

Interest in the Short Interest: The Rise of Private Sector Data

Minjae Kim
University of Texas at Austin
Minjae.Kim@mcombs.utexas.edu

John McInnis
University of Texas at Austin
John.McInnis@mcombs.utexas.edu

Wuyang Zhao
University of Texas at Austin
Wuyang.Zhao@mcombs.utexas.edu

September 2022

Abstract

The extent to which a stock is sold short (the “short interest”) is currently required to be disclosed twice per month, but regulators have consistently expressed a desire to increase this frequency. Meanwhile, short interest and lending data from private, third-party vendors has arisen to meet investor demand for short selling information on a daily basis. We find that this daily private-sector data is a strong predictor of the bimonthly regulatory short interest disclosure and investors appear to react to this daily data. However, we find that investors continue to react to the regulatory disclosure, suggesting that the market reaction is not fully preempted by private-sector data. Consistent with this finding, we show that the market appears to under-react to the information content of the private-sector data and the magnitude of preemption is fairly small. Although access to this data is relatively expensive, we find no evidence that retail investors are harmed in their trades around daily changes in the private-sector short interest. Overall, our findings suggest that increasing investor awareness of, or access to, private-sector short interest data may be a less-costly alternative to mandating increased frequency of regulatory short interest. We contribute to short selling literature by studying the interplay of private-sector and regulatory solutions in enhancing short selling transparency.

We appreciate helpful comments from workshop participants at the University of Texas at Austin. We are grateful to the S3 Partners for generously providing the data. We thank Sam Elphick, Shan Gao, and Jason Yang from IHS Markit, and Bob Sloan, and particularly Gregg Vier from S3 Partners for insightful discussion on the institutional details related to the data. We acknowledge the financial support from the Department of Accounting at UT Austin. All errors remain our own.

1 Introduction

Short selling has long been a controversial form of trading (e.g., Sloan 2009). While research generally documents its benefits to market efficiency (e.g., Saffi and Sigurdsson 2011; Boehmer and Wu 2013), market participants have expressed concern that short sellers might manipulate the market through abusive trading such as a “bear raid” (e.g., SEC 2008; Soros 2009; Khanna and Mathews 2012). The potential benefits to market quality and the concern over manipulation have been the two primary considerations for short selling regulations since 1930s (e.g., SEC 2014).

One key dimension of this regulatory framework is mandated disclosure of short selling activities (i.e., SEC 2021a; SEC 2022). Currently, the Financial Industry Regulatory Authority (“FINRA”) collects data on the short interest (i.e., the number of firm’s shares sold short) in individual securities through its broker-dealer members. This information is then published twice per month by the exchanges in which individual stocks are listed, with a delay of eight business days following the settlement date on which short interest is captured. Because this disclosure regime lacks timeliness in facilitating price discovery and detecting potentially abusive short selling, there has been regulatory pressure to alter these requirements over the past two decades. For example, following the Great Recession, the Dodd-Frank Act of 2010 required the SEC to consider the costs and benefits of real-time short position reporting. Most recently, following the “meme stock” craze partially aimed at short sellers, FINRA (2021) requested comments on increasing the frequency of regulatory short interest disclosures to a daily or weekly basis.

Meanwhile, over the past two decades, private data vendors have emerged that sell estimates of the daily short interest at the individual stock level to meet the demand for more timely short selling data (see, e.g. FINRA 2021). While they do not conduct or cite empirical analyses,

regulators argue that these private-sector data are not suitable substitutes for regulatory disclosure due to incompleteness and measurement error, both of which may render the data inaccurate. Moreover, as those vendors are only available through subscription, the rise of private access might put retail investors in a disadvantaged position. Given this regulatory attention, it is surprising that there is no existing research studying the informativeness of private-sector short interest data. We intend to fill this void in the literature by asking three inter-related questions in this paper. First, is private-sector short interest data informative to investors? Second, do these private-sector data preempt the market reaction to regulatory disclosures? Finally, are retail investors harmed by institutions' more-timely access to private-sector short interest data?

Answering these questions is important for the debate over increasing the frequency of regulatory short interest disclosure. If private sector data can predict and largely preempt regulatory disclosure without hurting retail investors, moving to a mandatory high frequency disclosure regime may be suboptimal, as the reporting costs would likely exceed the informational benefits. On the other hand, if daily disclosures are of low quality, do not preempt regulatory disclosure, and/or hurt retail investors, then it is possible that calls for higher frequency regulatory disclosure may have merit. Further, the answers to these research questions are *ex ante* unclear, due to two factors. First, as the SEC (2014) and FINRA (2021) argue, the private data are estimates, and may be incomplete and inaccurate in capturing actual short interest. Second, the private access might be too expensive for many investors to access (the datasets we examine cost between \$9,000 and \$100,000 per year for non-academic purposes). Therefore, it is unclear how informative private data are and whether they help preempt market reaction to regulatory disclosure.

