSS) is the volume of off-exchange retail short selling volume obtained on FINRA scaled by stock total trading volume. In regressions related to fundamentals, I additionally include dividend yield (DIV), firm age (LAGE) and general attention (ATT) as control variables. Dividend yield (DIV) is the percentage of dividends paid to shareholders over the stock's total market capitalization. Firm age is calculated as the as the natural logarithm of the number of years since the firm was first covered by CRSP. General attention is computed as the daily trading volume divided by the average trading volume of the previous year (252 trading days) and then aggregated to monthly frequency (Barber and Odean, 2000). All the explanatory variables are winsorized at the 1% and 99% levels and standardized at the end of each month.

[Insert Table 2.1 here]

Table 2.1 presents the summary statistics. In Panel A, the mean, median, standard deviation, 25th percentile and 75th percentile of the main variables are reported. The indicator variable of *consistent* trading has an average value of 6.7% for the long side and 3.8% for the short side, suggesting that there are 10.5% of the time when hedge funds and retail investors trade in a consistent way. As comparison, the indicator variable of *opposite* trading has an average value of 15.4% when the two scenarios are combined. It means that situations when hedge funds and retail investors trade in the opposite direction occur around 15% of the time. Overall, retail investors and hedge funds are more likely to trade in the opposite direction. Furthermore, considering changes in hedge fund holdings and short interest to capture a more comprehensive hedge fund trading as in Jiao, Massa, and Zhang (2016), I find that there are 7.1% of the time when hedge funds exhibit strong buying signal and also 7.1% of the time when they intensively sell. Panel B reports the Spearman correlation matrix for these variables, which imply relations between main variables and firm characteristics. I further report characteristics variables in different trading samples in Table A1.

2.3 Combined trading signal and stock performance

Results from Section 2.3.1 conform to the informativeness of trades from both hedge funds and retail investors in the aggregate level. Given that both groups predict stock returns, I interact their trading signals and test subsequent return patterns using portfolio analysis and Fama and MacBeth (1973) regressions in the remaining sub-sections.

2.3.1 Univariate portfolio sorting

Recent evidence of retail trading, especially off-exchange retail order flow, proves its strong stock return predictability beyond liquidity provision (e.g., Kaniel, Liu, Saar, and Titman, 2012; Kelley and Tetlock, 2013; Boehmer, Jones, Zhang, and Zhang, 2021). Such finding is in contrast with extant literature that regards individual investors as noise traders who make systematic mistakes on investment (e.g., Barber and Odean, 2000; Frazzini and Lamont, 2008; Barber, Odean, and Zhu, 2009). The informational role of hedge funds is less controversial. Prior literature portraits these institutions as the most sophisticated investors, in terms of the strong predictive power of both their trades (e.g., Ackermann, McEnally, and Ravenscraft, 1999; Kosowski, Naik, and Teo, 2007; Aggarwal and Jorion, 2010; Cao, Liang, Lo, and Petrasek, 2018) and short interest level (e.g., Asquith, Pathak, and Ritter, 2005; Karpoff and Lou, 2010; Dechow, Hutton, Meulbroek, and Sloan, 2001; Desai, Ramesh, Thiagarajan, and Balachandran, 2002; Diether, Lee, and Werner, 2009; Boehmer, Huszar, and Jordan, 2010). In this sub-section, I further examine whether the out-of-sample performance of these two groups remains significant.

At the end of each month, I rank all sample stocks based on short interest (SR) or retail order imbalance (OIB) or change in hedge fund holdings (Δ HF holding) and sort them into quartile portfolios. The portfolios are held for one month. The equal-weighted holding period returns on each quartile as well as on the long-short portfolio – that longs stocks in the highest quartile and shorts stocks in the lowest quartile – are reported. I calculate corresponding Fama and French (1993) three-factor alphas, Carhart (1997) alphas, and adjusted alphas using Fama and French (2015) five factors plus momentum factor plus liquidity factor for each portfolio.

[Insert Table 2.2 here]

Table 2.2 presents the abnormal returns using univariate portfolio sorting method. As shown in Panel A, a long-short strategy based on short interest generates a significant Fama-French three-factor alpha of 1.14% (t-stat = 6.38) on a monthly basis. The Carhart alpha equals 1.08% with a t-statistic of 6.44, and the adjusted alphas using Fama-French five factors + momentum factor + liquidity factor equals 1.03% (t-stat = 5.96). The abnormal returns in all specifications decrease monotonically from the lowest short interest quartile to the highest short interest quartile. Moreover, both the lowest and the highest quartile exhibit significant predictability, consistent with studies that suggest a low (high) level of short interest contains positive (negative) information. In Panel B and C, the abnormal returns for long-short portfolios remain statistically significant. However, compare with the returns shown in Panel A, the magnitude of returns decrease by half. Also, these returns do not monotonically increase along the legs. For example, in Panel B, the Fama-French three-factor alpha on the lowest quartile of retail order imbalance is -0.16% and further declines to -0.25% in Quartile 2, and then rises and becomes positive and significant as 0.35% per month in the highest quartile. The insignificant Carhart alpha in the lowest quartile together with the significantly positive alpha in the highest quartile indicates that retail investors are more informative when purchasing stocks rather than when selling stocks.[16]

2.3.2 Double sorts on hedge fund trading and individual trading

By interacting trading signals from hedge funds and individual investors, who are likely to be informed in different aspects of the stock value, this paper aims to investigate the information diversity between two parties and their aggregated informational role in the stock market. A bunch of theoretical papers aggregate informational signals from different groups of investors (e.g., Grossman, 1976; Diamond and Verrecchia, 1981; Admati and Pfleiderer, 1987). In general, there are two assumptions regarding the relationship between the nature of information from two groups, either to be complements or to be substitutes. As for the first type, Goldstein and Yang (2015) show that two types of traders are informed in different aspects of firm value and acquiring these two types of information can be complementary. Paul (1993), on the other hand, documents a substitution effect that arises from competition among two types of investors trading on the same type of information. In this paper, I propose two hypotheses that the informational signals from hedge funds and retail investors are complements or substitutes, where both groups of investors are assumed to be informed to some extent. I also propose a third hypothesis that institutions are smart traders while individual investors are noise traders who do not possess valuable information.

The *Complementarity Hypothesis* posits individual investors as smart traders who may have access to novel information sources that hedge funds either do not have or are constrained to trade on, due to concerns such as conflict of interest or

[16] As I exclude retail short selling, which are proved to be informative, from the general retail order imbalance, the short leg of portfolio sorted on OIB does not exhibit significant return. Including retail short selling in OIB calculation yields significant and negative return in the short leg.

binding regulation. On the other hand, the *Substitutability Hypothesis* assumes smart retail investors and hedge funds obtain the same type of information and both make investments on such information. Under the former (latter) hypothesis, combining trading signals from individuals and hedge funds yields (does not yield) superior stock performance than checking each signal unconditionally.[17] As for the *Noise trader Hypothesis*, individual investors are assumed to make investment mistakes and hedge funds make profits serving as counterparty to individuals.

To disentangle these hypotheses, I first investigate the subsets of stocks for which hedge funds and retail investors show consistent trading direction. At the end of each month, I sort all sample stocks into quartiles based on retail order imbalance and independently into quartiles based on hedge fund trading, for which I use either short interest or change in hedge fund holdings as proxy. I examine their ex post abnormal returns, which reflect valuable information when it is subsequently revealed. The corresponding double sorting results are presented in Table 2.3.

[Insert Table 2.3 here]

In both Panel A and Panel C of Table 2.3, stocks with *consistent* trading between hedge funds and individual investors exhibit strong return predictability, regardless of whether I use short interest or hedge fund holding changes to proxy for hedge fund trading. In Panel A, stocks with intensive retail buying (OIB in the highest 25% of the sample) and with short interest in the bottom 25% of the sample earn a significant Carhart alpha as 89 bps (t-stat = 4.76) next month. For the other side of *consistent* trading, i.e., retail investors intensively sell (OIB in the lowest 25% of the sample) and short interest falls in the top 25%

[17] The *Complementarity Hypothesis* (*Substitutability Hypothesis*) corresponds to "aggregation" ("unlocking") effect in Admati and Pfleiderer (1987).

of the sample, the Carhart alpha is -63 bps (t-stat = -2.96). Using change of hedge fund holdings to replace short interest, I find similar results in Panel C. The monthly Carhart alpha for stocks with both OIB and Δ HF holding in the highest (lowest) 25% of the sample is significant and as high as 47 bps (-30 bps), with a t-statistic of 4.43 (-2.31).

To alleviate the concern that the return premiums from double sorts are driven by trading from one group of investors, I further construct benchmark stocks as comparison. If information sources from individuals and hedge funds are complements, stocks with *consistent* trading direction are likely to exhibit a larger return premium compared to benchmark stocks. Alternatively, the return premium from stocks with *consistent* trading direction is no larger than that of the benchmark stocks, if information sources from individuals and hedge funds are substitutes. For benchmark stocks with only retail trading, I sort on their retail order imbalance and treat the highest (lowest) quartile as the long (short) side, i.e., "R_LONG" ("R_SHORT"). For benchmark stocks with retail trading in Panel B, where I use short interest to proxy for hedge fund trading, I restrict their short interest to be within 0.1% distance from its cross-sectional median, in order to reduce the influence from hedge funds as much as possible. Similarly, I remove stocks that have change of hedge fund holdings falling outside of the 0.1% distance from its cross-sectional median, to construct benchmark stocks with retail trading in Panel D. In the same spirit, I construct benchmark stocks with only hedge fund trading without influence from retail traders, i.e., "HF_LONG" and "HF_SHORT", by removing those whose retail order flow is is outside of [-0.1, 0.1], in both Panel B and Panel D. Interestingly, benchmark stocks that are constructed from either retail side or institution side yield small and insignificant return, for both the long leg and the short leg.

The above results for stocks with *consistent* trading between hedge funds and individual investors corroborate the *Complementarity Hypothesis*. That is, trading signal from either hedge funds or retail investors provides indispensable value beyond the information contained in other group of investors. It is thus necessary to consider the bundle of both signals when analyzing informed trading. Next, I investigate stocks for which hedge funds and retail investors show opposite trading directions, i.e., either individuals intensively buy stocks that hedge funds regard as overvalued or individuals intensively sell stocks that hedge funds treat as undervalued. In Panel A and C of Table 2.3, the abnormal returns of stocks with *opposite* trading are small and statistically insignificant. This finding rules out the *Noise trader Hypothesis* that portraits hedge funds as smart traders and retail investors as noise traders, as hedge funds do not earn any abnormal return when trading against individuals. Overall, results from Table 2.3 support the *Complementarity Hypothesis* and are also in line with the rational expectations equilibrium proposed by Diamond and Verrecchia (1981), where trading from each group of investors contains both information and noise and the noise portion is greatly reduced when combing the two trading signals together.

2.3.3 Additional control on two sides of hedge fund trading

Still, it is possible that the aforementioned complementary information solely come from hedge funds, as Jiao, Massa, and Zhang (2016) suggest the signals from change in hedge fund holdings and short interest complement each other in firm-level information. In this scenario, information from individual investors does not contribute to stocks' informational content beyond hedge fund trading. To address this concern, I examine stock performance of *consistent* and *opposite* trading (using short interest to proxy for hedge fund trading) conditioning on whether hedge fund holdings are increase or decrease contemporaneously, where an increase (decrease) in Δ HF holding means change in hedge fund holding is in the highest (lowest) quartile.

[Insert Table 2.4 here]

The main results in Table 2.4 remain robust with the additional control, as both long side and short side of *consistent* portfolios earn significant return premiums. It is important to note that even when hedge fund holdings increase simultaneously, "CON_SHORT" portfolio earns a negative return, though its magnitude and significance drops. Similarly, when hedge fund holdings decrease simultaneously, "CON_LONG" portfolio earns significantly positive return. Therefore, incorporating informational signal from individuals indeed adds significant value to hedge funds trading. Furthermore, I conduct Fama-MacBeth regressions with main variables of interest and variables controlling for changes in hedge fund holdings and short interest simultaneously.

[Insert Table 2.5 here]

Table 2.5 presents the Fama-MacBeth regression results. To specify, the indicator of the long (short) leg of *consistent* trading between individuals and hedge funds "CON_LONG" ("CON_SHORT") equals one if the stock has retail order imbalance in the highest (lowest) quartile and has short interest in the lowest (highest) quartile, and equals zero otherwise. As for indicators related to *opposite* trading, "OPP_1" ("OPP_2") equals one if the stock has both retail order imbalance and short interest in the lowest (highest) quartile, and equals zero otherwise. To control for contemporaneous changes in hedge fund holdings and short interest, I follow Jiao, Massa, and Zhang (2016) to construct "HF_LONG" ("HF_SHORT") that equals one if the stock has changes in hedge fund holdings in the highest (lowest) quartile and has changes in short interest in the lowest (highest) quartile, and equals zero otherwise. I further include variables controlling for firm size, book-to-market, momentum, short-term reversal, retail investor attention, turnover, institutional ownership, as well as retail short selling, in columns (2) and (3). In all specifications, stocks with *consistent* trading between hedge funds and individual investors exhibit strong return predictability even controlling for changes in hedge fund holdings and short interest simultaneously. In contrast, stocks with *opposite* trading do not yield significant returns. Also, returns from the long leg of *consistent* trading is much larger than those earned from the hedge fund long-side demand, and returns from the short leg of *consistent* trading are comparable in magnitude with those from the hedge fund short-side demand.

Results from Table 3 to 5 overall highlight the importance to observe trading signals from different investor groups when identifying informed trading. More importantly, they indicate that combining information signals that solely come from institutional investors is not enough, even considering the most sophisticated investors like hedge funds. Potential explanation could be that trading signals from institutional investors are related to one type of firm fundamental, while informed signals about other types of fundamentals may come from different investor groups (e.g., Grossman, 1976; Diamond and Verrecchia, 1981; Goldstein and Yang, 2015). Here, trading signal from individuals indeed provides indispensable information beyond hedge fund trading. When trading opposite to individuals, hedge funds cannot yield any significant returns. Also from retail investor side, negligence toward the signal from institutions would largely affect their stock performance. Therefore, it is important to incorporate signals from both investor groups to identify valuable firm-level information.

2.3.4 Information asymmetry and consistent trading

According to prior studies, returns are positively related to ex-ante information asymmetry for informed trading (e.g., Diamond and Verrecchia, 1991; Verrecchia, 2001; O'Hara, 2003; Easley and O'hara, 2004). If *consistent* trading from individual investors and hedge funds reflects valuable information that is yet absorbed by the market, the corresponding return premium is expected to be higher for stocks with larger information asymmetry. Therefore, I sort stocks on their ex-ante information asymmetry level, and examine whether the return premiums become larger for stocks with a higher level of information asymmetry.

[Insert Table 2.6 here]

Table 2.6 reports Fama-MacBeth regression results for subsamples of stocks that are constructed based on different proxies of ex-ante information asymmetry, which are market capitalization, firm age, and turnover (e.g., Llorente, Michaely, Saar, and Wang, 2002; Zhang, 2006). At the end of each month, all sample stocks are sorted into halves based on an information asymmetry proxy. They are then divided into quartiles using double sorting on retail order imbalance and short interest within each subsample. I measure firm size as the firm's market capitalization, firm age as the number of years since the firm was first covered by CRSP, and turnover as trading volume divided by shares outstanding. In all specifications, I find that return premiums of stocks with *consistent* trading are generally larger for firms with smaller size, younger age, and lower turnover. The return premiums are significant in the remaining sub-samples as well, which suggests that my main findings are not likely to be driven by these firm characteristics. The findings here are in contrast to Gromb and Vayanos (2010), who suggest a higher level of mispricing created by the irrational buying decisions of individual investors.

2.3.5 Return Persistence

As for the significant returns earned from *consistent* trading between hedge funds and retail investors, an alternative explanation apart from informed trading could be price pressure hypothesis. For example, Chordia and Subrahmanyam (2004) point out that persistent order flow could lead to high stock returns. Hence, I plot the cumulative abnormal returns for stocks with *consistent* trading along the holding period.

[Insert Figure 2.2 here]

Figure 2.2 presents the cumulative DGTW-adjusted returns of both sides of *Consistent* trading over the following twelve months. In Panel (a), the cumulative DGTW-adjusted returns in the long leg of *Consistent* trading increase over time and exceed 2.9% in month t+12. The cumulative abnormal returns in Panel (b) decline sharply and drop below -5% in month t+12 for the short leg of *Consistent* trading. Overall, I find stocks with *consistent* trading continue to earn significant abnormal returns without return reversal.

2.4 Fundamentals, anomalies and stock mispricing

The previous results from portfolio sorting and Fama-MacBeth regressions show strong return predictability of the joint signal by retail trading and hedge fund trading. However, evidence of its informational content is indirect. In this section, I further investigate the fundamental predictability of stocks with *consistent* trading from two groups, and check whether it goes beyond the skill to interpret public news. I also examine the relation between *consistent* trading and stock anomalies, in order to find out whether such trading helps reduce stock mispricing.

2.4.1 Firm fundamental predictability

To check the informational content of *consistent* trading between two groups, I examine whether such trading predicts firm fundamentals. The first set of proxies for firm fundamentals is related to the future earnings. I use both standardized unexpected earnings (SUE) and the 3-day cumulative abnormal returns (CAR) around the upcoming earnings announcements to measure stocks' cash flow realizations. SUE is computed as the seasonal difference in adjusted earnings per share scaled by the month-end adjusted price. I use value-weighted market return to adjust returns when calculating CAR. The second set of fundamental proxy is analyst revision, which is computed as the difference in average analyst earnings estimates from the previous month divided by the stock price at the end of the previous month. Furthermore, I also check subsequent changes in returns on assets (ROA), where I calculate ROA using the income before extraordinary items divided by total assets. Lastly, I measure tone in public news, i.e., the average stock-level news tone (negative, neutral, or positive) from Thomson Reuters (TR) news source.

[Insert Table 2.7 here]

Table 2.7 reports monthly panel regressions of firm fundamentals on both the long and short leg of *consistent* trading between hedge funds and retail investors. As for control variables, I include firm size, book-to-market ratio, momentum, short-term reversal, abnormal trading volume that, retail investor attention, dividend yield, firm age, institutional Ownership, retail short selling, as well as hedge fund holdings in the regressions. All specifications include month fixed effects and standard errors are two-way clustered at the firm and the month level. I find that the long (short) leg of the *consistent* trading significantly predicts positive (negative) CAR and SUE of firms. For stocks with *consistent* buying, their upcoming CAR will be 5.3% higher and SUE will be 1.0% larger than those without *consistent* buying. Also for *consistent* selling, SUE will decrease by -1.3% and CAR will drop by -6.6%, compared to the remaining stocks in the sample. As for upcoming analyst revision, CON_LONG predicts positive analyst revisions, while CON_SHORT also forecasts positive analyst revisions, though the magnitude is much smaller. It is likely to be the case as prior studies suggest that analysts may bias their opinions by upward adjusting their estimates to avoid earnings disappointment (e.g., Chan, Karceski, and Lakonishok, 2007). Furthermore, the last two columns in Table 2.7 show that the long (short) side of *consistent* trading positively (negatively) predicts changes in ROA and public news tone (one as good news and minus one as bad news). The predictive power remain highly significant. Overall, these results corroborate the notion that *consistent* trading between hedge funds and retail investors reflect these investors' abilities to forecast firm fundamentals.

2.4.2 Decomposition of fundamental predictability

While studies on the informational role of hedge funds and retail investors suggest both groups possess private information to some extent (e.g., Massoud, Nandy, Saunders, and Song, 2011; Kelley and Tetlock, 2013; Qian and Zhong, 2018; Boehmer, Jones, Zhang, and Zhang, 2021), it remains uncertain whether *consistent* trading between two groups contains any private information. The aforementioned predictability of firm fundamentals could be due to these traders' superior ability to interpret public news information instead of trading on private information. Hence, I follow Boehmer et al. (2020) and decompose both long and short legs of *consistent* trading into trades related or not related to public news, in order to examine their informational content. Specifically, I use multiple variables to proxy for public news that are available to the general market. There are in total four indicator variables, which are related to earnings announcement, analyst recommendation change, analyst earnings forecast change, and public news from Thomson Reuters News source. In panel regressions, I include these indicator variables and variables of *consistent* trading, as well as their interaction terms.

[Insert Table 2.8 here]

Table 2.8 reports monthly panel regressions about future firm fundamentals. The dependent variables include standardized unexpected earnings (SUE), analyst revisions in earnings estimates (Revision), as well as changes in firm return on assets (ΔROA) in the following year. When there is no contemporaneous public event or news, stocks that are associated with the long leg of *consistent* trading lead to a 11.9 bps increase in future SUE, 21.6 bps upward revision in analyst forecast, and 0.5 bps increase in tone of news (more positive), all remain statistically significant. Similarly, stocks in the short leg of *consistent* trading negatively predict firm fundamentals.[18] Most coefficients on the interaction terms are insignificant, indicating that the informational source from *consistent* trading is not due to those investors' superior public news analysis. The regression results suggest that both legs of *consistent* trading contain information that goes beyond public news for predicting firm fundamentals.

Furthermore, I examine both legs of *opposite* trading in the fundamental predictability test shown in Table A2. By separating out contemporaneous public news or events, I find that the coefficients on residual terms are not significant. The results here are consistent with the previous findings about *opposite* trading's inability to predict future returns. When hedge funds and retail investors trade in the opposite directions, there is no privileged information contained in their aggregate trading. I next use the same decomposition method to run panel

[18]The t-statistic in column (5) is insignificant and small, consistent with the notion that analysts may upward adjust their estimates to avoid earnings disappointment. Therefore the predictability is not significantly negative.

regressions of return predictability and the results remain robust. Table A3 suggests that when excluding contemporaneous public events in the analysis, stocks in the long leg of *consistent* trading yield a 71 bps higher return and stocks in the short leg of *consistent* trading generate a 107 bps lower return. Again, the coefficients on interaction terms remain insignificant. Overall, it is likely that both legs of *consistent* trading contain some private information.

2.4.3 Consistent trading for anomaly stocks

There has been long-standing interest in the relation of investor groups and stock anomalies/mispricing. For example, Edelen, Ince, and Kadlec (2016) suggest that institutional investors tend to trade on the wrong side of anomaly strategies. A recent study by McLean, Pontiff, and Reilly (2020) study nine types of market participants and their trades with respect to stock anomalies. They find that retail investors trade against anomalies while hedge funds buy more stocks in anomaly-long instead of anomaly-short. However, an analysis on combined trading signal from different investor groups is scant in the literature.

To figure out whether *consistent* trading is on the right side of stock anomalies/mispricing, I examine their relation by checking the percentage of the number of stocks in both sides of *consistent* trading corresponding to different anomaly legs. Following Stambaugh, Yu, and Yuan (2012), I construct two composite mispricing scores (MGMT and PERF) by classifying net stock issuance, accrual, asset growth, and investment to assets into "MGMT" cluster, and including financial distress, medium-term momentum, gross profitability, and return on assets in "PERF" cluster. Furthermore, I use the mispricing measure (MISP) constructed by Stambaugh, Yu, and Yuan (2015) to proxy for stock-level mispricing. At the end of each month, I sort stocks into quintile portfolios based on their three-month historical average of each mispricing proxy and calculate the percentage of the number of stocks from both sides of *consistent* trading with respect to different legs of stock anomalies/mispricing.

[Insert Table 2.9 here]

Table 2.9 indicates that the percentage of the number of stocks that in the long (short) side of *consistent* trading significantly increases (decreases) from the overvalued portfolio to the undervalued portfolio. To specify, the time-series average of the cross-sectional mean of weight in CON_LONG portfolio increases by 1.62% for "MGMT" long-short portfolio, increases by 1.81% for "PERF" long-short portfolio, and increase by 0.88% for "MISP" long-short portfolio. Also, such ratio in CON_SHORT portfolio significantly decreases for all three long-short portfolios. Such finding indicates that when hedge funds and retail investors trade in the same direction, they jointly tilt their trades to be in line with well-known anomalies/mispricing. To supplement the portfolio analysis, I also conduct panel regressions for both long and short sides of *consistent* trading next month on the mispricing measure (MISP) and include standard control variables as well. As a lower mispricing score means the stock is more likely to be undervalued, the regression results in Table A4 suggest that the long leg of *consistent* trading targets more undervalued stocks while the short leg targets more overvalued stocks. As a result, *consistent* trading by these two group of traders can be regarded as smart trading.

2.4.4 Consistent trading and price efficiency

Section 2.4.3 shows that *consistent* trading by hedge funds and individual investors is in the same direction with anomaly prediction. I further test whether such trading improves price efficiency. Specifically, I examine subsequent anomaly returns based on aforementioned proxies for stocks in both long and short side of *consistent* trading. Anomaly/mispricing indices (MISP, MGMT, and PERF) are constructed as in Table 2.9. First, I sort stocks based on their retail order imbalance and short interest into "CON_LONG" or "CON_SHORT", or keep them in full sample. Next, I sort these stocks into quintile portfolios based on their three-month historical average of each anomaly/mispricing proxy, and calculate the corresponding long-short anomaly returns for each sample.

[Insert Table 2.10 here]

Table 2.10 reports the Carhart alphas for long-short anomaly returns in each sample regarding *consistent* trading, where the long-short anomaly returns are calculated as the average of difference between Carhart alphas of the long and short leg anomaly portfolios over one-year horizon. I indeed find that *consistent* trading greatly alleviate stock-level mispricing, as the large and significant Carhart alphas in the full sample have greatly attenuated and mostly become statistically insignificant for stocks in both sides of *consistent* trading. For example, the long-short alpha of "MISP" index is 0.88% (t-stat = 2.66) for stocks in the full sample. It drops to only 0.02% (t-stat = 0.05) among stocks in the long leg of *consistent* trading, and also becomes 0.68% (t-stat = 0.95) among stocks in the short leg of *consistent* trading. The results remain consistent when checking long-short anomaly returns in "MGMT" and "PERF". The evidence overall corroborates the arbitrager hypothesis that by trading in the same direction, hedge funds and individual investors jointly correct stock-level mispricing and improve stock efficiency.

2.4.5 Sub-sample tests

I conduct sub-sample tests to assess the robustness of my main results. Table A1 reports the time-series average of the cross-sectional means of multiple firm characteristics variables in different trading samples defined based on trading signals from hedge funds and/or retail investors. Comparing firm characteristics of portfolios constructed using *consistent* or *opposite* trading, as well as those with single trading signal, I show that return premiums earned through *consistent* trading are not driven by specific firm characteristics, such as low stock price, investor disagreement, or lottery features.

Lastly, Table A5 presents the Fama-MacBeth regressions of next-month stock returns on both long and short sides of *consistent* trading, for subsamples of stocks constructed based on sentiment or disagreement indices, which are BW sentiment index (Baker and Wurgler, 2006), PLS sentiment index (Huang et al., 2015), and Huang, Li, and Wang (2021) disagreement index. The results are in line with Stambaugh, Yu, and Yuan (2012) that during high-sentiment periods, stocks tend to be over-valued and thus are overpriced. In this scenario, the short leg of *consistent* trading invest on such mispricing and yields high returns subsequently. As overpricing during low-sentiment periods is less likely (Stambaugh, Yu, and Yuan, 2012), the short leg of *consistent* trading would not yield significant return. It is not surprising that the sentiment-related overpricing does not much affect the long leg of *consistent* trading. Another strand of literature shows that higher investor disagreement leads to higher average bias and more overvaluation (e.g., Miller, 1977; Diether, Malloy, and Scherbina, 2002; Chen, Hong, and Stein, 2002; Atmaz and Basak, 2018), thereby implying a lower return associated with disagreement level. Again, the results corroborates this explanation as the short leg of *consistent* trading, which targets overvalued stocks, yields significant and negative returns when the disagreement index is high while it generates insignificant returns when the disagreement level is low.

2.5 Conclusion

In this paper, I investigate the information diversity between hedge funds and retail investors through their combined trading. While prior studies focus on heterogeneity of trades that are from professional investors such as active mutual funds and hedge funds, my results highlight the importance to interact signals from retail investors and institutional investors when identifying firm-level information, given that these two groups tend to be informed about different aspects of the firm's value.

Using marketable retail order flow to proxy for retail trading and short interest/changes in hedge fund holdings to represent hedge fund trading, I document evidence in line with the hypothesis that the information possessed by individuals and institutions are distinct in nature and serve as complements. Stocks that reflect consistent trading between two groups exhibit strong return predictability without reversal. For smart institutions like hedge funds, when trading in the opposite direction to retail investors, they cannot yield any significant return, even in a one-year horizon. Furthermore, consistent trading between hedge funds and retail investors significantly predicts firm fundamentals, beyond the ability to process public news such as earnings announcements, analyst revisions and media news. Such trading is in the same direction with stock anomaly returns and helps alleviate stock-level mispricing.

Overall, the findings suggest combining signals that solely from institutional investors is incomplete, as there remain useful information in retail trading that provides indispensable value beyond hedge fund trading. Hence, it is necessary to consider the bundle of both signals when analyzing informed trading. One of the main contributions of this paper is to further the understanding of the information diversity between institutions and individual investors and their aggregated informational role in the stock market. Consistent with prior theoretical studies that focus on interaction between different types of investors to understand the corresponding market outcomes (e.g., Grossman and Stiglitz, 1980; Diamond and Verrecchia, 1981), I show that negligence toward the signal from either group would largely affect subsequent stock performance and price efficiency. My findings are also in line with the rational expectations equilibrium proposed by Diamond and Verrecchia (1981), where trading from each group of investors contains both information and noise and the noise portion is greatly reduced when combing the two trading signals together.

Figure 2.1: Retail Order Imbalance and Short Interest Over Time

This figure plots the time-series monthly average of retail order imbalance and short interest from January 2010 to December 2018. Retail order imbalance, which is shown on the left axis, is computed as the difference between the number of shares bought and sold by off-exchange retail investors, divided by the total number of shares traded by them. I remove shares that are shorted by retail short sellers in this calculation. Short interest is the number of shares shorted over the total shares outstanding and is shown on the right axis.

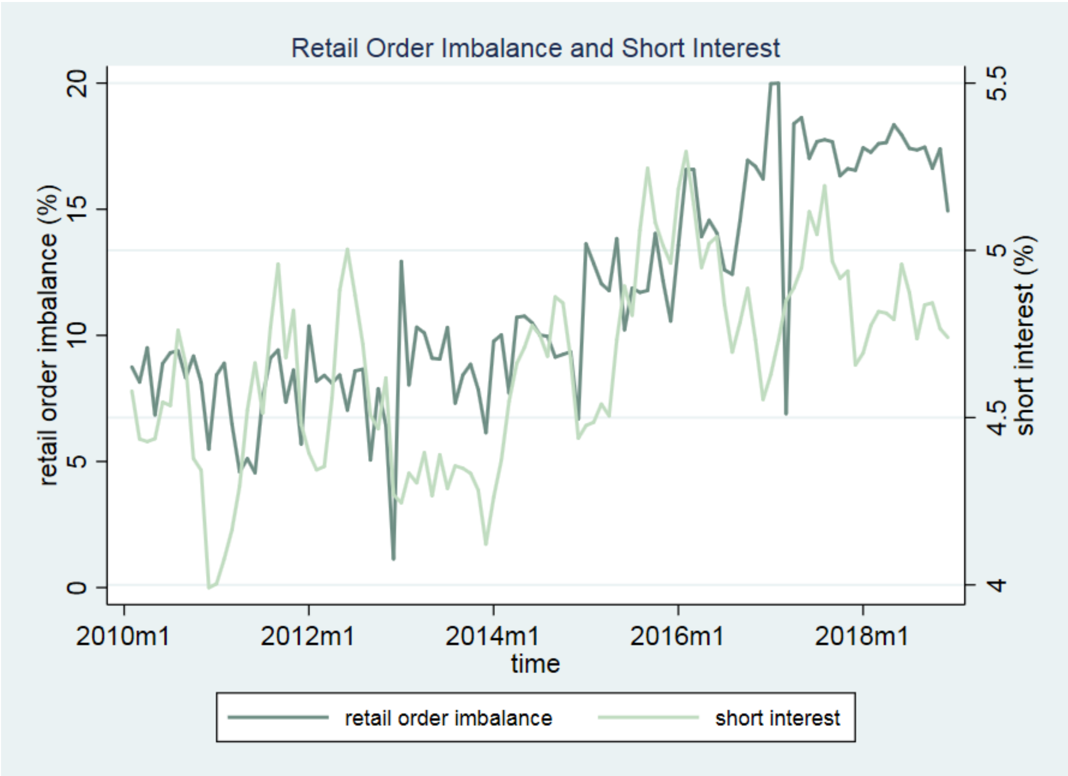
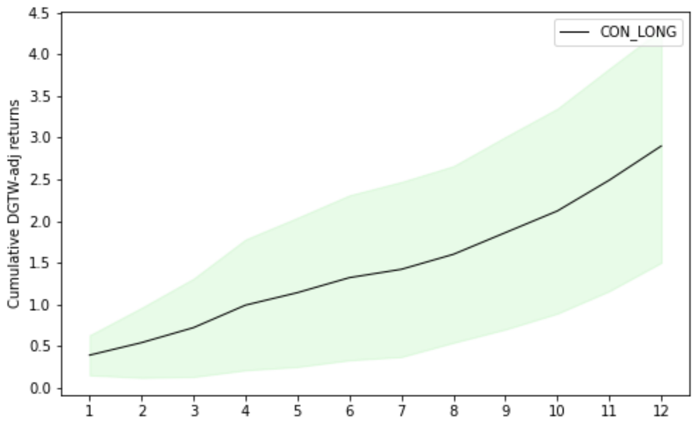


Figure 2.2: Cumulative Abnormal Returns based on *Consistent* Trading

This figure plots the cumulative DGTW-adjusted return (in percent) of both the long leg (Panel A) and the short leg (Panel B) of *consistent* trading between individuals and hedge funds. In Panel A, "CON_LONG" portfolio includes stocks that have retail order imbalance in the highest quartile and have short interest in the lowest quartile. In Panel B, "CON_SHORT" portfolio includes stocks that have retail order imbalance in the lowest quartile and have short interest in the highest quartile. The shaded areas in green correspond to the 90% confidence intervals (CI).

(a): Long side of *consistent* trading



(b): Short side of *consistent* trading

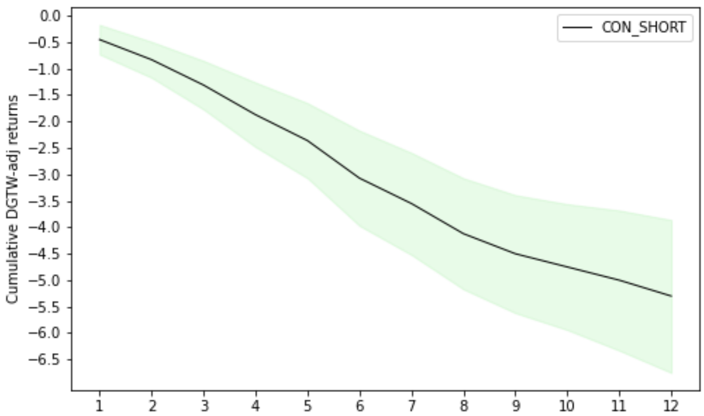


TABLE 2.1: Summary Statistics

This table reports summary statistics for the main variables. The sample period covers from January 2010 to December 2018. Panel A reports their mean, median, standard deviation, and 25% and 75% values. The indicator of the long leg of *consistent* trading between individuals and hedge funds (CON_LONG) equals one if the stock has retail order imbalance in the highest quartile and has short interest in the lowest quartile, and equals zero otherwise. The indicator of the short leg of *consistent* trading between individuals and hedge funds (CON_SHORT) equals one if the stock has retail order imbalance in the lowest quartile and has short interest in the highest quartile, and equals zero otherwise. There are two scenarios for *opposite* trading between individuals and hedge funds. The indicator of the first scenario (OPP_1) equals one if the stock has both retail order imbalance and short interest in the lowest quartile, and equals zero otherwise. The indicator of the second scenario (OPP_2) equals one if the stock has both retail order imbalance and short interest in the highest quartile, and equals zero otherwise. Following Jiao, Massa, and Zhang (2016), I define the indicator of the long leg of hedge fund trading (HF_LONG) to be one if the stock has changes in hedge fund holdings in the highest quartile and has changes in short interest in the lowest quartile, and to be zero otherwise. I also define the indicator of the short leg of hedge fund trading (HF_SHORT) to be one if the stock has changes in hedge fund holdings in the lowest quartile and has changes in short interest in the highest quartile, and to be zero otherwise. I then report statistics of key stock characteristics. Firm size (LSIZE) is the natural logarithm of market capitalization. Book-to-market ratio (BM) is the most recent fiscal year-end book value divided by the market capitalization. Momentum (MOM) is the past cumulative returns from month -12 to month -2. Short-term reversal (RET1) is the prior month's return. Retail investor attention (ASVI) is the log of weekly SVI minus the log of median SVI during the previous eight weeks, and is aggregated on a monthly basis. Institutional Ownership (IO) is the number of shares held by institutions (13F filings) in the most recent quarter, normalized by the number of shares outstanding. Panel B reports the Spearman correlation matrix for these variables. All the explanatory variables are winsorized at the 1% and 99% levels and standardized at the end of each month.

Panel A: Summary statistics of main variables

Variable	Mean	Median	St.Dev	p25	p75
CON_LONG	0.067	0.066	0.008	0.061	0.071
CON_SHORT	0.038	0.038	0.004	0.036	0.041
OPP_1	0.087	0.087	0.008	0.082	0.093
OPP_2	0.067	0.067	0.005	0.064	0.071
HF_LONG	0.071	0.071	0.005	0.067	0.074
HF_SHORT	0.071	0.071	0.006	0.068	0.074
LSIZE ($M)	4,387	4,390	1,080	3,740	5,359
BM	0.68	0.67	0.11	0.58	0.76
MOM	16.13%	16.58%	16.61%	4.70%	25.99%
RET1	1.09%	1.07%	4.33%	-1.37%	4.08%
ASVI	1.85%	1.69%	3.12%	-0.18%	3.59%
IO	58.63%	57.62%	4.05%	55.19%	62.72%

Table 2.1: (Cont.) Summary Statistics

Panel B: Spearman rank correlation

		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
CON_LONG	(a)	1											
CON_SHORT	(b)	0.22	1										
OPP_1	(c)	-0.75	-0.36	1									
OPP_2	(d)	-0.56	-0.34	0.47	1								
HF_LONG	(e)	0.14	0.09	-0.16	0.05	1							
HF_SHORT	(f)	-0.28	0.00	0.35	0.25	0.08	1						
LSIZE	(g)	-0.35	0.15	0.10	0.36	-0.01	0.14	1					
BM	(h)	0.32	-0.35	-0.13	-0.19	0.03	-0.19	-0.73	1				
MOM	(i)	-0.05	0.17	0.05	-0.10	-0.10	0.15	-0.45	-0.02	1			
RET1	(j)	0.12	-0.13	-0.18	-0.13	0.00	-0.10	-0.13	0.16	-0.09	1		
ASVI	(k)	0.24	-0.21	-0.03	-0.13	0.06	0.02	-0.65	0.54	0.22	0.06	1	
IO	(l)	-0.33	0.26	-0.01	0.29	0.00	0.18	0.89	-0.81	-0.22	-0.08	-0.62	1

Table 2.2: Abnormal Returns from Univariate Portfolio Sorting

This table reports monthly Fama and French (1993) three-factor alphas, Carhart (1997) alphas, adjusted alphas using Fama and French (2015) five factors plus momentum factor plus liquidity factor introduced in Pastor and Stambaugh (2003), for each of the quartile portfolios sorted by short interest (Panel A) or retail order imbalance (Panel B) or changes in hedge fund holdings (Panel C), as well as the corresponding long-short portfolios. At the end of each month, I rank all sample stocks based on short interest (SR) or retail order imbalance (OIB) or change in hedge fund holdings (Δ HF holding) and sort them into quartile portfolios. The portfolios are held for one month. The equal-weighted holding period returns on each quartile as well as on the long-short portfolio are reported. Short interest (SR) is the number of shares shorted over the total shares outstanding. Retail order imbalance (OIB) is computed as the difference between the number of shares bought and sold by off-exchange retail investors, divided by the total number of shares traded by them. I remove shares that are shorted by retail short sellers in this calculation. Change in hedge fund holdings (Δ HF holding) is calculated as the monthly change in number of shares held by hedge funds stated in their 13F filings. The sample period is from January 2010 to December 2018. All returns and alphas are in percent. The t-statistics are Newey-West adjusted.

Panel A: Portfolio returns sorted on SR

	FF3 alpha	t-stat	Carhart alpha	t-stat	FF5 + MOM + LIQ alpha	t-stat
Low	0.51	3.09	0.53	3.10	0.55	3.10
2	0.17	2.62	0.18	2.83	0.20	3.03
3	-0.27	-3.55	-0.24	-3.11	-0.20	-3.03
High	-0.64	-4.57	-0.56	-4.21	-0.47	-3.97
Low - High	1.14	6.38	1.08	6.44	1.03	5.96

Panel B: Portfolio returns sorted on OIB

	FF3 alpha	t-stat	Carhart alpha	t-stat	FF5 + MOM + LIQ alpha	t-stat
Low	-0.16	-1.64	-0.13	-1.22	-0.10	-0.91
2	-0.25	-2.32	-0.20	-1.82	-0.15	-1.42
3	-0.19	-2.18	-0.15	-1.71	-0.10	-1.24
High	0.35	4.81	0.37	5.18	0.39	5.57
High - Low	0.51	5.44	0.50	5.05	0.49	4.70

Panel C: Portfolio returns sorted on Δ HF holding

	FF3 alpha	t-stat	Carhart alpha	t-stat	FF5 + MOM + LIQ alpha	t-stat
Low	-0.26	-3.06	-0.22	-2.56	-0.17	-2.19
2	-0.28	-2.54	-0.24	-2.12	-0.22	-1.83
3	0.11	1.27	0.13	1.47	0.16	1.80
High	0.21	2.26	0.25	2.67	0.31	3.79
High - Low	0.47	6.51	0.47	6.47	0.48	6.49

Table 2.3: Double Sorts on Retail Trading and Hedge Fund Trading

This table reports the monthly Carhart (1997) alphas of portfolios independently sorted on retail trading and hedge fund trading. At the end of each month, I sort all sample stocks into quartiles based on retail order imbalance and independently into quartiles based on hedge fund trading, for which I use either short interest (Panel A and C) or change in hedge fund holdings (Panel C and D) as proxy. Panel A and Panel C report double sorting results using short interest (changes in hedge fund holdings) as proxy for hedge fund trading. In Panel B, the long-leg (short-leg) portfolio with *consistent* trading between individuals and hedge funds "CON_LONG" ("CON_SHORT") include stocks that have retail order imbalance in the highest (lowest) quartile and has short interest in the lowest (highest) quartile. The long-leg (short-leg) portfolio with hedge fund trading "HF_LONG" ("HF_SHORT") include stocks that have short interest in the lowest (highest) quartile, where I remove those with retail order imbalance that falls outside [-0.1,0.1]. The long-leg (short-leg) portfolio with retail trading "R_LONG" ("R_SHORT") include stocks that have retail order imbalance in the highest (lowest) quartile, where I remove those with short interest that falls outside 0.1% distance from its median. I use changes in hedge fund holdings to replace for short interest in Panel D and use the same method as in Panel B to choose stocks and calculate alphas. The sample period is from January 2010 to December 2018. All ratios are in percent. The t-statistics are in parentheses and are Newey-West adjusted.