We examine two private-sector data vendors that have risen to meet the demand for timelier short interest data. The primary dataset is IHS Markit ("Markit"), which measures the total cross-

broker demand in a security including all shares lent between banks and brokers and is available from 2006 onwards on a daily basis. Markit data has been used extensively in prior research (e.g., Beneish, Lee, and Nichols, 2015; Engelberg, Reed, and Ringgenberg, 2018). As a supplementary data source, we also use S3 Partners (“S3”), which measures estimated short interest based on a proprietary algorithm using data on trading positions and is available from 2015 onwards. S3 data has not been used in prior research to our knowledge, but has been frequently quoted in mainstream media such as Bloomberg (e.g., Wienberg 2022), Financial Times (e.g., Terazono 2022), and Wall Street Journal (e.g., Langley and Banerji 2022).

For our first research question, we evaluate the informativeness of private-sector short interest data in two ways. First, using a sample of short interest announcements from Compustat and daily short interest data from our two private vendors, we test the ability of changes in daily private-sector short interest to predict changes in FINRA short interest. We find that both private-sector data sources are highly predictive of changes in FINRA short interest, with Markit data explaining 21% and S3 data explaining 67% of the variation in FINRA short interest eight days in advance. Next, we examine the market reaction to the negative news contained in private-sector short interest data. We find the market reacts negatively to daily increases in Markit and S3 short interest. Specifically, a one-percentage-point increase in the daily short interest from private-sector sources is associated with a concurrent stock price decline of about 14-18 basis points (or about 35%-45% on an annualized basis). Taken together, these two sets of analyses show that private-sector short interest data are predictive of future regulatory short interest news and are informative to investors.

For our second research question, we examine whether the private-sector short interest preempts the market reactions to regulatory FINRA short interest disclosures. Since private-sector

short interest data are highly informative in predicting future regulatory disclosure and their disclosure is associated with concurrent daily returns, we expect that the market will price this information ahead of FINRA short interest disclosures. To test our prediction, we first examine the market reaction to *regulatory* short interest announcements during our sample period. We find that the market reacts significantly negatively to firms reporting large increases in regulatory short interest. This indicates that the private-sector data, although informative, does not fully preempt the reaction to the regulatory disclosure.

We next examine one possible explanation for this lack of preemption. We find that investors appear to significantly under-react to the information content of private sector short interest data. Specifically, daily short interest news predicts *future* returns around the FINRA short interest announcement date. In fact, we find that about half of the announcement-day stock return reaction to FINRA news can be predicted by previously announced daily short interest news from private-sector vendors (especially from S3). This suggests that although the daily private sector data is informative with respect to the short interest, investors do not appear to fully impound this information into stock prices on a timely basis. Next, we investigate the extent (if any) to which private-sector data preempts the market reaction to regulatory disclosures. We find evidence that the news in regulatory short interest is being impounded into prices more quickly due to the private-sector short interest data. However, the economic magnitude is fairly small. For example, we find that large increases in daily private sector short interest increases the intraperiod timeliness of returns (IPT) between FINRA short interest announcements dates by only about 5 to 8%.

Our finding that private data sources are informative and partially preemptive of regulatory disclosure raises an obvious question of fairness. Due to the cost of accessing private-sector short interest data, smaller retail investors are less likely to have access to these timely data sources and

may be put in a disadvantaged position relative to their institutional counterparts. In our last research question, we examine whether the existence of private data harms retail investors. We follow Boehmer et al. (2021) to identify marketable retail trades and to measure retail order imbalance, using TAQ data from 2009 to 2014 (due to data availability). We do *not* find evidence that retail investors' trading profitability is worsened when private-sector short interest data can better predict regulatory short interest news. One explanation could be that, because the market does not fully impound information in private-sector data into prices, retail investors are not as disadvantaged as they would have been had prices reflected such information more quickly. Overall, our results should assuage the concern that costly private-sector solution puts smaller retail investors in a disadvantaged position – a major motivation for regulators' consideration of expanding short interest disclosure.

While short interest disclosure is an important part of the regulatory effort to improve short selling transparency, it is not the only source of regulatory disclosure related to short selling. In additional analyses, we consider the daily disclosure of short sale trading volume started by FINRA in 2009 (Wang et al. 2020). As short volume captures the shorting flow in a certain *period* of time, it is fundamentally different from short interest, which captures shares in short positions at a *point* of time. We show that the daily short volume news contains rather limited information in predicting FINRA short interest news (i.e., R^2 is about 1%, much lower than the 20% for Markit and 60% for S3), and we confirm in untabulated analyses that our inferences are unchanged when explicitly controlling for short volume news.

This paper contributes to the short selling literature by examining the implications of private-sector providers of short selling information on regulatory short interest disclosure. Prior literature on short interest disclosure (e.g., Senchack and Starks, 1993; Pownall and Simko, 2005)

examines the informativeness of regulatory short interest disclosures in isolation without considering the potential substitutive effect of alternative sources of short interest data. More recent research primarily relies on private-sector data vendors (usually Markit) to study various dimensions of short selling (e.g., Beneish et al. 2015; Engelberg et al. 2018). We link those two streams of literature together and show that the private-sector short interest data can predict regulatory disclosure. To our knowledge, this study is the first to explicitly examine the interplay between the regulatory and private-sector short interest data sources.

More broadly, this paper adds to a small but fast-growing literature on the role of fintech in the financial markets (e.g., Goldstein et al., 2019). The rapid improvement of machine learning, big data, and artificial intelligence have profoundly changed the way we collect, analyze, interpret, disseminate, and trade on information in financial markets (Grennan and Michaely 2021). Prior research has documented various impacts of satellite images (e.g., Zhu 2020; Kang et al. 2021), transaction-level purchase data (Dichev and Qian 2022; Blankespoor et al. 2022), and granular GPS data (e.g., Noh et al. 2021). The private-sector vendors examined in this paper (Markit and S3) are also pioneers in applying new financial technologies to provide their clients with valuable information. We find evidence that, while these private-sector fintech firms exhibit the ability to predict regulatory disclosures and accelerate price discovery, regulatory disclosure continues to remain relevant even with the advent of accurate and more timely data sources.

Finally, this study has direct implications for regulatory decision making. Over the past decade, securities regulators have consistently expressed a desire to improve transparency around short selling activities, through increasing the frequency of regulatory short interest reporting (SEC 2014; FINRA 2021) and mandating detailed daily disclosures of securities lending activity (SEC 2021a). Regulators are well aware of the existence of private-sector data vendors that provide more

timely information, but are concerned that such data is incomplete, inaccurate, and too costly for retail investors. We find that the existing private-sector data can help predict regulatory disclosure, but investors do not appear to be fully aware of the information content of this private-sector data. One possible solution to this issue is increasing the frequency of mandatory short interest disclosures. Another possibility, which is perhaps less costly, is for regulators to increase investor awareness of these private sector data sources and/or pay these vendors to disseminate these data to a broader group of investors.

2 Prior literature and Institutional Background

Short selling is a unique form of trading in capital markets, and it is considered as risky and costly for various reasons ranging from unlimited downside to short squeezes. Prior literature shows that short sellers are typically sophisticated investors and overall concludes that short sellers accelerate price discovery and improve market efficiency (e.g., Bris, Goetzmann, and Zhu 2007; Saffi and Sigurdsson 2011; Boehmer and Wu 2013). However, the costs and risk of short selling also potentially motivate short sellers to engage in manipulative trading such as “bear raids” that cause the price to drop as soon as possible (e.g., SEC 2008; Soros 2009; Khanna and Mathews 2012).

Due to its controversial nature, short selling has led regulators since the 1930s to focus on how to balance information benefits against manipulative potential (e.g., SEC 2014). In particular, improving transparency (especially of the short interest) has become a key goal of short selling regulation. FINRA (The Financial Industry Regulatory Authority), a self-regulatory organization that regulates the broker-dealer industry, governs requirements for short interest reporting in the U.S. FINRA Rule 4560 requires member broker-dealers to maintain a record of total short positions in customer and proprietary firm accounts in all equity securities and to report this

information to FINRA. This information is then aggregated by FINRA at the firm level, reported to exchanges on which the securities are listed, and disseminated by the exchanges (or by FINRA for over-the-counter securities).¹ The disclosure was originally required once per month, until the SEC approved changes to increase the frequency of short interest reporting to twice per month beginning in September 2007 (Kahraman 2021).² FINRA member firms are currently required to report their short positions as of the settlement date on the 15th of each month (or the preceding business day) and as of the last business day of the month. FINRA compiles this data and provides it for publication on the 7th or 8th business day after the reporting settlement date at 4 p.m. ET. Therefore, we identify the announcement window of short interest information as the 8th or 9th day following the reporting settlement date. The timeline of settlement and announcement dates is presented in **Appendix A**.