Panel A: Abnormal returns sorted on OIB and SR

	Retail Order Imbalance				
SR	Low	2	3	High	High - Low
Low	0.30	0.57	0.31	0.89	0.60
	(1.47)	(2.76)	(1.72)	(4.76)	(3.97)
2	-0.05	0.09	0.15	0.52	0.58
	(-0.50)	(0.89)	(1.62)	(5.60)	(4.68)
3	-0.46	-0.30	-0.26	0.08	0.54
	(-3.80)	(-2.45)	(-2.42)	(0.91)	(3.78)
High	-0.63	-0.91	-0.57	-0.10	0.53
	(-2.96)	(-4.70)	(-3.43)	(-0.95)	(2.57)
High - Low	-0.93	-1.49	-0.89	-0.99	
	(-3.48)	(-6.89)	(-4.53)	(-4.80)	

Panel B: Comparison with benchmark stock performance (SR proxy)

(1)	(2)	(3)	(4)	(5)	(6)
CON_LONG	HF_LONG	R_LONG	CON_SHORT	HF_SHORT	R_SHORT
0.89	-0.14	-0.09	-0.63	0.09	-0.07
(4.76)	(-0.82)	(-0.52)	(-2.96)	(0.64)	(-0.57)
	(1) - (2)	(1) - (3)		(4) - (5)	(4) - (6)
	1.03	0.99		-0.72	-0.56
	(4.13)	(3.84)		(-2.85)	(-2.31)

Table 2.3: (Cont.) Double Sorts on Retail Trading and Hedge Fund Trading

Panel C: Abnormal returns sorted on OIB and Δ HF holding

	Retail Order Imbalance				
Δ HF holding	Low	2	3	High	High - Low
Low	-0.30	-0.38	-0.36	0.12	0.41
	(-2.31)	(-2.69)	(-3.05)	(1.22)	(2.98)
2	-0.51	-0.53	-0.44	0.35	0.85
	(-2.98)	(-2.79)	(-3.28)	(2.93)	(5.14)
3	0.11	-0.09	-0.13	0.50	0.39
	(0.87)	(-0.65)	(-1.14)	(4.64)	(3.14)
High	0.11	0.11	0.28	0.47	0.36
	(0.84)	(0.90)	(1.99)	(4.43)	(2.47)
High - Low	0.40	0.49	0.64	0.35	
	(2.93)	(4.72)	(5.36)	(2.64)	

Panel D: Comparison with benchmark stock performance (Δ HF holding proxy)

(1)	(2)	(3)	(4)	(5)	(6)
CON_LONG	HF_LONG	R_LONG	CON_SHORT	HF_SHORT	R_SHORT
0.47	-0.00	0.16	-0.30	0.17	0.04
(4.43)	(-0.01)	(1.19)	(-2.31)	(1.10)	(0.40)
	(1) - (2)	(1) - (3)		(4) - (5)	(4) - (6)
	0.47	0.31		-0.47	-0.34
	(3.10)	(1.81)		(-2.32)	(-2.02)

Table 2.4: Controlling for Hedge Fund Holding Changes

This table reports the Carhart (1997) monthly alphas of the long and short sides of *consistent* trading portfolio, using retail order imbalance and short interest to proxy for trading signals from retail investors and hedge funds respectively, with additionally controlling for hedge fund holding changes. The long-leg (short-leg) portfolio with *consistent* trading "CON_LONG" ("CON_SHORT") include stocks that have retail order imbalance in the highest (lowest) quartile and has short interest in the lowest (highest) quartile. An increase (decrease) in Δ HF holding means change in hedge fund holding is in the highest (lowest) quartile. The sample period is from January 2010 to December 2018. All ratios are in percent. The t-statistics are Newey-West adjusted.

Δ HF holding	Variable	Carhart alpha	t-stat
	CON_LONG	1.04	5.08
Increase	CON_SHORT	-0.38	-1.78
	diff	1.42	4.80
	CON_LONG	0.78	3.68
Decrease	CON_SHORT	-0.72	-3.35
	diff	1.50	4.97

Table 2.5: Fama-MacBeth Regressions of Future Returns

This table reports the Fama and MacBeth (1973) regressions of next-month stock returns on variables of interest. The indicator of the long (short) leg of *consistent* trading between individuals and hedge funds "CON_LONG" ("CON_SHORT") equals one if the stock has retail order imbalance in the highest (lowest) quartile and has short interest in the lowest (highest) quartile, and equals zero otherwise. There are two scenarios for *opposite* trading between individuals and hedge funds. The indicator of the first (second) scenario "OPP_1" ("OPP_2") equals one if the stock has both retail order imbalance and short interest in the lowest (highest) quartile, and equals zero otherwise. In columns (2) and (3), I include firm size (LSIZE), book-to-market ratio (BM), momentum (MOM), short-term reversal (RET1), retail investor attention (ASVI), turnover (TURN) , institutional ownership (IO), retail short selling (RSS) as control variables. I further add an indicator of the long (short) leg of hedge fund trading "HF_LONG" ("HF_SHORT") in column (3), which equals one if the stock has changes in hedge fund holdings in the highest (lowest) quartile and has changes in short interest in the lowest (highest) quartile, and equals zero otherwise (Jiao, Massa, and Zhang, 2016). The sample period is from January 2010 to December 2018. The t-statistics are in parentheses and are Newey-West adjusted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
Variable	RET1	RET1	RET1
CON_LONG	0.504***	0.667***	0.663***
	(3.41)	(4.41)	(4.44)
CON_SHORT	-0.454***	-0.413**	-0.420**
	(-3.26)	(-2.31)	(-2.36)
OPP_1	-0.037	0.163	0.156
	(-0.23)	(1.04)	(1.01)
OPP_2	0.104	0.095	0.118
	(0.88)	(0.99)	(1.24)
LSIZE		-0.016	-0.016
		(-0.33)	(-0.32)
BM		0.042	0.041
		(0.45)	(0.44)
MOM		0.293	0.298
		(1.47)	(1.49)
RET1		0.726	0.676
		(0.84)	(0.78)
ASVI		0.128	0.138
		(0.77)	(0.83)
TURN		-1.639***	-1.625***
		(-3.20)	(-3.20)
IO		1.051***	1.068***
		(5.37)	(5.48)
RSS		-6.783	-6.320
		(-1.57)	(-1.46)
HF_LONG			0.217**
			(2.36)
HF_SHORT			-0.440***
			(-5.41)
Observations	363,372	331,404	331,404
# of groups	107	107	107

Table 2.6: Information Asymmetry and Informed Trading

This table reports the Fama and MacBeth (1973) regressions from Table 5 for subsamples of stocks constructed based on different ex-ante information asymmetry, including market capitalization, firm age, as well as turnover ratio. The control variables are not reported here for brevity. At the end of each month, all sample stocks are sorted into halves based on an information asymmetry proxy, and within each subsample stocks are further sorted into quartiles based on retail order imbalance and independently into quartiles based on short interest. I measure firm size as the firm's market capitalization, firm age as the number of years since the firm was first covered by CRSP, and turnover as trading volume divided by shares outstanding. The sample period is from January 2010 to December 2018. The t-statistics are in parentheses and are Newey-West adjusted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Firm Size		Firm Age		Turnover	
	Small	Big	Young	Old	Low	High
	(1)	(2)	(3)	(4)	(7)	(8)
Variable	RET1	RET1	RET1	RET1	RET1	RET1
CON_LONG	0.756***	0.427**	0.803***	0.545***	0.422***	0.335***
	(4.20)	(2.01)	(3.93)	(2.93)	(3.43)	(2.67)
CON_SHORT	-0.699**	-0.297**	-0.635**	-0.312*	-0.455*	-0.371*
	(-2.15)	(-2.10)	(-2.02)	(-1.88)	(-1.80)	(-1.69)
Observations	165,030	166,376	156,201	175,205	165,880	165,526
# of groups	107	107	107	107	107	107

Table 2.7: Predicting Informational Events and Firm Fundamentals

This table reports monthly panel regressions of firm fundamentals and informational events on both the long and short leg of *consistent* trading between hedge funds and retail investors. In column (1), the dependent variable is cumulative abnormal returns (CAR) around next earnings announcement, and it is computed as market-adjusted returns upon earnings announcements over a $[-1, 1]$ windows. The dependent variable in column (2) is standardized unexpected earnings (SUE) on next earnings announcement. The dependent variables in column (3) is analyst revisions (Revision), which is computed as the difference in average analyst earnings estimates from the previous month divided by the stock price at the end of the previous month. In column (4), dependent variable is changes in firm return-on-assets (ΔROA) in the following year. The last dependent variable reported in column (5) is stock-level news tone (News Tone) from Thomson Reuters (TR) news source, which is the average content of news measure where news is classified as negative, neutral, or positive. I include firm size (LSIZE), book-to-market ratio (BM), momentum (MOM), short-term reversal (RET1), abnormal trading volume that is proxy for general investor attention (ATT), retail investor attention (ASVI), dividend yield (DIV), firm age (LAGE), institutional Ownership (IO), retail short selling (RSS), hedge fund holdings (HF_Holding) as control variables. In addition, I control for current-month news tone in column (5). The sample period is from January 2010 to December 2018. I include the month fixed effects in all specifications. Standard errors are two-way clustered at the firm and the month level. The t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Variable	CAR	SUE	Revision	ΔROA	News Tone
CON_LONG	0.058***	0.143***	0.118*	0.032***	0.025***
	(8.23)	(15.83)	(1.79)	(4.72)	(5.23)
CON_SHORT	-0.053***	-0.131***	0.064*	-0.028**	-0.016***
	(-5.61)	(-11.56)	(1.72)	(-2.25)	(-5.70)
LSIZE	-0.005***	0.035***	0.013*	-0.015***	-0.011***
	(-4.58)	(22.49)	(1.73)	(-12.74)	(-29.82)
BM	0.015***	-0.035***	0.113*	-0.030***	-0.009***
	(6.68)	(-12.15)	(1.79)	(-8.36)	(-12.65)
MOM	0.009*	0.400***	0.193	0.315***	0.027***
	(1.77)	(69.65)	(1.62)	(43.51)	(17.58)
RET1	0.052***	0.788***	0.159	0.593***	0.085***
	(2.58)	(35.61)	(0.41)	(23.37)	(14.30)
ATT	0.005***	-0.007***	-0.075*	0.012***	-0.001
	(2.66)	(-3.07)	(-1.72)	(6.87)	(-1.17)
ASVI	0.023***	0.138***	-0.010	0.029***	0.006*
	(2.83)	(13.12)	(-0.18)	(3.61)	(1.80)
DIV	-0.251***	-0.764***	2.402*	-0.189**	-0.507***
	(-2.64)	(-5.25)	(1.74)	(-2.14)	(-12.95)
LAGE	0.003	0.011***	-0.026*	0.025***	-0.002**
	(1.53)	(3.73)	(-1.74)	(10.87)	(-2.55)
IO	0.073***	0.061***	0.372*	0.071***	0.032***
	(9.81)	(5.91)	(1.78)	(8.91)	(12.54)
RSS	-1.895***	-2.226***	5.512	-2.596***	-1.371***
	(-7.77)	(-7.56)	(1.43)	(-7.72)	(-11.22)
HF_Holding	-0.023	-0.675***	0.197*	0.020	-0.084***
	(-0.91)	(-20.19)	(1.71)	(0.61)	(-9.58)
$NEWS_0$					0.218***
					(66.73)
Observations	308,748	287,564	273,151	325,981	162,065

Table 2.8: Decomposition of Fundamental Predictability

This table reports monthly panel regressions employed in Table 7, additionally including indicator variables related to public events or news, as well as their interaction with both legs of *consistent* trading indicators. The dependent variables include standardized unexpected earnings (SUE), analyst revisions in earnings estimates (Revision), as well as changes in firm return-on-assets (ΔROA) in the following year. There are four indicator variables that are related to contemporaneous public events. A dummy variable for earnings announcement (D_EA) equals one if the firm has an earnings announcement in current month and equals zero otherwise. A dummy variable for analyst earnings revision (D_EREV) equals one if any analyst changes her earnings forecast for the firm and equals zero otherwise. A dummy variable for analyst recommendation revision (D_RREV) equals one if any analyst changes her buy/sell recommendation for the firm and equals zero otherwise. A dummy variable for public news (D_NEWS) equals one if there is positive or negative news for the firm in current month and equals zero otherwise. The control variables are the same as in Table 7, which are not reported here for brevity. The sample period is from January 2010 to December 2018. I include the month fixed effects in all specifications. Standard errors are two-way clustered at the firm and the month level. The t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	SUE	Revision	ΔROA	SUE	Revision	ΔROA
CON_LONG	0.119***	0.216*	0.005***			
	(9.68)	(1.74)	(4.01)			
CON_SHORT				-0.130***	0.132	-0.006***
				(-6.75)	(1.55)	(-2.70)
CON_LONG * D_EA	0.029	-0.002	0.004***			
	(1.42)	(-0.02)	(2.93)			
CON_LONG * D_EREV	0.069***	-0.111	-0.005**			
	(2.62)	(-1.59)	(-2.24)			
CON_LONG * D_RREV	0.059	-0.045	0.000			
	(0.96)	(-1.40)	(0.07)			
CON_LONG * D_NEWS	0.018	-0.048	-0.002			
	(0.81)	(-0.71)	(-1.33)			
CON_SHORT * D_EA				0.034	0.032	0.003
				(1.22)	(0.35)	(1.03)
CON_SHORT * D_EREV				0.024	-0.132*	0.002
				(0.95)	(-1.65)	(0.56)
CON_SHORT * D_RREV				-0.003	-0.051	0.003
				(-0.06)	(-1.61)	(0.55)
CON_SHORT * D_NEWS				-0.066***	-0.000	-0.003
				(-4.19)	(-0.00)	(-1.49)
D_EA	0.011*	0.028	-0.001	0.011*	0.027	-0.001
	(1.91)	(0.31)	(-1.18)	(1.87)	(0.30)	(-1.00)
D_EREV	-0.081***	0.150	-0.002***	-0.082***	0.150	-0.002***
	(-14.21)	(1.60)	(-3.49)	(-14.40)	(1.60)	(-4.21)
D_RREV	0.004	0.067*	-0.000	0.004	0.067*	-0.000
	(0.40)	(1.84)	(-0.33)	(0.37)	(1.85)	(-0.57)
D_NEWS	0.096***	-0.026	0.004***	0.099***	-0.027	0.004***
	(27.62)	(-0.46)	(11.22)	(28.05)	(-0.45)	(11.45)
Observations	227,371	214,506	243,950	227,371	214,506	243,950

Table 2.9: Consistent Trading for Anomaly Stocks

This table reports monthly average ratio of both sides of *consistent* trading for quintile portfolios sorted by the anomaly/mispricing variables. I follow Stambaugh, Yu, and Yuan (2012) and classify net stock issuance, accrual, asset growth, and investment to assets into "MGMT" cluster, and include financial distress, medium-term momentum, gross profitability, and return on assets in "PERF" cluster. I further use the mispricing measure constructed by Stambaugh, Yu, and Yuan (2015) to proxy for stock-level mispricing "MISP". By sorting stocks into quintile portfolios based on their three-month historical average of each anomaly/mispricing proxy, I report percentage of the number of stocks that in the long (short) side of *consistent* trading. The sample period is from January 2010 to December 2018. The t-statistics for the long-short portfolios are in parentheses and are Newey-West adjusted.

	CON_LONG			CON_SHORT		
	MGMT	PERF	MISP	MGMT	PERF	MISP
Long	4.22	4.66	3.35	4.37	2.69	2.93
2	3.81	3.48	3.55	3.53	2.94	3.11
3	3.74	3.56	3.25	3.36	3.49	3.39
4	3.28	2.99	3.06	3.66	4.34	4.19
Short	2.60	2.86	2.47	4.63	6.11	5.24
Long - Short	1.62	1.81	0.88	-0.26	-3.42	-2.30
	(12.10)	(15.47)	(6.27)	(-2.20)	(-12.74)	(-17.05)

Table 2.10: Consistent Trading and Stock Mispricing

This table reports monthly Carhart (1997) alphas for anomaly/mispricing in full sample and in the long and short sides of *consistent* trading. "MISP", "MGMT", and "PERF" are constructed as in Table 9. "CON_LONG" ("CON_SHORT") include stocks that have retail order imbalance in the highest (lowest) quartile and have short interest in the lowest (highest) quartile. I first sort stocks based on their retail order imbalance and short interest into "CON_LONG" or "CON_SHORT", or keep them in full sample. By further sorting these stocks into quintile portfolios based on their three-month historical average of each anomaly/mispricing proxy, I then report corresponding long-short anomaly returns for each sample, which is calculated as the difference between the returns of the long and short leg portfolios over one-year horizon. The sample period is from January 2010 to December 2018. The t-statistics are Newey-West adjusted.

		L - S alpha	t-stat
MISP	Full sample	0.88	2.66
	CON_LONG	0.02	0.05
	CON_SHORT	0.68	0.95
MGMT	Full sample	1.03	3.77
	CON_LONG	0.55	1.25
	CON_SHORT	0.79	2.01
PERF	Full sample	1.98	3.21
	CON_LONG	0.24	0.04
	CON_SHORT	0.53	1.23

Table A1: Firm Characteristics in Different Sub-samples

This table reports descriptive statistics of firm characteristics in different sub-samples. "CON_LONG" ("CON_SHORT") includes stocks that have retail order imbalance in the highest (lowest) quartile and have short interest in the lowest (highest) quartile. "OPP_1" ("OPP_2") includes stocks that have both retail order imbalance and short interest in the lowest (highest) quartile. "R_LONG" ("R_SHORT") includes stocks that have retail order imbalance in the highest (lowest) quartile. "SR_LOW" ("SR_HIGH") includes stocks that have short interest in the lowest (highest) quartile. "HF_LONG" ("HF_SHORT") includes stocks that have change of hedge fund holding in the highest (lowest) quartile. "FULL" refers to the full sample. The time-series average of the cross-sectional means of the variables of interest are presented. Firm size (SIZE) is the market capitalization. Book-to-market ratio (BM) is the most recent fiscal year-end book value divided by the market capitalization. Turnover (Turnover) is the monthly trading volume divided by the number of shares outstanding, averaged over the past 12 months. Stock price (Price) is the current-month stock price. Momentum (MOM) is the past cumulative returns from month -12 to month -2. Dispersion of opinions from retail investors (R_dispersion) utilizes daily investor disagreement data provided by Cookson and Niessner (2020). Institutional Ownership (IO) is the number of shares held by institutions (13F filings) in the most recent quarter, normalized by the number of shares outstanding. The reported ratio of the number of lottery stocks to that of non-lottery stocks (Lott/Non-lott) for each sample is the time-series average, where I refer to Kumar (2009) for the definition of lottery stocks. The sample period is from January 2010 to December 2018.

	CON_LONG	OPP_1	R_LONG	SR_LOW	HF_LONG	FULL
Size ($M)	1,285	2,463	2,830	4,318	1,975	4,387
BM	1.02	1.02	0.68	0.98	0.88	0.68
Turnover (%)	4.05	5.13	15.03	5.56	7.97	17.20
Price ($)	17.2	14.6	34.5	18.6	16.8	29.4
Momentum (%)	12.33	13.35	15.36	14.06	21.09	16.13
R_dispersion (%)	11.63	12.49	17.07	14.43	13.92	18.79
IO (%)	28.8	28.9	60.8	31.8	43.0	58.6
Lott / Non-lott	0.43	0.51	0.17	0.47	0.46	0.23

	CON_SHORT	OPP_2	R_SHORT	SR_HIGH	HF_SHORT	FULL
Size ($M)	1,318	2,384	2,806	2,043	1,908	4,387
BM	0.66	0.50	0.80	0.56	0.56	0.68
Turnover (%)	27.29	28.21	12.63	32.08	33.58	17.20
Price ($)	17.6	39.9	19.6	28.5	29.2	29.4
Momentum (%)	6.91	15.54	13.35	14.25	15.62	16.13
R_dispersion (%)	19.22	20.96	14.85	23.38	23.37	18.79
IO (%)	62.2	80.6	49.2	71.3	78.5	58.6
Lott / Non-lott	0.35	0.09	0.31	0.22	0.19	0.23

Table A2: Decomposition of Fundamental Predictability for Opposite Trading

This table reports monthly panel regressions of future fundamentals on both legs of *opposite* trading indicators, additionally including indicator variables related to public events or news, as well as their interaction terms. The dependent variables include standardized unexpected earnings (SUE), analyst revisions in earnings estimates (Revision), as well as changes in firm return-on-assets (ΔROA) in the following year. There are four indicator variables that are related to contemporaneous public events. A dummy variable for earnings announcement (D_EA) equals one if the firm has an earnings announcement in current month and equals zero otherwise. A dummy variable for analyst earnings revision (D_EREV) equals one if any analyst changes her earnings forecast for the firm and equals zero otherwise. A dummy variable for analyst recommendation revision (D_RREV) equals one if any analyst changes her buy/sell recommendation for the firm and equals zero otherwise. A dummy variable for public news (D_NEWS) equals one if there is positive or negative news for the firm in current month and equals zero otherwise. The control variables are the same as in Table 8, which are not reported here for brevity. The sample period is from January 2010 to December 2018. I include the month fixed effects in all specifications. Standard errors are two-way clustered at the firm and the month level. The t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	(1) SUE	(2) Revision	(3) ΔROA	(4) SUE	(5) Revision	(6) ΔROA
OPP_1	-0.020	0.218	-0.000			
	(-1.61)	(1.00)	(-0.20)			
OPP_2				-0.022	-0.974	-0.003
				(-1.35)	(-1.00)	(-1.44)
OPP_1 * D_EA	-0.027	0.007	-0.003***			
	(-1.54)	(0.35)	(-2.70)			
OPP_1 * D_EREV	0.070***	-0.126	-0.003			
	(2.91)	(-0.99)	(-0.99)			
OPP_1 * D_RREV	-0.028	-0.062	0.004			
	(-0.63)	(-0.99)	(1.06)			
OPP_1 * D_NEWS	-0.001	-0.041	0.002			
	(-0.07)	(-0.98)	(1.56)			
OPP_2 * D_EA				0.011	0.435	0.002
				(0.93)	(1.00)	(0.94)
OPP_2 * D_EREV				-0.029	0.519	-0.002
				(-1.62)	(1.00)	(-1.30)
OPP_2 * D_RREV				0.008	0.166	0.002
				(0.23)	(0.99)	(0.95)
OPP_2 * D_NEWS				0.018**	0.640	-0.001
	(-0.07)	(-0.98)	(1.56)	(2.00)	(1.00)	(-0.43)
D_EA	0.011	0.027	-0.000	0.007	-0.005	-0.000
	(1.11)	(1.00)	(-0.46)	(0.79)	(-0.78)	(-0.82)
D_EREV	-0.018***	0.153	-0.002**	-0.011*	0.109	-0.002***
	(-2.81)	(1.00)	(-2.43)	(-1.82)	(0.99)	(-2.79)
D_RREV	-0.025***	0.068	0.000	-0.026***	0.050	-0.001
	(-2.68)	(1.00)	(0.25)	(-2.76)	(0.99)	(-0.49)
D_NEWS	0.052***	-0.025	0.003***	0.050***	-0.087	0.003***
	(11.30)	(-0.97)	(7.11)	(10.99)	(-1.00)	(5.98)
Observations	227,371	214,506	257,145	227,371	214,506	243,950

Table A3: Decomposition of Return Predictability

This table reports monthly panel regressions of next-month return on main variables of interest, additionally including indicator variables related to public events or news, as well as their interaction with both legs of *consistent* trading indicators. There are four indicator variables that are related to contemporaneous public events. A dummy variable for earnings announcement (D_EA) equals one if the firm has an earnings announcement in current month and equals zero otherwise. A dummy variable for analyst earnings revision (D_EREV) equals one if any analyst changes her earnings forecast for the firm and equals zero otherwise. A dummy variable for analyst recommendation revision (D_RREV) equals one if any analyst changes her buy/sell recommendation for the firm and equals zero otherwise. A dummy variable for public news (D_NEWS) equals one if there is positive or negative news for the firm in current month and equals zero otherwise. The control variables are the same as in Table 8, which are not reported here for brevity. The sample period is from January 2010 to December 2018. I include the month fixed effects in all specifications. Standard errors are two-way clustered at the firm and the month level. The t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
Variable	RET1	RET1	RET1	RET1
CON_LONG	0.550**	0.710***		
	(2.52)	(2.76)		
CON_SHORT			-0.695***	-1.073*
			(-2.87)	(-1.87)
CON_LONG * D_EA	0.292	0.275		
	(1.18)	(1.01)		
CON_LONG * D_EREV	-0.024	-0.329		
	(-0.09)	(-1.05)		
CON_LONG * D_RREV	-0.135	-0.591		
	(-0.24)	(-0.89)		
CON_LONG * D_NEWS		0.341		
		(1.17)		
CON_SHORT * D_EA			-0.507	-0.177
			(-1.15)	(-0.38)
CON_SHORT * D_EREV			0.530	0.594
			(1.23)	(1.14)
CON_SHORT * D_RREV			0.282	0.681
			(0.57)	(1.25)
CON_SHORT * D_NEWS				0.279
				(0.42)
D_EA	-0.150	-0.166	-0.122	-0.142
	(-1.32)	(-1.34)	(-1.11)	(-1.19)
D_EREV	-0.000	0.029	-0.032	-0.020
	(-0.00)	(0.32)	(-0.39)	(-0.23)
D_RREV	0.080	0.018	0.051	-0.026
	(0.91)	(0.18)	(0.57)	(-0.26)
D_NEWS		0.094		0.063
		(0.74)		(0.47)
Observations	331,404	261,344	331,404	261,344

Table A4: Consistent Trading on Stock Mispricing

This table reports monthly panel regressions of indicators for both long and short sides of *consistent* trading next month, i.e. CON_LONG1 and CON_SHORT1, on the mispricing measure (MISP), which is constructed by Stambaugh, Yu, and Yuan (2015) to proxy for stock-level mispricing. I include firm size (LSIZE), book-to-market ratio (BM), momentum (MOM), short-term reversal (RET1), retail investor attention (ASVI), turnover (TURN) , institutional ownership (IO), retail short selling (RSS), as well as hedge fund holdings (HF_HOLDING) as control variables. The sample period is from January 2010 to December 2018. I include the month fixed effects in all specifications. Standard errors are two-way clustered at the firm and the month level. The t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)
VARIABLES	CON_LONG1	CON_SHORT1
MISP	-0.390***	0.183***
	(-10.68)	(4.54)
LSIZE	-0.147***	-0.072***
	(-16.35)	(-12.27)
BM	0.058***	-0.010
	(9.98)	(-1.55)
MOM	-0.035***	-0.048***
	(-7.77)	(-9.16)
RET1	-0.014**	-0.002
	(-2.59)	(-0.43)
ASVI	0.014**	-0.013***
	(2.60)	(-4.67)
TURN	-0.078***	0.085***
	(-15.62)	(11.04)
IO	-0.162***	0.042***
	(-17.79)	(5.10)
RSS	0.079***	-0.044***
	(9.69)	(-9.99)
HF_HOLDING	0.016**	0.002
	(2.30)	(0.32)
Observations	211,646	211,646

TABLE A5: Sub-sample Tests

This table reports the Fama and MacBeth (1973) regressions of next-month stock returns on both long and short sides of *consistent* trading, for subsamples of stocks constructed based on sentiment or disagreement indices. At the end of each month, all sample stocks are sorted into halves based on an sentiment or disagreement index, and within each subsample stocks I further run Fama-MacBeth regressions on variables of interest. BW sentiment index data is from Baker and Wurgler (2006). PLS sentiment data is from Huang et al. (2015). Disagreement index data is from Huang, Li, and Wang (2021). The control variables include firm size (LSIZE), book-to-market ratio (BM), momentum (MOM), short-term reversal (RET1), retail investor attention (ASVI), turnover (TURN) , institutional ownership (IO), retail short selling (RSS), as well as an indicator variable of the long (short) leg of hedge fund trading "HF_LONG" ("HF_SHORT"), which equals one if the stock has changes in hedge fund holdings in the highest (lowest) quartile and has changes in short interest in the lowest (highest) quartile, and equals zero otherwise (Jiao, Massa, and Zhang, 2016). The sample period is from January 2010 to December 2018. The t-statistics are in parentheses and are Newey-West adjusted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	BW Sentiment		PLS Sentiment		Disagreement	
	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	RET1	RET1	RET1	RET1	RET1	RET1
CON_LONG	0.684***	0.576***	0.837***	0.441***	0.740***	0.517***
	(4.12)	(4.29)	(5.66)	(3.04)	(4.80)	(3.63)
CON_SHORT	-0.734***	-0.147	-0.694***	-0.204	-0.525**	-0.348
	(-3.84)	(-0.69)	(-5.10)	(-0.92)	(-2.05)	(-1.41)
LSIZE	-0.051	0.011	0.007	-0.043	-0.013	-0.026
	(-0.84)	(0.16)	(0.10)	(-1.13)	(-0.30)	(-0.44)
BM	-0.087	0.169	0.035	0.049	0.072	0.011
	(-1.02)	(1.24)	(0.36)	(0.40)	(0.50)	(0.09)
MOM	0.641***	-0.032	0.430**	0.184	-0.058	0.667***
	(3.88)	(-0.10)	(2.03)	(0.67)	(-0.19)	(3.19)
RET1	1.840**	-0.464	0.790	0.574	-0.035	1.403**
	(2.05)	(-0.34)	(1.59)	(0.50)	(-0.03)	(2.02)
ASVI	0.117	0.175	0.140	0.153	0.331**	-0.041
	(0.65)	(1.02)	(1.18)	(0.72)	(2.18)	(-0.17)
TURN	-2.296**	-0.994*	-2.136***	-1.185***	-1.272**	-2.012***
	(-2.65)	(-1.85)	(-3.20)	(-2.85)	(-2.04)	(-2.73)
IO	1.414***	0.714***	1.367***	0.782***	0.985***	1.138***
	(6.03)	(3.02)	(5.64)	(5.58)	(5.17)	(5.25)
RSS	0.234	-11.822***	-2.442	-8.955*	-7.203	-4.472
	(0.04)	(-2.98)	(-0.36)	(-1.91)	(-1.06)	(-0.72)
HF_LONG	0.042	0.379***	0.102	0.312*	0.339***	0.083
	(0.28)	(3.20)	(0.98)	(1.89)	(4.91)	(0.38)
HF_SHORT	-0.457***	-0.414***	-0.497***	-0.379***	-0.422***	-0.449***
	(-5.31)	(-4.76)	(-7.01)	(-4.47)	(-4.59)	(-6.58)
Observations	164,937	166,467	163,149	168,255	166,883	164,521
# of groups	53	54	51	56	54	53

Chapter 3

Can Retail Investors Learn from Insiders?

with Ekkehart Boehmer and Joe Zhang

3.1 Introduction

Facilitating price discovery in financial markets is an important function that market participants perform. The roles of these participants differ across trader types and the technologies that they control. While individually important, the interactions among different players are vital for the flow of information into prices. In this paper, we study how two special groups of traders, corporate insiders and retail traders, interact with each other and help improve the price discovery process.

Corporate insiders, by definition, have access to firm's private information ahead of other investors.[1] Previous studies have documented the informational role of insider trading. For example, Lakonishok and Lee (2001) find that insiders can predict stock returns in the cross-section. Several other papers also show that insiders trade on private information and earn significant returns.[2] In addition, Cohen, Malloy, and Pomorski (2012) separate insider trades into "opportunistic" trades and "routine" trades, and show that the return predictability mostly come from the "opportunistic" trades. The informational role of retail investors is less clear. Earlier studies tend to show retail investors as a group are uninformed

[1]The Securities and Exchange Act of 1934 defines insiders as officers, directors, and shareholders of 10% or more of any equity class of securities.

[2]For example, see Jaffe (1974), Seyhun (1986), Rozeff and Zaman (1988), Lin and Howe (1990), Bettis, Vickrey, and Vickrey (1997), Seyhun (1998), Lakonishok and Lee (2001), Jeng, Metrick, and Zeckhauser (2003), Marin and Olivier (2008), Jagolinzer (2009).

and irrational.[3] Some recent papers, on the other hand, point out that retail investors are informed and earn positive abnormal returns.[4]

As insider opportunistic trading is most likely information driven, to the extent that retail trading predicts stock returns, the interaction between insiders and retail investors provides a useful setting in disentangling the sources of the return predictability, which is often difficult to do. For example, if retail investors trade against insiders, it is likely that some of these traders are providing liquidity to insiders and may profit from that in the short term. If they follow insider trades, they are demanding liquidity, and their trading is likely driven by information content revealed from the insider trades. Examination of these potential sources of trading motivation and the corresponding price formation process would help us understand how information is impounded into stock prices.

In this paper, we study the joint trading patterns of insiders and retail investors and their impact on future stock returns. We then examine possible sources of any observed return predictability of these trades. Unlike most of the retail trading studies that use proprietary data from a few brokers, we follow Boehmer et al. (2021) to filter out daily retail trades from TAQ trade dataset. To identify insider trades, we use Thomson Reuters Insiders data (Form 4) and extract insider open market purchases and sales and examine the trading and stock return patterns at the daily and weekly level for both insiders and retail investors.

[3]Barber and Odean (2000) use household trading data and document that net returns earned by households are poor. Barber, Odean, and Zhu (2009) show that retail investor buying (selling) push prices too high (low) leading to subsequent reversal.

[4]For example, Kaniel, Saar, and Titman (2008) document positive excess returns in the month following intense individual buying and negative excess returns following intense individual selling. Boehmer et al. (2021) propose a novel method to separate out retail orders from TAQ, and they find that retail investors are well informed about firm news and are likely to hold private information. Other papers that document retail investors' informed trading are Kaniel et al. (2012), Kelley and Tetlock (2013), Barrot, Kaniel, and Sraer (2016), Fong, Gallagher, and Lee (2014) and etc.

Consistent with Boehmer et al. (2021), retail order imbalance is negative in general outside the insider trading event window (between the insider trade date and one day after the report date). Retail investors, however, tend to buy during the event window the stocks that insiders have purchased, and sell those that insider have sold, more than they usually sell. A Granger-type causality test indicates that retail investors follow informed insider purchases, and not the other way round. The evidence suggests that at least some retail investors are able to identify and follow insider trades in a timely fashion. Moreover, retail investors do not blindly follow insider trades. We follow Cohen, Malloy, and Pomorski (2012) to separate insider trades into "opportunistic" trades and "routine" trades. We find that retail investors tend to follow opportunistic insider purchases, but not routine insider purchases, during the event window. Further analysis suggests that retail investors keep buying stocks with opportunistic insider purchase for up to four quarters. To the extent that insider "opportunistic" trades reflect insider's private information, at least retail investors seem to be able to identify and learn from the actual informed trades, and act quickly.

One possible channel of learning is through retail investors' brokers. Li, Mukherjee, and Sen (2021) provide evidence that mutual fund managers gain informational advantage from affiliated brokers of corporate insiders. McLean, Pontiff, and Reilly (2021) argue that analysts and brokers at large investment banks also have incentives to "tip" retail investors. As a result, retail investors could learn about insider trades through their informed brokers. Moreover, with the advancement in technology and the abundance information, savvy retail investors can monitor and research about insider trades on their own. Indeed, we find an increase in the abnormal retail downloads of the insider trading filing Form 4 in the EDGAR database. Higher abnormal retail downloads during the event window are also associated with more retail buys in the following days. This evidence is most significant for stocks with opportunistic insider purchases.

There is however no clear distinction of retail investor trading pattern corresponding to opportunistic and routine insider sales. This result is not surprising, as the existing literature suggests that insider sales tend to be driven by liquidity and diversification reasons, whereas insider purchases are regarded as strong positive signals in stock values.[5] We hence focus on retail order flow around insider purchase window from now on.

While our results are consistent with retail investors following insider purchases because they learn from informed trades by insiders, it is however possible that insider trades simply increase the market attention to the stock. As investors have limited attention, the increased exposure of the stock prompts retail investors to buy more shares. Barber and Odean (2008) show that individual investors are net-buyers of attention-grabbing stocks. Da, Engelberg, and Gao (2011) further use Google search volume to proxy for the abnormal attention of retail investors, and they show transitory price pressures on those attention-grabbing stocks.[6] We follow Da, Engelberg, and Gao (2011) and manually collect the weekly Google Search Volume Index (SVI) to measure retail investor attention. We then examine how retail investors follow insider purchase for stocks with high abnormal SVI (ASVI) and low ASVI separately. While high ASVI stocks do experience larger retail investors trading volume in the insider trading event window, attention is not a driving force for retail investors to follow insider trades. When insiders buy opportunistically, retail investors tend to follow them, regardless of the level the investor attention. We also follow Barber and Odean (2008) and use abnormal trading volume to proxy for general investor attention. The results are similar: investor attention is positively correlated with retail investors' purchasing behavior, but it is not the main reason why retail investors

[5] Jeng, Metrick, and Zeckhauser (2003) find weak evidence on the profitability of insider sales. They find that insider purchase earn abnormal returns of more than 6% per year, while insider sales do not earn significant abnormal returns.

[6] Other papers such as Gervais, Kaniel, and Mingelgrin (2002) and Ben-Rephael, Da, and Israelsen (2017) also examine the investor attention effect.

follow insider trades.

Another possibility is that retail investors and corporate insiders make their trading based on the same set of information around the same time. For example, Kaniel et al. (2012) show that both purchases and sales of stocks by retail investors before the earnings announcements can predict returns after the announcements. To examine the effect of earnings as potential common source of information, we repeat the earlier analysis separately for stocks with or without earnings announcements within one month after the insider trading window. For those stocks with near future earnings announcements, we also examine those with positive and negative SUEs separately. The results show that retail investors follow opportunistic insider purchases regardless if there are earnings announcements. For the subsample with near future earnings announcements however, we do find that retail investors follow insider purchase more aggressively for stocks with positive future SUE than those with negative SUE, which is consistent with Kaniel et al. (2012). Taken together, these results suggest that some retail investors might trade in the same direction as the insiders because they share common information about future earnings. However, the near future earnings cannot be a driving force as retail investors still follow opportunistic insider purchase for the vast majority of the cases where there is no near future earnings announcements.

We also test analysts forecast revisions and recommendation updates as potential sources of common information. The results are consistent with those of the earnings announcements. Retail investors actively follow opportunistic insider purchases, regardless of whether there is recommendation upgrade/downgrade, or forecast up/down revision. If anything, when retail investors follow opportunistic insider purchases more aggressively, it is more likely to have near-future analyst downward revisions/recommendation downgrades than upward revisions/recommendation upwards. Collectively, these findings suggest that retail investors follow insider trades not simply because these trades catch their attention or because they share the same information with insiders. It is likely that they follow insider trades because of the private information content revealed by these trades.

If opportunistic insider trading does reveal private information that retail investors can identify, trading alongside insiders by retail investors should help expedite the price discovery process for the underlying stock. We test this hypothesis by examining the future returns of stocks traded by retail investors following insider buying. First, we provide out-of-sample evidence as in Cohen, Malloy, and Pomorski (2012) that the opportunistic trades by insiders earn higher returns than the routine trades. More importantly, for stocks with opportunistic insider purchases, our event-week portfolio test shows that the returns on those stocks that retail investors buy have higher market-adjusted returns than those that retail investors sell. Stocks with positive retail order flow earn abnormal return that is 10 basis points (t-stat = 3.55) higher next week than those with negative retail order flow. The cumulative return differences between these two groups remain significant for up to 18 weeks with no long-term return reversals. We also conduct calendar-week portfolio analysis. For each week, we form a long/short portfolio that longs the stocks with opportunistic insider purchases that are bought by retail investors, and shorts those that are sold by retail investors. The excess return on the long/short portfolio earns significantly positive CAPM and the Carhart four-factor alphas.

Next, we conduct predictive panel regressions of next-week stock returns on a dummy variable (Follow_Oppbuy) that equals one if retail investors choose to buy the stocks with opportunistic insider purchases during the same week. The coefficient on Follow_Oppbuy is positive and significant. Besides examining the sub-sample of stocks with opportunistic insider purchases, we also regress next-week stock returns on an opportunistic insider purchase dummy, a retail purchase dummy, an interaction term of the two, and a number of control variables in the full-sample. Importantly, controlling for insider purchases and retail purchases, the coefficient on the interaction term is positive and highly significant. The results of both sub-sample regression and full-sample regression are consistent with the hypothesis that retail investors learn from the opportunistic insider purchases, and their trading helps speed up price discovery by impounding the private information revealed by the insider trades into stock prices faster. These results pass a number of robustness tests. Moreover, we show that for stocks with opportunistic insider purchases, return differences between stocks retail investors purchase and those they sell are significantly larger for stocks with greater information symmetry (smaller stocks, stocks with higher idiosyncratic volatility and stocks with higher Amihud illiquidity), further supporting the information hypothesis.

Interestingly, when retail investors sell the stocks that insider have purchased, those stocks still have higher future returns, albeit not as high as those that bought by retail investors. This suggests that retail investors do not seem to have superior information during the insider trading window on top of what is possessed by the insiders. When they can learn from insider trades, they improve price discovery; when they fail to do so, they only delay the process.

The return predictability of retail trading is not driven by the liquidity provision. Because retail investors and insiders are in the same side of the trade during the event window, by definition retail investors are not providing liquidity to insiders. However, it is possible that they provide liquidity to other market participants who trade against them, and make profits from liquidity provision. To formally examine the liquidity provision hypothesis, we follow Boehmer et al. (2021) and decompose the retail trades into three components: one attributed to price pressure, one to liquidity provision, and the rest to information. We then repeat the panel regression analysis. The results show that only the coefficient on the information component is highly significant, and those on liquidity provision and temporary price pressure are statistically insignificant.

To further examine the efficiency gain of the retail trading following opportunistic insider purchase, we construct two additional commonly used price efficiency measures: the Lo and MacKinlay (1988) variance ratio and the Hou and Moskowitz (2005) price delay measure.[7] We repeat the above panel regressions, except that we replace the future returns with the absolute value of one minus the variance ratio and the price delay measure as the dependent variables. We find strong evidence of gain in efficiency measured by the variance ratio, but not by the price delay measure. Since the variance ratio mostly measures firm level information while the price delay measures the speed of the market information incorporated into prices, our tests suggest that the trading of retail investors helps impound the firm specific information into the stock prices, but not the market level information.

Several recent papers examine interaction between insider trading and retail trading. Mansi et al. (2021) show significant increase in insider opportunistic selling (buying) following increase (decrease) in retail attention. Stotz and Georgi (2012) obtain 1-year retail trades from a retail broker in Germany. They provide evidence of retail investors copying the trades of insiders, although retail investors' copying behavior yields insignificant future abnormal returns. Using NYSE data, Chung (2020) argues that a substantial portion of retail trading around the insider trading window is the trading by insiders themselves. Neither paper differentiates opportunistic insider trades from routine trades. As we explain later, retail trading in our sample is in the OTC market. With our identification methodology, insider trading is mostly excluded from the retail

[7]For other papers that use these measures, see Barnea (1974), Boehmer and Kelley (2009), and Boehmer and Wu (2013).

trading sample by construction. Removing the small amount of potential insider transactions do not affect the empirical results. Unlike the above papers, we show that retail investors not only able to follow opportunistic insider trades, but their trading can predict future returns beyond insider trading.

More generally, our paper contributes to the understanding of the interaction among different market players in the process of price discovery. While extensive studies have been done on insiders, institutional/retail investors, as well as short sellers, relatively few studies analyze the effects that these trader types have on other types and on how they interact with each other. Notably, Massa et al. (2015) examine how insiders' strategy changes when they need to compete with short sellers. Kelley and Tetlock (2013) control for insider sales when examining the retail shorting activity. These studies focus on the interaction between insiders and short sellers in the presence of negative information. Our paper examines the combined effect of insiders and retail investors, who are traditionally not seen as informed players. The results indicate that retail investors can be informed, by showing that they are able to identify the *positive* private information revealed by insider trades. Our paper is also related to Sias and Whidbee (2010), who show an inverse relation between insider trading and institutional demand during the same quarter and over the previous quarter. By comparison, we use comprehensive TAQ-based daily retail trading data, and show that retail investors trade alongside insiders during the insider trading window. Moreover, we show that retail investors follow insiders not for liquidity reasons, but to learn about insiders' private information from their opportunistic trades. By following insiders' informed trades, retail investors improve the price discovery process of the underlying stocks.

The rest of our paper is organized as follows. Section 3.2 provides the data description and summary statistics. Section 3.3 examines the trading pattern of retail investors around insider trades. Section 3.4 further explores the information content of retail trades around insider trades. Section 3.5 conducts additional robustness checks. Section 3.6 concludes the paper.

3.2 Data and Sample Construction

In this section we first illustrate various sources of the data used in our tests, and we then present the sample construction and summary statistics.