While FINRA mandated that short interest reporting frequency be increased from once per month to twice per month in 2007, regulatory debate continues regarding the frequency of the short interest reporting regime. Following the Great Recession of 2008, market regulators around the world have debated enhancing regulatory short sale disclosures. In the U.S., the 2010 Dodd-Frank Act required the SEC to consider the merits of real-time short interest reporting and to enhance public information about the lending or borrowing of securities. This debate has been renewed by events surrounding GameStop and other “meme” stocks, in which short sellers faced steep losses due to retail traders’ attempts to drive up the security’s price and profit from short sellers having

¹ Some prior papers have studied the market reactions to short interest announcements in the early years (Senchack and Starks 1993; Pownall and Simko 2005). Regulatory short interest is used in various prior studies to capture short selling activities (e.g., Brent et al. 1990; Dechow et al. 2001) and more recently to proxy for managers’ private bad news (Bao et al. 2019).

² Two papers explore the consequences of this change in short interest reporting regime. Kahraman (2021) studies whether the increased disclosure frequency of the SEC leads to improved market efficiency and finds some evidence that greater disclosure frequency improves price efficiency around earnings announcements. Hu (2019) finds that the increased short interest disclosure frequency led to greater voluntary disclosure around short interest announcements.

to cover their positions. House lawmakers asserted that the SEC did not satisfy the mandate under the Dodd-Frank Act to provide sufficiently greater short selling disclosure, and in November 2021, the SEC (2021) proposed mandating detailed daily disclosures of securities lending activity, which is a key component of short selling (Michaels and Lim 2021). FINRA publicized these considerations in June of 2021 when it requested comments on the merits of increasing the frequency of short interest disclosure to a weekly or a daily basis and reducing the delay of regulatory short interest announcements, among other proposed changes (FINRA 2021).

As discussed above, improving the transparency of the short selling and security lending market has attracted enormous regulatory attention. It is important to note, however, that private-sector vendors have also started to provide transparency into the short selling market alongside regulatory disclosures. Accordingly, when regulators such as the SEC and FINRA weigh the costs and benefits of mandating more disclosures on short selling activities, they often mention the existence of private-sector vendors who provide more timely data (e.g., SEC 2021a; FINRA 2021). However, regulators have largely characterized these data sources as “incomplete and inaccurate” and as a source of potential harm to smaller retail investors due to their costly access.³ On the other hand, the fact that there is such a demand for more timely information and investors are willing to pay a hefty price for the access of such private-sector vendor implies that those private-sector data are somewhat valuable.

³ For example, the SEC (2021a, page 20) states that “However, the available data are incomplete, as private vendors do not have access to pricing information that reflects all transactions. This in part, reflects the voluntary submission of transaction information by subscribers to vendors and is compounded by the unknown comparability of data due to, among other things, the variability of the transaction terms disseminated, as well as how those terms are defined. As no single vendor has information for all securities lending transactions that take place, some persons pay to subscribe to multiple vendors’ systems in order to capture as much of the currently available data as they determine to purchase, which can be expensive.”

Similarly, FINRA (2021, page 11) states that “Estimates of short interest at a daily frequency based on other sources such as securities lending data are available for purchase from vendors, but may be less accurate and are not freely available.”

Given the importance of private-sector data to the regulatory debate, it is surprising there is no explicit research examining the role of private-sector short interest data on the effectiveness of regulatory disclosure and on retail investors. As a result, we fill this void in the literature and examine the following three questions in this paper: First, are private-sector short interest data informative? If so, how much? Second, does private-sector short interest data preempt the information content in the FINRA regulatory disclosure? And third, are retail investors harmed by private-sector short interest data?

3 Data and Research Design

3.1 Private-Sector Data Vendors

We focus on two prominent private-sector data vendors that provide daily short interest data. IHS Markit (“Markit”) is a financial services provider that delivers, among other things, securities lending data at the daily level. Markit lending data includes more than \$25 trillion of global securities from lending programs of over 20,000 institutional funds and is sourced directly from industry participants that lend shares to each other (e.g., brokers and institutional funds). It has been extensively used by prior accounting and finance literature (e.g., Beneish et al. 2015; Engelberg et al. 2018). This data is available from 2006 on a daily basis. The variable of interest in the Markit data, “ValueOnLoan,” measures all shares on loan as reported by each institutional client of Markit. We scale this variable by market capitalization on the same day and use it as a proxy for the level of short interest, since stocks are most often borrowed for the purpose of shorting.⁴ As of July 2021, Markit costs a minimum of \$100,000 per year for non-academic customers.