3.2.1 Retail Trading Data

Our main data on retail investor trades come from TAQ trade dataset over the period from January 1, 2010 to December 31, 2018. According to Boehmer, Jones, Zhang, and Zhang (2021), retail orders that are internalized or executed by wholesalers are given a small amount of price improvement relative to the National Best Bid or Offer (NBBO). We follow their paper's price improvement measures to isolate retail investors' marke orders from institutional orders. Specifically, we identify a transaction as a retail buy if the subpenny price is between 60 and 100 basis points, and identify a retail sell if the subpenny price is between 0 and 40 basis points. We then aggregate retail buy volume, the number of retail buy trades, retail sell volume, and the number of retail sell trades for each stock on each trading day.[8] [9]

To construct retail investors' directional trades for each stock on each trading day, we follow Boehmer et al. (2021) and define the order imbalance in terms

[8]The sample period does not include years before 2010 because the subpenny trade practice only stabilizes after 2009. Also, because the SEC's tick size pilot program (TSPP) from 2016 to 2018 might affect the practice of subpenny price improvements unevenly in the cross section, we 1) exclude the TSPP period in the sample and re-run our main tests, and 2) construct TSPP as a dummy variable and interact it with our main variables in panel regressions. We find that the results during the pilot period is not materially different from the earlier sample. We omit the s for brevity.

[9]For more details, refer to Boehmer et al. (2021).

of both the trading volume $oibvol_{i,t}$ and the number of trades $oibtrd_{i,t}$, as the following:

$$oibvol_{i,t} = \frac{indbvol_{i,t} - indsvol_{i,t}}{indbvol_{i,t} + indsvol_{i,t}} \tag{1}$$

$$oibtrd_{i,t} = \frac{indbtrd_{i,t} - indstrd_{i,t}}{indbtrd_{i,t} + indstrd_{i,t}} \tag{2}$$

where $indbvol_{i,t}$ ($indsvol_{i,t}$) is the number of shares of stock i bought (sold) by retail investors on day t, and $indbtrd_{i,t}$ ($indstrd_{i,t}$) is the number of buy (sell) trades of stock i on day t.

3.2.2 Insider Trading Data

We obtain insider trading data from the Thomson Reuters Insider Filing database. According to the Securities and Exchange Act of 1934, open market trades by corporate insiders should be reported to the SEC within 10 days after the end of the month in which they took place. In 2002, the ten-day deadline was changed to a two-day deadline instead. However, as shown in Table A1, about 6.3% of insiders report their trading after the 2-day deadline. In our sample, we only use insider trades that are reported within the 2-day deadline.[10]

Corporate insiders from the database include company officers, directors, and beneficial owners of more than 10% of a company's stock. We extract the SEC's Form 4 data during the sample period from January 2010 to December 2018. We focus on open market purchases and sales by insiders, and exclude option exercises and private transactions.

Our data include both the transaction date and the SEC filing date (when the insider trading information is available to the general public). In our main

[10]Our results remain consistent when we include trades with reporting lags larger than 2 business days.

analysis, we define the insider trading event window as the days within the insider trading date and one day after the filing date, including both ends.[11]

Chung (2020) provides one potential explanation for the change of retail trading imbalance around the insider trading events – a large portion of these retail trading is actually the trading by insiders themselves. We rule out this explanation in our setting. While Chung (2020) obtains retail trading volume from NYSE, our sample includes only off-exchange retail trades. Moreover, in our sample, 94.81% (86.58%) of the open market purchases (sales) by insiders cannot be attributed to retail trading. The reason is that either the transaction price of an insider trading does not satisfy the price improvement algorithm, or the trading volume of an insider transaction already exceeds that of daily retail trading in total. For the remaining small portion of insider transactions that may be counted as retail trading, we remove them from our sample and the empirical results are not affected.

3.2.3 Other Types of Data

The accounting variables and the earnings announcement data are obtained from Compustat. We obtain analyst forecast data from the Institutional Brokers' Estimate System (IBES), and data on institutional holdings from Thomson Reuters (13F).

Additionally, we download IP search volume data from the SEC EDGAR log file database, which include internet search traffic for EDGAR filings.[12] Each log entry includes the IP address of the requesting user, time stamp of the request, Central Index Key (CIK) of the company that filed the form, as well

[11]We include one day after the filing date in the event window as there may be time lag of investors' responses. For robustness, we also examine our main tests when defining the event window that ends at the filing date, or ends at two days after the filing date, and the results remain consistent. The s are omitted for brevity.

[12]As the EDGAR log files cover from February 14, 2003 through June 30, 2017, our analysis with EDGAR search ends in mid-2017.

as the accession number that identifies the specific filing type. Following Drake, Roulstone, and Thornock (2015), we download the Master Index files from the SEC website and then match them with the log files based on the accession number to obtain both the filing type and filing date for each entry. As we are interested in the insider filing type, we only keep those entries with form 4 filings. We then exclude the records with index specifications to remove redundancies. Furthermore, we follow Drake, Roulstone, and Thornock (2015) and Chi and Shanthikumar (2018) to remove entries where the IP address has more than five requests per 60-second interval or more than 1,000 requests per day, so that the remaining records are likely to be employed by retail investors, who usually would not use automated web crawler programs to search and download files. As a result, for each stock on each day, we count the number of unique IP addresses searching for Form 4 filings and then subtract its prior sixty-day average value to obtain retail abnormal EDGAR search (*A_Search*).

We also obtain Google's Search Volume Index (SVI) and construct abnormal retail investor attention, similar to Da, Engelberg, and Gao (2011). We manually collect the weekly Search Volume Index for each stock ticker using a web-scraping technique, and extract stock tickers (TICKER) that appear in our main sample. We delete tickers with a generic meaning such as "ALL", "B", and "GPA" manually.[13] We collect weekly SVI from 5,524 distinct firm tickers during April 2009 through December 2018. We then merge SVI statistics with stocks in our main sample. Following Da, Engelberg, and Gao (2011), we use the abnormal search volume index (ASVI) as the proxy for the retail investor abnormal attention. We calculate ASVI as the log of SVI during the week minus the log of median SVI for the previous eight weeks. We also use the abnormal trading volume (ATT) as the proxy for general investor attention, following Barber and Odean (2008). We divide each stock's daily trading volume by its

[13]These generic-meaning tickers would cause ambiguity and create more noise.

average trading volume in the previous year (252 trading days), and then take the weekly average to get ATT.

We merge our retail and insider transaction data with the stock-level characteristic variables. Our sample contains common stocks (CRSP share codes 10 and 11) listed on NYSE, AMEX, and NASDAQ. We exclude stocks with price less than 1 dollar. We winsorize all the continuous variables at 1% and 99% level. The sample period is from January 1, 2010 to December 31, 2018.

3.2.4 Summary Statistics

1 shows summary statistics for the sample. We report firm characteristics for the retail trading sample as well as the insider trading sample. We divide the sample into stocks with different types of insider trades, and then further divide each sub-sample based on whether retail investors trade on the same side with insiders or on the opposite side.

[Insert Table 3.1 here]

Panel A of Table 3.1 presents summary statistics for stocks traded by insiders and retail investors. We report the main variables in weekly frequency. For retail order imbalance measures, the mean and median values for both the volume imbalance (*oibvol*) and the trade imbalance (*oibtrd*) are negative, with the mean of *oibvol* -0.021 and the median -0.013, and the mean of *oibtrd* -0.016 and the median -0.004. These statistics are consistent with Boehmer et al. (2021), in which the mean retail order imbalance is also negative. Insider trades vary a lot in terms of the number of shares traded. While the weekly median of the insider buy volume is 10,000 shares and the weekly median of the insider sale is 20,000, their weekly mean values reach 233,000 and 228,000, respectively. The 75^{th} percentile of both variables are 42,120 and 62,770. These figures suggest that some insiders trade with abundant blocks of shares.[14] Moreover, the number of insider purchases has a weekly average of 4.6 (times) and a weekly median of 2. The number of insider sales has a weekly average of 5.1 and a weekly median of 2.

Panel B and Panel C of Table 3.1 report characteristics of stocks purchased and sold by insiders, respectively. For each panel, we further group stocks into those that retail investors trade on the same side as insiders (Follow), and those that retail investors trade on the opposite side (Not-Follow). We compare stock size, book-to-market ratio, past 6-month momentum, return reversal, turnover, idiosyncratic volatility as well as investor abnormal attention for stocks in sub-samples. However, retail investors' trading decisions do not depend on specific stock characteristics. If we compare stock characteristics among insider trading sub-samples, we find that stocks insiders sell tend to have higher momentum (both past 6-month cumulative return and prior month return) than those bought by insiders. The contrarian trading pattern is consistent with prior literature (Piotroski and Roulstone, 2005; Cohen, Malloy, and Pomorski, 2012).[15]

3.2.5 Retail Order Flow around the Insider Trading Event Window

We define the insider trading event window as the days from the insider trading date until one day after the insider filing date. The U.S. Securities and Exchange Commission (SEC) requires that corporate insiders report their open market trades within 2 days after the trading date. This 2-day deadline has replaced the previous 10-day deadline since 2002. To rule out the effect of nearby insider trades in our event window for each stock, we cluster the

[14] As these large block trades are more likely to catch attention from regulators and other investors, we expect that they contain less information. Our results are robust after removing these large block trades.

[15] In this paper, we focus on what drives retail investors' trading on the same side as insiders. Therefore, we compare stock characteristics between Follow and Not-Follow groups in each insider universe.

adjacent insider trades (within 5 days) into one group.[16] Also, it is essential to exclude the possibility that insider trades in our sample are classified as OTC retail trading. Chung (2020) obtain NYSE retail trading volume and indicates it contains trades from insiders. However, even if we assume all the insider trades are executed by OTC wholesalers or via internalization, there are at most 4.9% of the total insider trades may be included in retail OTC volume.[17] Removing them from retail trades does not affect our main results. We report retail order imbalance within 5 trading days around the event window with Newey-West adjusted standard errors.

[Insert Table 3.2 here]

Table 3.2 provides summary statistics of retail order imbalance around event window. Consistent with 1, retail order imbalance is negative for days before and after the insider purchase event window. However, during the insider trading window, the aggregate retail investors tend to buy the stocks that insiders have purchased. Specifically, Panel A shows that *oibvol* increases by 129 percent from the day before to the days in the insider purchase window, and *oibtrd* increases by 146 percent during the same period. Panel B shows that retail investors tend to sell the stocks that insiders have sold, more than they usually sell, during the event window.[18] Between the day before and during the insider sale window, *oibvol* decreases by 14 percent and *oibtrd* decreases by 5 percent.

[16] Clustering is necessary because adjacent insider trades is common in our sample, where 48.6% of the sample have adjacent insider trades. For robustness, we also employ 10-day clustering and 20-day clustering, and the results remain similar.

[17] Using the price improvement measure in Boehmer et al. (2021) to identify retail trades, we find more than 85% of the insider trades in our sample cannot be classified as retail trading. For the remaining insider trades that may be counted in retail trading, we further exclude the insider buy (sell) transactions with volume exceeding retail buy (sell) volume on the same trading day.

[18] Given the negative-skewed nature of retail order imbalance, which means retail investors sell their shares overall, we still capture a more negative retail order imbalance during the insider sale event window.

Next, we differentiate between informed and uninformed insider trades, as in Cohen, Malloy, and Pomorski (2012). They identify insiders as either opportunistic (informed) or routine (uninformed), based on past trading patterns. Interestingly, retail investors' trades as a group are consistent with opportunistic insiders' trades, and not with routine insiders' trades. For the opportunistic insider trades, retail investor order imbalance during the event window is 0.9 percent in terms of *oibvol*, and 1.1 percent in terms of *oibtrd*, both statistically significant. For the routine insider trades, the corresponding *oibvol* is -1.5 percent and *oibtrd* -0.6 percent. These results suggest that aggregate retail investors do not blindly follow insider trades. Rather, they tend to follow informed insider trades more than they follow uninformed ones. There is no clear distinction of retail investor trading pattern for opportunistic and routine sale conditions. The result is not surprising, as most papers show that insider sales are usually driven by liquidity and diversification reasons, where insider purchases are regarded as strong positive signals in stock values. For example, Jeng, Metrick, and Zeckhauser (2003) find that insider purchases earn abnormal returns of more than 6% per year, while insider sales do not earn significant abnormal returns. Hence, from now on we focus on retail order flow around insider purchase window.

[Insert Figure 3.1 here]

Figure 3.1 presents retail order imbalance, *oibvol* and *oibtrd*, from 20 trading days before the insider trading event window to 20 trading days after the event window. Consistent with earlier results, there is a spike for the insider purchase event window from one day before, and a decrease for the insider sale event window from one day before. During the event window, order imbalance is positive for the opportunistic purchase sample but negative for the routine purchase universe, shown in both *oibvol* and *oibtrd*. The order imbalance in the routine sample is also more volatile than that in the opportunistic sample.[19]

[Insert Table 3.3 here]

Results in Table 3.2 suggest that retail investors learn about insider trades, especially opportunistic insider trades, and act in a timely way. While it is difficult to pin point the exact (likely multiple) channel of their learning, we examine one specific source of insider trading information that retail investors can have access to – the SEC's EDGAR database. We report the abnormal retail downloads of the insider Form 4 filing from EDGAR, as described in the earlier section. Panel A of Table 3.3 shows that there is a surge in the retail downloads of Form 4 filings during the event window for insider purchases, with the average abnormal EDGAR downloads at 2.36. And this surge is entirely for stocks with opportunistic insider purchases at 2.59. The corresponding abnormal downloads for routine purchases are actually negative. The average abnormal retail downloads from EDGAR for insider sales are much smaller at 0.8, and there is no significant difference for the abnormal EDGAR downloads between opportunistic insider sales and routine sales.

Panel B of Table 3.3 shows the retail trading on stocks with insider trades. We sort stocks based on the ranking of the abnormal EDGAR downloads, and compare the difference in retail trading between those stocks in the top quintile of the abnormal EDGAR downloads and those in the bottom quintile. There is significant difference in retail trading between the top and the bottom quintile during the event window and 2 days after the filing date, and the difference remain positive for the next 5 trading days. For example, *oibvol* during the event window for the top EDGAR search quintile is 2.50%, while that for the bottom quintile is -1.35%. The difference between the top and the bottom quintile is

[19]It could be the reason that retail investors regard routine trades as uninformative, and they tend to trade more arbitrarily than in the opportunistic sample.

3.85%, statistically significant at the 1% level. We further make the comparison separately for stocks with opportunistic insiders buys and routine insider buys. For stocks with opportunistic insider buys, *oibvol* for the stocks in the top quintile of the abnormal EGDAR downloads during the event window is 2.30%. The difference in retail trading between the top and the bottom quintile is 2.70%, statistically significant at the 5% level. For stocks with routine insider buys, *oibvol* for the stocks in the top quintile of the abnormal EGDAR downloads during the event window is smaller at 0.6% and statistically insignificant. Overall, results in Table 3.3 provide direct evidence that some retail investors do learn about insider trades through EDGAR search, and downloading the insider filing forms has effect on their trading decisions.

3.3 Retail Trading Patterns around Insider Trades

In this section we examine the retail weekly trading pattern for stocks that insiders bought. We then examine potential sources that may explain these trading patterns.

3.3.1 Baseline Results

In our main empirical analysis, we use weekly-frequency data to reduce microstructure noise. We also run daily-frequency tests in the appendix and the results all remain unchanged. We employ multivariate analysis of event-week retail order imbalance on insider trading indicators. Specifically:

$$Oib_{i,t} = a + b * Ins_{i,t} + c * X_{i,t-1} + \epsilon_{i,t} \quad (3)$$

We regress stock i's retail order imbalance in week t ($\mathrm{Oib_{i,t}}$) on an insider trading dummy variable in the same week ($\mathrm{Ins_{i,t}}$), and control for other stock characteristics ($\mathrm{X_{i,t\text{-}1}}$). We use both variables (*oibvol* and *oibtrd*) as proxies for $\mathrm{Oib_{i,t}}$. $\mathrm{Ins_{i,t}}$ refers to a dummy variable for an insider purchase, an opportunistic insider purchase, an opportunistic insider sale, a routine insider purchase, or a routine insider sale. The dummies equal one if the stock in week t has such an insider trade. To check the lead-lag relation where retail investors may follow insider trades, we employ another setting as follows, where the dependent variable is stock i's retail order imbalance in week *t+1* ($\mathrm{Oib_{i,t+1}}$):

$$Oib_{i,t+1} = a + b * Ins_{i,t} + c * X_{i,t-1} + \epsilon_{i,t} \tag{4}$$

The control variables $\mathrm{X_{i,t\text{-}1}}$ include firm size (LSIZE), book-to-market ratio (LBM), turnover (TURN), idiosyncratic volatility (IVOL), the past 6-month momentum (MOM), the prior month return (RET1), the prior week return (RET1W), investor general attention (ATT), as well as retail investor attention (ASVI). Specifically, LSIZE is the natural logarithm of market capitalization. LBM is the natural logarithm of the most recent fiscal year-end book value divided by the market capitalization. MOM is the past cumulative returns from the month -7 to the month -2. RET1 is the prior month's return. TURN is the monthly trading volume divided by the average number of shares outstanding over the past 12 months. IVOL is the standard deviation of the residuals from the Fama-French three-factor regressions of daily stock excess returns over the previous 6 months. ATT is the weekly average of the daily trading volume divided by the average trading volume in the previous year (252 trading days). ASVI is the weekly abnormal change in Google Search Volume Index. We run weekly panel regressions for all stocks with insider trades during the week, with week fixed effect and two-way (firm and week) clustered standard errors.

The results are reported in Table 3.4. Consistent with Figure 3.1, retail investors trade on the same side with insiders when insiders purchase the stock. Furthermore, retail investors tend to follow informed insider purchases (opportunistic buys) rather than uninformed ones (routine buys). Models (1)-(2) show the results when the dependent variable is *oibvol*. The coefficient on the insider buy dummy (Buy) is 0.125 with a t-statistics of 10.45, and the coefficient on the opportunistic insider buy dummy (Opp_buy) is 0.104 with a t-statistics of 4.42. The coefficient on the routine buy dummy (Rou_Buy) is -0.002 and statistically insignificant.

[Insert Table 3.4 here]

Results in Model (1) and Model (2) suggest that stocks with insider purchases have 12.5% higher retail *oibvol* in the current week compared to stocks with insider sales, and stocks with opportunistic insider purchases have 10.4% higher retail *oibvol* in the current week compared to stocks without opportunistic insider purchase. Model (3) and Model (4) show the results when the dependent variable is *oibtrd*. Stocks with insider purchases have 11.3% higher retail *oibtrd* in the current week compared to stocks with insider sales, and stocks with opportunistic insider purchases have 11.8% higher retail *oibtrd* in the current week compared to stocks without opportunistic insider purchase. Results in Models (5)-(8) for retail order imbalance in week t+1 remain consistent with the contemporaneous tests, although the coefficients on main independent variables decrease in magnitude.

Overall, the regression results indicate that retail order imbalances are consistent with insider purchases in the event week, especially with informed insider purchases. This finding is in contrast to previous findings that retail investors make systematic investment mistakes (e.g. Barber and Odean, 2000). More importantly, we show that retail investors as a whole do not blindly follow insider trades. Instead, their trading patterns around the insider trading window suggest that they tend to follow informed insider trades.

3.3.2 Retail Investor Attention

While our initial results appear consistent with retail investors following informed insider trades, other reasons may explain the observed pattern. It is plausible that insider trading increases retail attention at the aggregate level. A number of papers argue that investors' attention is limited when they are selecting stocks.[20] Barber and Odean (2008) show that retail investors are net buyers of attention-grabbing stocks, while it is not the case for selling because retail investors only sell what they own (short-sale is not common for retail investors). Da, Engelberg, and Gao (2011) further use Google search volume to proxy for the attention of retail investors, and show transitory price pressures on those attention-grabbing stocks.

We control for potential effect of investor attention by considering investor general attention (ATT, Barber and Odean, 2008) and retail investor attention (ASVI, Da, Engelberg, and Gao, 2011) in Table 3.4. Coefficients on both investor attention variables are positive and significant. However, after controlling for the general attention and retail investor attention, the coefficients of the main independent variables remain significant.

We also run the panel regressions separately for stocks with high and low investor attention. We use Abnormal Search Volume Index (ASVI) as the proxy for retail investor attention. We assign ASVI to be high when it is among the highest 33.3% of the sample, and low when it is among the lowest 33.3% of the sample. We then run regressions for the sub-samples $\mathrm{ASVI_t}$ = High and $\mathrm{ASVI_t}$ = Low separately, and report the coefficient estimates for Opp_Buy and Rou_Buy in 5, as well as coefficient differences between Model (1) and Model (3), and between Model (2) and Model (4).

[20]See, Kahneman (1973) and Hirshleifer and Teoh (2003).

[Insert Table 3.5 here]

The key message from Table 3.5 is that the coefficients on Opp_Buy remain significantly positive for both high ASVI and low ASVI stocks, while the coefficients on Rou_Buy stay insignificant for both cases. Specifically, when the retail order imbalance is defined as *oibvol*, the coefficient on Opp_Buy is 7.8% (t-stat = 3.01) for the high attention sub-sample and 9.5% (t-stat = 3.58) for the low attention sub-sample. Similarly, when the retail order imbalance is defined as *oibtrd*, the coefficient on Opp_Buy is 10.3% (t-stat = 4.43) for the high attention sub-sample and 8.2% (t-stat = 3.44) for the low attention sub-sample. The coefficients on Rou_Buy become negative and insignificant when ASVI is low.

These results suggest that elevated investor attention is not a driving force for retail investors to follow insider trades. When there are insider purchases, retail investors tend to follow them, regardless of the level of the investor attention. For informed trades (opportunistic buys), retail investors follow them anyway; for uninformed trades (routine buys), increase in attention is affecting retail investor trading direction to some extent, although not statistically significant. These findings suggest that retail investors do not simply trade stocks due to higher attention.[21]

3.3.3 Common Sources of Information - Earnings Announcement

Another potential explanation of retail trading pattern is that retail investors and corporate insiders share the same set of information around the same time when making their trading decisions. Earnings announcement, which is critical to assess the fundamental value of a firm, is an important source of

[21] We also use investor general attention (ATT) to separate the sample and find similar results.

information for most investors including small retail investors. For example, Lee (1992) and Frazzini and Lamont (2021) find evidence of net small buys on the earnings announcement date and immediately after the event. Kaniel et al. (2012) observe that individuals are net sellers at the time of the earnings announcement and several days after the event, and net buyers before the event. Besides, they show that purchases (sales) of stocks by retail investors before the earnings announcements can predict returns after the announcements.

In our context, if retail investors and insiders share common (private) information about future earnings, it could explain why they trade in the same direction around the same time. To examine this possibility, we divide the sample of stocks into two sub-samples, those with near-future earnings news and those without. Specifically, we define the upcoming earnings announcement dummy $\mathrm{EA}_{[t+1,\ t+4]} = 1$ if there is an earnings announcement for the stock within 4 weeks after the insider trading event, and 0 otherwise. We then repeat the panel regressions for each sub-sample separately.

Panel A of Table 3.6 presents the regression results based on whether there is any earnings announcement in the upcoming month. Results show that retail investors follow opportunistic insider purchases regardless of any earnings announcement. When there is an upcoming earnings announcement, retail investors tend to follow opportunistic insider purchases. Specifically, the coefficient on Opp_Buy is 8.3% with the t-statistic of 2.02 when the retail order imbalance is defined as *oibtrd*, although for *oibvol* the corresponding coefficient is 2.4% with the t-statistic of 0.53. When there is no upcoming earnings announcement, the coefficient on Opp_Buy is 11.1% with the t-statistic of 7.68 when the retail order imbalance is defined as *oibtrd*, and for *oibvol* the corresponding coefficient is 11.3% with the t-statistic of 6.98. It is possible retail investors learn about earnings information ex-ante, and trade correspondingly. But for the majority of our cases when there is no earnings news, retail investors still follow insider trades. As comparison, retail investors do not follow routine insider purchases in either scenario.

[Insert Table 3.6 here]

Furthermore, Kaniel et al. (2012) show that intense retail investors buying (selling) predicts large positive (negative) abnormal returns on and after earnings announcement. Thus, we further divide the sub-sample of $EA_{[t+1,\ t+4]} = 1$ into cases when SUE >0 and SUE <0. SUE is defined as the three-day abnormal return around the earnings announcement event. Panel B reports the regression results based on whether there is a positive or negative SUE.

We do find that retail investors follow insider purchases more aggressively for stocks with positive future SUE than those with negative future SUE, which appears consistent with Kaniel et al. (2012). Specifically, when the retail order imbalance is measured as *oibtrd*, the coefficient on Opp_Buy is 0.187 (t-stat = 3.03) for SUE >0, and the corresponding coefficient is 0.048 (t-stat = 0.74) for SUE <0. The difference in the coefficient estimates in these two scenarios is however not statistically significant. The results are similar when the retail order imbalance is measured as *oibvol*. As for routine-purchased stocks, retail investors tend to purchase these stocks when positive SUE is expected, and they sell when negative SUE is expected. The difference between the two scenarios for the Rou_Buy is 30.0% (t-stat = 2.33) when the retail order imbalance is measured as *oibtrd*, and it is 49.9% (t-stat = 3.58) when the retail order imbalance is measured as *oibvol* .

Overall, results in this sub-section suggest that some retail investors might have information about future earnings. However, the upcoming earnings information is not a driving force for retail investors to follow informed insider trades, as they follow insider purchases for the vast majority of the cases when there is no near future earnings announcement.

3.3.4 Common Sources of Information - Analyst Revision

Another source of common information could be the near-future analyst forecast revision or recommendation change. Prior literature documents that (changes in) analyst recommendations yield abnormal future returns (e.g. Womack, 1996; Barber, Lehavy, Mcnichols, and Trueman, 2001; Jegadeesh, Kim, Krische, and Lee, 2004). Moreover, Malmendier and Shanthikumar (2007) and Mikhail, Walther, and Willis (2007) show that analyst recommendations are treated literally by retail investors.

To examine the effect of analyst recommendation changes/forecast revisions on retail trading patterns following insider purchasing, we repeat the above analysis, except that we use analyst recommendation changes and forecast revisions as the conditioning variables to divide the sample. We report the regression results for the sub-samples of recommendation upgrades and downgrades in the next month separately. We define the sub-sample as "Rec Upgrade" when the change in analyst recommendations within the next month after the insider trading week is positive. We define the sub-sample as "Rec Downgrade" when the change in analyst recommendations within the next month after the insider trading week is negative. The change in analyst recommendations is calculated as the next month consensus recommendation minus the consensus on the same stock in the past month. Table 3.6, Panel C reports the regression results conditioning on analyst recommendation change. Retail investors tend to follow informed (opportunistic) trades regardless of recommendation upgrades/downgrades, indicating that information about analyst revision do not drive the retail trading pattern following insider purchases. As for coefficients of routine buys, we find that retail investors do not follow routine purchases in both scenarios. Especially for *oibtrd*, the coefficient on Rou_Buy is -12.8% with a t-statistics of -1.89 when there is an upcoming downgrade.

We further report the regression results for the sub-samples of stocks with analyst earnings forecast up revision and down revision in the next month separately. We define Forecast Up Revision as when the change of analyst EPS forecast is positive. We define Forecast Down Revision as when the change of analyst EPS forecast is negative. The change of analyst EPS forecast is calculated as the next month EPS Median Estimate averaged across all analysts minus the same average forecast for the same stock in the past month. Results in Panel D also show that retail investors follow opportunistic insider purchases regardless if there is any analyst forecast revision. Specifically, coefficients on Opp_Buy remain significantly positive in both scenarios, for both measures of the retail order imbalance *oibvol* and *oibtrd*. Coefficients on Rou_Buy is insignificant in all cases. Overall, neither the analyst forecast revision nor the recommendation change provides an explanation for retail investors following opportunistic insider purchase. Retail investors follow insider trades not because they share the common information with insiders. It is likely that they follow insider trades because of the private information content revealed from these informed trades, which is not yet incorporated into prices.

3.4 Price Discovery of Retail Trades around Insider Trades

As shown in the previous section, the trading pattern of retail investors, i.e. following informed insider trades during the event window, is likely due to their observing information from insider trades. Prior literature document that traders who possess private information help promote price efficiency, by moving stock prices closer to their fundamental values (e.g. Kyle, 1985; Diamond and Verrecchia, 1987). If retail investors learn about private information revealed by insider trading, they should help expedite the price discovery process of the underlying stocks by trading alongside insiders.

In this section, we employ portfolio analysis and panel regressions of future returns to test the above hypothesis. We then decompose the retail trades into three components to further examine whether results in the previous sections are driven by information, liquidity provision, or price pressure. We also examine the effect of information asymmetry on the return premiums by retail trading following insider purchases. Lastly, we provide evidence on the price efficiency gain using variance ratio and price delay measure.

3.4.1 Portfolio Returns

If retail investors expedite price discovery by following opportunistic insider purchases, we would expect high future returns for stocks they purchase. We construct event-week portfolios based on whether retail investor purchase or sell the stocks that opportunistic insider have bought during the event week. The portfolio "Follow" include stocks with positive *oibvol* during the event week, and "Not-Follow" include those with negative *oibvol* during the event week. We use CRSP daily value-weighted market returns to calculate the average weekly market return, and subtract it from weekly stock return to get the weekly market-adjusted return for each portfolio. Table 3.7 reports next-week returns as well as cumulative returns for the next 4 weeks, 12 weeks, 18 weeks, 24 weeks and 52 weeks, respectively.

[Insert Table 3.7 here]

For stocks with opportunistic insider purchases, those that retail investors have bought yield 10 bps higher returns the next week than those that retail investors have sold. The weekly cumulative return difference between the two groups remains significantly positive until week 18, peaking in week 12 at 22 bps before gradually decreasing. There is no future return reversal, inconsistent with temporary price pressure from the retail trading or retail investors providing liquidity. Results here indicate that some retail investors, by observing insider information and trading alongside them, help expedite price discovery to its fundamental value for insider-purchased stocks.

[Insert Figure 3.2, Figure 3.3 here]

Figure 3.2 plots the cumulative abnormal returns (market-adjusted and size-BM-adjusted) of the Follow and Not-Follow portfolios. Both Panel A and B show that the cumulative abnormal return difference between the Follow and Not-Follow portfolios increases in the first 12 weeks and then decreases over time, indicating a converging trend between the two groups. Panel C and D show that for stocks with routine insider purchases, the corresponding cumulative return difference is negative over time, which is consistent with routine-purchased stocks containing no valuable information (Cohen, Malloy, and Pomorski, 2012).

For stocks with opportunistic and routine purchases, we further construct a long-short portfolio (Follow – Not-Follow) and plot its cumulative abnormal returns in Figure 3.3. Consistent with results from Table 3.7, for stocks with opportunistic insider purchases, the long-short portfolio earns statistically significant positive abnormal returns up to 4 months. For comparison, the long-short portfolio earns insignificant cumulative returns over time for stocks with routine insider purchases.[22]

[22]We also check for 1-year horizon, and find that for stocks with opportunistic insider purchases, the long-short portfolio earns positive cumulative return without return reversal. For stocks with routine insider purchases, the long-short portfolio's cumulative return starts to become negative from week 30.

We conduct calendar-week portfolio analysis as an additional test. Each week for the stocks with opportunistic insider purchases, we construct the Follow – Not-Follow portfolio as defined before. We rebalance the portfolio either every week or every four weeks. For each portfolio, we calculate the equal-weighted weekly returns, the CAPM alphas, the Fama and French (1993) three-factor alphas, and the Carhart (1997) four-factor alphas.

[Insert Table 3.8 here]

Table 3.8 reports the weekly alphas for the retail Follow portfolio and Not-Follow portfolio, as well as those for the long-short portfolio (Follow - Not-Follow). Specifically, the equal-weighted portfolio that goes long Follow portfolio and goes short Not-Follow portfolio earns a CAPM alpha of 7.3 bps (t-stat = 2.01) and a Carhart alpha of 7.9 bps (t-stat = 2.15) the next week. Both the CAPM alpha and Carhart alpha remain significant until week 4, when the CAPM alpha becomes 3.6 bps per week (t-stat = 1.70) and the Carhart alpha becomes 3.6 bps per week (t-stat = 1.68).

Collectively, evidence in this section suggests that by taking advantage of information revealed from informed insiders, retail investors help expedite the price discovery process for the underlying stocks. By purchasing these stocks, retail investors move the stock prices to their fundamental values faster than those who sell the stocks.

3.4.2 Panel Regressions

We next conduct panel regressions to explain the retail trades' information content around insider trades. We use two settings for the empirical design. We regress one-week-ahead stock return on the opportunistic insider purchase dummy (Opp_Buy), the routine insider purchase dummy (Rou_Buy), as defined in 4, and the retail investor purchase dummy (Retail_Buy). Retail_Buy equals 1 if retail investors buy the stock, and 0 if they sell the stock. We also include interaction terms Opp_Buy * Retail_Buy and Rou_Buy * Retail_Buy. We then run the following panel regression:

$$Ret_{i,t+1} = a + b * Opp\_Buy_{i,t} + c * Rou\_Buy_{i,t} + d * Retail\_Buy_{i,t} + e * Opp\_Buy_{i,t} * Retail\_Buy_{i,t} + f * Rou\_Buy_{i,t} * Retail\_Buy_{i,t} + g * X_{i,t} + \epsilon_{i,t} \quad (5)$$

We include firm size, book-to-market ratio, turnover, idiosyncratic volatility, momentum, short-term reversal, prior week return, investor attention as well as retail investor attention as control variables. The regression results are reported in Table 3.9.

[Insert Table 3.9 here]

Model (1) in Table 3.9 shows the return predictability of opportunistic insider buys and routine insider buys. It provides out-of-sample evidence of Cohen, Malloy, and Pomorski (2012) about the informational role of different insider types. Cohen, Malloy, and Pomorski (2012) conduct portfolio tests to examine next-month abnormal returns for stocks with opportunistic and routine insider trades. Similarly, our weekly panel regressions show that stocks bought by opportunistic insiders earn significantly higher returns the next week, while stocks bought by routine insiders do not earn significantly higher returns the next week. Specifically, a stock with opportunistic insider purchase (Opp_Buy) has 13 bps (t-stat = 7.25) higher return the next week compared to the remaining sample. On the other hand, Rou_Buy predicts a 3 bps higher next-week return with a t-statistics of 0.80.

Model (2) shows that stocks bought by retail investors earn significantly higher returns the next week than those sold by retail investors. The coefficient on the Retail_Buy dummy is 2 bps (t-stat = 6.47), consistent with Boehmer et al. (2021). The magnitude of Retail_Buy coefficient is only about 15.4% of the magnitude of Opp_Buy coefficient in Model (1). It is not surprising that retail trades are not as informative as informed insider trades. What are the sources of information for retail investors? They may be adept at incorporating insider-trading information, or have information that is unrelated to insider information. We test these alternatives and report results in Model (3) of Table 3.9.

We include the interaction terms Opp_Buy * Retail_Buy and Rou_Buy * Retail_Buy in Model (3). For stocks with opportunistic insider purchases, those that retail investors buy earn higher returns next week, with the coefficient on the interaction term Opp_Buy * Retail_Buy equals 7 bps (t-stat = 2.04). The statistically and economically significant coefficient on Opp_Buy * Retail_Buy suggests that retail investors learn from the opportunistic insider purchases, and they help impound the private information revealed from the informed insider trades into stock prices. By comparison, the coefficient on the interaction term Rou_Buy * Retail_Buy is not significant. As expected, retail investors do not earn significantly higher returns by following routine trades, which contain no valuable information.

In the second regression setting, we focus on the stocks with opportunistic insider purchases. Specifically, we run the following panel regressions of next-week returns on Follow_Oppbuy, controlling for other stock characteristics ($\mathrm{X_{i,t}}$):

$$Ret_{i,t+1} = a + b * Follow\_Oppbuy_{i,t} + c * X_{i,t} + \epsilon_{i,t} \tag{6}$$

Follow_Oppbuy equals 1 if retail investors buy stocks in the event week following opportunistic insider purchases, i.e., positive *oibvol*, and equals 0 if they sell stocks in the event week following opportunistic insider purchases, i.e., negative *oibvol*.

The result is shown in Model (4) of Table 3.9. For stocks with opportunistic insider purchases, those that retail investors buy earn 7 bps higher return next-week. Again, the result is consistent with that retail investors observing insider information and trading along with insiders. Overall, both portfolio analysis and panel regressions in this section indicate that retail investors learn from informed (opportunistic) insider purchases and their trading alongside insiders helps speed up price discovery.

3.4.3 Decomposing Retail Order Imbalance around Insider Trades

As we discussed in the previous section, there are alternative explanations for the return premium of stocks bought by retail investors following opportunistic insider purchases. One potential explanation is that retail investors possess information that is orthogonal to insiders' information. Our earlier results on retail trading controlling for earnings announcements, analysts forecast revision and recommendation updates in Section 3.3 are against this hypothesis. Moreover, portfolio tests in this section show that when retail investors sell stocks that opportunistic insider have purchased, these stocks still have higher future returns, albeit not as high as those that bought by retail investors. It seems that retail investors do not have superior information during the insider trading window on top of that possessed by insiders.

Prior literature documents that retail traders benefit from their liquidity provision (e.g. Kaniel, Saar, and Titman, 2008; Kaniel, Liu, Saar, and Titman, 2012). In our context, because retail investors trade in the same direction with insiders, they are not providing liquidity to insiders. However, it is possible that retail investors provide liquidity to other market participants who trade against them, and earn profits from the liquidity provision.

Apart from liquidity provision explanation, price pressure may also lead to the above return premium in the short-run. For example, Chordia and Subrahmanyam (2004) point out the persistence of order flow which results in the predictability of future stock returns. They denote this phenomenon as the "price pressure hypothesis". By decomposing the retail order imbalance into three components related to price pressure, liquidity provision and information, Boehmer et al. (2021) show that nearly half of the predictive power of the retail order imbalance comes from price pressure (order imbalance persistence), and most of the rest comes from the information component.

We follow Boehmer et al. (2021) and conduct a two-step decomposition of retail order imbalance into three components: Retail_Buy_Persistence, Retail_Buy_Contrarian and Retail_Buy_Other.[23] In the first step, we estimate the following regression model:

$$Retail\_Buy_{i,t} = a_t + b_t * Retail\_Buy_{i,t-1} + c_t * Ret_{i,t-1} + \epsilon_{i,t} \tag{7}$$

For each week t, we obtain the cross-sectional means of a, b and c as $\widehat{a_{\mathrm{t}}}$, $\widehat{b_{\mathrm{t}}}$ and $\widehat{c_{\mathrm{t}}}$, respectively. We then compute Retail_Buy_Persistence as $\widehat{b_t}$ * Retail_Buy$_{\mathrm{i,t-1}}$. We compute Retail_Buy_Contrarian as $\widehat{c_t}$ * Ret$_{\mathrm{i,t-1}}$. The residual part from the first-stage regression is denoted as Retail_Buy_Other, which is likely driven by informational content. In the second stage, we run panel regressions similar to Table 3.9, except that we replace Retail_Buy with the three components that we compute in the first stage. We include the same set of control variables as in Table 3.9. We include week fixed effects for all models. Standard errors are two-way clustered at the firm and the week level in all models.

[23] We also follow Kaniel et al. (2012) decomposition method and find information part mostly explains our above results. However, Boehmer et al. (2021) methodology suits our scenario better as it also involves retail order imbalance.

Table 3.10 reports the second-stage regression results of the decomposition analysis. Model (1) shows that for general retail buys without opportunistic insider purchases, the abnormal returns from retail purchasing come from both the persistence (price pressure) component (Retail_Buy_Persistence predicts a 10 bps positive return next week with t-stat = 5.31) and the residual (information) component (Retail_Buy_Other predicts a 2 bps positive return next week with t-stat = 5.72), where the magnitude of price pressure component is 5 times as large as information component. However, for stocks with opportunistic insider purchases, the coefficients on the interaction terms of the three components with Opp_Buy indicate that the return predictability mostly come from the information component. The interaction term Opp_Buy * Retail_Buy_Other predicts a 7 bps positive return next week, with a t-statistic of 2.06. The magnitude (7 bps) is much larger than that of the general retail trading (2 bps), indicating that the information content is an essential driver for the return predictability of the retail trading following opportunistic insider purchases.

[Insert Table 3.10 here]

For comparison, in Model (2) we also examine the predictive power for Rou_Buy, with all else equal, and find that neither of the three components explains the return premium. Our results remain consistent when we include both Opp_Buy and Rou_Buy and corresponding interaction terms in Model (3).

In Model (4), we restrict the sample to stocks with opportunistic insider purchases only, and interact Follow_Oppbuy dummy (whether retail investors buy or sell these stocks) with the three retail buy components as constructed above. The t-statistics for the interaction terms Follow_Oppbuy * Retail_Buy_Persistence and Follow_Oppbuy * Retail_Buy_Contrarian show insignificance. By comparison, the coefficient on the interaction term Follow_Oppbuy * Retail_Buy_Other predicts a 1.7% positive future return with a t-statistic of 2.55.

Overall, the two-step decomposition of the retail order imbalance on stocks with opportunistic insider purchases shows that the return predictability does not come from price pressure or liquidity provision. Instead, it most likely comes from the information that retail investors learn from insider trades.

3.4.4 Information Asymmetry and Stock Returns

Information asymmetry between informed and uninformed investors affect stock return. O'Hara (2003) and Easley and O'hara (2004) use information-risk models to demonstrate that returns are positively related to information asymmetry when investors rely more on private information. Coupled with limits to arbitrage theory, as price deviates from fundamental value,[24] informed investors could earn return premiums with long persistence. If retail investors indeed help impound private information into prices by trading along with informed insiders, we expect this effect to be stronger for stocks with higher ex-ante information asymmetry. In this sub-section, we sort stocks on their ex ante information asymmetry, and examine whether the return premiums become larger when the ex-ante information asymmetry is higher.

For stocks with opportunistic insider purchases, we first separate them into the Follow portfolio and the Not-Follow portfolio, based on whether retail investors follow insiders as defined in Table 3.1. We then categorize stocks into three levels of information asymmetry. For each level, we compute the difference in the cumulative market-adjusted returns between the Follow portfolio and the Not-Follow portfolio, up to 24 weeks. We use three proxies for the ex-ante information asymmetry. The first proxy is firm size, which is defined as the logarithm of the market capitalization.[25] Prior literature shows that small firms usually have higher information uncertainty as they are less diversified and have

[24]See, for example, Long et al. (1990) and Shleifer and Vishny (1997).

[25]The use of firm size as a proxy is common. It is justified in papers such as Atiase (1985), Bamber (1987) and Llorente et al. (2002).

less information available than large firms. Additionally, we use idiosyncratic volatility to proxy for information asymmetry, as Pastor and Veronesi (2003) show a positive relation between information asymmetry and idiosyncratic return volatility. Lastly, we use Amihud (2002) illiquidity as the third proxy for information asymmetry.

[Insert Table 3.11 here]

Table 3.11 reports the effect of retail investors following opportunistic insider purchases on market-adjusted returns for stocks with different levels of information asymmetry. Results show that the cumulative return differences are larger for stocks with smaller size, higher idiosyncratic volatility, or higher Amihud illiquidity. For instance, from the lowest idiosyncratic volatility tercile to the highest tercile, the 4-week cumulative abnormal return of the long-short (Follow – Not-Follow) portfolio increases by 61 bps.

The above findings further support our hypothesis that by trading alongside with informed insiders, retail investors help impound valuable information into the underlying stock prices, and such effects are stronger for stocks with higher ex-ante information asymmetry. In contrast to Gromb and Vayanos (2010) who document the irrational buying decisions of individual investors creating more mis-pricing, our paper indicates that individual investors are not necessarily irrational, but help improve price efficiency through informed trading.

3.4.5 Testing Price Efficiency Using Variance Ratio and Price Delay Measure

In this sub-section, we further examine the efficiency gain of retail trading via two commonly-used measures: the Lo and MacKinlay (1988) variance ratio and the Hou and Moskowitz (2005) price delay measure.

According to efficient pricing models, relative informational efficiency can be measured as how closely transaction prices resemble a random walk. Given information arrival and market frictions, price trajectory deviate from efficient prices temporarily. Testing the informational efficiency of prices can be done by examining how fast those deviations disappear over time.[26] Following Lo and MacKinlay (1988), we first use variance ratios (long-term to short-term return variances) to measure price efficiency. VR(n,m) represents variance ratio of the m-day return variance per unit time divided by the n-day return variance per unit time estimated, using daily returns next month. We compute |1- VR(n,m)| to examine the gap between the actual and the efficient prices in either direction. The null hypothesis is that |1- VR(n,m)| equals 0 in an efficient market: we expect that the retail trading negatively predicts |1- VR(n,m)| if they improve the price efficiency. We test the variance ratios in terms of (1, 10) and (1, 20) days. We repeat the previous panel regressions, except that we use variance ratios as the dependent variables. The results are presented in Table 3.12, Models (1)-(4).

[Insert Table 3.12 here]

The first 4 Models in Table 3.12 show that the coefficient of interaction term Opp_Buy * Retail_Buy is significantly negative. In addition, Follow_Oppbuy significantly predicts a negative next-month |1- VR| for stocks with opportunistic insider purchases. The coefficient on Opp_Buy * Retail_Buy is -0.0145 (t-stat = -1.82) for |1- VR(1, 10)| and -0.0459 for |1- VR(1, 20)|. The coefficient on Follow_Oppbuy is -0.0169 (t-stat = -2.05) for |1- VR(1, 10)| and -0.0561 for |1- VR(1, 20)|. By comparison, the coefficient on Rou_Buy * Retail_Buy is statistical insignificant. The tests on the variance ratio suggest that following opportunistic

[26]See, for example, Barnea (1974), Lo (2004) and Chordia, Roll, and Subrahmanyam (2005) documenting about random walk, adaptive market hypothesis and astute traders.

insider purchases, retail investors improve informational efficiency on the stocks they buy.

We next conduct panel regressions analysis of the price delay measure. The price delay measures the relative price efficiency by quantifying the speed of the adjustment to the market-wide information (Morck, Yeung, and Yu, 2000; Griffin, Kelly, and Nardari, 2010; Saffi and Sigurdsson, 2010; Boehmer and Wu, 2013). We adopt Hou and Moskowitz (2005)'s measure to estimate how quickly prices incorporate the information retail investors learn from insider trades. Note that the information applied in price delay measure is of market type whereas the variance ratio mostly measures firm-level information, and thus the two sources of information differ fundamentally. As our data is in daily frequency as opposed to Hou and Moskowitz (2005)'s monthly frequency, we compute the price delay measure by calculating the R^2 based on daily market return and stock return as follows:

$$r_{i,t} = a_i + b_i * R_{m,t} + \sum_{n=1}^{5} R_{m,t-n} + \epsilon_{i,t} \tag{8}$$

$\mathrm{R_{m,t}}$ ($\mathrm{r_{i,t}}$) represents the daily market return (stock return). We calculate price delay measure as 1 - [(R^2(restricted model) / R^2(unrestricted model))], where we compute R^2(restricted model) by restricting the coefficients on lagged market returns to be zero. A larger price delay indicates that the stock incorporates market-wide information with a slower speed.

Results for the price delay regressions are shown in Models (5)-(6) in 12. The corresponding coefficient on Opp_Buy * Retail_Buy in the full sample is negative but insignificant (coeff. = -0.4369 with t-stat = -0.35). The magnitude of the coefficient on Follow_Oppbuy is slightly larger (coeff. = -1.3602) but only marginally significant at the 10% level. These results show little gain in price efficiency measured by the price delay. As we mentioned earlier, the variance ratio mostly measures the firm-level information incorporated into prices, where the price delay measures the speed of the market information incorporated into prices. The empirical results in this sub-section indicate that the trading of retail investors alongside informed insiders helps impound firm specific information, rather than the aggregate market-wide information, into stock prices faster.

3.4.6 Retail and Institutional Trading Patterns in Longer Horizons

We further examine whether retail investors trade on the same direction with insiders in longer horizons. Prior studies show that insider purchases yield significant and persistent returns. Jeng, Metrick, and Zeckhauser (2003) find that insider purchases earn cumulative abnormal returns of more than 6% over the subsequent 100 trading days. Cohen, Malloy, and Pomorski (2012) document that the 12-month event-time return on the opportunistic long-short portfolio (opportunistic buys - opportunistic sells) is more than 5%, with no future reversal. We provide evidence in Section 3.4.1 that stocks purchased by opportunistic insiders earn positive and significant cumulative returns up to one year. A natural question is whether retail investors continue to trade in the same direction as insiders during the following months, and whether they can potentially benefit from the persistent positive returns.

We first examine the longer-term retail trading pattern. We employ the same econometric specification and the set of control variables as in Section 3.3.1, except that we replace the dependent variable with retail order imbalances in each of the subsequent four quarters.[27]

[Insert Table 3.13 here]

[27]We also check retail order imbalance in each of the subsequent twelve months, and the result remains similar. We do not report it here for brevity.

The results are reported in Panel A of Table 3.13. During the quarter with the insider trading event, retail investors trade in the same direction as insiders in general (model 1), and the same as opportunistic insiders (model 2) if they make purchases. Models (1)-(2) show positive and significant coefficients on Buy and Opp_buy, respectively. Further, retail investors during the following year keep buying the stocks that insiders have bought. The coefficients on Buy (models 3, 5, 7, and 9) remain statistically significant, although much smaller in magnitude, in the following four quarters. Meanwhile, for stocks with opportunistic insider purchases, retail investors generally buy them in the subsequent year, but the coefficients on Opp_buy (models 4, 6, 8, and 10) are no longer statistically significant.

Next, we examine institutional trading subsequent to insider purchases. We regress the change in institutional ownership of a stock on insider trading dummy variables (main variables of interest: Buy and Opp_Buy) for that stock. We measure the change in institutional ownership in each of the four quarters following insider purchases. Panel B of Table 3.13 indicates that for stocks bought by insiders in general (model 1) and opportunistic insiders (model 2), institutional investors trade on the opposite direction with these insiders in the contemporaneous quarter. The coefficient on Buy is -0.5% (t-stat = -3.28) and the coefficient on Opp_Buy is -0.5% as well (t-stat = -2.76). We further find that institutional investors keep (short) selling the stocks that insiders have bought in the subsequent year. According to models 3, 5, 7, and 9, the Buy coefficients are all negative, but their magnitudes decrease over time and become statistical insignificant after two quarters. After opportunistic insider purchases, institutional investors sell them for up to subsequent three quarters. The magnitudes are larger for the following two quarters (models 4 and 6) compared to those in the general insider purchasing case (models 3 and 5). After the subsequent two quarters, the coefficients on Opp_buy (models 8 and 10) are no longer statistically significant.[28]

Taken together, this sub-section provides empirical evidence that retail investors keep following (opportunistic) insider purchases over a longer horizon. On the other hand, besides contemporaneously providing liquidity to inside buyers, institutions continue to trade on the opposite side of (opportunistic) insiders in the subsequent year.

3.5 Additional Robustness Checks

3.5.1 Granger-type Causality Test

The reverse causality concern is that insiders may learn from the retail order imbalance. We run Granger-type causality test to address the direction of causality. We first regress the retail order imbalance of stock i in week t on the 1-week-lagged insider trading dummies. We then regress the insider trading dummies of stock i in week t on the 1-week-lagged retail order imbalance, both during the insider-trading event window. We run Fama-MacBeth regressions with Newey-West 5 lags adjusted and include similar control variables as in Table 3.4. The results of Models (1)-(3) in Table A2 are consistent with those in Table 3.4, where more insider purchases are associated with higher next-week retail order imbalance *oibvol*.[29] Specifically, the dummy l_buy predicts a 7.3% increase in the retail order flow next week, with a t-statistic of 3.60. The dummy l_opp_buy predicts a 14.8% increase in the retail order flow next week (t-stat = 3.01) whereas the dummy l_rou_buy has a negative and insignificant coefficient. Results from Models (1)-(3) further corroborate the idea that retail investors follow insider trades. In particular, they follow informed insider trades

[28] Cohen, Malloy, and Pomorski (2012) find that institutional investors buy those insider-purchased stocks in the following quarter. They aggregate the insider trading at the quarterly level. Our regressions are at the weekly level, and our sample period is after their sample ends.

[29] We also use *oibtrd* as (in)dependent variable and find similar results. We do not report it for brevity.

(l_opp_buy) instead of the uninformed ones (l_rou_buy). Moreover, large and significant coefficients on the lagged order flow indicate that the retail order flow has high autocorrelation, consistent with Boehmer et al. (2021).

In Table A2 Models (4)-(6), we regress the insider trading dummy (Buy in (4), Opp_Buy in (5) and Rou_Buy in (6)) on the 1-week-lagged order imbalance (l_oibvol) to examine the reverse causality. Coefficients on l_oibvol are negative in all models, indicating that the buying decision of insiders is not significantly affected by past retail order imbalance. We also find high autocorrelation of opportunistic purchases and routine purchases, as adjacent insider trades are common in our sample. Overall, results in this sub-section suggest that retail investors follow informed insider purchases, not the other way round.

3.5.2 Sub-sample Tests

We conduct sub-sample tests to evaluate the robustness of our results. Specifically, we examine the effect of retail investors following opportunistic insider purchases on stocks with different characteristics. We include book-to-market ratio, prior month return, Google ASVI, institutional ownership as well as earnings surprise SUE as conditioning variables to divide the sample of stocks into three levels.[30] We further separate the sub-sample stocks into the Follow portfolio and the Not-Follow portfolio, based on whether retail investors follow insiders as defined in Table 3.1. We then compute the difference in the cumulative market-adjusted returns between the Follow portfolio and the Not-Follow portfolio, up to 24 weeks, in each sub-sample. The results are presented in Table A3. We find that the return premium of Follow − Not-Follow is not concentrated in stocks with certain types.

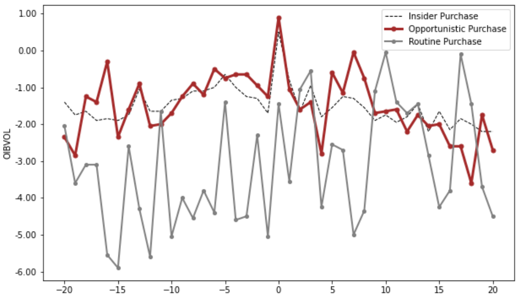
[30] Prior literature documents that these stock characteristics may affect stock returns significantly. See, for example, Fama and French (1995), Da, Engelberg, and Gao (2011), Edelen, Ince, and Kadlec (2016), Kaniel et al. (2012), and Doyle, Lundholmn, and Soliman (2006).

3.6 Conclusion

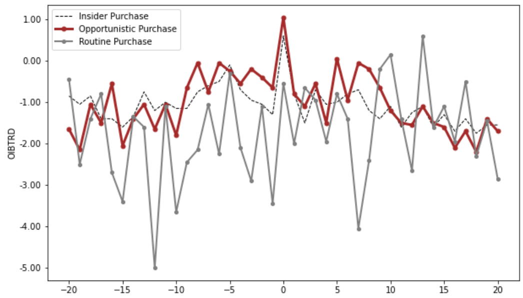
Using the comprehensive TAQ data and following the approach in Boehmer et al. (2021), we identify daily retail trading, and document evidence consistent with informed trading of retail investors. Retail investors buy more aggressively right after the insiders' opportunistic purchases, but not after the insiders' routine purchases. There is an increase in the abnormal retail downloads of the Form 4 filings from the EDGAR databases for the opportunistic insider purchases. The retail purchases are also higher with higher abnormal Form 4 downloads by retail investors. These trading patterns cannot be explained by the increased attention on these stocks, or by the shared common information between retail investors and corporate insiders, such as information on upcoming earnings announcements, analyst forecast revisions, or updates on analysts recommendation. Retail investors keep following (opportunistic) insider purchases in subsequent four quarters as well.

Moreover, for stocks with opportunistic insider purchases, those bought by retail investors experience significantly higher returns than those sold by retail investors, up to 18 weeks. In panel regressions of stock returns on an opportunistic insider purchase dummy, a retail purchase dummy, and the interaction of the two, the coefficient on the interaction term is positive and statistically significant. A long/short portfolio strategy that longs the retail-buy stocks and shorts the retail-sell stocks for stocks with opportunistic insider purchases generates significant alphas. We further decompose retail trading into three components: price pressure, liquidity provision, and information. Only the information component significantly predicts future stock returns, and this effect is stronger for stocks with greater information uncertainty. Collectively, these results suggest that retail trading helps improve price efficiency by impounding information revealed from insider trading into stock prices. Further analysis suggests that retail trading following opportunistic insider purchases lowers future variance ratio, but not the price delay measure. To the extent that variance ratio measures the level of information efficiency at the firm level and that price delay measures at the market level, our results suggest that price discovery and efficiency gain through retail trading are mostly about the firm level information, rather than the market level information.

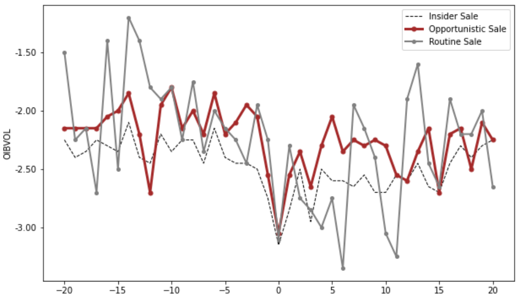
There is an ongoing debate with mixed evidence on the role of retail investors in financial markets, particularly regarding their informational role. Using the interaction with insider trading, we examine one specific source of information for retail investors. We show that at least some retail investors learn about private information revealed by insiders' opportunistic purchases. As retail investors face fewer restrictions relative to institutional investors and have fewer concerns over risk exposure, they can act promptly upon this information. Their trading predicts future returns and helps expedite price discovery.



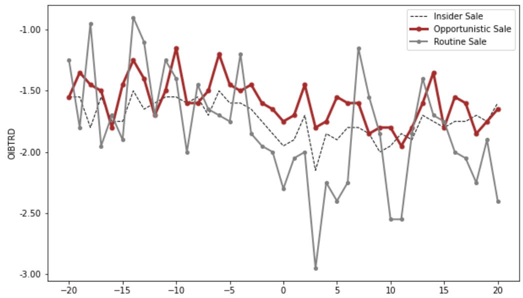
Panel A: *oibvol* around Insider Purchase Universe



Panel B: *oibtrd* around Insider Purchase Universe



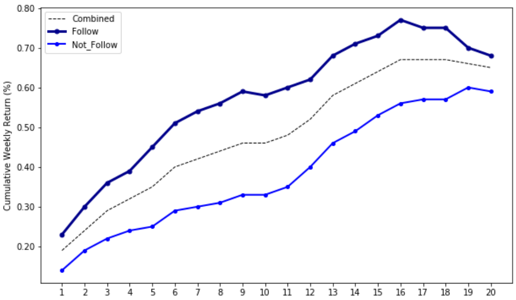
Panel C: *oibvol* around Insider Sale Universe



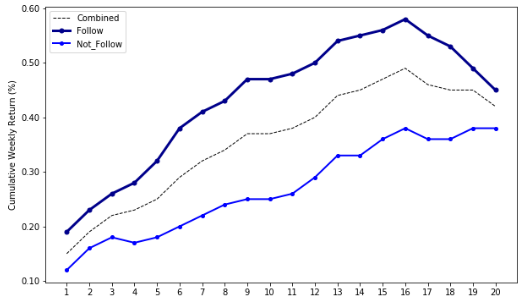
Panel D: *oibtrd* around Insider Sale Universe

FIGURE 3.1: Daily Retail Order Imbalance around Insider-Trading Event Windows

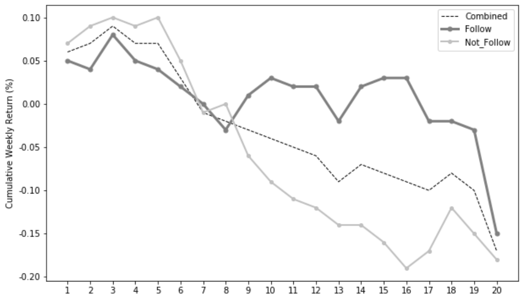
This figure presents the retail order imbalance, both *oibvol* and *oibtrd* around the insider trading window. Panel A reports *oibvol* in insider purchase universe. Panel B reports *oibtrd* in insider purchase universe. Correspondingly, Panel C and Panel D report order imbalance in insider sale universe. Each node is Newey-West adjusted and we use clustering method to make adjacent insider trades (within 5 days) as one trade.



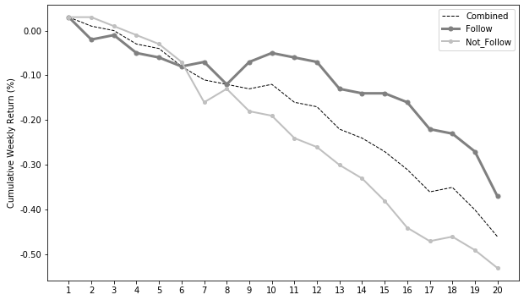
Panel A: Cumulative Mkt-adj Returns: Opportunistic Purchase



Panel B: Cumulative Size-BM-adj Returns: Opportunistic Purchase



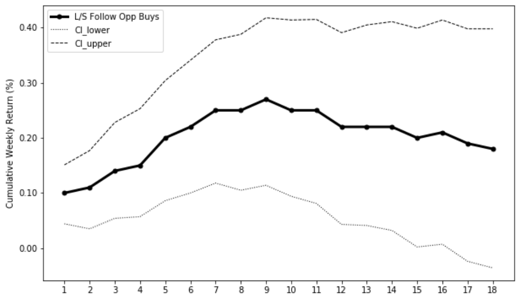
Panel C: Cumulative Mkt-adj Returns: Routine Purchase



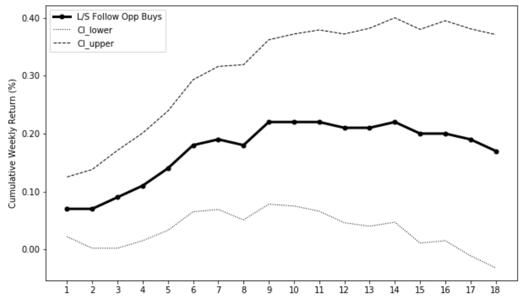
Panel D: Cumulative Size-BM-adj Returns: Routine Purchase

FIGURE 3.2: Cumulative Returns of Retail Investor Portfolios

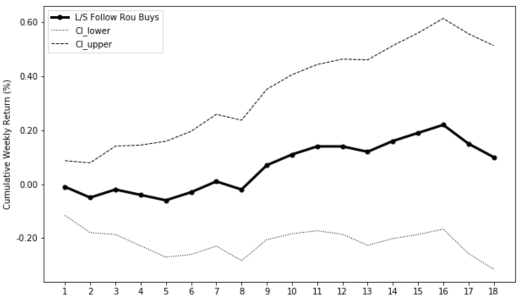
This figure presents the cumulative adjusted returns for stocks with opportunistic insider and routine insider purchases, up to 20 weeks, based on whether retail investors trade on the same side (Follow or Not-Follow) with opportunistic or routine insiders. Specifically, we divide the opportunistic/routine sample into two sub-samples based on whether retail investor purchase or sell the stock in the event week (from the insider purchase date to one day after the SEC filing date). We define Follow group to be the stocks with positive retail order imbalance *oibvol* during the event week, and Not-Follow group to be the stocks with negative retail order imbalance *oibvol* during the event week. Note that label named 'Combined' refers to the overall cumulative returns for the sub-sample, i.e. regardless of the sign of the retail order flow. We use CRSP daily value-weighted market returns to calculate the average weekly market return, and subtract it from weekly stock return to get the weekly market-adjusted return. We also construct 5*5 weekly size-BM benchmark portfolio return to get size-BM-adjusted return. Panel A reports the cumulative market-adjusted returns for stocks with opportunistic insider purchases, and Panel B reports the cumulative size-BM-adjusted returns for stocks with opportunistic insider purchases. Panel C and Panel D present the cumulative returns for stocks with routine insider purchases.



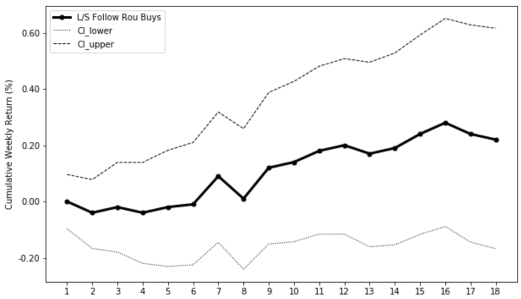
Panel A: Cumulative Mkt-adj Returns: Opportunistic Purchase



Panel B: Cumulative Size-BM-adj Returns: Opportunistic Purchase



Panel C: Cumulative Mkt-adj Returns: Routine Purchase



Panel D: Cumulative Size-BM-adj Returns: Routine Purchase

Figure 3.3: Long-Short Portfolio Returns in Insider Purchase Universe

This figure presents long-short portfolios of cumulative adjusted returns for stocks with opportunistic and routine insider purchases, up to 18 weeks, between retail investor investment choice (Long Follow portfolio and Short Not-Follow portfolio). Specifically, we divide the opportunistic/routine sample into two sub-samples based on whether retail investor purchase or sell the stock in the event week (from the insider purchase date to one day after the SEC filing date). We define Follow group to be the stocks with positive retail order imbalance *oibvol* during the event week, and Not-Follow group to be the stocks with negative retail order imbalance *oibvol* during the event week. 95% confidence interval is shown in this figure, including upper and lower limit. We use CRSP daily value-weighted market returns to calculate the average weekly market return, and subtract it from weekly stock return to get the weekly market-adjusted return. We also construct 5*5 weekly size-BM benchmark portfolio return to get size-BM-adjusted return. Panel A reports the cumulative market-adjusted returns for stocks with opportunistic insider purchases, and Panel B reports the cumulative size-BM-adjusted returns for stocks with opportunistic insider purchases. Panel C and Panel D present the cumulative returns for stocks with routine insider purchases.

TABLE 3.1: Summary Statistics

This reports summary statistics for retail order imbalance, insider trades, and firm characteristics over the sample period from Jan 2010 to Dec 2018. Panel A presents summary statistics for stocks that are traded by insiders or retail investors. All variables are calculated in weekly frequency. The retail order imbalance variable (*oibvol*) is computed as the difference between the number of shares bought and sold by retail investors, divided by the total number of shares traded by retail investors. Another order imbalance variable (*oibtrd*) is computed as the difference between the number of buy trades and sell trades by retail investors, divided by the total number of trades by retail investors. The number of shares purchased by insiders and the number of shares sold by insiders are presented in thousands. Panel B and Panel C present characteristics of stocks purchased and sold by insiders, respectively. The sample is further separated into sub-samples of stocks with insider purchase, insider opportunistic purchase, insider routine purchase, insider sale, insider opportunistic sale, and insider routine sale. In each sub-sample, we further divide the stocks based on whether retail investors trade on the same side (Follow) or the opposite side (Not-Follow) with insiders. Specifically, we follow Cohen, Malloy, and Pomorski (2012) and define a routine trader as the insider who has placed a trade in the same calendar month for at least three years in the past, and an opportunistic trader as the insider who has traded for at least three years in the past, but does not have an obvious discernible pattern. Panel B and C present characteristics of stocks traded by these various subgroups. Firm size (LSIZE) is the natural logarithm of market capitalization. Book-to-market ratio (LBM) is the natural logarithm of the most recent fiscal year-end book value divided by the market capitalization. Momentum (MOM) is the past cumulative returns from month-7 to month-2, in percent. Short-term reversal (RET1) is the prior month's return, in percent. Turnover (TURN) is the monthly trading volume divided by number of shares outstanding, averaged over the past 12 months, in percent. Idiosyncratic volatility (IVOL) is the standard deviation of the residuals from the Fama and French (1993) three-factor regressions of daily stock excess returns over the previous 6 months, in percent. Investor attention (ATT) is the daily trading volume divided by the average trading volume in the previous year (252 trading days), and then take the weekly average (Barber and Odean, 2008).

Panel A: Summary Statistics for Insider Trades and Retail Trades

Variable	Mean	Std. Dev.	P25	Median	P75
oibvol	-0.021	0.275	-0.156	-0.013	0.114
oibtrd	-0.016	0.232	-0.133	-0.004	0.103
Number of shares purchased by insiders	233	7,220	2	10	42
Number of shares sold by insiders	228	3,379	6	20	63
Number of insider purchase trades	4.6	13.8	1	2	4
Number of insider sale trades	5.1	12.8	1	2	5

Table 3.1: (Cont.) Summary Statistics

Panel B: Characters of Stocks (Mean Value) Purchased by Insiders

	Insider Purchase		Routine Purchase		Opportunistic Purchase	
	Follow	Not-Follow	Follow	Not-Follow	Follow	Not-Follow
LSIZE	12.75	12.76	13.09	13.22	13.06	13.16
LBM	-0.56	-0.54	-0.47	-0.39	-0.52	-0.50
MOM	3.21	3.30	6.12	6.90	5.87	5.42
RET1	-1.41	-0.82	0.64	1.02	-1.09	-0.53
TURN	15.60	15.54	11.18	10.90	16.36	16.3
IVOL	1.31	2.59	1.76	1.74	2.31	2.18
ATT	1.53	1.44	1.14	1.27	1.52	1.52

Panel C: Characters of Stocks (Mean Value) Sold by Insiders

	Insider Sale		Routine Sale		Opportunistic Sale	
	Follow	Not-Follow	Follow	Not-Follow	Follow	Not-Follow
LSIZE	14.47	14.45	15.16	15.10	15.02	15.06
LBM	-1.14	-1.18	-1.37	-1.39	-1.16	-1.21
MOM	14.52	15.29	8.45	9.79	11.43	12.25
RET1	3.32	3.32	1.65	1.36	2.62	2.64
TURN	20.85	21.88	22.92	22.26	20.41	20.94
IVOL	1.90	1.97	1.68	1.70	1.60	1.63
ATT	1.17	1.25	1.03	1.09	1.10	1.14

Table 3.2: Retail Order Imbalance around Insider Trading Window

This presents retail investor order imbalance around the event window over the sample period from Jan 2010 to Dec 2018. We define the event window as the days from insider trading date to one day after the SEC filing date of this trade. We report retail order imbalance 5 trading days around the event window. The retail order imbalance variable (*oibvol*) is computed as the difference between the number of shares bought and sold by retail investors, divided by the total number of shares traded by retail investors. Another order imbalance variable (*oibtrd*) is computed as the difference between the number of buy trades and sell trades by retail investors, divided by the total number of trades by retail investors. The sample is further separated into sub-samples of stocks with insider purchase, insider opportunistic purchase, insider routine purchase, insider sale, insider opportunistic sale, and insider routine sale. Specifically, we follow Cohen, Malloy, and Pomorski (2012) and define a routine trader as the insider who has placed a trade in the same calendar month for at least three years in the past, and an opportunistic trader as the insider who has traded for at least three years in the past, but does not have an obvious discernible pattern. Panel A presents the results for insider purchase universe and Panel B presents those for insider sale universe. To rule out the effect from nearby insider trades in our event window, we cluster the adjacent insider trades (within 5 days) for each stock into one group. We report the estimated mean with t-statistics (in parentheses). The standard errors are Newey-West adjusted. The mean values reported in this are in percent.

Panel A: Retail Order Imbalance around Insider Purchase Event Window

Retail Order Imbalance	-5	-4	-3	-2	-1	Event Window	1	2	3	4	5
All Insider Purchase Event Window											
oibvol	-0.7	-1.0	-1.3	-1.3	-1.7	0.5	-0.8	-1.7	-1.0	-1.8	-1.6
t-stat	(-1.98)	(-3.07)	(-3.75)	(-3.86)	(-5.90)	(2.59)	(-2.43)	(-4.92)	(-2.73)	(-5.28)	(-5.34)
oibtrd	-0.1	-0.7	-1.0	-1.1	-1.3	0.6	-0.8	-1.5	-0.7	-1.1	-1.0
t-stat	(-0.37)	(-2.49)	(-3.41)	(-3.63)	(-5.39)	(3.76)	(-2.86)	(-5.21)	(-2.40)	(-3.62)	(-4.03)
Insider Purchase Opportunistic Event Window											
oibvol	-0.8	-0.7	-0.7	-1.0	-1.3	0.9	-1.1	-1.6	-1.4	-2.8	-0.6
t-stat	(-1.10)	(-0.95)	(-0.95)	(-1.38)	(-2.06)	(2.28)	(-1.58)	(-2.29)	(-1.99)	(-4.04)	(-0.94)
oibtrd	-0.3	-0.6	-0.2	-0.4	-0.7	1.1	-0.8	-1.1	-0.6	-1.5	0.1
t-stat	(-0.42)	(-0.95)	(-0.35)	(-0.70)	(-1.28)	(3.25)	(-1.42)	(-1.85)	(-0.94)	(-2.53)	(0.13)
Insider Purchase Routine Event Window											
oibvol	-1.4	-4.6	-4.5	-2.3	-5.1	-1.5	-3.6	-1.1	-0.6	-4.3	-2.6
t-stat	(-0.86)	(-2.88)	(-2.72)	(-1.38)	(-3.57)	(-1.57)	(-2.24)	(-0.64)	(-0.29)	(-2.58)	(-1.82)
oibtrd	-0.3	-2.1	-2.9	-1.1	-3.5	-0.6	-2.0	-0.7	-1.0	-2.0	-0.8
t-stat	(-0.22)	(-1.63)	(-2.13)	(-0.81)	(-2.94)	(-0.70)	(-1.54)	(-0.47)	(-0.62)	(-1.39)	(-0.70)

Panel B: Retail Order Imbalance around Insider Sale Event Window

Retail Order Imbalance	-5	-4	-3	-2	-1	Event Window	1	2	3	4	5
All Insider Sale Event Window											
oibvol	-2.4	-2.5	-2.5	-2.5	-2.8	-3.2	-2.9	-2.5	-3.0	-2.5	-2.6
t-stat	(-19.86)	(-20.01)	(-19.08)	(-18.92)	(-24.21)	(-47.82)	(-22.41)	(-18.74)	(-22.49)	(-20.07)	(-24.45)
oibtrd	-1.6	-1.6	-1.7	-1.8	-1.9	-2.0	-1.9	-1.7	-2.2	-1.9	-1.9
t-stat	(-15.61)	(-15.46)	(-15.57)	(-15.74)	(-19.54)	(-34.79)	(-17.98)	(-15.51)	(-19.65)	(-17.68)	(-21.25)
Insider Sale Opportunistic Event Window											
oibvol	-2.2	-2.1	-2.0	-2.1	-2.6	-3.1	-2.6	-2.4	-2.7	-2.3	-2.1
t-stat	(-11.04)	(-10.17)	(-9.03)	(-9.34)	(-13.53)	(-28.25)	(-11.82)	(-10.53)	(-12.13)	(-10.97)	(-11.37)
oibtrd	-1.5	-1.5	-1.5	-1.6	-1.7	-1.8	-1.7	-1.5	-1.8	-1.8	-1.6
t-stat	(-8.53)	(-8.72)	(-8.30)	(-8.53)	(-10.54)	(-19.15)	(-9.71)	(-7.89)	(-9.92)	(-10.16)	(-10.26)
Insider Sale Routine Event Window											
oibvol	-2.2	-2.3	-2.5	-2.0	-2.3	-3.1	-2.3	-2.8	-2.9	-3.0	-2.8
t-stat	(-4.10)	(-4.10)	(-4.20)	(-3.18)	(-4.23)	(-10.74)	(-3.95)	(-4.48)	(-4.79)	(-5.64)	(-5.75)
oibtrd	-1.8	-1.2	-1.9	-2.0	-2.0	-2.3	-2.1	-2.0	-3.0	-2.3	-2.4
t-stat	(-4.13)	(-2.61)	(-3.73)	(-3.88)	(-4.62)	(-9.11)	(-4.21)	(-4.00)	(-6.09)	(-5.01)	(-6.02)

Table 3.3: Retail Abnormal EDGAR Search and Corresponding Trades around Insider Trading Window

This presents retail investors' abnormal EDGAR search as well as their order imbalance corresponding to high or low EDGAR search scenarios, all around the event window and over the sample period from Jan 2010 to Jun 2017. We define the event window as the days from insider trading date to one day after the SEC filing date of this trade. We report retail abnormal EDGAR search 5 trading days around the event window, and retail order imbalance 5 trading days after the event window. Panel A presents the retail investors' abnormal EDGAR search around insider trading event window. For each stock on each day, we count the number of unique IP addresses searching for Form 4 filings and then subtract its prior sixty-day average value to obtain *A_Search*. Panel B presents retail order imbalance at the event window and in the following 5 days, corresponding to high or low retail EDGAR search scenarios. During the event window, stocks in our sample are ranked into quintiles based on retail *A_Search*. We assign stocks within the top *A_Search* quintile into *High A_Search* group, and those within the bottom *A_Search* quintile into *Low A_Search* group. The retail order imbalance variable (*oibvol*) is computed as the difference between the number of shares bought and sold by retail investors, divided by the total number of shares traded by retail investors. Another order imbalance variable (*oibtrd*) is computed as the difference between the number of buy trades and sell trades by retail investors, divided by the total number of trades by retail investors. The sample is further separated into sub-samples of stocks with insider purchase, insider opportunistic purchase, insider routine purchase, insider sale, insider opportunistic sale, and insider routine sale. To rule out the effect from nearby insider trades in our event window, we cluster the adjacent insider trades (within 5 days) for each stock into one group. We report the estimated mean with t-statistics (in parentheses). The standard errors are Newey-West adjusted. The mean values reported in this are in percent.

Panel A: Retail Abnormal EDGAR Search around Insider Trading Event Window

A_Search	-5	-4	-3	-2	-1	Event	1	2	3	4	5
All Purchase	-2.04	-2.08	-2.12	-2.16	-2.11	2.36	-1.35	-1.51	-1.72	-1.81	-1.73
Opp Purchase	-2.02	-2.10	-2.16	-2.17	-2.16	2.59	-1.41	-1.55	-1.74	-1.86	-1.72
Rou Purchase	-4.58	-4.72	-4.68	-4.78	-4.72	-0.81	-4.22	-4.45	-4.50	-4.51	-4.28
All Sale	-2.09	-2.18	-2.31	-2.39	-2.37	0.80	-1.82	-1.99	-2.15	-2.16	-2.01
Opp Sale	-2.14	-2.28	-2.38	-2.43	-2.40	0.80	-1.78	-1.98	-2.15	-2.15	-2.04
Rou Sale	-1.84	-2.07	-2.12	-2.35	-2.38	1.03	-1.76	-1.92	-2.05	-2.12	-1.88

Table 3.3: (Cont.) Retail Abnormal EDGAR Search and Corresponding Trades around Insider Trading Window

Panel B: Retail Order Imbalance in Different EDGAR Search Scenarios

Retail OIB		Event Window	1	2	3	4	5
		All Insider Purchase Event Window					
oibvol	High A_Search	2.50	1.55	-1.30	-0.25	-0.65	0.10
		(6.77)	(2.44)	(-1.91)	(-0.34)	(-0.90)	(0.13)
	Low A_Search	-1.35	-2.60	-2.90	-1.05	-1.20	-2.35
		(-3.08)	(-3.44)	(-3.80)	(-1.32)	(-1.60)	(-3.54)
	Diff.	3.85	4.15	1.60	0.80	0.60	2.45
		(6.70)	(4.20)	(1.55)	(0.75)	(0.55)	(2.70)
oibtrd	High A_Search	2.35	1.10	-0.20	-0.20	0.20	0.15
		(7.23)	(1.90)	(-0.32)	(-0.35)	(0.35)	(0.31)
	Low A_Search	-0.95	-2.05	-1.80	-1.00	-0.85	-0.80
		(-2.60)	(-3.20)	(-2.88)	(-1.53)	(-1.30)	(-1.43)
	Diff.	3.35	3.15	1.60	0.80	1.05	0.95
		(6.70)	(3.65)	(1.85)	(0.90)	(1.20)	(1.25)
		Insider Purchase Opportunistic Event Window					
oibvol	High A_Search	2.30	2.25	-2.40	-0.50	-3.00	0.20
		(3.05)	(1.84)	(-1.74)	(-0.32)	(-2.03)	(0.18)
	Low A_Search	-0.40	-0.90	-3.60	-0.15	-0.50	-0.90
		(-0.47)	(-0.57)	(-2.29)	(-0.08)	(-0.30)	(-0.62)
	Diff.	2.70	3.15	1.20	-0.35	-2.50	1.10
		(2.30)	(1.60)	(0.55)	(-0.15)	(-1.15)	(0.60)
oibtrd	High A_Search	2.90	1.35	-0.75	0.65	-1.20	0.55
		(4.38)	(1.27)	(-0.60)	(0.49)	(-0.94)	(0.51)
	Low A_Search	0.50	-1.30	-2.05	0.50	-0.20	0.80
		(0.63)	(-0.96)	(-1.54)	(0.36)	(-0.14)	(0.67)
	Diff.	2.45	2.65	1.30	0.15	-1.00	-0.25
		(2.40)	(1.55)	(0.70)	(0.10)	(-0.55)	(-0.15)
		Insider Purchase Routine Event Window					
oibvol	High A_Search	0.60	0.25	3.65	-7.75	-2.85	2.00
		(0.27)	(0.07)	(0.93)	(-1.82)	(-0.69)	(0.56)
	Low A_Search	-4.80	-6.10	-6.35	3.75	-7.00	-4.00
		(-2.21)	(-1.63)	(-1.86)	(0.93)	(-1.96)	(-1.36)
	Diff.	5.40	6.35	10.00	-11.50	4.15	6.00
		(1.70)	(1.25)	(1.95)	(-1.95)	(0.75)	(1.30)
oibtrd	High A_Search	0.95	1.40	4.70	-1.10	1.65	4.00
		(0.48)	(0.47)	(1.38)	(-0.30)	(0.44)	(1.42)
	Low A_Search	-3.15	-1.35	-3.65	-2.05	-5.30	-2.10
		(-1.76)	(-0.47)	(-1.33)	(-0.66)	(-1.85)	(-0.82)
	Diff.	4.05	2.75	8.35	0.95	6.95	6.10
		(1.55)	(0.65)	(1.95)	(0.20)	(1.50)	(1.60)

Table 3.4: Following Insider Trading: Panel Regressions

This reports results of panel regressions of retail order imbalance of stock i in week t on the same-week insider trading dummies (Models (1) - (4)) and retail order imbalance of stock i in week $t+1$ on the prior-week insider trading dummies (Models (5) - (8)), all in the insider-trading universe. We construct insider trading dummy variables as an insider buy dummy (Buy), an opportunistic buy dummy (Opp_Buy), an opportunistic sell dummy (Opp_Sell), a routine buy dummy (Rou_Buy), and a routine sell dummy (Rou_Sell). The dummies equal one if the stock in week t has such an insider trade. In Model (1) and Model (2), the dependent variable is $oibvol_t$, as defined in 1. The dependent variable in Model (3) and Model (4) is $oibtrd_t$, as defined in 1. The dependent variables in Models (5) - (8) with subscript $t+1$ represent the corresponding one-week-ahead order imbalance. The main independent variables include Buy in Model (1), (3), (5) and (7), Opp_Buy, Rou_Buy, Opp_Sell, and Rou_Sell in Model (2), (4), (6) and (8). We include firm size (LSIZE), book-to-market ratio (LBM), turnover (TURN), idiosyncratic volatility (IVOL), momentum (MOM), short-term reversal (RET1), prior week return (RET1W), investor attention (ATT) as well as retail investor attention (ASVI) as control variables. The sample period is from January 2010 to December 2018. We include week fixed effects in all specifications. Standard errors are two-way clustered at the firm and the week level. *p<0.1; **p<0.05, ***p<0.01.

Variable	(1) $oibvol_t$	(2) $oibvol_t$	(3) $oibtrd_t$	(4) $oibtrd_t$	(5) $oibvol_{t+1}$	(6) $oibvol_{t+1}$	(7) $oibtrd_{t+1}$	(8) $oibtrd_{t+1}$
Opp_Buy		0.10***		0.12***		0.07***		0.07***
		(4.42)		(5.78)		(3.48)		(3.58)
Rou_Buy		-0.00		0.03		-0.01		0.05
		(-0.03)		(0.48)		(-0.18)		(0.80)
Opp_Sell		-0.00		-0.01		0.00		-0.01*
		(-0.57)		(-0.96)		(0.10)		(-1.66)
Rou_Sell		-0.01		-0.02		-0.03*		-0.03*
		(-0.53)		(-1.07)		(-1.92)		(-1.79)
LSIZE	0.02***	0.02***	0.02***	0.02***	0.01***	0.01***	0.02***	0.018***
	(5.92)	(4.69)	(6.56)	(5.72)	(4.16)	(3.35)	(5.44)	(4.86)
LBM	-0.03***	-0.02***	-0.03***	-0.02***	-0.02***	-0.02***	-0.03***	-0.02***
	(-4.82)	(-3.79)	(-4.77)	(-3.98)	(-4.39)	(-3.70)	(-4.89)	(-4.31)
TURN	0.08***	0.07***	0.02	0.01	0.10***	0.08***	0.01	-0.002
	(3.43)	(2.68)	(0.78)	(0.25)	(4.08)	(3.51)	(0.30)	(-0.07)
IVOL	3.41***	3.85***	2.13***	2.50***	1.98***	2.29***	1.41***	1.67***
	(6.71)	(7.53)	(4.39)	(5.14)	(4.35)	(5.00)	(3.07)	(3.64)
MOM	-0.02	-0.04***	0.01	-0.00	-0.01	-0.02	0.02*	0.01
	(-1.53)	(-2.70)	(0.86)	(-0.14)	(-0.59)	(-1.45)	(1.77)	(0.92)
RET1	-0.08**	-0.12***	-0.06**	-0.10***	-0.02	-0.04	0.02	-0.01
	(-2.51)	(-3.76)	(-2.11)	(-3.21)	(-0.50)	(-1.34)	(0.58)	(-0.25)
RET1W	-2.34***	-2.66***	-2.09***	-2.35***	-1.13***	-1.35***	-1.37***	-1.575***
	(-7.70)	(-8.73)	(-7.56)	(-8.48)	(-3.77)	(-4.49)	(-4.85)	(-5.60)
ATT	0.05***	0.05***	0.06***	0.06***	0.03***	0.03***	0.04***	0.038***
	(12.04)	(12.45)	(16.38)	(16.69)	(7.95)	(8.27)	(11.34)	(11.66)
ASVI	0.01*	0.01*	0.00	0.00	0.00	0.00	0.00	0.00
	(1.87)	(1.90)	(1.41)	(1.40)	(0.10)	(0.12)	(1.03)	(0.98)
Buy	0.13***		0.11***		0.09***		0.09***	
	(10.45)		(9.43)		(7.09)		(7.07)	
Constant	-0.56***	-0.48***	-0.55***	-0.48***	-0.42***	-0.37***	-0.47***	-0.42***
	(-10.71)	(-9.17)	(-9.73)	(-8.64)	(-8.19)	(-7.19)	(-8.33)	(-7.50)
Obs	94,665	94,665	94,665	94,665	94,012	94,012	94,012	94,012
R-squared	0.026	0.024	0.030	0.028	0.021	0.020	0.026	0.025

TABLE 3.5: Retail Order Flow around Insider Trades - Investor Attention

This reports results of panel regressions of retail order imbalance of stock i in week t on the same-week insider trading dummies in the insider-trading universe. We construct insider trading dummy variables as an opportunistic buy dummy (Opp_Buy), an opportunistic sell dummy (Opp_Sell), a routine buy dummy (Rou_Buy), and a routine sell dummy (Rou_Sell). The dummies equal one if the stock in week t has such an insider trade. In Models (1) and (3), the dependent variable is *oibvol*, as defined in 1. The dependent variable in Models (2) and (4) is *oibtrd*, as defined in 1. As for main independent variables, we use Opp_Buy as opportunistic buy dummy and Rou_Buy as routine buy dummy. We assign ASVI to be high when it is among the highest 33.3% of the sample, and low when it is among the lowest 33.3% of the sample. We run regressions for the sub-samples $\mathrm{ASVI_t}$ = High and $\mathrm{ASVI_t}$ = Low separately. We report the coefficient estimates for Opp_Buy and Rou_Buy as well as coefficient differences between Model (1) and Model (3), and between Model (2) and Model (4). We include firm size (LSIZE), book-to-market ratio (LBM), turnover (TURN), idiosyncratic volatility (IVOL), momentum (MOM), short-term reversal (RET1), prior week return (RET1W) as well as investor attention (ATT) as control variables. We do not report coefficients on control variables for brevity. The sample period is from January 2010 to December 2018. *p<0.1; **p<0.05, ***p<0.01.