⁴ Markit’s “ValueOnLoan” can deviate from FINRA short interest due to the data generation process. Most importantly, “ValueOnLoan” captures only the borrowing that happens across brokers, not within brokers. For instance, if a customer in Brokerage A wants to borrow shares of Apple to short, then Brokerage A may look within its internal

S3 Partners (“S3”) is a financial data provider and service vendor. S3 produces research that is regularly quoted by financial publications including Barron’s, The Wall Street Journal, and The Financial Times. S3 has received enormous attention for providing daily updates of short interest for specific securities including GameStop and certain SPACs. We are among the first to use S3 data in academic research, but the usage by public press gives credence to its legitimacy. By serving as a technology/data partner to prominent institutions (e.g., Bloomberg and Nasdaq), S3 collects data on trading positions including shorts and uses a proprietary algorithm to compute and provide daily updates to firm-level short interest. This data is available from 2015. A Bloomberg terminal is required to access the S3 short interest data, which is available through an add-on. As of July 2021, S3 data costs \$9,000 per year for non-academic customers. Importantly, S3 explicitly benchmarks their short interest figures to the semi-monthly regulatory disclosures in their estimation process. Thus, we expect it to closely track the FINRA short interest.

3.2 Sample and Key Variables

For our main sample construction, we begin with all FINRA short interest available in Compustat from 2009 – 2020 and merge this dataset with CRSP to match shares outstanding. We limit our sample period to start from 2009 for three reasons: 1) Markit data is available at a daily basis starting from 2006; 2) FINRA short interest was announced once a month prior to September 2007 and twice a month afterwards; and 3) during the financial crisis (2007-2008), market reactions and changes in short interest were particularly volatile and may unduly affect inferences

inventory (e.g., from an internal derivative desk of a prime brokerage client’s long positions) to see if this demand can be met internally. If so, these shares on loan would not be reported to Markit. On the other hand, if Brokerage A does not have enough GameStop shares in its inventory to lend out, it will need to reach out to Brokerage B to facilitate the trade. In this case, these demands are reported to Markit and counted as a part of “*ValueOnLoan*.” In contrast, the mandatory short interest captures even short positions that are enabled by shares lent within the brokerage.

for the broader period. For tests including S3 data, we further limit the sample to 2016 – 2020 since S3 data prior to that period was not well populated.

Our key variables are related to the information content in regulatory FINRA short interest announcements and in private-sector short interest datasets. For the regulatory disclosure, we calculate the short interest ratio as shares held short divided by total shares outstanding, and calculate the change in short interest ratio ($Dsiratio_FINRA_{[Prior_Treg, Treg]}$) by taking the difference in short interest ratios from the prior to the current regulatory disclosure. “*Treg*” denotes the date of regulatory disclosure by FINRA twice per month, and in so doing we distinguish it from “*t*” that denotes the daily trading date as we use later. Conceptually, this measure serves as our proxy for the amount of news that is disclosed in the short interest announcement. As discussed previously, FINRA publishes on the 7th or 8th business day after the reporting settlement date at 4 p.m. ET. Thus, the market would first have time to react to this news as of the 8th or 9th business day after the settlement date. Of these two dates, we choose the first date as the date of announcement and treat the two-day window as the announcement window. As such, the short selling information would become available both on the date of announcement and on the day after. **Table 1 Panel A** presents the sample selection process. After removing duplicates, merging with CRSP, and removing observations with missing variables, we have 1,057,752 FINRA short interest announcements from 2009 to 2020 for common stocks. **Panel B** presents that the distribution by year is relatively stable across all 12 years in the sample. For our analyses, we further require these FINRA short interest announcements to have matching data from Markit and/or S3, as shown in **Table 1 Panel A** and discussed further below.

For the information content contained in private-sector short interest data, we define $Dsiratio_Private-Sector$ representing a vector of variables over a 10-day window from

$Dsiratio_Private-Sector_{[Treg-9, Treg-8]}$ through $Dsiratio_Private-Sector_{[Treg-19, Treg-18]}$, in which each variable represents the daily change in short interest relative to the date of FINRA short interest announcement. Note that for the daily private-sector data, we use a lag of at least 8 days prior to the regulatory announcement date, because as we discuss above, the regulatory disclosure measures short interest on an 8-day lag. As another measure, we also examine $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$, which captures the total change in short interest between the previous and current FINRA settlement dates. These variables are measured using both Markit and S3 data. **Appendix A** presents a timeline of our variable measurement. We present descriptive statistics in **Table 2**. $Dsiratio_FINRA_{[Prior_Treg, Treg]}$ appears evenly distributed around the mean and median of approximately zero, with the spread between the 10th and 90th percentile being approximately 0.011. $Dsiratio_Private-Sector_{[Treg-9, Treg-8]}$ and $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ follow a similar symmetric distribution around zero as expected. Throughout the paper, we winsorize all continuous variables (except returns as dependent variables) at 1% and 99% by short interest announcement cycle. **Appendix B** presents variable descriptions.