		$\mathbf{ASVI_t}$ = High		**$\mathbf{ASVI_t}$ = Low**			
		(1)	(2)	(3)	(4)	(1) - (3)	(2) - (4)
Variable		oibvol	oibtrd	oibvol	oibtrd	oibvol	oibtrd
Opp_Buy	Coeff.	0.078***	0.103***	0.095***	0.082***	-0.017	0.021
	t-stat	(3.01)	(4.43)	(3.58)	(3.44)	(-0.47)	(0.64)
Rou_Buy	Coeff.	0.038	0.048	-0.014	-0.024	0.052	0.072
	t-stat	(0.74)	(1.03)	(-0.22)	(-0.43)	(0.65)	(1.00)
Obs.		32,364		31,159			

Table 3.6: Retail Order Flow around Insider Trades - Common Information

This reports results of panel regressions of retail order imbalance of stock i in week t on the same-week insider trading dummies in the insider-trading universe. We construct insider trading dummy variables as an opportunistic buy dummy (Opp_Buy), an opportunistic sell dummy (Opp_Sell), a routine buy dummy (Rou_Buy), and a routine sell dummy (Rou_Sell). The dummies equal one if the stock in week t has such an insider trade. In Model (1) and Model (3), the dependent variable is *oibvol*, as defined in 1. The dependent variable in Model (2) and Model (4) is *oibtrd*, as defined in 1. As for main independent variables, we use Opp_Buy as opportunistic buy dummy and Rou_Buy as routine buy dummy. Panel A and B identify earnings news as one common information source. Panel C and D regard analyst information as another common source. Specifically, Panel A reports the regression results based on whether there is any upcoming earnings announcement $\mathrm{EA}_{[\mathrm{t+1,\ t+4}]}$ in the next month (week t+1 to week t+4). $\mathrm{EA}_{[\mathrm{t+1,\ t+4}]}$ equals 1 if there is earnings announcement event, and 0 otherwise. We further divide the sub-sample of $\mathrm{EA}_{[\mathrm{t+1,\ t+4}]} = 1$ into cases when SUE >0 and SUE <0. SUE is defined as three-day abnormal return around the earnings announcement event. Panel B reports the regression results based on whether there is a positive or negative SUE. Panel C reports the regression results for the sub-samples of recommendation upgrade and downgrade in the next month separately. We define the sub-sample as "Rec Upgrade" when the change of analyst recommendations is positive. We define the sub-sample as "Rec Downgrade" when the change of analyst recommendations is negative. The change of analyst recommendations is calculated as the next month consensus recommendation minus its value for the same stock one month ago. In Panel D, we report the regression results for the sub-samples of analyst earnings forecast up revision and down revision in the next month separately. We define the sub-sample as "Forecast Up" when the change of analyst EPS forecast is positive. We define the sub-sample as "Forecast Down" when the change of analyst EPS forecast is negative. The change of analyst EPS forecast is calculated as the next month EPS Median Estimate averaged across all analysts minus its value for the same stock one month ago. We report the coefficient estimates for Opp_Buy and Rou_Buy as well as coefficient differences between Model (1) and Model (3), and between Model (2) and Model (4). We include firm size (LSIZE), book-to-market ratio (LBM), turnover (TURN), idiosyncratic volatility (IVOL), momentum (MOM), short-term reversal (RET1), prior week return (RET1W) as well as investor attention (ATT) as control variables. We do not report coefficients on control variables for brevity. The sample period is from January 2010 to December 2018. *p<0.1; **p<0.05, ***p<0.01.

Table 3.6: (Cont.) Retail Order Flow around Insider Trades - Common Information

Panel A: Conditions based on Upcoming Earnings Announcement Events

		$\mathbf{EA_{[t+1,\ t+4]} = 1}$		$\mathbf{EA_{[t+1,\ t+4]} = 0}$			
		(1)	(2)	(3)	(4)	(1) - (3)	(2) - (4)
Variable		oibvol	oibtrd	oibvol	oibtrd	oibvol	oibtrd
Opp_Buy	Coeff.	0.024	0.083**	0.113***	0.111***	-0.089*	-0.028
	t-stat	(0.53)	(2.02)	(6.98)	(7.68)	(-1.84)	(-0.65)
Rou_Buy	Coeff.	-0.164**	-0.038	-0.013	0.001	-0.151**	-0.040
	t-stat	(-2.51)	(-0.65)	(-0.35)	(0.04)	(-2.01)	(-0.59)
Obs.		12,538		81,860			

Panel B: Conditions based on Upcoming Good/Bad Earnings News

		SUE >0		**SUE <0**			
		(1)	(2)	(3)	(4)	(1) - (3)	(2) - (4)
Variable		oibvol	oibtrd	oibvol	oibtrd	oibvol	oibtrd
Opp_Buy	Coeff.	0.137**	0.187***	0.046	0.048	0.091	0.139
	t-stat	(2.00)	(3.03)	(0.68)	(0.74)	(0.94)	(1.56)
Rou_Buy	Coeff.	0.195**	0.207**	-0.304***	-0.093	0.499***	0.300**
	t-stat	(2.01)	(2.36)	(-3.04)	(-0.98)	(3.58)	(2.33)
Obs.		6,080		5,826			

Panel C: Upcoming Analyst Recommendation Update

		Rec Upgrade		**Rec Downgrade**			
		(1)	(2)	(3)	(4)	(1) - (3)	(2) - (4)
Variable		oibvol	oibtrd	oibvol	oibtrd	oibvol	oibtrd
Opp_Buy	Coeff.	0.063**	0.094***	0.126***	0.099***	-0.063	-0.005
	t-stat	(2.23)	(3.59)	(4.39)	(3.75)	(-1.57)	(-0.13)
Rou_Buy	Coeff.	-0.001	0.037	-0.100	-0.128*	0.099	0.164*
	t-stat	(-0.02)	(0.59)	(-1.37)	(-1.89)	(1.00)	(1.79)
Obs.		27,261		27,256			

Panel D: Upcoming Analyst Forecast Revision

		Forecast Up		**Forecast Down**			
		(1)	(2)	(3)	(4)	(1) - (3)	(2) - (4)
Variable		oibvol	oibtrd	oibvol	oibtrd	oibvol	oibtrd
Opp_Buy	Coeff.	0.087***	0.110***	0.139***	0.136***	-0.052	-0.026
	t-stat	(3.83)	(5.34)	(6.18)	(6.64)	(-1.63)	(-0.89)
Rou_Buy	Coeff.	-0.036	0.009	-0.033	0.037	-0.002	-0.028
	t-stat	(-0.74)	(0.20)	(-0.68)	(0.83)	(-0.03)	(-0.45)
Obs.		48,600		37,738			

Table 3.7: Cumulative Market-adjusted Returns Based on Retail Order Flow within Opportunistic Insider Purchase Window

This reports the cumulative market-adjusted stock returns up to 52 weeks after opportunistic insider purchase, based on whether retail investors trade on the same side with opportunistic insiders (Follow or Not-Follow). The bottom row reports the difference between these two groups in the future cumulative market-adjusted returns. Specifically, we divide the sample into two sub-samples based on whether retail investor purchase or sell the stock that opportunistic insider bought in the event week (from the opportunistic insider buy date to one day after the SEC filing date of this trade). The sub-sample "Follow" include stocks with positive retail order imbalance *oibvol* during the event week, and "Not-Follow" include those with negative retail order imbalance *oibvol* during the event week. We present next-week returns as well as cumulative returns for next 4 weeks, 12 weeks, 18 weeks, 24 weeks and 52 weeks, respectively. We use CRSP daily value-weighted market returns to calculate the average weekly market return, and subtract it from weekly stock return to get the weekly market-adjusted return. Standard errors are clustered at the week level. The returns are shown in percent. *p<0.1; **p<0.05, ***p<0.01.

Classification		Time Periods					
		Week 1	Week 1-4	Week 1–12	Week 1-18	Week 1–24	Week 1-52
Follow	Mean	0.23	0.39	0.62	0.75	0.69	0.89
	t-stat	(11.35)	(8.73)	(8.06)	(9.35)	(6.89)	(6.37)
Not-Follow	Mean	0.14	0.24	0.40	0.57	0.52	0.79
	t-stat	(7.55)	(6.25)	(5.56)	(7.55)	(5.13)	(5.83)
Follow – Not-Follow	Mean	0.10***	0.15***	0.22**	0.18*	0.18	0.10
	t-stat	(3.55)	(3.16)	(2.36)	(1.65)	(1.48)	(0.50)

Table 3.8: Calendar-Week Portfolio Analysis

This reports the weekly alphas for the retail Follow portfolio and Not-Follow portfolio, as well as those for the long-short portfolio (Follow - Not-Follow). For the sub-sample of stocks with opportunistic insider purchases, we further divide it into two sub-samples based on whether retail investor purchase or sell those stocks during the event week (from the opportunistic insider buy date to one day after the SEC filing date of this trade). The Follow portfolio includes stocks with positive retail order imbalance *oibvol* in the event week, and the Not-Follow portfolio includes stocks with negative retail order imbalance *oibvol* in the event week. We rebalance the portfolio every week (Panel A) or every four weeks (Panel B). For each portfolio, we calculate the equal-weighted weekly returns, the CAPM alphas, the Fama and French (1993) three-factor alphas, and the Carhart (1997) four-factor alphas. The sample period is from January 2010 to December 2018. The returns and alphas are shown in percent. *p<0.1; **p<0.05, ***p<0.01.

Panel A: Equal-Weighted Portfolio Returns for Next Calendar Week

	Follow	Not-Follow	Follow – Not-Follow
Raw Return	0.268***	0.195***	0.079**
	(7.13)	(5.76)	(2.16)
CAPM Alpha	0.201***	0.133***	0.073**
	(6.75)	(4.90)	(2.01)
Carhart Alpha	0.211***	0.140***	0.079**
	(7.48)	(5.36)	(2.15)

Panel B: Equal-Weighted Portfolio Weekly Returns for 4-Week Holding Period

	Follow	Not-Follow	Follow – Not-Follow
Raw Return	0.175***	0.114***	0.052**
	(5.07)	(3.78)	(2.45)
CAPM Alpha	0.097***	0.054**	0.036*
	(3.77)	(2.41)	(1.70)
Carhart Alpha	0.106***	0.065***	0.036*
	(4.35)	(3.18)	(1.68)

Table 3.9: Return Predictability of Retail Purchases Following Insider Purchases: Panel Regressions

This reports the panel regression results of one-week ahead return of stock i from week t (Ret_{t+1}) on insider trading and retail trading indicators. In the first three models, we include three dummy variables: opportunistic insider purchase (Opp_Buy), routine insider purchase (Rou_Buy), as defined in 4, and retail investor purchase (Retail_Buy). Retail_Buy equals 1 if retail investors buy the stock, and 0 if they sell the stock. We also include interaction terms Opp_Buy * Retail_Buy and Rou_Buy * Retail_Buy. Model (4) only includes stocks with opportunistic insider purchases. We include an additional dummy variable Follow_Oppbuy, which equals 1 if retail investors buy the stock during the week, and 0 otherwise. In all models we include firm size (LSIZE), book-to-market ratio (LBM), turnover (TURN), idiosyncratic volatility (IVOL), momentum (MOM), short-term reversal (RET1), prior week return (RET1W), investor attention (ATT) as well as retail investor attention (ASVI) as control variables. The sample period is from January 2010 to December 2018. We include week fixed effects in all models. Standard errors are two-way clustered at the firm and the week level in all models. *p<0.1; **p<0.05, ***p<0.01.

Variable	(1) RET_{t+1}	(2) RET_{t+1}	(3) RET_{t+1}	(4) RET_{t+1}
LSIZE	0.0001***	0.0001***	0.0001***	0.0002*
	(5.01)	(4.93)	(5.00)	(1.69)
LBM	-0.0004***	-0.0004***	-0.0004***	-0.0002
	(-8.97)	(-8.93)	(-8.95)	(-0.80)
TURN	-0.0017***	-0.0017***	-0.0017***	-0.0025*
	(-6.49)	(-6.49)	(-6.48)	(-1.82)
IVOL	0.0737***	-0.0074	-0.0075	-0.0075
	(2.94)	(-1.55)	(-1.58)	(-1.57)
MOM	-0.0003	-0.0003	-0.0003	-0.0019*
	(-1.59)	(-1.56)	(-1.55)	(-1.89)
RET1	-0.0008*	-0.0008*	-0.0008*	0.0028
	(-1.77)	(-1.78)	(-1.76)	(1.23)
RET1W	-0.0096**	-0.0095**	-0.0093**	-0.0119
	(-2.46)	(-2.42)	(-2.38)	(-0.60)
ATT	0.0001***	0.0001***	0.0001***	0.0003*
	(3.83)	(3.76)	(3.65)	(1.73)
ASVI	0.0001***	0.0001***	0.0001***	-0.0002
	(6.66)	(6.61)	(6.62)	(-1.26)
Opp_Buy	0.0013***		0.0009***	
	(7.25)		(3.83)	
Rou_Buy	0.0003		0.0005	
	(0.80)		(0.89)	
Retail_Buy		0.0002***	0.0002***	
		(6.47)	(6.29)	
Opp_Buy * Retail_Buy			0.0007**	
			(2.04)	
Rou_Buy * Retail_Buy			-0.0004	
			(-0.75)	
Follow_Oppbuy				0.0007**
				(1.98)
Constant	-0.0014***	-0.0015***	-0.0015***	-0.0027
	(-3.52)	(-3.64)	(-3.72)	(-1.38)
Observations	597,422	596,168	596,168	3,840
R-squared	0.189	0.189	0.189	0.292

Table 3.10: Decomposition of Retail Order Imbalance

This reports the second-stage regression results of a two-stage retail order imbalance decomposition analysis. We follow Boehmer et al. (2021) and decompose retail order imbalance into persistence, contrarian, and other (information) components. The decomposition is done through two-stage regressions. For the first stage, we estimate the following regression model:

$$\mathrm{Retail\_Buy}^{i}{}_{t} = a_t + b_t{*}\mathrm{Retail\_Buy}^{i}{}_{t\text{-}1} + c_t{*}\mathrm{Ret}^{i}{}_{t\text{-}1} + \epsilon^{i}{}_{t}$$

Retail_Buy_Persistence is then defined as $\widehat{b_t} * Retail\_Buy^{i}{}_{t\text{-}1}$ and Retail_Buy_Contrarian is defined as $\widehat{c_t} * Ret^{i}{}_{t\text{-}1}$, where $\widehat{b_t}$ and $\widehat{c_t}$ are the estimated coefficients of the first-stage regression. The residual part from the first-stage regression is denoted as Retail_Buy_Other. In the second stage, we run panel regressions similar to 9, except that we replace Retail_Buy with the three components that are computed in the first stage. We include the same set of control variables as in 9. The coefficient estimates for those variables are not reported for brevity. We include week fixed effects for all models. Standard errors are two-way clustered at the firm and the week level in all models. *p<0.1; **p<0.05, ***p<0.01.

Variable	(1) RET_{t+1}	(2) RET_{t+1}	(3) RET_{t+1}	(4) RET_{t+1}
Retail_Buy_Persistence	0.0010***	0.0011***	0.0011***	0.0088
	(5.31)	(5.40)	(5.34)	(0.51)
Retail_Buy_Contrarian	-0.0051	-0.0050	-0.0051	0.0340
	(-1.16)	(-1.16)	(-1.17)	(1.12)
Retail_Buy_Other	0.0002***	0.0002***	0.0002***	0.0078
	(5.72)	(5.91)	(5.73)	(0.44)
Opp_Buy	0.0012***		0.0012***	
	(5.01)		(4.97)	
Rou_Buy		0.0006	0.0006	
		(1.39)	(1.24)	
Opp_Buy * Retail_Buy_Persistence	-0.0000		0.0000	
	(-0.02)		(0.01)	
Opp_Buy * Retail_Buy_Contrarian	0.0069		0.0065	
	(0.56)		(0.53)	
Opp_Buy * Retail_Buy_Other	0.0007**		0.0007**	
	(2.06)		(2.07)	
Rou_Buy * Retail_Buy_Persistence		-0.0046	-0.0048	
		(-1.28)	(-1.33)	
Rou_Buy * Retail_Buy_Contrarian		0.0385	0.0372	
		(1.58)	(1.54)	
Rou_Buy * Retail_Buy_Other		-0.0003	-0.0004	
		(-0.54)	(-0.59)	
Follow_Oppbuy				-0.0171
				(-0.90)
Follow_Oppbuy * Retail_Buy_Persistence				0.0013
				(1.62)
Follow_Oppbuy * Retail_Buy_Contrarian				0.0069
				(0.30)
Follow_Oppbuy * Retail_Buy_Other				0.0170**
				(2.55)
Controls	Yes	Yes	Yes	Yes
Observations	597,422	596,168	596,168	3,840
R-squared	0.189	0.189	0.189	0.292

TABLE 3.11: Cumulative Market-adjusted Returns on Stocks Traded by Retail Investors Following Insider Trading: The Effect of Information Asymmetry

This reports the effect of retail investors following opportunistic insider purchases on market-adjusted returns for stocks with different levels of information asymmetry. We use firm size (Panel A), idiosyncratic volatility (Panel B) as well as Amihud illiquidity (Panel C) as proxies for information asymmetry. In each panel we further separate the sample stocks into Follow portfolio and Not-Follow portfolio, based on whether retail investors follow insiders as defined in 1. In each panel, we categorize the sample of stocks into three levels of information asymmetry. For each level, we compute the difference in the cumulative market-adjusted returns between Follow portfolio and Not-Follow portfolio, up to 24 weeks. We use CRSP daily value-weighted market returns to calculate the average weekly market return, and subtract it from weekly stock return to get the weekly market-adjusted return. Standard errors are clustered at the week level. The returns are shown in percent. *p<0.1; **p<0.05, ***p<0.01.

Panel A: Sort by Firm Size

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Small	Follow – Not-Follow	Mean	0.26**	0.56***	0.57**
		t-stat	(2.43)	(2.73)	(1.98)
Mid-Cap	Follow – Not-Follow	Mean	0.16**	0.17	0.23
		t-stat	(2.08)	(1.28)	(1.24)
Large	Follow – Not-Follow	Mean	0.03	-0.04	-0.20
		t-stat	(0.48)	(-0.32)	(-1.02)
Small – Large	Diff. in Diff.	Mean	0.23*	0.60**	0.77**
		t-stat	(1.73)	(2.53)	(2.16)

Panel B: Sort by Idiosyncratic Volatility

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Low	Follow – Not-Follow	Mean	0.08*	0.22***	0.22*
		t-stat	(1.67)	(2.65)	(1.88)
Mid	Follow – Not-Follow	Mean	0.08	0.07	0.08
		t-stat	(1.55)	(0.72)	(0.55)
High	Follow – Not-Follow	Mean	0.38**	0.57*	0.49
		t-stat	(2.19)	(1.72)	(1.15)
High – Low	Diff. in Diff.	Mean	0.61**	0.30	0.25
		t-stat	(1.97)	(0.73)	(0.35)

Panel C: Sort by Amihud Illiquidity

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Low	Follow – Not-Follow	Mean	0.02	0.03	-0.22
		t-stat	(0.35)	(0.22)	(-1.15)
Mid	Follow – Not-Follow	Mean	0.20**	0.08	0.28
		t-stat	(2.34)	(0.52)	(1.41)
High	Follow – Not-Follow	Mean	0.24**	0.52***	0.48*
		t-stat	(2.30)	(2.65)	(1.74)
High – Low	Diff. in Diff.	Mean	0.22*	0.49**	0.69**
		t-stat	(1.66)	(2.17)	(2.10)

Table 3.12: Improvement of Informational Efficiency: Variance Ratio and Price Delay Measure

This reports the regression results of proxies for informational efficiency on the insider and retail trading dummies in insider-trading event week. In Models (1)-(4), the dependent variables are $|1 - VR(n,m)|$, where VR(n,m) represents variance ratios of the m-day return variance per unit time divided by the n-day return variance per unit time estimated, using daily returns next month. In Models (5)-(6), the dependent variables are Prc_dalay, which is the monthly price delay measure of Boehmer and Wu (2013). We include dummy variables for opportunistic insider purchase (Opp_Buy), routine insider purchase (Rou_Buy), retail investor purchase (Retail_Buy), all as defined in the previous s, as well as interaction terms Opp_Buy * Retail_Buy and Rou_Buy * Retail_Buy. We also include the dummy variable Follow_Oppbuy as defined in 9. We include firm size (LSIZE), book-to-market ratio (LBM), turnover (TURN), idiosyncratic volatility (IVOL), momentum (MOM), short-term reversal (RET1), prior week return (RET1W), investor attention (ATT) as well as retail investor attention (ASVI) as control variables. The sample period is from January 2010 to December 2018. We include week fixed effects in all models. Standard errors are two-way clustered at the firm and the week level in all models. *p<0.1; **p<0.05, ***p<0.01.

Variable	(1) $\lvert 1-VR(1,10)\rvert$	(2) $\lvert 1-VR(1,10)\rvert$	(3) $\lvert 1-VR(1,20)\rvert$	(4) $\lvert 1-VR(1,20)\rvert$	(5) Prc_delay	(6) Prc_delay
LSIZE	-0.005***	-0.002	0.006***	0.027***	-0.564*	0.234
	(-20.90)	(-0.64)	(8.94)	(3.44)	(-1.82)	(0.68)
LBM	-0.000	0.002	-0.006***	0.002	-0.211	0.140
	(-0.89)	(0.42)	(-4.80)	(0.13)	(-1.24)	(1.07)
TURN	-0.015***	-0.053**	0.051***	-0.074	-0.337	-1.002
	(-6.40)	(-2.02)	(8.08)	(-0.99)	(-0.29)	(-1.03)
IVOL	0.191***	0.233	0.749***	4.878***	-28.770	23.871
	(5.14)	(0.55)	(7.64)	(4.23)	(-0.92)	(0.86)
MOM	-0.011***	0.003	-0.034***	0.001	-0.440	-0.825
	(-8.09)	(0.18)	(-9.74)	(0.01)	(-0.82)	(-1.09)
RET1	-0.006*	0.053	-0.031***	0.187*	0.022	2.761*
	(-1.74)	(1.38)	(-3.52)	(1.86)	(0.01)	(1.81)
RET1W	-0.085**	-0.518	-0.007	-0.630	-8.812	-6.001
	(-2.48)	(-1.39)	(-0.08)	(-0.65)	(-0.25)	(-0.81)
ATT	-0.008***	-0.010***	-0.012***	-0.006	0.847	-0.075
	(-23.92)	(-2.58)	(-12.75)	(-0.59)	(1.18)	(-0.50)
ASVI	-0.000	-0.003	-0.001	0.003	0.148	-0.272
	(-0.86)	(-1.01)	(-1.28)	(0.41)	(0.69)	(-1.13)
Opp_Buy	0.002		0.016		-0.719	
	(0.39)		(1.07)		(-0.80)	
Rou_Buy	0.011		-0.010		1.125	
	(0.87)		(-0.30)		(1.15)	
Retail_Buy	-0.000		-0.001		-0.545	
	(-0.12)		(-0.49)		(-0.65)	
Opp_Buy * Retail_Buy	-0.015*		-0.046**		-0.437	
	(-1.82)		(-2.19)		(-0.35)	
Rou_Buy * Retail_Buy	-0.011		-0.035		-0.701	
	(-0.58)		(-0.74)		(-0.62)	
Follow_Oppbuy		-0.017**		-0.056**		-1.360*
		(-2.05)		(-2.55)		(-1.69)
Constant	0.394***	0.360***	0.539***	0.166	8.513	-2.011
	(101.80)	(8.60)	(52.58)	(1.50)	(1.56)	(-0.49)
Observations	596,896	3,853	596,866	3,853	498,951	3,163
R-squared	0.031	0.145	0.074	0.148	0.001	0.472

TABLE 3.13: Following Insider Trading: Longer Horizons

This reports results of panel regressions of both retail order imbalance and changes of institutional ownership of stock i from quarter t to quarter $t+4$ on the insider trading dummies in quarter t in the insider-trading universe. We construct insider trading dummy variables same as in 4. In Panel A, the dependent variables are $oibvol_t$, $oibvol_{t+1}$, $oibvol_{t+2}$, $oibvol_{t+3}$, $oibvol_{t+4}$, which are the quarterly retail order imbalance in the subsequent four quarters, for the stocks that insiders purchased. In Panel B, the dependent variables are ΔIO_t, ΔIO_{t+1}, ΔIO_{t+2}, ΔIO_{t+3}, ΔIO_{t+4}, which are the quarterly changes of institutional ownership in the subsequent four quarters, for the stocks that insiders purchased. The main independent variables include Buy in odd-number models, Opp_Buy, Rou_Buy, Opp_Sell, and Rou_Sell in even-number models. We include firm size (LSIZE), book-to-market ratio (LBM), turnover (TURN), idiosyncratic volatility (IVOL), momentum (MOM), short-term reversal (RET1), prior week return (RET1W), investor attention (ATT) as well as retail investor attention (ASVI) as control variables. The sample period is from January 2010 to December 2018. We include week fixed effects in all specifications. Standard errors are two-way clustered at the firm and the week level. *p<0.1; **p<0.05, ***p<0.01.

Panel A: Retail Order Imbalance

Variable	(1) $oibvol_t$	(2) $oibvol_t$	(3) $oibvol_{t+1}$	(4) $oibvol_{t+1}$	(5) $oibvol_{t+2}$	(6) $oibvol_{t+2}$	(7) $oibvol_{t+3}$	(8) $oibvol_{t+3}$	(9) $oibvol_{t+4}$	(10) $oibvol_{t+4}$
Opp_Buy		0.051***		0.004		0.019		0.012		0.006
		(3.38)		(0.23)		(1.16)		(0.69)		(0.36)
Rou_Buy		0.002		-0.007		-0.009		0.025		-0.001
		(0.06)		(-0.17)		(-0.24)		(0.62)		(-0.04)
Opp_Sell		-0.003		-0.005		0.000		0.004		0.006
		(-0.56)		(-0.82)		(0.08)		(0.61)		(1.00)
Rou_Sell		-0.005		-0.018*		-0.020*		-0.018		-0.019
		(-0.53)		(-1.68)		(-1.72)		(-1.56)		(-1.57)
LSIZE	0.017***	0.015***	0.016***	0.016***	0.015***	0.014***	0.013***	0.012***	0.012***	0.012***
	(5.91)	(5.15)	(5.84)	(5.78)	(5.13)	(5.03)	(4.38)	(4.26)	(4.19)	(3.98)
LBM	-0.019***	-0.016***	-0.016***	-0.016***	-0.018***	-0.017***	-0.025***	-0.025***	-0.023***	-0.022***
	(-4.10)	(-3.35)	(-3.47)	(-3.33)	(-3.62)	(-3.48)	(-5.12)	(-4.99)	(-4.50)	(-4.26)
TURN	0.107***	0.095***	0.077***	0.074***	0.065***	0.062**	0.053**	0.051**	0.066***	0.063***
	(5.05)	(4.38)	(3.58)	(3.43)	(2.71)	(2.57)	(2.07)	(2.00)	(2.74)	(2.60)
IVOL	2.611***	2.940***	2.169***	2.239***	1.654***	1.750***	0.924**	1.022**	1.075**	1.202***
	(6.10)	(6.77)	(5.04)	(5.18)	(3.66)	(3.89)	(2.04)	(2.28)	(2.47)	(2.78)
MOM	-0.009	-0.020*	0.013	0.009	0.020*	0.017	0.033***	0.030***	0.031***	0.028**
	(-0.86)	(-1.95)	(1.19)	(0.89)	(1.82)	(1.54)	(2.95)	(2.72)	(2.83)	(2.50)
RET1	-0.048**	-0.077***	0.012	0.004	0.019	0.011	0.024	0.016	0.057**	0.048**
	(-2.17)	(-3.48)	(0.54)	(0.20)	(0.86)	(0.50)	(1.05)	(0.71)	(2.55)	(2.17)
RET1W	0.229	0.137	-0.106	-0.134	-0.004	-0.033	0.199	0.168	0.167	0.127
	(1.37)	(0.82)	(-0.62)	(-0.79)	(-0.02)	(-0.20)	(1.19)	(1.01)	(0.96)	(0.73)
ATT	0.025***	0.026***	0.011***	0.011***	0.010***	0.010***	0.011***	0.011***	0.003	0.003
	(9.76)	(10.10)	(4.59)	(4.67)	(4.33)	(4.44)	(4.41)	(4.53)	(1.09)	(1.21)
ASVI	0.001	0.001	0.002	0.002	0.001	0.001	0.001	0.001	0.004**	0.004**
	(0.44)	(0.50)	(0.98)	(0.95)	(0.72)	(0.71)	(0.59)	(0.59)	(2.11)	(2.12)
Buy	0.074***		0.017*		0.020**		0.019*		0.019*	
	(7.34)		(1.76)		(1.97)		(1.71)		(1.83)	
Constant	-0.458***	-0.412***	-0.419***	-0.409***	-0.382***	-0.374***	-0.347***	-0.338***	-0.351***	-0.340***
	(-10.12)	(-9.18)	(-9.68)	(-9.55)	(-8.86)	(-8.70)	(-7.83)	(-7.72)	(-7.76)	(-7.53)
Observations	57,490	57,490	57,206	57,206	56,858	56,858	56,060	56,060	53,368	53,368
R-squared	0.055	0.052	0.042	0.042	0.040	0.040	0.044	0.044	0.044	0.044

Panel B: Changes in Institutional Ownership

Variable	(1) ΔIO_t	(2) ΔIO_t	(3) ΔIO_{t+1}	(4) ΔIO_{t+1}	(5) ΔIO_{t+2}	(6) ΔIO_{t+2}	(7) ΔIO_{t+3}	(8) ΔIO_{t+3}	(9) ΔIO_{t+4}	(10) ΔIO_{t+4}
Opp_Buy		-0.005***		-0.005***		-0.004**		-0.001		0.001
		(-2.76)		(-2.95)		(-2.07)		(-0.54)		(0.61)
Rou_Buy		-0.002		-0.004		0.001		0.002		-0.001
		(-0.56)		(-1.44)		(0.45)		(0.63)		(-0.17)
Opp_Sell		-0.004***		-0.002***		-0.003***		-0.002**		-0.003***
		(-5.16)		(-2.70)		(-3.92)		(-2.50)		(-2.86)
Rou_Sell		-0.004***		-0.003**		-0.002		-0.001		-0.001
		(-3.64)		(-2.24)		(-1.62)		(-1.10)		(-1.01)
LSIZE	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.001***	-0.001***
	(-5.12)	(-4.23)	(-5.89)	(-5.33)	(-7.83)	(-7.24)	(-5.94)	(-5.50)	(-4.99)	(-4.49)
LBM	-0.003***	-0.003***	-0.003***	-0.003***	-0.004***	-0.004***	-0.003***	-0.003***	-0.002***	-0.002***
	(-3.37)	(-3.96)	(-4.84)	(-5.28)	(-6.19)	(-6.64)	(-4.19)	(-4.50)	(-3.51)	(-3.78)
TURN	-0.002	-0.001	-0.009*	-0.008*	-0.010***	-0.010***	-0.009**	-0.009**	-0.011***	-0.011**
	(-0.29)	(-0.10)	(-1.85)	(-1.74)	(-2.78)	(-2.66)	(-2.31)	(-2.23)	(-2.61)	(-2.57)
IVOL	0.401***	0.360***	0.283***	0.260***	0.234***	0.213***	0.243***	0.226***	0.181***	0.167***
	(5.29)	(4.79)	(4.38)	(4.08)	(4.10)	(3.78)	(3.69)	(3.49)	(2.81)	(2.63)
MOM	0.018***	0.018***	0.017***	0.017***	0.009***	0.009***	0.005**	0.006***	0.004**	0.004**
	(7.49)	(7.81)	(8.29)	(8.55)	(4.80)	(4.91)	(2.53)	(2.61)	(2.47)	(2.54)
RET1	0.038***	0.040***	0.015***	0.016***	0.015***	0.015***	0.014***	0.015***	0.008**	0.008**
	(7.84)	(8.14)	(3.92)	(4.30)	(4.44)	(4.67)	(3.21)	(3.39)	(2.16)	(2.26)
RET1W	0.057	0.065	0.028	0.032	0.033	0.036	0.013	0.016	0.019	0.022
	(1.35)	(1.54)	(0.96)	(1.09)	(1.31)	(1.45)	(0.41)	(0.49)	(0.72)	(0.81)
ATT	0.004***	0.004***	0.002***	0.002***	0.001**	0.001**	0.001**	0.001**	-0.000	-0.000
	(9.39)	(9.22)	(4.79)	(4.65)	(2.45)	(2.35)	(2.37)	(2.28)	(-0.14)	(-0.21)
ASVI	-0.001*	-0.001**	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-1.95)	(-2.15)	(-1.20)	(-1.30)	(-1.42)	(-1.56)	(-1.37)	(-1.47)	(-1.13)	(-1.24)
Buy	-0.005***		-0.003***		-0.002*		-0.001		-0.000	
	(-3.28)		(-3.01)		(-1.80)		(-0.94)		(-0.06)	
Constant	0.023***	0.019***	0.027***	0.024***	0.031***	0.030***	0.025***	0.023***	0.020***	0.019***
	(4.02)	(3.41)	(5.25)	(4.86)	(6.92)	(6.62)	(4.85)	(4.65)	(4.38)	(4.17)
Observations	50,610	50,610	50,421	50,421	50,263	50,263	50,032	50,032	49,382	49,382
R-squared	0.067	0.067	0.072	0.072	0.070	0.070	0.052	0.052	0.045	0.045

TABLE A1: Lag in Days between Insider Trading Date and SEC Filing Date

This presents summary statistics for the number of days between insider trading and their reporting the trades to the SEC.

Lag (in days)	0	1	2	3	4	5	>5
Number of Observations	127,082	380,011	368,907	16,900	5,316	3,019	33,577
% in sample	13.5%	40.7%	39.5%	1.8%	0.6%	0.3%	3.6%
		93.7%			6.3%		

Table A2: Granger-type Causality Test

This runs the Granger-type causality test by regressing retail order imbalance of stock i in week (oibvol) t on the 1-week lagged insider trading dummies (l_buy, l_opp_buy, l_rou_buy), as well as regressing insider trading dummies of stock i in week t (Buy, Opp_Buy, Rou_Buy) on the 1-week lagged retail order imbalance (l_oibvol), both in the insider-trading event window. We construct 1-week lagged insider trading dummy variables as a lagged insider buy dummy (l_buy), a lagged opportunistic buy dummy (l_opp_buy), a lagged opportunistic sell dummy (l_opp_sell), a lagged routine buy dummy (l_rou_buy), and a lagged routine sell dummy (l_rou_sell). The dummies equal one if the stock in week *t-1* has such an insider trade. We control for 1-week lagged dummies (l_prior_ins, l_prior_opp, l_prior_rou) indicating insider, opportunistic insider, and routine insider trades in week *t-2*. We include lagged firm size (l_lsize), lagged book-to-market ratio (l_lbm), lagged turnover (l_lturn), lagged idiosyncratic volatility (l_ivol), lagged momentum (l_mom), lagged short-term reversal (l_ret1), lagged prior week return (l_ret1w), lagged investor attention (l_att) as well as lagged retail investor attention (l_asvi) as control variables. The sample period is from January 2010 to December 2018. We use Fama-MacBeth regressions with Newey-West 5 lags adjusted. *p<0.1; **p<0.05, ***p<0.01.

Variable	(1) oibvol	(2) oibvol	(3) oibvol	(4) Buy	(5) Opp_Buy	(6) Rou_Buy
l_oibvol	0.213***	-0.043	0.213***	-0.012	-0.003	-0.000
	(12.20)	(-0.17)	(13.15)	(-0.85)	(-1.15)	(-0.45)
l_buy	0.073***			0.664***		
	(3.60)			(53.64)		
l_opp_buy		0.148***			0.623***	
		(3.01)			(28.24)	
l_rou_buy			-0.014			0.467***
			(-0.42)			(15.63)
l_lsize	0.017***	-0.018	0.014**	-0.025***	-0.008***	-0.001***
	(2.82)	(-0.57)	(2.30)	(-4.55)	(-7.78)	(-4.92)
l_lbm	-0.007	-0.022**	-0.020**	0.046***	0.008***	0.001
	(-0.63)	(-2.42)	(-2.42)	(5.69)	(5.33)	(0.52)
l_turn	0.073*	0.522	0.075**	-0.084***	-0.018**	-0.009**
	(1.67)	(1.35)	(2.19)	(-2.72)	(-2.56)	(-2.48)
l_ivol	3.972***	-25.471	3.527***	2.138***	0.223	-0.057
	(4.09)	(-0.93)	(3.56)	(3.95)	(1.21)	(-1.08)
l_mom	0.006	0.321	-0.027*	-0.068***	-0.008	0.006
	(0.33)	(1.02)	(-1.71)	(-4.72)	(-1.21)	(1.30)
l_ret1	0.004	-1.749	-0.024	-0.215***	-0.034***	-0.011
	(0.06)	(-1.07)	(-0.40)	(-7.37)	(-3.23)	(-1.42)
l_ret1w	-0.417	-7.736	-0.652	-1.145***	-0.205*	0.015
	(-0.89)	(-1.04)	(-1.33)	(-4.05)	(-1.83)	(0.19)
l_att	-0.000	-0.079	0.011**	0.001	0.001	-0.001***
	(-0.00)	(-0.93)	(2.26)	(0.31)	(0.74)	(-3.22)
l_asvi	0.001	0.152	0.008	-0.006	-0.000	0.000
	(0.15)	(1.05)	(1.29)	(-1.12)	(-0.22)	(0.88)
l_prior_ins	0.011	-0.117	0.013	-0.041***	-0.019***	-0.005***
	(0.82)	(-0.93)	(0.87)	(-3.68)	(-10.33)	(-2.92)
l_prior_opp		-0.103			0.018***	
		(-1.07)			(3.97)	
l_prior_rou			0.001			0.024***
			(0.04)			(5.31)
Constant	-0.447***	0.286	-0.417***	0.540***	0.155***	0.029***
	(-4.62)	(0.43)	(-4.40)	(6.74)	(10.98)	(7.76)
Observations	90,753	90,753	90,753	90,756	90,756	90,756
R-squared	0.190	0.195	0.191	0.471	0.382	0.472
# groups	456	456	456	456	456	456

Table A3: Cumulative Market-adjusted Returns on Stocks Traded by Retail Investors Following Insider Trading: Sub-sample Tests

This reports the effect of retail investors following opportunistic insider purchases on market-adjusted returns for stocks with different levels of characteristics. We focus on firm book-to-market ratio (Panel A), prior month return (Panel B), Google ASVI (Panel C), institutional ownership (Panel D) as well as earnings surprise SUE (Panel E). In each panel we further separate the sample stocks into Follow portfolio and Not-Follow portfolio, based on whether retail investors follow insiders as defined in 1. In each panel, we categorize the sample of stocks into three levels of stock characteristics. For each level, we compute the difference in the cumulative market-adjusted returns between Follow portfolio and Not-Follow portfolio, up to 24 weeks. We use CRSP daily value-weighted market returns to calculate the average weekly market return, and subtract it from weekly stock return to get the weekly market-adjusted return. Standard errors are clustered at the week level. The returns are shown in percent. *p<0.1; **p<0.05, ***p<0.01.

Panel A: Sort by Book-to-market Ratio

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Low	Follow – Not-Follow	Mean	0.32***	0.04	-0.03
		t-stat	(2.60)	(0.20)	(-0.11)
Mid	Follow – Not-Follow	Mean	0.13**	0.25**	0.27*
		t-stat	(1.96)	(2.02)	(1.67)
High	Follow – Not-Follow	Mean	0.09	0.31	0.37
		t-stat	(0.80)	(1.55)	(1.31)
High – Low	Diff. in Diff.	Mean	0.23	0.27	0.40
		t-stat	(-1.41)	(0.93)	(0.98)

Panel B: Sort by Prior Month Return

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Low	Follow – Not-Follow	Mean	0.26**	0.43**	0.70**
		t-stat	(2.37)	(2.13)	(2.45)
Mid	Follow – Not-Follow	Mean	0.14*	0.20	0.11
		t-stat	(1.71)	(1.35)	(0.56)
High	Follow – Not-Follow	Mean	0.17*	0.36*	0.01
		t-stat	(1.77)	(1.73)	(0.04)
High – Low	Diff. in Diff.	Mean	-0.09	-0.08	-0.69*
		t-stat	(-0.60)	(-0.27)	(-1.73)

Table A3: (Cont.) Cumulative Market-adjusted Returns on Stocks Traded by Retail Investors Following Insider Trading: Sub-sample Tests

Panel C: Sort by Google ASVI

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Low	Follow – Not-Follow	Mean	0.07	0.36	0.60*
		t-stat	(0.53)	(1.48)	(1.85)
Mid	Follow – Not-Follow	Mean	0.29**	0.66***	0.61*
		t-stat	(2.20)	(2.63)	(1.93)
High	Follow – Not-Follow	Mean	0.23*	0.52**	0.36
		t-stat	(1.75)	(2.28)	(1.08)
High – Low	Diff. in Diff.	Mean	0.16	0.16	-0.24
		t-stat	(0.87)	(0.47)	(-0.52)

Panel D: Sort by Institutional Ownership

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Low	Follow – Not-Follow	Mean	0.15	0.28	0.27
		t-stat	(1.46)	(1.63)	(1.08)
Mid	Follow – Not-Follow	Mean	0.18**	0.14	0.09
		t-stat	(2.36)	(0.97)	(0.41)
High	Follow – Not-Follow	Mean	0.13*	0.23*	0.24
		t-stat	(1.76)	(1.76)	(1.43)
High – Low	Diff. in Diff.	Mean	-0.02	-0.05	0.02
		t-stat	(-0.14)	(-0.22)	(-0.07)

Panel E: Sort by Earnings Surprise (SUE)

Classification			Time Periods		
			Week 1 – Week 4	Week 1 – Week 12	Week 1 – Week 24
Negative SUE	Follow – Not-Follow	Mean	0.18**	0.30*	0.08
		t-stat	(1.97)	(1.78)	(0.37)
No SUE	Follow – Not-Follow	Mean	0.12*	0.15	0.20
		t-stat	(1.80)	(1.20)	(1.26)
Positive SUE	Follow – Not-Follow	Mean	0.23**	0.40*	0.42
		t-stat	(2.33)	(1.84)	(1.41)
Positive – Negative	Diff. in Diff.	Mean	0.04	0.10	0.34
		t-stat	(0.35)	(0.38)	(0.93)

Chapter 4

Trust, but Verify: The Economics of Scams in Initial Coin Offerings

with Kenny Phua, Chishen Wei, and Gloria Yu

> *"We embrace new technologies, but we also want investors to see what fraud looks like. I encourage investors to do their diligence and ask questions."*
>
> — Former SEC Chairman Jay Clayton on the *HoweyCoin* ICO[1]

4.1 Introduction

Frauds and financial scams are estimated to exceed U.S. $5 trillion annually. There are also significant psychic costs as victims often suffer depression, grief, shame, and suicidal thoughts.[2] While these welfare losses are substantial, empirical evidence on financial scams is relatively scarce. Data is often difficult to obtain because scammers work to evade detection and victims are often reluctant to step forward. To study the economics of scams, we exploit a unique setting from the market for initial coin offerings (ICOs) of cryptocurrencies. An ICO is a form of crowdfunding for a blockchain/cryptocurrency project. The ICO market has grown rapidly with almost no investor protection rules and mostly voluntary, unverified disclosures. ICOs have also become notorious for scams and frauds (Howell, Niessner, and Yermack, 2020). Investors' enthusiasm for ICOs, however, has not waned. ICOs have continued to successfully raise capital, with an

[1]In 2018, the U.S. Securities and Exchange Commission (SEC) created a mock ICO on `HoweyCoins.com` to educate the general public on the prevalence of scams and to urge investors to perform due diligence before investing in cryptocurrencies.

[2]See for example: Gee and Button (2019) and Button, Lewis, and Tapley (2009)

estimated U.S. $50 billion dollars raised through 2020 (PriceWaterhouseCoopers, 2020).

This paper investigates scams in the ICO market and demonstrates how malicious issuers target naïve investors. To perform our analysis, we collect 13 months of point-in-time snapshots of self-reported ICO data from five leading ICO listing websites. Listing websites are aggregators of past, current, and upcoming ICOs for prospective investors and are distinct from cryptocurrency exchanges. Manual collection is necessary for two important reasons. First, ICO data have no centralized repository and are scattered across listing websites. Second, the self-reported data are not reliable (Lyandres, Palazzo, and Rabetti, 2021) and can change over time. To analyze how an ICO was initially marketed to investors, we require data at the point-in-time in which the ICO occurred. For example, Figure 1 shows snapshots of the AdHive ICO on three websites. The `ICODrops` website reported a hardcap amount of $17,490,000, but `ICOBench` and `ICORating` reported amounts of $12,000,000. Discrepancies also occur in ICO ending dates, accepted payment methods, and total amounts raised.

[Insert Figure 4.1 here]

The case of the AdHive ICO is not special. In our sample of 5,935 ICOs, 34% of tokens have discrepancies at their first appearances. The prevalence of these discrepancies is somewhat puzzling. Why would so many issuers—who supposedly have the technical expertise to launch sophisticated blockchain projects—fail to accurately report ICO information on listing websites? A discrepancy implies that the issuer has misrepresented the offering because at least one of the reported material facts must be untrue. Investors may develop different perceptions and opinions of the offering depending on which website they happen to visit. Ideally, we would verify the accuracy of ICO information on listing websites against legal records. But, ICOs often avoid regulatory purview by sidestepping standard filing requirements for securities. Fewer than 1% of ICOs in our sample register with the SEC, although most are likely to qualify as securities under the Howey test (Gensler, 2021).

To better understand why misrepresentations are so prevalent, we model the behavior of a malicious ICO issuer who faces a pool of naïve and astute investors. Naïve investors are unsophisticated. They are unable to conduct proper due diligence and are likely to fall for an ICO scam. In contrast, astute investors carefully evaluate the offering and eventually refrain from funding it. Both naïve and astute investors may consume the issuer's time and resources by asking for more information or raising questions on public forums. From the issuer's perspective, astute investors are undesirable targets because they ultimately do not fund the scam. Indeed, a common tactic used by online vigilantes to disrupt tech-support scammers is to pose as victims and hold tedious, unfruitful conversations. In a recent interview, Kitboga (alias) said, *"[...] important for everyone to know [...] how much these scammers hate when you ask questions"*.[3] The former SEC chairman Jay Clayton also encouraged prospective investors to ask questions to ICO issuers (SEC, 2018). Because investor types are unobservable ex-ante, the issuer would optimally wish to screen out astute investors.

We hypothesize that malicious ICO issuers use misrepresentations, along with other suspicious actions, to screen out astute investors and target naïve investors. Astute investors will notice the cross-site discrepancies and immediately dismiss the offering without consuming the issuer's time and resources. In contrast, naïve investors overlook the cross-site discrepancies and remain viable victims of the ICO scam. Ultimately, the investors who remain are likely to be naïve investors—the ideal targets of the malicious issuer. For centuries, this

[3]Source: https://www.newsweek.com/laughter-death-threats-meet-kitboga-youtuber-exposing-tech-support-scams-938384

screening strategy has been observed in various advanced fee financial scams.[4] A modern example is the Nigerian Prince email hoax, which solicits potential victims to send money to a fictitious Nigerian Prince in exchange for a large fortune. This incredible narrative is crafted to repel discerning individuals and target the most gullible victims (Herley, 2012). Using decentralized immutable blockchain data, we shed new light on this classic financial fraud scheme.

Our main finding is that ICOs with misrepresentations are significantly more likely to be scams. To identify ICO scams, we collect crowdsourced scam events from `DeadCoin.com` and corroborate these records with reports from news articles, message boards, and regulatory authorities. Estimates from our hazard regressions reveal that the presence of at least one misrepresentation more than triples the odds of a ICO scam. At the intensive margin, an additional misrepresentation raises the odds of a scam by 14.0%. To sharpen our analysis, we focus on misrepresentations of basic characteristics (i.e., ticker, start/end dates, duration of fundraising, country of origin, countries from which investors are banned, and acceptable payment modes). Such misrepresentations are a potent screen for investor naïvety because these characteristics are fundamental in performing basic due diligence and do not require investor expertise. Consistent with this view, we find that the odds of a scam increase by 24.0% per unit of such misrepresentations.

To assess our screening mechanism more carefully, we extract data from the Ethereum blockchain. First, we find the Ethereum block height corresponding to 10 days after the end date of every ICO. Next, we gather data on token holdings and transaction activities from wallets that hold its tokens as at that block height. Using these data, we characterize the sophistication of the typical token holders in ICOs and test whether misrepresentations are associated with

[4] For example, Eugène François Vidocq, a French private investigator, detailed in his 1832 memoirs a scam known as the "letters of Jerusalem". The scammer typically solicits the victims' (financial) help to recover fictitious treasures.

lower investor sophistication. Consistent with this view, wallets that hold tokens of misrepresented ICOs (i) have lower portfolio values, (ii) are less diversified, and (iii) are less active. Overall, these findings lend further credence to our interpretation that malicious issuers use misrepresentations to screen for naïve investors.

An alternative interpretation of our findings is that misrepresentations are unintentional mistakes made by careless issuers. We design three sets of tests to evaluate this potential explanation. First, if the underlying motives are nefarious, we expect regulatory scrutiny to reduce the use of misrepresentations. Consistent with this prediction, ICOs launched shortly after news of regulatory action in the cryptocurrency markets have significantly fewer misrepresentations. Second, misrepresentations may occur when low quality issuers fail to accurately market their listing. To the extent that these issuers have lower quality blockchain projects, misrepresentations should be negatively associated with ICO quality. However, using disclosure practices (Bourveau et al., 2021) and fundraising outcomes as proxies for ICO quality, we find no differences in quality between misrepresented and non-misrepresented ICOs.

Third, we apply network analysis to detect suspicious patterns of misrepresentation behavior among ICO issuers. Short of conducting interviews with scammers, we cannot directly observe the true intentions behind misrepresentation behavior. But, we can examine whether the strategic use of misrepresentations leaves suspicious footprints throughout the ICO ecosystem. For this analysis, we exploit the prevalence of ICO advisors who are hired by issuers to launch token offerings. These advisors often work on multiple ICOs. If misrepresentation behavior is learned or passed through common advisors, the network position of an ICO should be related to its misrepresentation behavior.[5] Consistent with this prediction, we find that ICOs with higher Katz centrality in the network have more misrepresentations. Surprisingly, we find that advisors of misrepresented ICOs are not penalized, but obtain more subsequent advisory opportunities. This finding suggests that there exists many malicious issuers who solicit the services of such advisors. Overall, our evidence indicates that misrepresentation behavior is systemic within the ICO ecosystem.

To complement their use of misrepresentations, malicious issuers may conduct other suspicious actions to target naïve investors. First, such issuers may use celebrity endorsements to attract less sophisticated investors. Consistent with the warnings of the SEC, we find that celebrity endorsements are strongly associated with ICO scam risk. Second, we conjecture that passive web traffic arising from paid advertisements, referral links, and search engines reflects visits from less sophisticated individuals. Using data on web traffic flows in our sample period, we find that malicious issuers prefer to promote their ICOs on listing websites with higher passive web traffic. These findings suggest that malicious ICO issuers use a variety of tactics to attract naïve investors to their offerings. Nevertheless, we find that misrepresentations retain a distinct predictive effect on ICO scam risk.

Finally, we perform a welfare analysis of the financial losses from ICO scams in our sample. A key challenge in identifying financial scams is the reluctance of victims to report losses. Thus, many scams go unreported and undetected. To overcome this partial observability problem, we use detection-controlled estimation (DCE) methods (Feinstein, 1990) and estimate that the total financial losses exceed U.S. $12 billion in our sample. As many as 40% of ICOs in our sample may be scams, but most go undetected. These large estimates imply that

[5] Ballester, Calvó-Armengol, and Zenou (2006) show that when there are strategic complementarities in behavior, such as learning or social norms, agents who are more central in a network exhibit a higher level of this behavior.

more stringent regulations and stronger enforcement actions may be justified to protect investor welfare.

Our paper contributes to a recent literature on the controversies surrounding cryptocurrencies (Yermack, 2015). For example, Griffin and Shams (2020) find that Tether, a digital currency pegged to the U.S. dollar, is used to manipulate bitcoin prices. Li, Shin, and Wang (2021); Dhawan and Putniņš (2022) document choreographed pump-and-dump trading schemes in cryptocurrencies. Studies also find evidence of wash trading that artificially boosts trading volumes on crypto-exchanges (Aloosh and Li, 2019; Cong et al., 2020).[6] A distinguishing feature of our study is the focus on the initial offering stage. While suspicions of ICO scams are widespread, evidence to date is relatively scarce. Using point-in-time data, we provide evidence on how unscrupulous actors target naïve investors and estimate the size of scams in the crypotcurrency market.

Our study also builds on Lyandres, Palazzo, and Rabetti (2021), who document the limitations of available ICO data and the ways to characterize data quality. We find the data quality contains key information on likelihood of a scam. Thus, our findings add a new perspective to existing studies that analyze the determinants of ICO success (Benedetti and Kostovetsky, 2021; Deng, Lee, and Zhong, 2018; Dittmar and Wu, 2019; Howell, Niessner, and Yermack, 2020). Our findings may also be of interest to recent theoretical work on ICOs, which links token development to value and utility (Cong, Li, and Wang, 2020; Sockin and Xiong, 2020).

[6] Aloosh and Li (2019) exploit individual accounts on the Mt. Gox crypto-exchange for direct evidence. Cong et al. (2020) applies Benford's Law to identify wash trading patterns for 29 exchanges.

4.2 ICO overview

An ICO allows entrepreneurs to raise capital via cryptographically secured tokens. Typically, an issuer resorts to an ICO when other sources of capital (e.g., venture capital and private equity) are prohibitively expensive or inaccessible. Thus, an ICO is a risky crowdfunding operation, in which the issuer sells tokens that will serve as the payment medium for the products or services of the start-up. There are several stages in the ICO process. First, the issuer creates fundraising campaign materials. Next, the issuer sets pricing terms and markets the offering on listing websites. Finally, if the financing goals of the ICO are met, the issuer then creates and distributes tokens to the investors.

4.2.1 Fundraising campaign: Listing websites

The fundraising campaign entails (i) producing a whitepaper, (ii) hosting a website to provide additional information, (iii) maintaining an active social media presence, and (iv) listing the token on ICO listing websites. A whitepaper describes the goals, objectives, and development milestones of the project. But, whitepapers often lack details of business operations and rarely contain financial disclosures.

To list an ICO on a listing website, the issuer directly submits token information on the website and awaits approval. Listings are typically free, but for an additional fee, the website can feature and promote the ICO. The issuer may also hire advisors to advertise and market the ICO. These advisors usually have technical or marketing expertise, and may alleviate information asymmetry between the issuer and potential investors. However, celebrities with little or no blockchain expertise are also employed as advisors to promote the ICO. The SEC has warned that celebrity endorsements are often associated with ICO scams.

4.2.2 ICO pricing and listing on secondary markets

The pricing structure of ICOs are often opaque. On listing websites, issuers advertise a subscription price to the general public. But, many ICOs invite privileged investors to an earlier presale offering. While details on the presale pricing structure are not publicly available, Fahlenbrach and Frattaroli (2020) find that presales offer a significant discount to the subsequent public offering price. Presale funding rounds are controversial. They may signal strong demand from informed investors, but are also used to manipulate the sentiments of the general public. The SEC has also warned that presales are often associated with ICO scams.

The issuer may set funding goals in the ICO. The softcap is the minimum amount of funds raised to continue the project. An issuer may also specify a hardcap, which is the maximum number of tokens that can be sold in the ICO. The hardcap limits the amount of funds that can be raised in the ICO. If the softcap is met and the project is successful, the issuer will create and distribute the tokens to investors. Subsequently, investors may trade the tokens in the secondary market or use the tokens for its utility (e.g., access products or services funded by the ICO). Investors tend to have short holding periods and flip the tokens on cryptocurrency exchanges (Fahlenbrach and Frattaroli, 2020).

4.2.3 Regulatory environment

The ICO regulatory environment differs across countries. Some countries impose outright bans on ICOs (e.g., China and South Korea), while other countries adopt regulatory guidelines (e.g., Australia and the United States). The SEC of the United States uses the Howey Test framework to determine whether a digital asset qualifies as a security.[7] Specifically, a digital asset is a security

[7]See, https://www.sec.gov/corpfin/framework-investment-contract-analysis-digital-assets

if (i) there is an investment of money and (ii) expectation of profits; (iii) the investment of money is in a common enterprise; and (iv) any profit comes from the efforts of a promoter or third party. The SEC Chairman Gary Gensler and his predecessor Jay Clayton believe most ICOs pass the Howey Test and are hence subject to U.S. securities laws.

Issuers of security tokens can register with SEC via form S-1 or apply for registration exemptions. Although most ICOs should arguably be classified as security offerings, fewer than 100 tokens in our sample are registered with the SEC potentially due to the high compliance costs. For exemptions, regulation D applies if funds are raised from only accredited investors; Regulation A and A+ apply if funds are raised from a broader set of investors but the offering is less than $50 million; and issuers can also make token sales under Regulation Crowdfunding.

4.2.4 Are misrepresentations a violation of securities law?

ICOs classified as security offerings are subject to the Rule 10b-5, which specifies the conditions for securities fraud as follows[8]:

> It shall be unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce, or of the mails or of any facility of any national securities exchange, (a) To employ any device, scheme, or artifice to defraud, (b) To make any untrue statement of a material fact or to omit to state a material fact necessary in order to make the statements made, in the light of the circumstances under which they were made, not misleading, or (c) To engage in any act, practice, or course of business which

[8] Rule 10b-5 is issued by the SEC under section 10(b) of the Securities Exchange Act of 1934.

operates or would operate as a fraud or deceit upon any person, in connection with the purchase or sale of any security.

Misrepresentations of ICO characteristics are necessarily an untrue statement of material fact (violation of part (b)) because at least one of the reported characteristics is false. Moreover, such untrue statements can potentially mislead investors. If misrepresentation are purposely used to commit fraud or deceit, then they would also violate part (c) of the rule. As of January 2021, the SEC has taken regulatory actions against 68 ICOs and cryptocurrency offerings. The judgments from these regulatory actions totaled U.S. $99.8 million, of which U.S. $88.9 million were refunds and U.S. $10.9 million were penalties. Additionally, 20 securities class action lawsuits have been filed against ICO issuers. However, websites that aggregate ICO information voluntarily reported by issuers have minimal disclosure requirements and are lightly regulated.

4.3 Main hypothesis

We develop a model to analyze how malicious issuers use cross-website discrepancies of the ICO attributes to screen for naïve investors. Our model shares similarities with frameworks that analyze the prevalence of other scams and hoaxes in cyberspace (e.g., Herley, 2012).

4.3.1 The issuer's classification problem

There are three periods in the model. The malicious ICO issuer faces a mass of m investors, of which there are n naïve investors and $m - n$ astute investors. Individual investor types are ex ante unobservable. The key difference between the investor types is that naïve investors may not) fund the ICO scam while astute investors will not. We define d_i to be the number of misrepresentations tolerated by an investor i, above which the investor immediately dismisses an

ICO scam. Some naïve investors could have lower d than astute ones. But, on average, naïve investors are more tolerant of misrepresentations such that the average d for naïve investors is higher than their astute counterparts. We structure the following description of our model around Figure 4.2.

[Insert Figure 4.2 here]

In period one, the issuer sets the number of misrepresentations d^*, which acts as a cutoff (screen) for investors who are viable targets. In forming this targeting strategy, the issuer faces the risk of classification errors. For a given d^*, the fraction of naïve investors who immediately dismisses the ICO scam is $F_{d|\text{type}}(d^* \mid \text{naïve})$. Conversely, the fraction of naïve investors who remain viable targets to the scam is the complementary conditional cumulative distribution function $\bar{F}_{d|\text{type}}(d^* \mid \text{naïve}) = 1 - F_{d|\text{type}}(d^* \mid \text{naïve})$. Likewise, the fraction of astute investors targeted is $\bar{F}_{d|\text{type}}(d^* \mid \text{astute})$. Because $\bar{F}(\cdot)$ is monotonically decreasing in d, a higher (lower) d^* leads the issuer to target lower (higher) fractions of both naïve and astute investors.

In period two, any remaining investor (i.e., those that have not dismissed the scam) may request more information from the issuer or raise questions about the ICO on public forums such as Reddit, Twiiter, and Bitcointalk. The public nature of these forums implies that the issuer cannot avoid these costs by ignoring investor queries without raising suspicion. Without loss of generality, the malicious issuer incurs a constant cost C per remaining investor (both astute and naïve) that reflects the time and resources needed to address questions.

In the final period, naïve investors ultimately fund the scam and while astute investors do not. Targeting a naïve investor yields the issuer a net profit $G = Q - C$, where Q is the gross proceeds from the scam. Whereas, an astute investor refrains from funding the scam, hence yielding the issuer a net loss C.

Astute investors are undesirable because they consume resources but provide no financial rewards to the issuer. The issuer's expected profits $\mathbb{E}(\Pi)$ can be expressed as a function of d^*.

$$\mathbb{E}(\Pi) = m\left[z \cdot \bar{F}_{d|\text{type}}(d^* \mid \text{naïve}) \cdot G - (1-z) \cdot \bar{F}_{d|\text{type}}(d^* \mid \text{astute}) \cdot C\right], \quad \text{where } z = n/m \tag{4.1}$$

It is instructive to examine an indiscriminate targeting strategy that abandons the screening strategy. The issuer targets all investors by choosing $d^* = 0$, thereby setting $\bar{F}_{d|\text{type}}(\cdot) = 1$. Imposing these constraints and $\mathbb{E}(\Pi) > 0$, we obtain equation (4.2). When $C > 0$, equation (4.2) implies that that an indiscriminate targeting strategy is profitable if the fraction of naïve investors is greater than the ratio $C/(C + G)$. For example, suppose 1% of investors are naïve and G = \$1,000, then C can at most be $0.01/(1 - 0.01) \times \$1{,}000 = \$10.10$ per investor. Indiscriminate targeting can also be profitable in the special case of $C = 0$. However, this case is unlikely given the threat of reputation loss and regulatory scrutiny, and resources required to entertain investors' queries. Finally, targeting all investors is also profitable in the technical case of $G \to \infty$, which is patently unrealistic. The prevalence of misrepresented ICOs suggests that the above conditions are unmet in our sample.

$$z = \frac{n}{m} > \frac{C}{G + C} \tag{4.2}$$

4.3.2 Misrepresentations as a screening device

We examine tradeoffs implied by the targeting strategies. Figure 4.3 presents probability density plots of d, conditional on investor types—astute (black) and naïve (red). Shaded areas in black and red represent the complementary conditional cumulative distributions $\bar{F}_{d|\text{type}}(d^* \mid \text{astute})$ and $\bar{F}_{d|\text{type}}(d^* \mid \text{naïve})$, respectively. In Subfigure 4.3a, the malicious issuer adopts a conservative targeting strategy by choosing a high number of misrepresentations (high d^*). Because $\bar{F}(\cdot)$ is monotonically decreasing in d, the conservative strategy avoids many costly astute investors. However, the issuer necessarily forgoes many profitable naïve investors in the population. In Subfigure 4.3b, the issuer sets an aggressive targeting strategy by choosing a low d^*. While this strategy captures more naïve investors, it also retains more costly astute investors hence eroding the issuer's profits. Thus, the issuer needs to strike a balance between extremely conservative and aggressive targeting strategies.

[Insert Figure 4.3 here]

The above exercise conveys the intuition for (i) why misrepresentations are so widespread, and (ii) how they are used as a screening device. To complete our analysis, we formalize the intuition from Figure 4.3. Given that d^* affects the quantities of naïve and astute investors being targeted, we can solve for the optimal targeting strategy (henceforth, OTP) of the malicious issuer. Using the chain rule, the issuer maximizes profits in equation (4.1) by choosing d^* such that:[9]

$$\partial \bar{F}(d^* \mid \text{naïve}) / \partial \bar{F}(d^* \mid \text{astute}) = \frac{1-z}{z} \cdot \frac{C}{G} \tag{4.3}$$

Under the OTP, equation (4.3) prescribes the rate of naïve investors targeted per astute investor. This rate is a function of z, C, and G. For example, suppose the issuer believes that there are many naïve investors (high z). Then, the OTP prescribes a low rate, which translates to an aggressive targeting strategy (see, Subfigure 4.3b). If the issuer has an inferior technology to entertain investors' queries (high C), then the issuer optimally chooses a higher rate that is achieved by a higher and more conservative d^*. Above all, issuers cannot observe the

[9] We first write the first order conditions of $\mathbb{E}(\Pi)$ with respect to $\bar{F}(d^* \mid \text{naïve})$ and $\bar{F}(d^* \mid \text{astute})$: $\partial \mathbb{E}(\Pi)/\partial \bar{F}(d^* \mid \text{naïve}) = zGm$ and $\partial \mathbb{E}(\Pi)/\partial \bar{F}(d^* \mid \text{astute}) = (1-z)Cm$. Next, we use chain rule to express the OTP as a function of z, C, and G: $\partial \bar{F}(d^* \mid \text{naïve})/\partial \bar{F}(d^* \mid \text{astute}) = (1-z)Cm/zGm = (1-z)/z \cdot C/G$.

parameters—z, C, and G—and may form heterogeneous beliefs about them. In turn, these heterogeneous beliefs may lead to heterogeneity in misrepresentation behavior across our sample ICOs.

In the context of our above analyses, we discuss two candidate explanations of ICO misrepresentations. First, the malicious issuer is unlikely to use misrepresentations to maximize investor interest. If that were the goal, misrepresentations are counterproductive because cross-site verification of ICO information is easy. Put differently, maximizing investor interest is like an overly aggressive targeting strategy, attracting too many costly astute investors who eventually balk at funding the ICO. Second, 34% of ICOs have misrepresentations at their first appearances in our sample. The sheer number of misrepresented ICOs makes issuers' carelessness an unsatisfactory explanation. Absent the screening mechanism and intention to scam, it is puzzling that so many issuers fail to accurately provide ICO information on listing websites.

Instead, we propose that the malicious issuer uses misrepresentations to screen for investor sophistication. Because investor sophistication is unobservable, a good strategy is to get naïve investors to self-identify. ICO misrepresentations will induce suspicions in all but the most naïve investors. Any astute investor who performs due diligence would recognize the misrepresentations and ignore the ICO. Those who remain are the naïve investors—the ideal targets of the malicious issuer.[10] The issuer increases her odds of profitability by targeting naïve investors and repelling their astute counterparts. Having established the

[10]The use of misrepresentations as a screening device in ICO scams has parallels with other notorious scams such as the advance-fee scams. The advance-fee scammer promises prospective victims in e-mails a large sum of money in return for a small upfront administrative fee. These e-mails often contain grammatical errors and use outlandish language. In some cases, the emails also tell an incredible story, in which the scammer impersonates a member of the Nigerian royal family. The inclusion of these tell-tale signs is not accidental but strategic (Herley, 2012). Astute people, who could waste the scammer's time and resources, recognize these signs and ignore the emails. Whereas, only the most gullible victims would respond to the emails, hence self-identifying their gullibility to the scammer.

modus operandi of malicious issuers, we hypothesize that ICO misrepresentations predict scam risk.

4.4 Data and descriptive statistics

This section describes our data collection process, defines the main variables, and presents the descriptive statistics of our sample.

4.4.1 Data sources

We systematically collect point-in-time ICO data from five major websites that aggregate ICO listings—(i) ICOBench (ii) ICOCheck (iii) ICOData (iv) ICODrops (v) ICORating. We select these five listing websites based on (i) their popularity reported by Alexa Traffic Rank on August 15th 2018, (ii) the number of ICOs covered, and (iii) the technical feasibility of scraping the websites.[11] On the 15th of every month from August 2018 to August 2019, we scrape ICO data from these five websites. In total, we have 13 data collection events and a time-series of ICO characteristics for every ICO-website pair. Because ICO identifying information may vary across websites, we manually cross-check all ICOs and designate a set of unique identifiers to every ICO in our sample. To resolve residual conflicts in our collected data, we hand-check other Internet sources. Thus, we alleviate concerns of variation in ICO names, misspellings, and name changes. Overall, our sample contains 5,935 matched ICOs.[12]

We collect ICO scam allegations from a prominent crowdsourced anti-fraud project hosted on `DeadCoins.com`. The `DeadCoins` website curates a list of ICOs that are alleged scams, alongside a summary of every scam and corresponding

[11] Based on the Alexa Traffic Rank on November 30th 2018, Lyandres, Palazzo, and Rabetti (2021) obtain ICO data from `ICOBench`, `ICODrops`, `ICORating`, `ICOMarks`, and `ICOData`. We replace `ICOMarks` with `ICOCheck` for the latter two considerations.

[12] The numbers of unique ICOs covered by the listing websites are: `ICORating` (4,166), `ICOBench` (4,021), `ICOData` (1,896), `ICODrops` (625), and `ICOCheck` (580).

information sources. Reasons behind scam allegations include charges by regulators for fraudulent activities, cancellation by exchanges, obvious technical flaws, disappearance of ICO issuers, and prolonged inactivity. For example, the Shopin token was marked as "dead" (i.e., inactive) on `Deadcoins` following a SEC complaint. Subsequently, the founders and company behind the Shopin token were charged with securities fraud and violations of registration processes.

To mitigate concerns of false positivity, we corroborate every `Deadcoin` scam allegation with several media sources.[13] First, we check whether the ICO is reported by regulatory authorities (e.g., SEC, DoJ). Second, we search on Factiva for press coverage (e.g., news articles, website articles, journal articles) of the ICO scam. Third, we search popular online forums and social media (e.g., Reddit, Cryptocompare) for mentions of the ICO scam. We admit an alleged ICO scam into our sample only if it is found on at least one of the above three media channels. In total, we match 115 ICO scams to our sample.

We collect regulatory filings (Form D, Form 1-A, and Form C) of ICOs that are available on the SEC EDGAR database. We search the database using the keywords "token", "ICO", "initial coin offering", "coin", and "crypto". We then manually determine whether every filing is ICO-related. We first read the filing document and check whether it pertains to an initial coin offering or other types of offering. If this information is not stated, we then use the firm name written in the document combined with the keywords "ICO", "offering", "token" to perform a search on SEC EDGAR. All else failing, we use the names of persons (i.e., founders, CEOs, and directors) in the filing combined with the above keywords to perform another search on SEC EDGAR. In our sample, 77, two, and eight ICOs have filed for a Form D, Form 1-A, and Form C, respectively.

[13]Notably, the `Deadcoin` website also prominently displays a form to contest scam allegations.

4.4.2 Variables

Our key independent variable is the *misrep* of an ICO—the total number of cross-website discrepancies of 13 commonly reported characteristics at its first appearance in our sample.[14] Figure 4.4 visualizes the proportion of ICOs with at least one cross-website discrepancy by these characteristics at first appearances in our sample. The most common misrepresented characteristic is *whitelist* (36.9%). Other commonly misrepresented characteristics are *start date* (25.9%), *end date* (26.12%), *presale* (20.7%), and *banned* (16.6%). Misrepresentations in *softcap*, *ticker*, and *country* are uncommon.

[Insert Figure 4.4 here]

In our empirical tests, we control for a suite of variables that describes the fundraising structure and regulatory environment of an ICO. The following control variables are coded as indicators that switch on if the ICO has the corresponding features. An ICO is *banned* if it is banned by at least one regulatory authority. A *whitelist* allows an ICO issuer to limit the sale of tokens to a selected group of registered investors. An ICO can hold a *presale* round to sell tokens before the public fundraising campaign is set up. The *hardcap* is the upper limit on the number of tokens that can be sold in an ICO. The *softcap* is the minimum amount of funds that must be raised in an ICO, or else funds are returned to investors and the project is discontinued. We control for payment options in the ICO with *accept BTC* (*ETH*, *USD*). The last indicator is *SEC filing*, which switches on if the ICO has regulatory filings with the SEC. The remaining control variables are continuous. The *duration* of an ICO is the length of its fundraising period in days. Finally, the *enforcement* and *disclosure* indices

[14] The 13 characteristics used to construct *misrep* are *banned*, *whitelist*, *presale*, *hardcap*, *softcap*, *accept BTC*, *accept ETH*, *accept USD*, *ticker*, *start date*, *end date*, *duration*, and *country*.

from La Porta et al. (2000) control for the regulatory environment in the ICO's country of registration.

4.4.3 Descriptive statistics

Table 4.1 reports summary statistics of our sample. Panel A reports that the average ICO has 1.28 *misrep*, and 34% of ICOs have at least one *misrep*. 95% of ICOs are banned in at least one country, which is unsurprising as ICOs are illegal in several countries (e.g., China, Egypt, Morocco). About half of ICOs impose selectivity in their investor clientele or fundraising structures; 55% of ICOs have an investor *whitelist*, and 47% of them have *presale* rounds. Most ICOs (70%) have a *hardcap* in their fundraising structures, but only a minority (26%) have a *softcap*. ETH (USD) is the most (least) popular payment currency among ICO issuers. Fewer than 1% of ICOs in our sample have regulatory filings with the SEC. The fundraising period for the average (median) ICO is 54 (37) days. Panel B reports the Pearson pairwise correlations among our variables. Our key variable *misrep* is weakly correlated with most variables, except for *presale* (0.31), *hardcap* (28%), and *accept ETH* (31%).

[Insert Table 4.1 here]

Table 4.2 reports differences in ICO scam rates and characteristics between (i) ICOs with at least one *misrep* and (ii) ICOs with no *misrep*. We observe significant differences across the two groups. ICOs with at least one *misrep* are more likely to incur a scam allegation (4% vs. 1%). Such ICOs also have weaker governance—they are less likely to have a investor *whitelist* (46% vs. 60%) and are more likely to hold a *presale* funding round (68% vs. 36%). These ICOs are also more likely to have salient attributes that imply limited supply—misrepresented ICOs have shorter fundraising periods (*duration* of 48 days vs. 58 days) and are more likely to have a *hardcap* (89% vs. 60%). Misrepresented ICOs also accept a wider range of payment options.

[Insert Table 4.2 here]

4.5 Misrepresentations and ICO scams

We design two tests of our hypothesis that malicious ICO issuers use misrepresentations to screen for naïve investors. First, we perform survival analysis to examine whether misrepresented ICOs are more likely to be scams. Second, to assess our screening mechanism more carefully, we extract data from the Ethereum blockchain to characterize the sophistication of investors who hold tokens in misrepresented ICOs.

4.5.1 Survival analysis: ICO scam risk

We perform survival analysis to test the hypothesis that ICOs with more misrepresentations are more likely to be scams. Our objective is to track the survival time of an ICO—the time elapsed between its entry into our sample and occurrence of a scam allegation. There are three notable features of our empirical setting that are well accommodated by survival analysis. First, ICOs can enter and exit our sample at different points in time. Second, we only have information about which ICOs survive (i.e., remain in our sample) at any point in time. An ICO exits our sample when it incurs a scam allegation. Otherwise, it is right-censored. Right-censoring occurs if an ICO (i) becomes unlisted on listing websites, or (ii) survives till the end of our 13-month observation window without a scam allegation.[15] Third, survival times usually do not have normal distributions.

[15]Right-censored observations are not necessarily cleared of scams.

We plot the proportion of surviving ICOs—the survival function $S(t)$—with respect to survival time t. First, we sort ICOs by their *misrep* into four groups. Where r_t is the number of surviving and uncensored ICOs instantaneously before time t, and f_t is the number of ICOs that incur scam allegations, we next compute the survival function within every group:

$$S(t) = \begin{cases} \dfrac{(r_t - f_t)}{r_t} \times S(t-1), & \text{for } t > 0 \\ 1, & \text{for } t = 0 \end{cases} \tag{4.4}$$

Figure 4.5 shows that all four groups begin with $S(0) = 1$ because our sample precludes ICOs that are known to be scams. As time progresses, the survival functions of all four groups decline as ICO scams are flagged on the `DeadCoin` website. However, we find that the survival function in the high-*misrep* group declines most quickly. In comparison, the decline in survival function of the low-*misrep* group is substantially slower. This difference in trends is first evidence that *misrep* is positively associated with the incidence of ICO scams.

[Insert Figure 4.5 here]

We now estimate the effect of *misrep* on the incidence of ICO scams with Cox regression models. Where $h(t) = -\frac{\delta}{\delta t} \log S(t)$ is the expected hazard that denotes the rate of ICO scams conditional on survival up to time t, and $h_0(t)$ is the baseline hazard when all covariates equal zero, we estimate specification (4.5).

$$h(t) = h_0(t) \exp\left(\beta_1 \, misrep + \mathbf{X}^\top \boldsymbol{\beta}\right) \tag{4.5}$$

The vectors $\mathbf{X}$ and $\boldsymbol{\beta}$ represent vectors of control variables and their corresponding estimated coefficients, respectively. For ease of interpretation, we express estimated coefficients as hazard ratios. A hazard ratio that equals one implies that an increase in the covariate has no effect on the hazard of ICO scams. If the hazard ratio is above (below) one, then the covariate is associated with an increase (decrease) in the hazard of ICO scams.

[Insert Table 4.3 here]

Our estimates in Table 4.3 show that ICOs with higher *misrep* are more likely to be scams. Column 1 shows that the presence of *misrep* more than triples ($t = 5.46$) the hazard ratio of ICO scams. At the intensive margin, we find in column 2 that an additional *misrep* is associated with a 25.3% ($t = 6.71$) rise in hazard of ICO scams. We further add coverage quartile fixed effects and stratify our ICOs by their calendar-quarter cohorts in column 3.[16] These augmentations address two concerns. First, the coverage fixed effects alleviate the concern that *misrep* is mechanically driven by the number of websites that an ICO is listed on. Second, the stratification allows ICOs to have cohort-specific baseline hazards $h_0(t)$—this absorbs heterogeneity in hazard of ICO scams across cohorts. In this augmented specification, we find that an additional *misrep* increases the hazard of ICO scams by 14.0%. ($t = 2.18$). To add color to our findings, we focus on misrepresentations in a subset of basic ICO characteristics.[17] Basic ICO characteristics are salient, requires little expertise to understand, and should be fundamental to investors' due diligence. In column 4, we find that an additional $misrep^{basic}$ increases the hazard of ICO scams by 24.0% ($t = 4.86$).[18] This finding reinforces our screening hypothesis—investors who fail to notice discrepancies in the most basic ICO characteristics likely also fail to perform due diligence. Thus, such discrepancies are particularly potent screens for investor sophistication.

[16]Coverage is the number of listing websites that an ICO is listed on. Two ICOs are in the same cohort if their ICO start dates are in the same calendar quarter.

[17]Basic ICO characteristics are *ticker*, *country*, *banned*, *start date*, *end date*, *duration*, and acceptable payment modes. Nonbasic ICO characteristics are *softcap*, *hardcap*, *whitelist*, and *presale*.

[18]In contrast, we find in untabulated results that misrepresentations of nonbasic characteristics has a negligible predictive effect ($-3.3\%, t = 0.40$) on ICO scam risk.

Overall, we find that misrepresentations of ICO attributes on listing websites are a powerful ex-ante predictor of scams. Consistent with our screening hypothesis, the predictive effect is primarily driven by misrepresentations of basic ICO information. Our findings suggest that simple cross-website verification of ICO attributes is an effective form of due diligence for prospective investors.

4.5.2 Misrepresentations and wallet characteristics

To assess our screening mechanism more carefully, we extract data from the Ethereum blockchain.[19] This blockchain is a digitally distributed, decentralized, public ledger of all transactions that occur on the Ethereum network. This means that we can observe token holdings and transaction activities of cryptocurrency wallets (henceforth, wallets) on the network. For every ICO, we use these data to characterize the sophistication of wallet-users who hold its tokens. Thereafter, we examine the relation between misrepresentations in an ICO and the sophistication of its typical token holder.

We provide details on the data collection process and how we measure the sophistication of wallet-users. First, we find the contract addresses of our sample ICOs by manually matching them by name and ticker on the website `Etherscan.io`.[20] For every ICO, we then find the Ethereum block height corresponding to 10 days after its end date. Next, by querying the contract address of an ICO in the `Covalent Unified` Application Programming Interface (API), we find the wallet addresses that hold its tokens as at its corresponding block height. For this analysis, we focus on the top 100 wallet addresses of every ICO by token holdings, and exclude an ICO if it has fewer than 30 wallets holding its tokens. Finally, we again query the `Covalent Unified` API to extract granular

[19]Most ICO tokens adopt the ERC-20 (Ethereum Request for Comments 20) standard, which facilitates interoperability with other tokens on the Ethereum network.

[20]Every ICO token has a unique contract address on the Ethereum blockchain. The Internet Appendix contains further details of our matching process.

data on 110,607 wallets holding tokens of 1,996 ICOs.[21]

$$\log\left(sophistication\right) = \alpha + \beta_1 \mathbb{1}(misrep > 0) + \mathbf{X}^\top \boldsymbol{\beta} \tag{4.6}$$

For every ICO, we characterize the sophistication of its typical token holder. Specifically, we compute—at the wallet level—(i) total portfolio value (in U.S. dollars) of all tokens held, (ii) number of distinct tokens held, and (iii) number of transactions. Then, we aggregate these measures at the ICO level by taking their medians to obtain *value*, *diversity*, and *activity*, respectively. We make three conjectures about naïve investors. First, to the extent that wealth positively correlates with sophistication, they have lower wallet values. Second, they are more reckless or uninformed, so they diversify less by holding fewer distinct ICO tokens. Third, they have weaker technical or trading expertise, so they make fewer transactions. Thus, we expect investor sophistication to correlate positively with *activity*, *diversity*, and *age*. To test whether malicious issuers successfully use misrepresentations to screen for naïve investors, we estimate Poisson regressions in specification (4.6).

[Insert Table 4.4 here]

Our results in Table 4.4 suggest that investors who hold tokens of misrepresented ICOs are less sophisticated. The dependent variable is one of *value*, *diversity*, and activity. The key independent variable is $\mathbb{1}(misrep > 0)$—an indicator that switches on if the ICO has at least one *misrep* at its first appearance in our sample. Our models include ICO calendar-quarter cohort fixed effects, and standard errors are clustered by these cohorts. For ease of interpretation, we express estimated coefficients as incidence rate ratios. Column 1 indicates that

[21] Of the full sample of 5,935 ICOs, we unambiguously matched 4,611 ICOs to their contract addresses on `Etherscan.io`. The remaining attrition is due to our requirement that the ICO token must have at least 30 holders at 10 days after the end date.

the typical investor in misrepresented ICOs have a 40.1% ($t = 2.61$) lower wallet *value*. This result supports our view that misrepresented ICOs tend to attract less sophisticated investors. In column 2, we find that switching on $\mathbb{1}(misrep > 0)$ is associated with a 19.7% ($t = 2.88$) decline in *diversity*. This finding suggests that token holders in misrepresented ICOs are more reckless and less financially savvy, pointing to investor naïvety. Column 3 shows that $\mathbb{1}(misrep > 0)$ is associated with a 9.0% ($t = 2.62$) decrease in transaction *activity*. Thus, wallets that hold misrepresented ICO tokens likely belong to naïve investors who—due to their weaker expertise or inexperience—make fewer transactions.

Overall, our findings suggest that malicious issuers successfully use misrepresentations to screen for naïve investors. Wallets that hold tokens of misrepresented ICOs have characteristics associated with a lack of investor sophistication—they are less wealthy, less diversified, and less active. A caveat of our findings here is that a single person may control multiple wallets. However, it is unclear how this feature necessarily biases our findings.

4.6 Are misrepresentations unintentional mistakes?

The evidence in the previous section indicates that misrepresented ICOs are more likely to be scams. This finding is consistent with our main hypothesis that malicious issuers use misrepresentations to screen for naïve investors. Nevertheless, it is difficult to know the true motives behind misrepresentation behavior. An alternative explanation is that ICO misrepresentations could simply be unintentional mistakes. We design three sets of tests to address this explanation. First, we focus on the misrepresentation behavior of ICOs launched shortly after news of regulatory actions taken by U.S. authorities. Second, we examine the relation between misrepresentations and ICO quality. Third, we apply network analysis to assess systematic patterns of misrepresentation behavior in the ICO ecosystem.

4.6.1 Regulatory action and misrepresentations

We examine whether the threat of regulatory action deters the use of misrepresentations by malicious ICO issuers. To test the deterrence effect, we begin by collecting news of regulatory actions taken by the U.S. authorities. As Appendix A.I shows, these regulatory actions primarily involve ICO fraud and conflicts of interest. None of these actions targets inaccurate disclosures on listing websites. Under our screening hypothesis, the prospect of costly regulatory scrutiny should deter malicious issuers from using misrepresentations. Alternatively, under the unintentional-mistakes explanation, regulatory scrutiny should have no effect on misrepresentations.

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 news + \mathbf{X}^\top \boldsymbol{\beta} \tag{4.7}$$

$$\log\left(misrep\right) = \alpha + \beta_1 news + \mathbf{X}^\top \boldsymbol{\beta} \tag{4.8}$$

We construct two variables based on the timings of regulatory news releases and the first appearances of ICOs in our sample. First, the indicator variable $\mathbb{1}(\textit{regulatory action})$ switches on if regulatory news is released in the calendar month prior to the first appearance of the ICO in our sample. Second, *regulatory intensity* is the number of regulatory news articles released one month prior to the first appearance of the ICO in our sample. Subsequently, we test how these variables affect the use of ICO misrepresentations. We estimate logistic regressions in specification (4.7). The outcome variable in this specification is $\mathbb{1}(misrep > 0)$, an indicator that equals one if the ICO has at least one misrepresentation at its first appearance in our sample. The term p is the corresponding probability that $\mathbb{1}(misrep > 0)$ switches on. Because *misrep* is a strictly non-negative quantity, we also estimate Poisson regressions in specification (4.8). The vectors $\mathbf{X}$ and $\boldsymbol{\beta}$ represent vectors of control variables and their corresponding estimated coefficients, respectively.

[Insert Table 4.5 here]

Our results in Table 4.5 show that ICOs that shortly follow regulatory news have fewer misrepresentations. We estimate logistic (Poisson) regressions in columns 1 and 2 (3 and 4). where the dependent variable is $\mathbb{1}(misrep > 0)$ ($misrep$). Estimated coefficients in the first (last) two columns are expressed as odds (incidence rate) ratios. On the extensive margin, we find in column 1 that the odds of an ICO using misrepresentations is 46.0% ($t = 3.23$) lower following releases of regulatory news. On the intensive margin in column 2, we find that the release of an additional regulatory news article decreases the odds of a misrepresented ICO in the next month by 20.5% ($t = 2.13$). Our results in columns 3 and 4 corroborate the view that the threat of regulatory action deters misrepresentation behavior. In column 3, we find that ICOs launched after releases of regulatory news have 35.6% ($t = 3.90$) fewer *misrep*. Column 4 shows that, following the release of an additional news article, ICOs have 16.2% ($t = 2.91$) fewer misrepresentations.

Overall, we find that the threat of regulatory action is correlated with misrepresentation behavior. Thus, misrepresentations are unlikely to be unintentional mistakes. Rather, there likely are elements of malice and criminality in the use of misrepresentations. Remarkably, we find a link between regulatory news and misrepresentation behavior although our sample of news articles does not mention the latter. Our preferred interpretation is that the threat of regulatory scrutiny deters malicious issuers from the strategic use of ICO misrepresentations. Under our screening model framework, this effect amounts to the issuer adopting a more aggressive targeting strategy, which may hurt the profitability of the ICO scam.

An alternative interpretation of our empirical results is that malicious issuers merely delay the launches of their ICO scams. If malicious issuers tactically time their launches, we expect *misrep* to have a stronger predictive effect on ICO scam risk when the threat of regulatory scrutiny is weaker. To assess this interpretation, we estimate Cox regressions of ICO scams on the interaction terms $\mathbb{1}(\textit{regulatory action}) \times \textit{misrep}$ and $\textit{regulatory intensity} \times \textit{misrep}$.[22] We find that the loadings on these interaction terms are statistically insignificant. Thus, our findings in Table 4.5 are unlikely to reflect an issuer timing effect.