4 Results

4.1. Testing RQ1: The Informativeness of Private-Sector Short Interest Disclosures

Our first research question asks whether private-sector short interest data is informative. To answer this question, we focus on the daily news in private-sector short interest data and examine whether it: (1) can predict future FINRA short interest, and (2) is associated with concurrent stock returns on a daily basis as it is announced.

4.1.1 Private-sector short interest data and upcoming FINRA short interest

We first examine whether news disclosed in daily private-sector short interest can predict upcoming FINRA short interest news. Specifically, we regress the change in FINRA short interest

($Dsiratio_FINRA_{[Prior_Treg, Treg]}$) on daily changes in short interest disclosed by S3 and Markit in the 10 trading days leading up to the settlement date on which FINRA short interest is measured. As discussed above, the settlement date at which short interest is measured is 8 days prior to the FINRA announcement date (as presented in the timeline in **Appendix A**), and is therefore notated with the subscript “ $Treg-8$.” In other words, we examine whether the daily changes in private-sector data for each day within a regulatory settlement cycle (i.e., roughly two weeks) can predict the news contained in the upcoming FINRA announcement.

We run the following regressions and examine the R^2 and coefficients to see how well private-sector short interest data can predict regulatory disclosures. Note all regressions in this paper are at the firm-time level. As a result, we omit the firm-indicator i in all equations for brevity.

$$Dsiratio_FINRA_{[prior_Treg, Treg]} = \alpha + \sum_{t=8, \dots, 18} \beta_{[Treg-t-1, Treg-t]} * Dsiratio_Private-Sector_{[Treg-t-1, Treg-t]} + \varepsilon_{Treg} \quad (1)$$

Results of Equation (1) are presented in **Table 3**. Column 1 (3) reports the relation between the daily news from Markit (S3) short interest data in a regulatory settlement cycle and news from the upcoming regulatory short interest disclosure. We find that the news contained in both Markit and S3 data in a regulatory settlement cycle can predict changes in FINRA short interest, with all regressors loading significantly at the 1% or higher level. Further, we find that S3 overall does a much better job of predicting FINRA short interest with an R^2 of 67.2% compared to Markit’s R^2 of 21.7%. This is hardly surprising, as S3 actively benchmarks to FINRA short interest as described in Section 2.⁵

⁵ To confirm that this difference is not due to differences in time frame, we examine the explanatory power of Markit and S3 using only observations in which both sets of data are available. Our inferences hold very similarly, with S3 having an R^2 of 67.5% and Markit having an R^2 of 21.1%.

As an alternative way to capture the overall information content contained in the private-sector short interest data, we aggregate all daily news in a regulatory settlement cycle together by using $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ as the sole regressor in the regression. Columns 2 and 4 report the results using Markit and S3 data, respectively. The inferences are the same as those in Columns 1 and 3 that (1) both private-sector solutions predict news in upcoming regulatory short interest disclosure, and (2) S3 has much better predictive power than Markit data. Overall, our results provide strong evidence that a significant portion of the news contained in regulatory short interest disclosures can be predicted by private-sector short interest data.

4.1.2 Private-sector short interest data and concurrent stock returns

Next, we examine the informativeness of private-sector short interest data by studying its association with current returns. As short sellers are sophisticated investors, and an increase in the short interest conveys bad news about a firm's fundamentals, prior literature has documented a robust negative relationship between increases in short interest and concurrent returns, as predicted by Diamond and Verrecchia (1987). To the extent that the private-sector short interest data captures meaningful short selling activities, we would expect that the increase in daily short interest would be associated with negative announcement returns.