4.6.2 Misrepresentations and ICO quality

Misrepresentations may simply be unintentional mistakes. Suppose low quality issuers fail to exert the necessary effort to accurately market their offerings on listing websites. Then, to the extent that such issuers produce poorer blockchain projects, *misrep* should be negatively associated with ICO quality. High quality ICOs may choose higher levels of voluntary disclosure to signal their quality and separate themselves from low-quality ICOs (Bourveau et al., 2021). First, ICO issuers may voluntarily disclose the source code of their smart contracts on blockchain explorer services such as `Etherscan.io`. Second, issuers may also post on `Etherscan` the security audits of their source code.

To test whether misrepresentations merely reflect poor ICO/issuer quality, we examine the relation between *misrep* and the code disclosure practices of ICOs. To operationalize this test, we define the indicator $\mathbb{1}(\textit{code posted})$ to equal one if the ICO discloses its source code on `Etherscan.io` and equals zero otherwise. Likewise, the indicator $\mathbb{1}(\textit{code audited})$ switches on if the ICO posts a security audit of its source code on `Etherscan.io`. We estimate logistic regressions

[22]The Internet Appendix contains detailed results of these estimations.

following specification (4.9). The term p is the probability that $\mathbb{1}(\textit{code posted})$ (or, $\mathbb{1}(\textit{code audited})$) switches on. The vectors $\mathbf{X}$ and $\boldsymbol{\beta}$ represent vectors of control variables and their corresponding estimated coefficients, respectively. For ease of interpretation, we express estimated coefficients as odds ratios.

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 \, \textit{misrep} + \mathbf{X}^\top \boldsymbol{\beta} \tag{4.9}$$

Our results in Table 4.6 suggest that ICO quality is not significantly different between misrepresented and non-misrepresented ICOs. In column 1, we find that an additional *misrep* is associated with 1.6% ($t = 0.31$) lower odds of the ICO disclosing its code on `Etherscan.io`. This finding fails to support the idea that misrepresentations are a reflection of issuer quality and are unintentional mistakes. Column 2 shows a weak relation between *misrep* and odds of the ICO posting a security audit of its source code ($+1.1\%, t = 0.26$). Again, this pattern is inconsistent with the alternative story that misrepresentations point to lower ICO quality.

As a robustness check, we adopt a market-based measure of ICO quality in column 3. Suppose that the market places a lower value on low-quality blockchain projects. Then, under the alternative explanation, we should observe that misrepresented ICOs attract less funds. Because the amount of funds *raised* is a strictly non-negative quantity, we estimate a Poisson regression in column 3. Here, we find that the link between *misrep* and the amount of funds raised in the ICO campaign is statistically insignificant ($+5.8\%, t = 1.04$). This finding is also inconsistent with a quality-based explanation of ICO misrepresentations.

[Insert Table 4.6 here]

Overall, we find that measures of ICO quality do not significantly vary with

ICO misrepresentations. Thus, our findings reject the view that misrepresentations are merely unintentional mistakes, reflecting low issuer quality. Instead, our results thus far point to the strategic motives of issuers to target naïve investors with misrepresented ICO information.

4.6.3 Systematic patterns of misrepresentation behavior

To further substantiate our view that the use of misrepresentation is strategic, we apply network analysis to assess unusual patterns of this behavior among ICO issuers. If misrepresentations are intentionally and strategically deployed, they should leave systematic footprints throughout the ICO ecosystem. Specifically, we examine whether ICO advisers (henceforth, advisors) play a role in promoting misrepresentation behavior. Advisors are hired by ICO issuers to provide technical, marketing, and economic expertise. About 60% of ICOs in our sample hire an advisor. Advisors are also controversial—some have been convicted of illegal touting and tax evasion, while others have allegedly failed to perform basic due diligence on client ICOs.

Because advisors often work on multiple ICOs, they could play a role in promoting misrepresentation behavior. We hypothesize that misrepresentation behavior is correlated among ICOs that share common advisors. This correlation could arise from strategic complementarities that are typical in criminal behavior (e.g., Ballester, Calvó-Armengol, and Zenou, 2006). Complementarities in misrepresentation behavior can materialize in two ways. First, there is no formal way to learn the effective use of misrepresentations as a screening device. So, malicious issuers may have to learn from their peers via common advisors who convey know-how about the use of ICO misrepresentations. This learning channel implies that a malicious issuer's payoffs from misrepresentations are higher with technological transfers from other issuers of misrepresented ICOs. Second, misrepresentation behavior may be viewed as an acceptable norm among ICOs that share common advisors. An issuer who observes the use of misrepresentations by other issuers may infer that this behavior is commonplace, In response, the issuer is likely to use more misrepresentations, which symmetrically leads other issuers to the same inference and to do likewise.

We formalize the above hypothesis in a simple network model of misrepresentations with strategic complementarities. Appendix A.II contains details of this model. Our model predicts that—in a network of ICOs linked by common advisors—ICOs with higher Katz centrality in the network exhibit more *misrep*. We empirically test this prediction. To construct the ICO network, we manage to match 2,110 advisors with 2,271 ICOs using data extracted from the `ICOBench` listing website.[23] In this network, we link two ICOs if they share at least one common advisor. We present a circular layout of this network in Figure 4.6. ICOs are arranged according to their *misrep* on the circumference of the circle. As we move along the circumference in the clockwise direction, the ICOs have more *misrep*. Lines inside the circle represent links between ICOs. We observe that ICOs with more *misrep* tend to locate in regions with higher densities of links. Generally, such ICOs are also more central in the network.

[Insert Table 4.7 here]

To examine the relation between Katz centrality and *misrep* more rigorously, we estimate Poisson regressions in Table 4.7. Estimated coefficients are presented as incidence rate ratios. Consistent with our model predictions, column 1 shows that a 10% increase in Katz centrality is associated with a 4.6%

[23]This test has a smaller sample because we must exclude ICOs that either have no advisors or are unlinked to any ICOs.

(t = 2.27) rise in *misrep*.[24] Next, we conjecture that transmissions of misrepresentation behavior is stronger between two ICOs if they share more common advisors. Thus, we also construct a weighted ICO network, in which links are weighted by the number of common advisors. In column 2, we find a quantitatively similar effect using weighted links—a 10% increase in Katz centrality is associated with a 5.4% rise (t = 2.17) in *misrep*. In the next two columns, we use as our key independent variable an indicator $\mathbb{1}$(*high centrality*) that switches on if an ICO has an above-median Katz centrality. Columns 3 and 4 report that central ICOs have 6.1% (t = 1.96) and 6.7% (t = 2.25) higher *misrep* than peripheral ICOs, respectively.

Our empirical results in Table 4.7 support predictions from our network model—central ICOs have more misrepresentations. Due to strategic complementarities, we find systematic patterns of misrepresentation behavior among advisor-linked ICOs. In additional tests, we show that advisors of misrepresented ICOs are not penalized, but obtain more subsequent advisory opportunities. This surprising finding suggests that there exists many malicious issuers who solicit the services of such advisors. The Internet Appendix contains details and results of these tests. Overall, while advisors are valuable information and service intermediaries in the ICO market, some may play a role in the promotion of malignant behaviors.

4.7 Other suspicious actions

While malicious ICO issuers use misrepresentations to target naïve investors, such issuers may also engage in other suspicious actions to screen for investor sophistication. We collect data on two examples of such actions—celebrity

[24] We calculate this economic magnitude as follows: $\log(1.1) \times (1 - 1.485) = 0.046$.

endorsements and choice of listing websites—and test their predictive effects on ICO scam risk.

First, the U.S. SEC warns on an investor education website that celebrity endorsements of ICOs are prominent red flags of investment scams.[25] Celebrity endorsements may be a potent screening device because naïve investors are likely to act on financial advice offered on social media, particularly when it comes from famous individuals. To collect data on celebrity endorsements, we conduct web searches using combinations of keywords: "celebrity"/"promoter"/"influencer" and "ICO"/"initial coin offering"/"token". Next, we read all relevant search results and identify ICOs that are promoted by celebrities. To ensure completeness of our search efforts, we also search for the same combinations of keywords on the Factiva database. Our sample includes celebrities who span the entertainment, sports, business and media sectors.

Second, most ICOs are promoted on multiple, but not all, listing websites. We examine whether malicious issuers choose listing websites based on the characteristics of their web traffic. Using data from SEMrush—a web traffic analytics vendor—we measure the quantities of passive and active web traffic in each of the five listing websites. Specifically, passive web traffic counts visitors referred to a listing website via paid advertisements, third-party referral links, and search engines. Whereas, active web traffic counts visitors who access a listing website by directly typing its Uniform Resource Locator (URL) in browsers or through the use of saved browser bookmarks. Then, we define the *web traffic ratio* of an ICO as the ratio of passive traffic to active traffic, aggregated across the listing websites that list it in the month prior to its start date. We conjecture that active web traffic reflects a purposeful and targeted pattern of information acquisition, which is typical of more sophisticated investors.

[25]Source: `https://www.investor.gov/ico-howeycoins`

[Insert Table 4.8 here]

To test whether celebrity endorsements and strategic choices of listing websites predict ICO scams, we estimate Cox regressions in Table 4.8. We express estimated coefficients as hazard ratios. The key independent variable in column 1 is $\mathbb{1}(celebrity)$—an indicator that switches on if an ICO is endorsed by a celebrity. Here, we find that the scam risk of an ICO with a celebrity endorsement is more than 25 times ($t = 10.64$) that of an ICO without one. This finding supports the warning issued by the SEC that celebrity endorsements are red flags of investment scams. In column 2, we examine whether celebrity endorsements subsume the predictive effect of *misrep* on ICO scam risk. They do not. While $\mathbb{1}(celebrity)$ remains a strong predictor of ICO scam risk, we find that an additional *misrep* raises the odds of a scam by 14.5% ($t = 2.04$). This result suggests that misrepresentations and celebrity endorsements are distinct screening devices in the malicious issuer's repertoire. Because only a minority of ICOs are endorsed by celebrities, keeping a lookout for misrepresentations remains incrementally useful.

Column 3 shows that a unit increase in *web traffic ratio* is associated with a 26.5% ($t = 2.23$) higher odds of an ICO scam. This pattern suggests that malicious issuers strategically choose listing websites that receive a relatively larger share of passive web traffic. Through the lens of our theoretical framework in Section 4.3, this strategic choice has a similar effect to choosing an investor mass with a higher density z of naïve investors. In turn, a higher z increases the issuer's expected profits, ceteris paribus. In column 4, we find that *misrep* remains a positive and statistically significant predictor of ICO scam risk. Thus, misrepresentations have a screening effect incremental to that from the strategic choice of listing websites.

Overall, to complement their use of misrepresentations, malicious issuers may use other strategies to target naïve investors. We find that celebrity endorsements and the choice of listing websites are two such strategies. Nevertheless, misrepresentations a distinct predictive effect on ICO scam risk. To identify ICO scams, investors could use simple cross-site verification—alongside these red flags—to look for misrepresentations.

4.8 Partial observability of ICO scams

We account for the partial observability of ICO scams and discuss its econometric implications. Specifically, we face an inherent data limitation—our sample of ICO scams detected on the `DeadCoins` website may be incomplete. First, we discuss and address incomplete detection of ICO scams. Next, we estimate the proportion of ICOs that are scams, including those that go undetected. Finally, we discuss welfare effects from our findings.

4.8.1 Detection controlled estimation

To motivate our discussion, consider this scenario: (i) Unsophisticated ICO scams tend to have more misrepresentations, and (ii) such scams are more prone to detection on the `DeadCoins` website. Two econometric issues ensue. First, we may overestimate the effect of *misrep* on ICO scam risk because we cannot directly observe the sophistication of ICO scams. Second, we may underestimate the prevalence of ICO scams because we inadequately detect sophisticated scams. By reducing ICO scams, tighter regulations may improve investor welfare. However, these improvements must be balanced against the cost of regulations. Thus, the socially optimal level of regulations is a function of the prevalence of ICO scams, which we need to carefully assess.

To account for incomplete detection, we use detection controlled estimation (DCE) methods (Wang, Winton, and Yu, 2010; Comerton-Forde and Putniņš, 2014; Foley, Karlsen, and Putniņš, 2019). In our DCE model, we simultaneously estimate a system of two equations: one models ICO scams, while the other models detection conditional on the occurrence of ICO scams. Thereafter, we estimate the DCE model using the maximum likelihood method. The Internet Appendix contains full details of our DCE model and a derivation of its likelihood function.

To identify our DCE model, we require instrumental variables that are uniquely associated with either the scam or detection stage. In selecting our instruments, we hypothesize that malicious issuers opportunistically perform ICOs during periods of strong sentiment in cryptocurrency markets to capture more funds. Operationally, we measure market sentiment with *BTC returns* (*BTC search*), which is the cumulative returns of Bitcoin (cumulative Google Trends search volume index of the word "Bitcoin") in the 30 days prior to ICO start dates.

Both instruments are arguably unassociated with detection probabilities for three reasons. First, to the extent that detection is idiosyncratic (i.e., ICO-specific), our Bitcoin-based measure of marketwide sentiment should be orthogonal to detection probabilities. Second, if ICO scams were primarily detected on the basis of our sentiment-timing mechanism, then we should expect detection to be quick. However, we find that several months elapse between the end date of the average ICO scam and its subsequent detection on the `DeadCoins` website. Third, we manually verify that reasons behind scam allegations on the `DeadCoins` website do not allude to sentiment-timing.

[Insert Table 4.9 here]

Table 4.9 reports estimates from our DCE models. Estimated coefficients are expressed as odds ratios. The first two columns belong to Model A, which uses *BTC search* and *BTC returns* as instruments in the scam stage. We find in column 1 that one standard deviation increases in *BTC search* and *BTC returns* raise the odds of ICO scams by 62.8% ($t = 4.74$) and 41.9% ($t = 4.63$), respectively.[26] This pattern supports our idea that malicious issuers opportunistically time their ICOs to ride on periods of strong sentiment in cryptocurrency markets. Crucially, misrepresentations continues to predict ICO scams—an additional *misrep* increases the odds of ICO scams by 11.3% ($t = 6.16$). Column 2 shows that an ICO scam with more *misrep* is more likely to be detected, suggesting that misrepresentations also draw scrutiny from market participants.

As a robustness check, we set up Model B, which uses *altcoin search* (i.e., Google Trends search volume index for the word "ICO") and *altcoin returns* as instruments in the scam stage. These instruments are constructed similarly to our Bitcoin-based instruments, but are based on alternative coins—all cryptocurrencies excluding Bitcoin. Using *altcoin search* and *altcoin returns* as instruments, our results in columns 3 and 4 are also consistent with our prior conclusions. ICOs coinciding with stronger sentiment in the alt-coin market are subsequently more likely to be scams.[27] In addition, we continue to find that misrepresented ICOs are more likely to be scams and detected as such.

4.8.2 Welfare analysis of ICO scams

Using estimates from our DCE models, we fit the models in columns 1 and 3 of Table 4.9 to probabilistically identify ICO scams. To obtain an empirical distribution of the proportion of probable scams, we perform a stratified bootstrap (`DeadCoins` sample vs. all other ICOs) over 500 iterations. In every iteration, we re-estimate our DCE models and re-compute the proportion of probable scams.

[26] We calculate economic magnitudes in column 1 as follows. $\sigma(\textit{BTC search}) = 20.95$; $20.95 \times (1.030 - 1) = 0.6285$. $\sigma(\textit{BTC returns}) = 29.4\%$; $29.4\% \times (2.428 - 1) = 41.98\%$.

[27] We calculate economic magnitudes in column 3 as follows. $\sigma(\textit{altcoin search}) = 20.93$; $20.93 \times (1.023 - 1) = 0.4814$. $\sigma(\textit{altcoin returns}) = 82.7\%$; $82.7\% \times (1.362 - 1) = 29.94\%$.

Model A and Model B in Table 4.9 estimate that 38.6% ($\hat{\sigma} = 29.0\%$) and 40.4% ($\hat{\sigma} = 26.8\%$) of ICOs in our sample are scams, respectively. Thus, there are potentially many ICO scams that are undetected. For additional context, the ICO advisory firm Satis Group estimates in an industry report that 78% of ICOs are scams (Dowlat, 2018).[28]

We discuss welfare considerations from our empirical exercise. Should policymakers be concerned about harm to ICO investors? This is an important question, to which there is no obvious answer. On one hand, the potential financial losses to ICO investors are substantial based on a back-of-envelope calculation. On average, an ICO raises U.S. \$5.07 million in our sample. Suppose 40% of the 5,935 ICOs are scams. Then, ICO investors may be facing a loss of U.S. \$5.07 million $\times$ 0.4 $\times$ 5,935 = U.S. \$12.03 billion. Thus, given the prevalence of ICO scams, more stringent regulations and enforcement—although costly—may be justified to protect investors. Specifically, because misrepresentations remain a powerful predictor of ICO scams, educating investors to perform simple cross-site verification of ICO characteristics may yield large benefits.

On the other hand, individuals may view risky ICO investments and traditional gambling devices in the same light.[29] For example, the U.S. Census Bureau reports that state-administered lottery funds alone generated U.S. \$76.4 billion in sales in 2018. To the extent that the average skewness-loving individual substitutes between ICO investments and traditional gambling devices, the net welfare loss to her from ICO scams would be smaller. From this perspective, more choices of gambling devices offered by the multitude of ICOs on the market may even increase individual welfare.

Overall, while our paper is agnostic on the net welfare effects, our estimated scale of ICO scams and its associated financial impact may inform cost-benefit

[28]The Satis Group report uses a smaller and earlier sample and a different definition of ICO scams.

[29]Anecdotal evidence from social media, such as the Reddit forums, supports this consideration.

tradeoffs of future regulatory policies. Specifically, given the role of misrepresentations in ICO scams, investments in regulatory scrutiny of ICO listing websites and in investor education could be particularly beneficial.

4.9 Conclusions

In this paper, we analyze how malicious issuers target naïve investors in ICO scams. Using point-in-time snapshots of data extracted from ICO listing websites, we find widespread cross-site discrepancies in ICO characteristics. The results suggest that malicious ICO issuers strategically use cross-site misrepresentations to screen for naïve investors. Astute investors conduct due diligence and immediately dismiss the ICO scam. However, naïve investors overlook these misrepresentations, fall for the scam, and eventually fund the ICO. Ultimately, the investors who remain are likely to be naïve—the ideal targets of the malicious issuer. Our evidence indicates that the use of misrepresentations is nefarious—an additional misrepresentation raises the hazard of ICO scams by 14.0%. This effect is concentrated in the misrepresentations of basic ICO characteristics that are fundamental to investors' due diligence. Using wallet information from the Ethereum blockchain, we find that cryptocurrency wallets holding tokens of misrepresented ICOs (i) have less total values, (ii) are less diversified, and (iii) are less active. These patterns support our view that malicious issuers (successfully) use misrepresentations to screen for naïve or unsophisticated investors.

We find that ICO misrepresentations are unlikely to be unintentional mistakes. First, the threat of regulatory scrutiny deters the use of misrepresentations. This finding implies that there are likely to be elements of malice and criminality in the use of misrepresentations. Second, misrepresented ICOs and their non-misrepresented counterparts do not have significantly different disclosure practices and fundraising outcomes. To the extent that issuer quality is positively correlated with these proxies, our findings are inconsistent with a quality-based explanation. Third, we use network analysis to show that misrepresentation behavior is likely to be deliberate in the ICO ecosystem. We present a simple network model that captures complementarities (e.g., learning and social norms) in misrepresentation behavior. Due to complementarities facilitated by advisors, the model predicts that ICOs with higher Katz centrality use more misrepresentations. Our empirical results support this prediction. Furthermore, we find that advisors of misrepresented ICOs are more likely to obtain subsequent advisory opportunities. Thus, culpable advisors, instead of being penalized, can continue to promote malignant behaviors in the ICO ecosystem.

A welfare analysis of the financial losses from ICO scams in our sample shows that around 40% of ICOs are potentially scams, but most go undetected. Based on this estimate, the financial losses to ICO investors due to ICO scams could exceed U.S. $12 billion. Against the backdrop of these estimates, more stringent regulations and stronger enforcement actions may be justified to protect investor welfare. We believe that increased regulatory scrutiny of ICO listing websites may be particularly beneficial. Regulators can also educate the general public on how these frauds are conducted by bringing attention to red flags such as misrepresentations. Even in an environment with limited regulations and investor protection, simple and low-cost due diligence can help investors avoid scams. Specific to our setting, our analysis also highlights two important issues hindering the adoption of ICOs as a financing vehicle—(i) unreliability of self-reported ICO information and (ii) widespread scams.

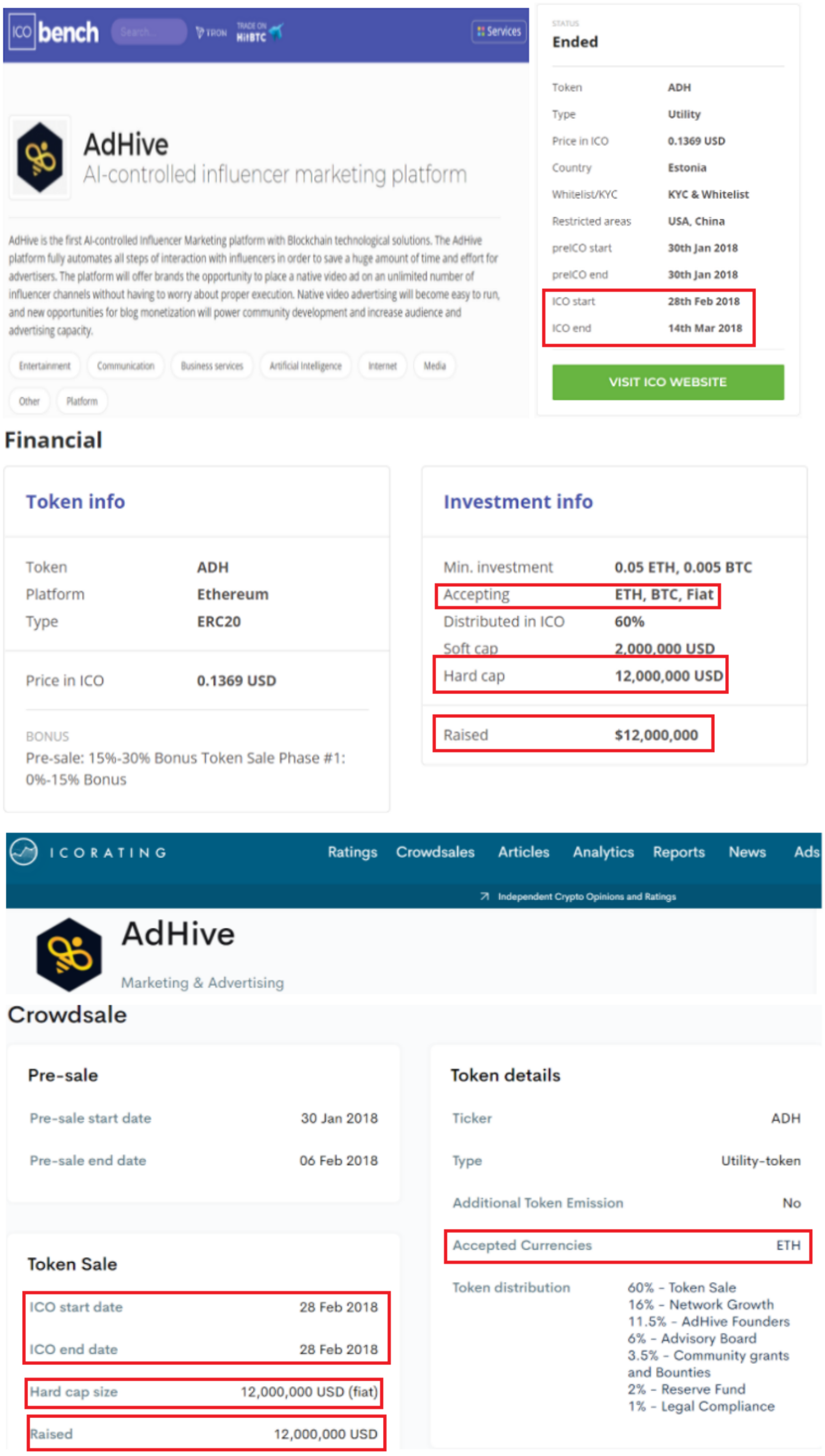


FIGURE 4.1: This figure presents screenshots of the AdHive ICO information pages on three ICO listing websites—`ICOBench.com`, `ICORating.com`, and `ICODrops.com`.

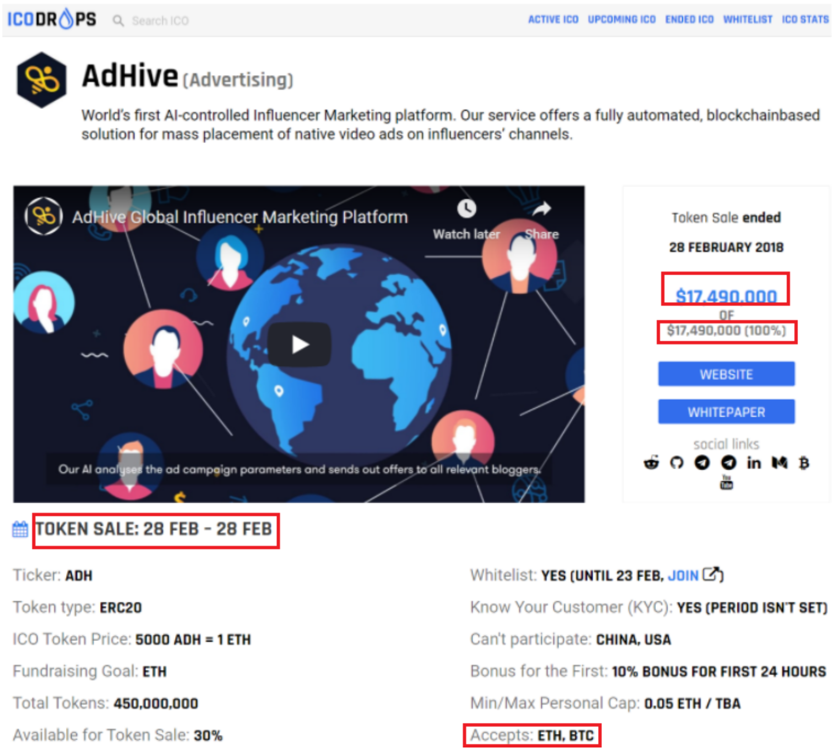
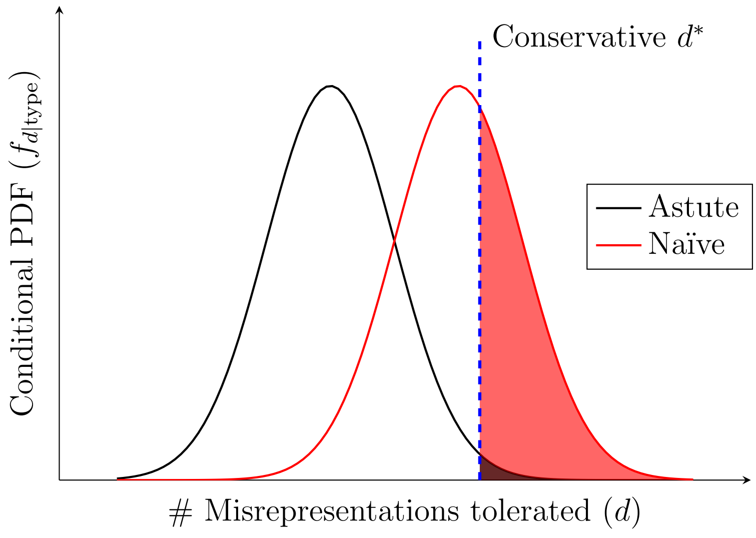


FIGURE 4.1: (continued)

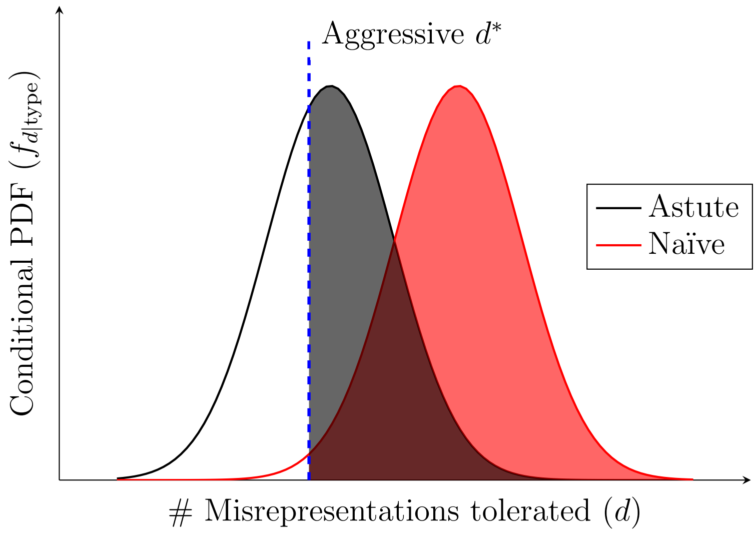
FIGURE 4.2: This figure visualizes the three periods of the model described in Section 4.3. The ICO launches in Period (1), and the issuer selects d^* in the targeting strategy. Some naïve and astute investors immediately dismiss the ICO. The remaining investors are targeted. In Period (2), these targeted investors impose costs on the issuer by seeking additional information and asking questions on public forums. In Period (3), only naïve investors proceed to fund the completed ICO scam.

Astute investors ultimately refrain from funding the scam.

Issuer selects d^* in targeting strategy	Targeted investors impose costs	Gross funding proceeds
• Targeted investors $mz \cdot \bar{F}_{d\mid\text{type}}(d^* \mid \text{naïve})$ $m(1-z) \cdot \bar{F}_{d\mid\text{type}}(d^* \mid \text{astute})$ • Dismissed investors $mz \cdot F_{d\mid\text{type}}(d^* \mid \text{naïve})$ $m(1-z) \cdot F_{d\mid\text{type}}(d^* \mid \text{astute})$	$mz \cdot \bar{F}_{d\mid\text{type}}(d^* \mid \text{naïve}) \times C$ $m(1-z) \cdot \bar{F}_{d\mid\text{type}}(d^* \mid \text{astute}) \times C$	$mz \cdot \bar{F}_{d\mid\text{type}}(d^* \mid \text{naïve}) \times Q$
Period (1): ICO launches	Period (2): ICO in progress	Period (3): ICO completes



(A) Conservative targeting strategy



(B) Aggressive targeting strategy

FIGURE 4.3: This figure presents probability density plots of d, conditional on two investor types—astute (black) and naïve (red). Shaded areas in black and red represent the complementary conditional cumulative distributions $\bar{F}_{d|\text{type}}(d^* \mid \text{astute})$ and $\bar{F}_{d|\text{type}}(d^* \mid \text{naïve})$, respectively. Subfigures 4.3a and 4.3b visualize a conservative targeting strategy (high d^*) and an aggressive targeting strategy (low d^*), respectively.

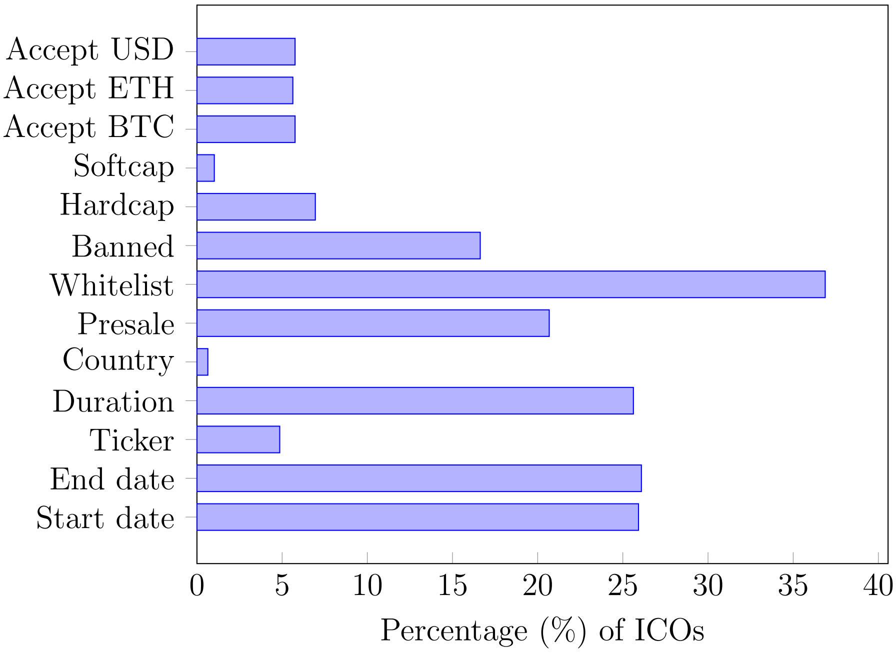


FIGURE 4.4: This figure presents the proportion of ICOs with at least one cross-website discrepancy in a particular characteristic at first appearances in our sample.

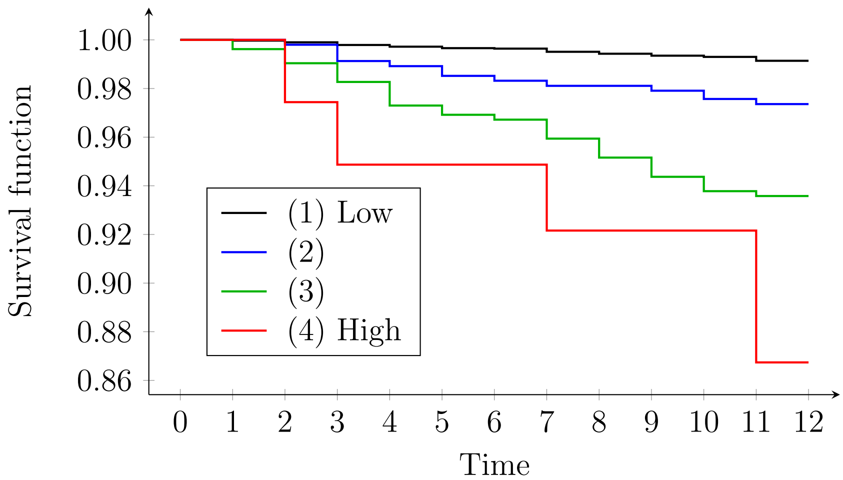


FIGURE 4.5: This figure presents the survival functions of ICOs in our sample. We assign every ICO into one of four groups based on its number of cross-website discrepancies in its characteristics at its first appearance in our sample (*misrep*). The x-axis is the time-to-event—months elapsed from the time of entry into our sample. The y-axis is the groupwise proportion of ICOs that are not identified as scams on `DeadCoin.com` (i.e., survive) at a given time.

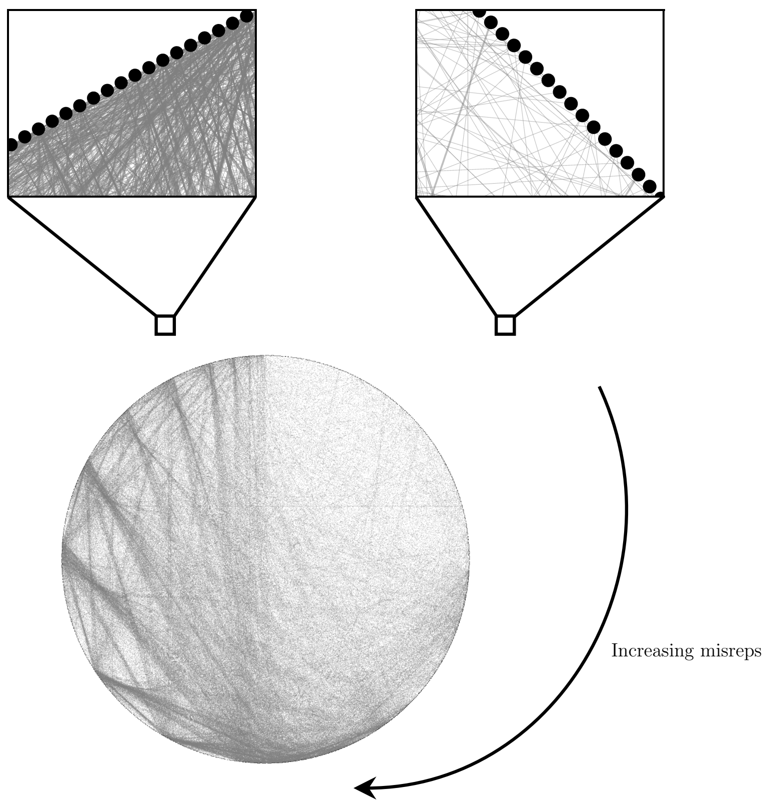


FIGURE 4.6: This figure presents a circular layout of the advisor-linked ICO network described in Section 4.6.3. The ICOs are arranged according to their *misrep* on the circumference of the circle. The ICO at the 12 o'clock position has the fewest *misrep*. As we move along the circumference in the clockwise direction, the ICOs have more *misrep*. Lines inside the circle represent network links between ICOs.

Table 4.1: Descriptive statistics

This table presents descriptive statistics of our sample at the ICO level. The variables presented in this table are extracted from the first appearance of each ICO in our 13-month observation window. Panel A reports the summary statistics of ICO characteristics and the misrepresentation measures. Panel B presents Pearson pairwise correlations between variables. Section 4.4.2 contains definitions of variables presented in this table.

Panel A. Summary statistics

	N	μ	σ	p10	p50	p90
Misrep	5,960	1.26	2.16	0	0	4
$\mathbb{1}_{\text{Misrep}>0}$	5,960	0.34	0.48	0	0	1
Banned	5,960	0.95	0.22	1	1	1
Whitelist	5,960	0.55	0.50	0	1	1
Presale	5,960	0.47	0.50	0	0	1
Hardcap	5,960	0.70	0.46	0	1	1
Softcap	5,960	0.26	0.44	0	0	1
Accept BTC	5,960	0.28	0.45	0	0	1
Accept ETH	5,960	0.58	0.49	0	1	1
Accept USD	5,960	0.10	0.30	0	0	0
SEC filing (%)	5,960	0.89	9.38	0	0	0
Enforcement	5,960	0.26	0.42	0	0	1
Disclosure	5,960	1.20	1.23	0	0.73	2.92
Duration (days)	5,960	54.38	50.25	15	37	109

Panel B. Pairwise correlations

		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
Misrep	(a)												
Banned	(b)	−0.01											
Whitelist	(c)	−0.07	0.10										
Duration	(d)	−0.12	−0.04	−0.03									
Presale	(e)	0.31	−0.01	0.17	−0.06								
Hardcap	(f)	0.28	0.02	−0.20	−0.06	0.12							
Softcap	(g)	0.03	−0.04	0.06	0.10	0.16	0.36						
Accept BTC	(h)	0.16	0.00	0.08	0.06	0.24	0.09	0.18					
Accept ETH	(i)	0.31	0.01	0.14	−0.01	0.43	0.17	0.16	0.44				
Accept USD	(j)	0.05	0.00	0.07	0.06	0.15	0.07	0.13	0.38	0.23			
SEC filing	(k)	0.04	0.00	0.02	0.01	0.04	0.02	0.01	0.04	0.03	0.05		
Enforcement	(l)	0.11	−0.02	−0.04	−0.01	0.05	0.05	0.07	−0.01	0.02	0.02	−0.03	
Disclosure	(m)	0.13	−0.11	−0.04	0.02	0.04	0.02	0.08	−0.03	−0.01	0.01	0.06	0.31

Table 4.2: Differences in means

This table presents differences in ICO scam rates and characteristics between misrepresented ICOs and non-misrepresented ICOs. Column (1) contains ICOs with at least one misrepresentation. Column (2) contains ICOs with no misrepresentations. We report differences in means (Δ) and their associated t-statistics. Section 4.4.2 contains definitions of variables presented in this table.

	(1)	(2)	$\Delta_{(1)-(2)}$	t
ICO scam	0.04	0.01	0.03	6.88
Banned	0.95	0.95	−0.01	0.90
Whitelist	0.46	0.60	−0.15	10.96
Presale	0.68	0.36	0.32	25.15
Hardcap	0.89	0.60	0.29	27.58
Softcap	0.29	0.25	0.04	3.16
Accept BTC	0.39	0.22	0.16	12.99
Accept ETH	0.80	0.46	0.34	28.82
Accept USD	0.12	0.09	0.04	4.21
SEC filing (%)	1.21	0.72	0.49	1.79
Duration (days)	47.71	57.91	−10.20	8.29
Enforcement	0.33	0.22	0.11	9.52
Disclosure	1.44	1.07	0.37	11.11

(1): ICOs with at least one misrepresentation
(2): ICOs with no misrepresentations

Table 4.3: Misrepresentations and ICO scams

This table presents estimates from Cox regressions. Estimated coefficients are expressed as hazard ratios. The failure event in these regressions is *ICO scam*. An ICO triggers the event if the `DeadCoin` site identifies it as a scam. Otherwise, it is right-censored. The key independent variables in our regressions are *misrep*, $\mathbb{1}(misrep > 0)$, and $misrep^{basic}$. The *misrep* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The indicator $\mathbb{1}(misrep > 0)$ equals one if the ICO has at least one *misrep*, and equals zero otherwise. The $misrep^{basic}$ of an ICO is the number of cross-site discrepancies of its basic characteristics at its first appearance in our sample. Section 4.4.2 contains variable definitions. Some models contain coverage-quartile fixed effects and are stratified by ICO cohorts. Standard errors in some models are clustered by ICO cohorts. t-statistics are reported in parentheses.

Event: ICO scam

	(1)	(2)	(3)	(4)
$\mathbb{1}$(Misrep > 0)	3.740			
	(5.46)			
Misrep		1.253	1.140	
		(6.71)	(2.18)	
Misrep^{basic}				1.240
				(4.86)
Banned	0.992	0.984	1.015	1.015
	(0.02)	(0.04)	(0.04)	(0.04)
Whitelist	1.439	1.196	1.402	1.470
	(1.71)	(0.85)	(1.47)	(1.85)
Duration	0.999	1.000	1.000	1.000
	(0.45)	(0.18)	(0.08)	(0.00)
Presale	0.951	0.881	0.967	1.020
	(0.22)	(0.54)	(0.21)	(0.15)
Hardcap	1.709	1.653	1.619	1.625
	(1.76)	(1.62)	(1.72)	(1.93)
Softcap	0.873	0.879	0.985	0.951
	(0.61)	(0.58)	(0.12)	(0.35)
Accept BTC	1.355	1.331	1.291	1.210
	(1.36)	(1.27)	(1.14)	(0.84)
Accept ETH	1.024	1.081	1.159	1.066
	(0.10)	(0.30)	(0.61)	(0.26)
Accept USD	1.224	1.238	1.287	1.316
	(0.68)	(0.72)	(0.80)	(0.82)
Enforcement	0.635	0.643	0.625	0.603
	(1.77)	(1.72)	(1.95)	(2.11)
Disclosure	0.934	0.939	0.922	0.907
	(0.83)	(0.77)	(1.29)	(1.59)
SEC filing	0.674	0.587	0.559	0.552
	(0.39)	(0.53)	(0.78)	(0.76)
# ICOs	5,935	5,935	5,935	5,935
Cohort strata	N	N	Y	Y
Coverage-quartile FE	N	N	Y	Y
Clustered SE	N	N	Y	Y

Table 4.4: On-chain analysis: Misrepresentations and wallet characteristics

This table presents estimates from Poisson regressions. Estimated coefficients are expressed as incidence rate ratios. For every ICO, we first analyze individual cryptocurrency wallets that hold its tokens at 10 days after the ICO end date. Next, we compute wallet characteristics by extracting data from the Ethereum blockchain. Finally, we aggregate wallet-level measures at the ICO level by taking medians. The dependent variables *value* (column 1), *diversity* (column 2), and *activity* (column 3). The *value* of an ICO is the median portfolio value (in U.S. dollars) of wallets that hold its tokens. The *diversity* of an ICO is the median number of distinct tokens held in wallets that hold its tokens. The *activity* of an ICO is the median number of blockchain transactions performed by wallets that hold its tokens. The key independent variable in our regressions is $\mathbb{1}(misrep > 0)$. The *misrep* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The indicator $\mathbb{1}(misrep > 0)$ equals one if the ICO has at least one *misrep*, and equals zero otherwise. Section 4.4.2 contains variable definitions. Models contain ICO cohort fixed effects. Standard errors are clustered by ICO cohorts. t-statistics are reported in parentheses.