To test this prediction, we sample all daily observations from CRSP and retain samples from 2009 – 2020 and 2016 – 2020 with matching Markit and S3 data, respectively. It is worth noting that a decrease in short interest does not necessarily mean good news, because it could be driven by a decrease in lendable shares (Beneish et al. 2015), which could in turn be driven by bad news. Indeed, as Kahraman (2021) shows in her Figure 2, while firms with big increases in short interest witness significant price declines, firms with big decreases in short interest also witness negative returns (although insignificant). We take two approaches to circumvent this non-linearity

issue. First, we focus on the subsample of short interest announcements reporting increases in short interest. Second, we define a vector of indicators ($Big_Inc_Private-Sector_{[t-1, t]}$) that take the value of one if the daily change in private-sector short interest ratio is in the top decile.⁶ We make sure that the changes in short interest are available to investors on that day for both Markit and S3.⁷ We separately regress $AbRet_t$, which is the daily return adjusted by a size-decile-matched portfolio (Sloan 1996), on the $Dsratio_Private-Sector_{[t-1, t]}$ and $Big_Inc_Private-Sector_{[t-1, t]}$ as in the following equation:

$$AbRet_t = \alpha + \beta * Dsratio_Private-Sector_{[t-1, t]} \text{ or } Big_Inc_Private-Sector_{[t-1, t]} + \varepsilon_t \quad (2)$$

Table 4 presents the results of Equation (2). We find strong evidence that increases in the daily private-sector short interest are associated with negative returns. This pattern holds for both Markit and S3, and based on both a continuous variable of raw increase and an indicator variable of large increase in short interest. Specifically, the coefficient of $Dsratio_Private-Sector_{[t-1, t]}$ is -0.1377 (-0.1777) in Column 1 (3), indicating that a one-percentage-point increase in short interest change from Markit (S3) is associated with 14 (18) basis point lower return on the trading date when the news becomes available, which is equivalent to a 35% (45%) annualized return. The coefficient of $Big_Inc_Private-Sector_{[t-1, t]}$ is -0.0005 (-0.0003) in Column 2 (4), indicating that being in the top decile of short interest increase relative to the prior trading day is associated with a 5 (3) basis point abnormal return, which is equivalent to a 12.6% (7.6%) annualized return. Taken together, those results in Table 4 show that private-sector short interest data is associated with

⁶ Note it is also common in the prior literature to focus on large increases in short interest to show the informativeness of short interest (i.e., Desai et al. 2002; Pownall and Simko 2005).

⁷ The daily Markit shares on loan data is disclosed to subscribers at 8am EST on the following date (e.g., the shares on loan data for Apple on March 15 would be disclosed to Markit subscribers at 8am EST on March 16). We adjust the data date forward by one day to ensure that the data is actually available to investors on the daily data date in our sample. Daily S3 short interest data is available on the data date. However, it is possible that the market might react to the private-sector news ahead of their official release in subscribers' data feed. For example, those big brokers (such as Goldman) who are the data sources of Markit, know the loan information one day in advance. As a result, their clients or proprietary trading desk can trade on such information inferred from the big increase in loans.

daily announcement returns, providing further support that this data is informative to market participants.

4.2 Testing RQ2: Private data preempting the information contents in regulatory disclosure

The prior section presents evidence that private-sector short interest data can (1) predict regulatory FINRA short interest news, and (2) is associated with concurrent stock returns. Those two findings lead to a natural question of whether the daily private-sector short interest news preempts the market reaction to bi-weekly regulatory disclosure. If the private-sector short interest news fully preempts the regulatory disclosures, there should be no market reaction to the latter. On the other hand, if there is no preemption, or only partial preemption, investors should still react to the bi-weekly regulatory disclosures. Partial preemption could occur if the regulatory disclosures contain relevant information beyond private-sector disclosures (which seems likely), and/or investors underreact to the daily private-sector disclosures. We examine preemption directly below.

4.2.1 The Market reactions to the regulatory disclosure

We begin by confirming whether there are significant market reactions to regulatory short interest disclosures in our sample period.⁸ Figure 1 presents graphical evidence. Specifically, we focus on the subsample reporting the top-decile change in short interest ratio (i.e., $Dsiratio_FINRA_{[prior_Treg, Treg]}$), and examine the average cumulative abnormal returns for the [-5, +5] window around short interest regulatory announcements. If regulatory short interest disclosures are informative and not being fully preempted by private-vendor data, then one should

⁸ Note Senchack and Starks (1993) find a negative market reaction to unexpectedly high increases short-interest using a sample from 1980 to 1986, Pownall and Simko (2005) and Kahraman (2021) document a similar finding for a larger sample of firms from 1989 to 1998, and from 2003 to 2012, respectively. However, it is still important for us to confirm this finding for our more recent sample because, as Kahraman (2021) points out, “market reactions to short-interest announcements might be different in more recent periods, for instance, as a result of the availability of more information on short-selling activity.”

observe a significant decrease in the abnormal returns following announcement of large increases in short interest. We find this to be the case, with roughly -20bps combined reaction during the announcement window [0, +1]. The negative reaction appears largely concentrated in the announcement window, as would be expected if this reaction is driven by the announcement. This result confirms the pattern documented in the prior literature using earlier data that regulatory short interest announcements are informative (Senchack and Starks 1993; Pownall and Simko 2005) and not being fully preempted (e.g., Kahraman 2021).