	(1)	(2)	(3)
Dependent variable:	Value	Diversity	Activity
$\mathbb{1}$(Misrep > 0)	0.399	0.803	0.910
	(2.61)	(2.88)	(2.62)
Banned	14.899	1.093	0.863
	(2.78)	(0.57)	(2.51)
Whitelist	1.080	0.995	1.076
	(0.23)	(0.04)	(1.37)
Duration	0.992	0.998	0.998
	(2.38)	(1.93)	(5.49)
Presale	0.602	0.812	0.946
	(0.80)	(1.27)	(0.85)
Hardcap	2.019	0.860	1.001
	(1.90)	(1.82)	(0.02)
Softcap	1.233	1.042	0.965
	(0.57)	(0.28)	(0.92)
Accept BTC	1.802	1.012	0.917
	(0.89)	(0.14)	(1.57)
Accept ETH	1.605	0.982	0.951
	(1.26)	(0.10)	(0.90)
Accept USD	0.325	0.802	0.793
	(0.91)	(0.73)	(2.22)
Enforcement	1.010	1.031	0.999
	(0.03)	(0.23)	(0.01)
Disclosure	1.032	0.961	0.973
	(0.22)	(1.06)	(2.34)
SEC filing	0.000	0.666	0.962
	(12.23)	(1.39)	(0.17)
# ICOs	1,996	1,996	1,996
Cohort FE	Y	Y	Y
Clustered SE	Y	Y	Y

Table 4.5: Regulatory action and misrepresentations

Columns 1 and 2 (3 and 4) of this table present estimates from logistic (Poisson) regressions. Estimated coefficients in columns 1 and 2 (3 and 4) are expressed as odds (incidence rate) ratios. The dependent variable in columns 1 and 2 is $\mathbb{1}(misrep > 0)$—an indicator that equals one if the ICO has at least one cross-site discrepancies of its characteristics at its first appearance in our sample, and equals zero otherwise. The dependent variable in columns 3 and 4 is *misrep*. The misrep of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The key independent variables are $\mathbb{1}(regulatory\ action)$ and *regulatory intensity*. The variable $\mathbb{1}(regulatory\ action)$ is an indicator that equals one if regulatory news is released within the calendar month prior to the first appearance of the ICO in our sample, and equals zero otherwise. The variable *regulatory intensity* is the number of regulatory news articles released within the calendar month prior to the first appearance of the ICO in our sample. Section 4.4.2 contains variable definitions. Models contain ICO cohort fixed effects. Standard errors are clustered by ICO cohorts. t-statistics are reported in parentheses.

Dependent variable: Misrep

	(1)	(2)	(3)	(4)
Dependent variable:	$\mathbb{1}$(Misrep > 0)		Misrep	
$\mathbb{1}$(Regulatory action)	0.540		0.644	
	(3.23)		(3.90)	
Regulatory intensity		0.795		0.838
		(2.13)		(2.91)
Banned	0.763	0.772	0.921	0.926
	(1.53)	(1.41)	(1.72)	(1.58)
Whitelist	0.509	0.506	0.940	0.938
	(4.42)	(4.47)	(1.49)	(1.53)
Duration	0.998	0.998	0.998	0.998
	(2.29)	(2.25)	(3.30)	(3.30)
Presale	4.270	4.277	2.425	2.432
	(8.71)	(8.71)	(8.99)	(9.03)
Hardcap	4.471	4.479	3.239	3.253
	(10.13)	(10.19)	(22.39)	(22.11)
Softcap	0.811	0.818	0.990	0.993
	(1.85)	(1.78)	(0.30)	(0.21)
Accept BTC	1.268	1.270	1.140	1.141
	(2.35)	(2.29)	(3.48)	(3.45)
Accept ETH	4.879	4.867	2.444	2.447
	(6.30)	(6.29)	(8.89)	(8.83)
Accept USD	0.881	0.882	0.996	0.997
	(1.23)	(1.23)	(0.11)	(0.10)
Enforcement	1.541	1.538	1.161	1.161
	(3.29)	(3.26)	(4.26)	(4.26)
Disclosure	1.373	1.373	1.146	1.147
	(4.21)	(4.19)	(6.73)	(6.76)
SEC filing	1.000	1.014	0.990	0.995
	(0.00)	(0.04)	(0.08)	(0.04)
# ICOs	5,935	5,935	5,935	5,935
Cohort FE	Y	Y	Y	Y
Clustered SE	Y	Y	Y	Y

Table 4.6: Misrepresentations and ICO quality

This table present estimates from logit (columns 1 and 2) and Poisson (column 3) regressions. Estimated coefficients are expressed as odds ratios (incidence rate ratios) in columns 1 and 2 (column 3). The dependent variables are $\mathbb{1}(\textit{code posted})$, $\mathbb{1}(\textit{code audited})$, and *raised*. The indicator $\mathbb{1}(\textit{code posted})$ equals one if the ICO posts the source code of its smart contract on `Etherscan.io` and equals zero otherwise. The indicator $\mathbb{1}(\textit{code audited})$ equals one if the ICO posts a security audit of its source code on `Etherscan.io` and equals zero otherwise. The variable *raised* is the amount of capital (in U.S. dollars) raised by the ICO. The key independent variables in our regressions is *misrep*. The *misrep* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. Section 4.4.2 contains variable definitions. Models contain cohort fixed effects. The sample sizes here are smaller than those in Table 4.3 because of data limitations. Standard errors are clustered by ICO cohorts. t-statistics are reported in parentheses.

	(1)	(2)	(3)
Dependent variable:	$\mathbb{1}$(Code posted)	$\mathbb{1}$(Code audited)	Raised
Misrep	0.984	1.011	1.058
	(0.31)	(0.26)	(1.04)
Banned	1.419	0.940	0.948
	(0.74)	(0.19)	(0.25)
Whitelist	0.942	0.953	2.300
	(0.48)	(0.17)	(4.72)
Duration	0.998	0.996	1.004
	(1.07)	(1.45)	(0.56)
Presale	0.988	0.790	0.748
	(0.08)	(0.82)	(1.53)
Hardcap	1.313	1.664	0.891
	(2.64)	(2.53)	(0.35)
Softcap	0.853	0.865	0.800
	(1.50)	(0.71)	(1.20)
Accept BTC	1.200	1.312	0.816
	(0.89)	(1.30)	(0.85)
Accept ETH	1.035	1.265	1.625
	(0.29)	(0.92)	(1.66)
Accept USD	0.811	0.969	1.594
	(0.78)	(0.10)	(1.32)
Enforcement	1.062	0.847	0.734
	(0.55)	(0.76)	(2.52)
Disclosure	1.110	1.130	0.980
	(2.47)	(1.77)	(0.24)
SEC filing	0.299	1.00	1.182
	(1.40)	(0.00)	(0.53)
# ICOs	4,604	4,604	2,985
Cohort FE	Y	Y	Y
Clustered SE	Y	Y	Y

Table 4.7: Central ICOs and misrepresentations

This table presents estimates from Poisson regressions. Estimated coefficients are expressed as incidence rate ratios. The dependent variable is *misrep*. The *misrep* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The key independent variables are $\log(\textit{centrality})$ and $\mathbb{1}(\textit{high centrality})$. The variable $\log(\textit{centrality})$ is the log-transformed Katz centrality of the ICO. See Appendix A.II for details on the Katz centrality measure. The variable $\mathbb{1}(\textit{high centrality})$ is an indicator that equals one if the ICO has a higher Katz centrality than the median Katz centrality in the sample, and equals zero otherwise. See Appendix A.II for details on Katz centrality. Section 4.4.2 contains variable definitions. Models contain cohort fixed effects. Standard errors are clustered by ICO cohorts. t-statistics are reported in parentheses.

Dependent variable: Misrep

	(1)	(2)	(3)	(4)
Weighted links	N	Y	N	Y
log (Centrality)	1.485	1.567		
	(2.27)	(2.17)		
$\mathbb{1}$(High centrality)			1.061	1.067
			(1.96)	(2.25)
Banned	0.974	0.974	0.974	0.974
	(0.48)	(0.47)	(0.45)	(0.46)
Whitelist	1.134	1.134	1.133	1.133
	(1.85)	(1.85)	(1.82)	(1.82)
Duration	0.999	0.999	0.999	0.999
	(1.56)	(1.56)	(1.56)	(1.57)
Presale	1.590	1.591	1.588	1.587
	(7.47)	(7.49)	(7.60)	(7.62)
Hardcap	1.598	1.599	1.596	1.597
	(6.75)	(6.77)	(6.98)	(6.90)
Softcap	0.996	0.996	0.996	0.997
	(0.29)	(0.26)	(0.30)	(0.22)
Accept BTC	1.065	1.065	1.067	1.067
	(1.31)	(1.31)	(1.32)	(1.35)
Accept ETH	1.249	1.249	1.245	1.243
	(2.31)	(2.31)	(2.25)	(2.26)
Accept USD	1.033	1.034	1.036	1.035
	(0.76)	(0.77)	(0.81)	(0.79)
Enforcement	1.023	1.022	1.023	1.025
	(0.73)	(0.72)	(0.74)	(0.76)
Disclosure	1.001	1.001	1.000	1.000
	(0.08)	(0.08)	(0.02)	(0.02)
SEC filing	0.947	0.946	0.942	0.944
	(0.62)	(0.62)	(0.66)	(0.63)
# ICOs	2,271	2,271	2,271	2,271
Cohort FE	Y	Y	Y	Y
Clustered SE	Y	Y	Y	Y

Table 4.8: Other suspicious actions

This table presents estimates from Cox regressions. Estimated coefficients are expressed as hazard ratios. The failure event in these regressions is *ICO scam*. An ICO triggers the event if the `DeadCoin` site identifies it as a scam. Otherwise, it is right-censored. The key independent variables in our regressions are $\mathbb{1}(celebrity)$, *web traffic ratio*, and *misrep*. The indicator $\mathbb{1}(celebrity)$ equals one if an ICO is endorsed by a celebrity, and equals zero otherwise. To compute *web traffic ratio* of an ICO, we first classify web traffic to listing websites into two categories—passive and active. Passive web traffic counts visitors referred to a listing website via third-party referral links, paid advertisements, and search engines. Active web traffic counts visitors who access a listing website by directly typing its Uniform Resource Locator (URL) or through the use of saved browser bookmarks. Next, we define the *web traffic ratio* of an ICO as the ratio of passive traffic to active traffic, aggregated across the listing websites that list it in the month prior to its start date. The *misrep* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. Section 4.4.2 contains variable definitions. Models contain coverage-quartile fixed effects and are stratified by ICO cohorts. Standard errors in some models are clustered by ICO cohorts. *t*-statistics are reported in parentheses.

Event: ICO scam

	(1)	(2)	(3)	(4)
$\mathbb{1}$(Celebrity)	25.780	27.027		
	(10.64)	(9.37)		
Web traffic ratio			1.265	1.254
			(2.23)	(2.07)
Misrep		1.145		1.136
		(2.04)		(2.12)
Banned	1.062	1.058	1.059	1.048
	(0.18)	(0.16)	(0.18)	(0.14)
Whitelist	1.497	1.374	1.507	1.423
	(2.11)	(1.55)	(1.87)	(1.54)
Duration	0.999	0.999	1.000	1.000
	(0.21)	(0.23)	(0.02)	(0.01)
Presale	1.205	1.042	1.096	0.946
	(1.65)	(0.26)	(0.74)	(0.37)
Hardcap	1.758	1.574	1.820	1.651
	(2.77)	(1.88)	(2.67)	(1.85)
Softcap	0.917	0.919	0.947	0.940
	(0.48)	(0.48)	(0.39)	(0.45)
Accept BTC	1.376	1.361	1.339	1.303
	(1.38)	(1.38)	(1.30)	(1.17)
Accept ETH	1.110	1.034	1.383	1.286
	(0.38)	(0.14)	(1.14)	(0.98)
Accept USD	1.321	1.321	1.308	1.298
	(0.89)	(0.91)	(0.87)	(0.84)
Enforcement	0.592	0.582	0.582	0.586
	(2.20)	(2.18)	(2.13)	(2.09)
Disclosure	0.927	0.919	0.898	0.892
	(1.08)	(1.25)	(1.64)	(1.77)
SEC filing	0.591	0.602	0.550	0.560
	(0.81)	(0.67)	(0.93)	(0.77)
# ICOs	5,935	5,935	5,935	5,935
Cohort strata	Y	Y	Y	Y
Coverage-quartile FE	Y	Y	Y	Y
Clustered SE	Y	Y	Y	Y

Table 4.9: Partial observability of ICO scams

This table presents estimates from detection controlled estimation (DCE) models. Estimated coefficients are expressed as odds ratios. We simultaneously model the scam and detection processes of ICO scams. The instruments for the scam process in Model A (Model B) are *BTC search* and *BTC returns* (*altcoin search* and *altcoin returns*). The variable *BTC search* (*altcoin search*) is the cumulative search volume index of the word "Bitcoin" ("ICO") on Google Trends 30 days prior to the ICO start date. The variable *BTC returns* (*altcoin returns*) is the cumulative returns of Bitcoin (non-Bitcoin cryptocurrencies) 30 days prior to the ICO start date. Section 4.4.2 contains variable definitions. t-statistics are reported in parentheses.

Detection controlled estimation (DCE)

	(1)	(2)	(3)	(4)
	Model A		Model B	
	Scam	Detection	Scam	Detection
BTC search	1.030			
	(4.74)			
BTC returns	2.428			
	(4.63)			
Altcoin search			1.023	
			(5.20)	
Altcoin returns			1.362	
			(5.06)	
Misrep	1.113	1.110	1.130	1.116
	(6.16)	(6.32)	(6.65)	(6.60)
Banned	0.716	1.103	0.752	1.158
	(1.51)	(0.48)	(1.33)	(0.72)
Whitelist	2.902	0.842	1.786	0.915
	(4.24)	(1.74)	(4.13)	(0.95)
Duration	1.000	1.000	1.004	0.999
	(0.06)	(0.41)	(2.79)	(1.45)
Presale	0.499	1.157	0.344	1.214
	(3.58)	(1.40)	(4.33)	(1.80)
Hardcap	2.418	1.106	1.412	1.301
	(4.35)	(0.78)	(2.69)	(2.12)
Softcap	0.649	1.006	0.580	1.031
	(3.39)	(0.06)	(3.96)	(0.31)
Accept BTC	1.492	1.095	0.976	1.178
	(3.00)	(0.91)	(0.23)	(1.64)
Accept ETH	2.265	0.846	4.859	0.632
	(3.79)	(1.47)	(4.50)	(3.40)
Accept USD	2.043	0.967	0.625	1.260
	(3.46)	(0.25)	(2.60)	(1.64)
Enforcement	0.259	1.173	0.300	1.131
	(4.46)	(1.31)	(4.85)	(1.02)
Disclosure	1.230	0.958	1.377	0.917
	(3.45)	(1.20)	(4.21)	(2.30)
SEC filing	0.064	2.107	0.353	0.887
	(3.12)	(1.21)	(2.08)	(0.25)
# ICOs	5,935	5,935	5,935	5,935

A.I News of regulatory actions taken by U.S. authorities

Date	Title	News summary
16th Jun 2018	SEC: Fraud surrounds initial coin offerings, blockchain security notwithstanding.	SEC has a unit that monitors ICO scams.
21st Jun 2018	Members of the House will now be required to disclose bitcoin, other cryptocurrency holdings; Ethics Committee strongly encourage House members who are considering investing in an ICO to seek guidance.	Ethics Committee have taken actions to regulate House members in ICO investments.
27th Jun 2018	Facebook to accept cryptocurrency ads again; January's blanket ban is reversed, though crypto firms will have to get case-by-case approval.	Tech companies such as Facebook banned cryptocurrencies ads. Promotional efforts for cryptocurrencies have come under fire from federal and state regulators.
15th Aug 2018	Even free tokens face regulatory heat as coin offerings scrutinized; SEC punishes company that didn't sell any tokens, saying potential investors were misled about details of oil-drilling project.	The SEC punished a firm that did not sell any tokens to crack down on fraud in the market for initial coin offerings.
12th Sep 2018	SEC takes first action against hedge fund over cryptocurrency investments; In a separate case that's another first, agency penalizes brokers who ran an "ICO superstore".	The SEC fined a hedge fund manager who falsely advertised his cryptocurrency fund as the first regulated crypto-fund in the United States. Separately, the SEC also fined two men who ran a website that connects investors with initial coin offerings.
12th Sep 2018	Judge lets cryptocurrency fraud case go forward, in win for SEC; For first time a federal court weighs in on the government's jurisdiction over ICOs in a criminal case.	The SEC scored a victory in their crackdown on cryptocurrency fraud as a judge ruled that initial coin offerings are subject to U.S. securities laws.

(To be continued)

Date	Title	News summary
11th Oct 2018	SEC says stop ICOs that falsely claimed SEC approval.	SEC's complaint charges Blockvest and Ringgold with violating federal securities laws.
22nd Oct 2018	SEC suspends trading in company for making false cryptocurrency-related claims about SEC regulation and registration.	SEC suspended trading in the securities of a company for making false cryptocurrency-related claims.
16th Nov 2018	SEC settles enforcement actions over two initial coin offerings	Two startups agreed to comply with investor protection rules and offer money back to thousands of people who bought their digital tokens.
30th Nov 2018	Boxer Mayweather Jr., producer DJ Khaled agree to settle SEC crypto charges.	Celebrity endorsements of coin offerings may be illegal if the promoters fail to disclose the source and amount of their compensation.
21st May 2019	SEC obtains emergency order halting alleged diamond-related ICO Scheme targeting hundreds of investors.	SEC halted a Ponzi scheme, which was purportedly a cryptocurrency business.
5th Jun 2019	SEC challenges Canada firm's coin offering	SEC sued Kik for not providing investors with full and fair disclosure about its token and its business.

TABLE A.I.1: News of regulatory actions taken by U.S. authorities (Aug '18–Aug '19)

A.II Details of network model

Consider a set of ICOs $N = \{1, 2, \ldots, n\}$ that are members of a network g. In this network, two ICOs share a link if they are advised by at least one common advisor. Formally, for two ICOs i and j, we write:

$$g_{ij} = \begin{cases} 1, & \text{share a direct link} \\ 0, & \text{do not share a direct link or } i = j \end{cases} \tag{A.II.1}$$

A square symmetric matrix $\mathbf{G} = [g_{ij}]$ represents this network and tracks the direct links among ICOs. The matrix $\mathbf{G}$ is also known as the adjacency matrix of the network. We will also consider a weighted network, in which g_{ij}'s are not necessarily binary and can take on numeric weights.

We use Katz centrality, which measures the network prominence of an ICO as the weighted sum of walks that emanate from it to other ICOs in the network. This implies that we need to account for indirect links of ICOs. To track indirect links in networks, we use the k-th power of the adjacency matrix—$\mathbf{G}^k, k \in \mathbb{Z}$.[30] An element $g_{ij}^{[k]}$ in $\mathbf{G}^k$ gives the number of walks of length $k \geq 1$ from i to j in the network.[31] In the special case of $k = 0$, $\mathbf{G}^k$ is defined as the identity matrix $\mathbf{I}$.

To operationalize the Katz centrality measure, consider a matrix $\mathbf{M}$ that tracks the number of walks of all lengths between any two ICOs.

$$\mathbf{M} = \sum_{k=0}^{+\infty} \theta^k \mathbf{G}^k \quad \text{with element } m_{ij} = \sum_{k=0}^{+\infty} \theta^k g_{ij}^{[k]} \tag{A.II.2}$$

The term θ^k is the decay factor applied to walks of length k. Economically, the decay factor controls how much influence an ICO has on another ICO in the

[30] While an ICO plays its equilibrium number of misrepresentations based on its direct network neighbors', these neighbors respond to their own set of neighbors, so on and so forth. Thus, the equilibrium response of an ICO also depends on other indirectly linked ICOs.

[31] This is an established result in graph theory.

network. This influence increasingly wanes if two ICOs are further away (i.e., $k > 1$) from each other. We can also derive an equivalent expression of $\mathbf{M}$ below.[32]

$$\mathbf{M} = [\mathbf{I} - \theta \mathbf{G}]^{-1} \tag{A.II.3}$$

By definition, the Katz centrality of ICO i—denoted as $b_i(g, \theta)$—is the sum of elements of the i-th row in $\mathbf{M}$.

$$b_i(g, \theta) = \sum_{j=1}^{n} m_{ij} = \sum_{j=1}^{n} \sum_{k=0}^{+\infty} \theta^k g_{ij}^{[k]} \tag{A.II.4}$$

Following equation (A.II.3), the $(n \times 1)$ vector of Katz centralities is thus:

$$\mathbf{b}(g, \theta) = \mathbf{M} \cdot \mathbf{1} = [\mathbf{I} - \theta \mathbf{G}]^{-1} \cdot \mathbf{1} \tag{A.II.5}$$

We specify a linear-quadratic utility function of ICOs (or equivalently, issuers) that captures both ICO-specific and complementary components of misrepresentation behavior. This formulation is popular in network economics because it admits a tractable solution and cleanly characterizes the equilibrium as a function of network structure. For $\alpha_i > 0$ and $\theta > 0$, we write equation (A.II.6).

$$u_i(d_i, d_{-i}, g) = \alpha_i d_i - \frac{1}{2} d_i^2 + \theta \sum_{j=1}^{n} g_{ij} d_i d_j \tag{A.II.6}$$

Let us emphasize the perspective here—the network has already formed. The ICO observes its advisor-linked ICO peers and chooses d_i to maximize utility in equation (A.II.6). We are agnostic about the network formation process.

[32]Following equation (A.II.2), we first express $\mathbf{M} = \mathbf{I} + \theta\mathbf{G} + \theta^2\mathbf{G}^2 + \theta^3\mathbf{G}^3 + \ldots$. Next, we multiply this expression by $\theta\mathbf{G}$ to get $\theta\mathbf{G}\mathbf{M} = \theta\mathbf{G} + \theta^2\mathbf{G}^2 + \theta^3\mathbf{G}^3 + \ldots$. Finally, the difference of these expressions yields equation (A.II.3).

Rather, our model takes the network structure as given and focuses on the complementarities in misrepresentation behavior among ICOs.

We obtain a tractable solution of this game. The ICO network has formed, and the issuer chooses d_i to maximize utility. The first-order condition of equation (A.II.6) gives the following best-response function.

$$d_i^* = \alpha_i + \theta \sum_{j=1}^{n} g_{ij} d_j^*, \quad \forall i = 1, 2, \ldots, n \tag{A.II.7}$$

The best-response function can be equivalently expressed in matrix form. Solving equation (A.II.8) and using equation (A.II.5), we show that the Nash equilibrium vector $\mathbf{d}^*$ is proportional to the vector of Katz centralities $\mathbf{b}$.

$$\mathbf{d}^* = \boldsymbol{\alpha} + \theta \mathbf{G} \mathbf{d}^* = [\mathbf{I} - \theta \mathbf{G}]^{-1} \boldsymbol{\alpha} = \mathbf{M} \boldsymbol{\alpha} \tag{A.II.8}$$

References

Ackermann, Carl, Richard McEnally, and David Ravenscraft. 1999. "The Performance of Hedge Funds: Risk, Return, and Incentives." *Journal of Finance* 54 (3):833–874.

Admati, Anat R. and Paul Pfleiderer. 1987. "Viable Allocations of Information in Financial Markets." *Journal of Economic Theory* 43 (1):76–115.

Aggarwal, Rajesh K. and Philippe Jorion. 2010. "The performance of emerging hedge funds and managers." *Journal of Financial Economics* 96 (2):238–256.

Aloosh, Arash and Jiasun Li. 2019. "Direct evidence of Bitcoin wash trading." *Available at SSRN 3362153* .

Amihud, Yakov. 2002. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of Financial Markets* 5 (1):31–56.

Asquith, Paul, Parag A. Pathak, and Jay R. Ritter. 2005. "Short interest, institutional ownership, and stock returns." *Journal of Financial Economics* 78 (2):243–276.

Atiase, Rk. 1985. "Predisclosure Information, Firm Capitalization, and Security Price Behavior around Earnings Announcements." *Journal of Accounting Research* 23 (1):21–36.

Atmaz, Adem and Suleyman Basak. 2018. "Belief Dispersion in the Stock Market." *Journal of Finance* 73 (3):1225–1279.

Baker, Malcolm and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance* 61 (4):1645–1680.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. 2006. "Who's who in networks. Wanted: The key player." *Econometrica* 74 (5):1403–1417.

Bamber, Linda Smith. 1987. "Unexpected Earnings, Firm Size, and Trading Volume around Quarterly Earnings Announcements." *The Accounting Review* 62 (3):510–532.

Barber, B. M. and T. Odean. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance* 55:773–806.

———. 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.

Barber, B. M., T. Odean, and N. Zhu. 2009. "Do Retail Trades Move Markets?" *Review of Financial Studies* 22 (1):151–186.

Barber, Brad, Reuven Lehavy, Maureen Mcnichols, and Brett Trueman. 2001. "Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns." *Journal of Finance* 56 (2):531–563.

Barnea, Amir. 1974. "Performance Evaluation of New York Stock Exchange Specialists." *Journal of Financial and Quantitative Analysis* 9 (4):511–535.

Barrot, Jean Noel, Ron Kaniel, and David Sraer. 2016. "Are retail traders compensated for providing liquidity?" *Journal of Financial Economics* 120 (1):146–168.

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

Benedetti, Hugo and Leonard Kostovetsky. 2021. "Digital tulips? Returns to investors in initial coin offerings." *Journal of Corporate Finance* 66.

Bettis, Carr, Don Vickrey, and Donn W. Vickrey. 1997. "Mimickers of Corporate Insiders Who Make Large-Volume Trades." *Financial Analysts Journal* 53 (5):57–66.

Boehmer, Ekkehart, Zsuzsa R. Huszar, and Bradford D. Jordan. 2010. "The good news in short interest." *Journal of Financial Economics* 96 (1):80–97.

Boehmer, Ekkehart, Charles M. Jones, Juan Wu, and Xiaoyan Zhang. 2020. "What Do Short Sellers Know?" *Review of Finance* 24 (6):1203–1235.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang. 2021. "Tracking Retail Investor Activity." *Journal of Finance* forthcoming.

Boehmer, Ekkehart and Eric K. Kelley. 2009. "Institutional Investors and the Informational Efficiency of Prices." *Review of Financial Studies* 22 (9):3563–3594.

Boehmer, Ekkehart and Juan (Julie) Wu. 2013. "Short Selling and the Price Discovery Process." *Review of Financial Studies* 26 (2):287–322.

Bourveau, Thomas, Emmanuel De George, Atif Ellahie, and Daniele Macciocchi. 2021. "The role of disclosure and information intermediaries in an unregulated capital market: evidence from initial coin offerings." *Journal of Accounting Research* 60 (1):129–167.

Button, Mark, Chris Lewis, and Jacki Tapley. 2009. "A better deal for fraud victims." *National Fraud Authority, United Kingdom* .

Cao, Charles, Bing Liang, Andrew W Lo, and Lubomir Petrasek. 2018. "Hedge Fund Holdings and Stock Market Efficiency." *Review of Asset Pricing Studies* 8 (1):77–116.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52 (1):57–82.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok. 2007. "Analysts' Conflicts of Interest and Biases in Earnings Forecasts." *Journal of Financial and Quantitative Analysis* 42 (4):893–913.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein. 2002. "Breadth of Ownership and Stock Returns." *Journal of Financial Economics* 66 (2–3):171–205.

Chi, Sabrina and Devin Shanthikumar. 2018. "Do Retail Investors Use SEC Filings? Evidence from EDGAR Search." *Available at SSRN: https://ssrn.com/abstract=3281234* .

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam. 2005. "Evidence on the speed of convergence to market efficiency." *Journal of Financial Economics* 76 (2):271–292.

Chordia, Tarun and Avanidhar Subrahmanyam. 2004. "Order imbalance and individual stock returns: Theory and evidence." *Journal of Financial Economics* 72 (3):485–518.

Chung, Ling Tak Douglas. 2020. "Insider Trading and the Return Predictability of Retail Trading." *Available at SSRN: https://ssrn.com/abstract=3479686* .

Cohen, Lauren, Christopher Malloy, and Lukasz Pomorski. 2012. "Decoding Inside Information." *Journal of Finance* 67 (3):1009–1043.

Comerton-Forde, Carole and Tālis Putniņš. 2014. "Stock price manipulation: Prevalence and determinants." *Review of Finance* 18 (1):23–66.

Cong, Lin William, Xi Li, Ke Tang, and Yang Yang. 2020. "Crypto wash trading." *Available at SSRN 3530220* .

Cong, Lin William, Ye Li, and Neng Wang. 2020. "Tokenomics: Dynamic adoption and valuation." *Review of Financial Studies* 00:1–51.

Cookson, J. Anthony and Marina Niessner. 2020. "Why Don't We Agree? Evidence from a Social Network of Investors." *Journal of Finance* 75 (1):173–228.

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. "In Search of Attention." *Journal of Finance* 66 (5):1461–1499.

Dechow, Patricia M., Amy P. Hutton, Lisa Meulbroek, and Richard G. Sloan. 2001. "Short-sellers, fundamental analysis, and stock returns." *Journal of Financial Economics* 61 (1):77–106.

Deng, Xin, Yen Teik Lee, and Zhengting Zhong. 2018. "Decrypting coin winners: Disclosure quality, governance mechanism and team networks." *Available at SSRN 3247741* .

Desai, Hemang, S. Ramesh, Ramu Thiagarajan, and Bala V. Balachandran. 2002. "An Investigation of the Informational Role of Short Interest in the Nasdaq Market." *Journal of Finance* 57 (5):2263–2287.

Dhawan, Anirudh and Tālis Putniņš. 2022. "A new wolf in town? Pump-and-dump manipulation in cryptocurrency markets." *Review of Finance* :forthcoming.

Diamond, Douglas W. and Robert E. Verrecchia. 1981. "Information aggregation in a noisy rational expectations economy." *Journal of Financial Economics* 9 (3):221–235.

———. 1987. "Constraints on short-selling and asset price adjustment to private information." *Journal of Financial Economics* 18 (2):277–311.

———. 1991. "Disclosure, Liquidity, and the Cost of Capital." *Journal of Finance* 46 (4):1325–1359.

Diether, Karl, Christopher Malloy, and Anna Scherbina. 2002. "Differences of Opinion and the Cross Section of Stock Returns." *Journal of Finance* 57 (5):2113–2141.

Diether, Karl B., Kuan-Hui Lee, and Ingrid M. Werner. 2009. "Short-Sale Strategies and Return Predictability." *Review of Financial Studies* 22 (2):575–607.

Dittmar, Robert and Di (Andrew) Wu. 2019. "Initial coin offerings hyped and dehyped: An empirical examination." *Available at SSRN 3259182* .

Dowlat, Sherwin. 2018. "Cryptoasset market coverage initiation: Network creation." Tech. rep., Satis Group.

Doyle, JEFFREY T., RUSSELL J. Lundholmn, and MARK T. Soliman. 2006. "The Extreme Future Stock Returns Following I/B/E/S Earnings Surprises." *Journal of Accounting Research* 4:849–887.

Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock. 2015. "The Determinants and Consequences of Information Acquisition via EDGAR." *Contemporary Accounting Research* 32 (3):1128–1161.

Easley, David and Maureen O'hara. 2004. "Information and the Cost of Capital." *Journal of Finance* 59 (4):1553–1583.

Edelen, Roger M., Ozgur S. Ince, and Gregory B. Kadlec. 2016. "Institutional investors and stock return anomalies." *Journal of Financial Economics* 119 (3):472–488.

Fahlenbrach, Rüdiger and Marc Frattaroli. 2020. "ICO investors." *Financial Markets and Portfolio Management* .

Fama, Eugene F. and Kenneth R. French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33 (1):3–56.

———. 1995. "Size and Book-to-Market Factors in Earnings and Returns." *Journal of Finance* 50 (1):131–155.

———. 2015. "A five-factor asset pricing model." *Journal of Financial Economics* 116 (1):1–22.

Fama, Eugene F. and James D MacBeth. 1973. "Risk, return, and equilibrium: Empirical tests." *Journal of Political Economy* 81 (3):607–636.

Farrell, Michael, T. Clifton Green, Russell Jame, and Stanimir Markov. 2018. "The Democratization of Investment Research and the Informativeness of Retail Investor Trading." *Available at SSRN 3222841* .

Feinstein, Jonathan. 1990. "Detection controlled estimation." *Journal of Law and Economics* 33 (1):233–276.

Foley, Sean, Jonathan Karlsen, and Tālis Putniņš. 2019. "Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?" *Review of Financial Studies* 32 (5):1798–1853.

Fong, Kingsley Y. L., David R. Gallagher, and Adrian D. Lee. 2014. "Individual investors and broker types." *Journal of Financial and Quantitative Analysis* 49 (2):431–451.

Frazzini, Andrea and Owen A. Lamont. 2008. "Dumb money: Mutual fund flows and the cross-section of stock returns." *Journal of Financial Economics* 88 (2):299–322.

———. 2021. "The Earnings Announcement Premium and Trading Volume." *NBER Working Paper No. w13090* .

Gee, Jim and Mark Button. 2019. "The financial cost of fraud 2019: The latest data from around the world." .

Gensler, Gary. 2021. "Remarks before the Aspen Security Forum." *Public Statement* URL `https://www.sec.gov/news/public-statement/gensler-aspen-security-forum-2021-08-03`.

Gervais, Simon, Ron Kaniel, and Dan H. Mingelgrin. 2002. "The High-Volume Return Premium." *Journal of Finance* 56 (3):877–919.

Goldstein, Itay and Liyan Yang. 2015. "Information Diversity and Complementarities in Trading and Information Acquisition." *Journal of Finance* 70 (4):1723–1765.

Griffin, John and Amin Shams. 2020. "Is Bitcoin really untethered?" *Journal of Finance* 75 (4):1913–1964.

Griffin, John M., Patrick Kelly, and Federico Nardari. 2010. "Do Market Efficiency Measures Yield Correct Inferences? A Comparison of Developed and Emerging Markets." *Review of Financial Studies* 23 (8):3225–3277.

Griffin, John M. and Jin Xu. 2009. "How Smart Are the Smart Guys? A Unique View from Hedge Fund Stock Holdings." *Review of Financial Studies* 22 (7):2531–2570.

Gromb, Denis and Dimitri Vayanos. 2010. "Limits of Arbitrage." *Annual Review of Financial Economics* 2:251–275.

Grossman, Sanford J. 1976. "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information." *Journal of Finance* 31 (2):573–585.

Grossman, Sanford J. and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70 (3):393–408.

Herley, Cormac. 2012. "Why do Nigerian scammers say they are from Nigeria?" In *WEIS*.

Hirshleifer, David and Siew Hong Teoh. 2003. "Herd Behaviour and Cascading in Capital Markets: a Review and Synthesis." *European Financial Management* 9 (1):25–66.

Hou, Kewei and Tobias J. Moskowitz. 2005. "Market Frictions, Price Delay, and the Cross-Section of Expected Returns." *Review of Financial Studies* 18 (3):981–1020.

Howell, Sabrina, Marina Niessner, and David Yermack. 2020. "Initial coin offerings: Financing growth with cryptocurrency token sales." *Review of Financial Studies* 33 (9):3925–3974.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou. 2015. "Investor Sentiment Aligned: A Powerful Predictor of Stock Returns." *Review of Financial Studies* 28 (3):791–837.

Huang, Dashan, Jiangyuan Li, and Liyao Wang. 2021. "Are disagreements agreeable? Evidence from information aggregation." *Journal of Financial Economics* 141 (1):83–101.

Jaffe, Jeffrey F. 1974. "Special Information and Insider Trading." *Journal of Business* 47 (3):410–428.

Jagolinzer, Alan D. 2009. "SEC Rule 10b5-1 and Insiders' Strategic Trade." *Management Science* 55 (2):224–239.

Jegadeesh, Narasimhan, Joonghyuk Kim, Susan D. Krische, and Charles M. C. Lee. 2004. "Analyzing the Analysts: When Do Recommendations Add Value?" *Journal of Finance* 59 (3):1083–1124.

Jeng, Leslie A., Andrew Metrick, and Richard Zeckhauser. 2003. “Estimating the Returns to Insider Trading: A Performance-Evaluation Perspective.” *Review of Economics and Statistics* 85 (2):453–471.

Jiang, Hao and Zheng Sun. 2014. “Dispersion in beliefs among active mutual funds and the cross-section of stock returns.” *Journal of Financial Economics* 114 (2):341–365.

Jiao, Yawen, Massimo Massa, and Hong Zhang. 2016. “Short selling meets hedge fund 13F: An anatomy of informed demand.” *Journal of Financial Economics* 122 (3):544–567.

Kahneman, D. 1973. *Attention and effort.* Prentice-Hall.

Kaniel, R., G. Saar, and S. Titman. 2008. “Individual Investor Trading and Stock Returns.” *Journal of Finance* 63:273–310.

Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman. 2012. “Individual Investor Trading and Return Patterns around Earnings Announcements.” *Journal of Finance* 67 (2):639–680.

Karpoff, Jonathan M. and Xiaoxia Lou. 2010. “Short Sellers and Financial Misconduct.” *Journal of Finance* 65 (5):1879–1913.

Kelley, Eric K. and Paul C. Tetlock. 2013. “How Wise Are Crowds? Insights from Retail Orders and Stock Returns.” *Journal of Finance* 68 (3):1229–1265.

Kosowski, Robert, Narayan Y. Naik, and Melvyn Teo. 2007. “Do hedge funds deliver alpha? A Bayesian and bootstrap analysis.” *Journal of Financial Economics* 84 (1):229–264.

Kumar, Alok. 2009. “Who Gambles in the Stock Market?” *Journal of Finance* 64 (4):1889–1933.

Kyle, Albert S. 1985. “Continuous Auctions and Insider Trading.” *Econometrica* 53 (6):1315–1335.

La Porta, Rafael, Florencio Lopez-De-Silanes, Andrei Shleifer, and Robert Vishny. 2000. "Agency problems and dividend policies around the world." *Journal of Finance* 55 (1):1–33.

Lakonishok, Josef and Inmoo Lee. 2001. "Are Insider Trades Informative?" *Review of Financial Studies* 14 (1):79–111.

Lee, Charles M. C. 1992. "Earnings news and small traders: An intraday analysis." *Journal of Accounting and Economics* 15 (2-3):265–302.

Li, Frank Weikai, Abhiroop Mukherjee, and Rik Sen. 2021. "Inside Brokers." *Journal of Financial Economics* 141 (3):1096–1118.

Li, Tao, Donghwa Shin, and Baolian Wang. 2021. "Cryptocurrency pump-and-dump schemes." *Available at SSRN 3267041* .

Lin, Ji-Chai and John S Howe. 1990. "Insider Trading in the OTC Market." *Journal of Finance* 45 (4):1273–1284.

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang. 2002. "Dynamic Volume-Return Relation of Individual Stocks." *Review of Financial Studies* 15 (4):1005–1047.

Lo, Andrew W. 2004. "The Adaptive Markets Hypothesis." *Journal of Portfolio Management* 30 (5):15–29.

Lo, Andrew W. and A. Craig MacKinlay. 1988. "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." *Review of Financial Studies* 1 (1):41–66.

Long, J. Bradford De, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy* 98 (4):703–738.

Lyandres, Evgeny, Berardino Palazzo, and Daniel Rabetti. 2021. "ICO success and post-ICO performance." *Management Science* :forthcoming.

Malmendier, Ulrike and Devin Shanthikumar. 2007. "Are small investors naive about incentives?" *Journal of Financial Economics* 85 (2):457–489.

Mansi, Sattar, Lin Peng, Jianping Qi, and Han Shi. 2021. "Retail Investor Attention and Insider Trading." *Available at SSRN: https://ssrn.com/abstract=3831504* .

Marin, Jose M. and Jacques P. Olivier. 2008. "The Dog That Did Not Bark: Insider Trading and Crashes." *Journal of Finance* 63 (5):2429–2476.

Massa, Massimo, Wenlan Qian, Weibiao Xu, and Hong Zhang. 2015. "Competition of the informed: Does the presence of short sellers affect insider selling?" *Journal of Financial Economics* 118 (2):268–288.

Massoud, Nadia, Debarshi Nandy, Anthony Saunders, and Keke Song. 2011. "Do hedge funds trade on private information? Evidence from syndicated lending and short-selling." *Journal of Financial Economics* 99 (3):477–499.

McLean, David, Jeffrey Pontiff, and Christopher Reilly. 2021. "Retail investors and Analysts." *Working Paper* .

McLean, R. David, Jeffrey Pontiff, and Christopher Reilly. 2020. "Taking Sides on Return Predictability." *Available at SSRN 3637649* .

Mikhail, Michael B., Beverly R. Walther, and Richard H. Willis. 2007. "When Security Analysts Talk, Who Listens?" *Journal of Finance* 82 (5):1227–1253.

Miller, Edward M. 1977. "Risk, Uncertainty, and Divergence of Opinion." *Journal of Finance* 32 (4):1151–1168.

Morck, Randall, Bernard Yeung, and Wayne Yu. 2000. "The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?" *Journal of Financial Economics* 58 (1-2):215–260.

O'Hara, Maureen. 2003. "Presidential Address: Liquidity and Price Discovery." *Journal of Finance* 58 (4):1335–1354.

Pastor, Lubos and Robert F. Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy* 111 (3):642–685.

Pastor, Lubos and Pietro Veronesi. 2003. "Stock Valuation and Learning about Profitability." *Journal of Finance* 58 (5):1749–1789.

Paul, Jonathan M. 1993. “Crowding Out and the Informativeness of Security Prices.” *Journal of Finance* 48 (4):1475–1496.

Piotroski, Joseph D. and Darren T. Roulstone. 2005. “Do insider trades reflect both contrarian beliefs and superior knowledge about future cash flow realizations?” *Journal of Accounting and Economics* 39 (1):55–81.

PriceWaterhouseCoopers. 2020. “6th ICO/STO Report: A strategic perspective.” .

Qian, Hong and Zhaodong Zhong. 2018. “Do Hedge Funds Possess Private Information about IPO Stocks? Evidence from Post-IPO Holdings.” *Review of Asset Pricing Studies* 8 (1):117–152.

Rozeff, Michael S and Mir A Zaman. 1988. “Market Efficiency and Insider Trading: New Evidence.” *Journal of Business* 61 (1):25–44.

Saffi, Pedro A. C. and Kari Sigurdsson. 2010. “Price Efficiency and Short Selling.” *Review of Financial Studies* 24 (3):821–852.

SEC. 2018. “The SEC has an opportunity you won’t want to miss: Act now!” *Press Release* URL `https://www.sec.gov/news/press-release/2018-88`.

Seyhun, N. H. 1986. “Insiders’ profits, costs of trading, and market efficiency.” *Journal of Financial Economics* 16 (2):189–212.

———. 1998. *Investment Intelligence from Insider Trading.* Cambridge: MIT Press.

Shleifer, Andrei and Robert W. Vishny. 1997. “The Limits of Arbitrage.” *Journal of Finance* 52 (1):35–55.

Sias, Richard W. and David A. Whidbee. 2010. “Insider Trades and Demand by Institutional and Individual Investors.” *Review of Financial Studies* 23 (4):1544–1595.

Sockin, Michael and Wei Xiong. 2020. “A model of cryptocurrencies.” .

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan. 2012. "The short of it: Investor sentiment and anomalies." *Journal of Financial Economics* 104 (2):288–302.

———. 2015. "Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle." *Journal of Finance* 70 (5):1903–1948.

Stotz, Olaf and Dominik Georgi. 2012. "A logit model of retail investors' individual trading decisions and their relations to insider trades." *Review of Financial Economics* 21 (4):159–167.

Verrecchia, Robert E. 2001. "Essays on disclosure." *Journal of Accounting and Economics* 32 (1–3):97–180.

Wang, Tracy Yue, Andrew Winton, and Xiaoyun Yu. 2010. "Corporate fraud and business conditions: Evidence from IPOs." *Journal of Finance* 65 (6):2255–2292.

Womack, Kent L. 1996. "Do Brokerage Analysts' Recommendations Have Investment Value?" *Journal of Finance* 51 (1):137–167.

Yermack, David. 2015. "Is Bitcoin a real currency? An economic appraisal." In *Handbook of digital currency*. Elsevier, 31–43.

Zhang, X. Frank. 2006. "Information Uncertainty and Stock Returns." *Journal of Finance* 61 (1):105–137.