In addition to graphical evidence in Figure 1, we also use the same format as in Table 4 and Equation (2), and separately regress $CAR_{[Treg, Treg+1]}$, which is the cumulative abnormal returns (adjusted by size-decile portfolios) during the announcement window [0, +1], on $Dsiratio_FINRA_{[prior_Treg, Treg]}$ and $Big_Inc_FINRA_{[prior_Treg, Treg]}$, as in the following equation:

$$CAR_{[Treg, Treg+1]} = \alpha + \beta * Dsiratio_FINRA_{[prior_Treg, Treg]} \text{ or } Big_Inc_FINRA_{[prior_Treg, Treg]} + \varepsilon_{Treg} \quad (3)$$

Table 5 presents the results of Equation (3). We find that the coefficients on $Dsiratio_FINRA_{[prior_Treg, Treg]}$ and $Big_Inc_FINRA_{[prior_Treg, Treg]}$ are highly significantly negative in Columns 1 and 2, when we use the full sample from 2009 to 2020. The coefficient of -0.0025 (t = -8.242) in Column 2 means that firms with top-decile increases in short interest relative to the previous FINRA disclosure experience returns that are 25 basis points lower than other firms in the two days starting from the FINRA announcement date. This magnitude is comparable with the evidence in Figure 1. In Columns 3 and 4, we limit the sample to only 2016 to 2020 when we have S3 data, and all the inferences are the same with a somewhat smaller magnitude and lower significance. Taken together, these results confirm the strong negative market reactions to large increases in FINRA short interest disclosure.

4.2.2 Private-sector short interest data and investors' underreaction prior to regulatory disclosure

Next, we explore one reason for why the private-sector data does not fully preempt the market reaction to regulatory disclosure. To the extent that private-sector data is informative (as we have shown), it is possible that the market initially underreacts to the content of the private-sector data until the information is more widely disseminated through regulatory disclosure. We explicitly examine such possibility in this subsection in two ways.

First, we study the relation between news in private-sector short interest data and abnormal return around the upcoming FINRA short interest announcements. To the extent that the market has fully incorporated news contained in private-sector short interest data, we would not expect a significant relation between abnormal returns around FINRA announcements and prior news in private-sector short interest data (which was available 8 trading days prior to the regulatory announcement). **Table 6 Panel A** presents the results. As in earlier tables, we report results using the continuous variable for the subsample with short interest increases and using the large-increase indicator for the full sample for both Markit and S3. Across all four columns, we find that increases in short interest from private-sector data significantly predict negative returns in the upcoming FINRA short interest announcements, indicating that the market does not fully incorporate the news contained in private-sector data in a timely manner.⁹

Second, to better quantify the extent of market's underreaction to the news contained in private-sector short interest data, we explore how much of the abnormal returns to the FINRA

⁹ Using two-day placebo returns surrounding the FINRA announcement window as alternative dependent variables, we find that results for S3 become insignificant in the placebo periods. Using Markit data, the placebo returns in the two days following the announcement window are significantly (but less strongly) negatively correlated with increases in private-sector short interest, which may imply that our measure underestimates the extent to which future returns can be predicted using Markit data.

news could be predicted or explained by private-sector news. Specifically, we regress abnormal returns to FINRA announcement on the short interest news from FINRA ($Dsratio_FINRA_{[prior_Treg, Treg]}$) as in Table 5. We then label the fitted value as $Pred_CAR_{[Treg, Treg+1]}$, which represents the predicted cumulative abnormal return around the FINRA announcements given contemporaneously-released FINRA short interest news for that date. Then we regress $Pred_CAR_{[Treg, Treg+1]}$ on prior private-sector short interest news. We are interested in the R-square of this regression, as it tells us how much of the return reaction to FINRA news can be predicted or explained by previously-released private-sector data.

Table 6 Panel B reports the results. In Column 1 (2), we include only the news contained in Markit (S3) short interest data, and the R-Square is 16.2% (45.6%). The fact that the R-Square of using S3 is much higher than that of using Markit is intuitive, because the S3 data does a better job of tracking the FINRA short interest, as we show in Table 3. In Column 3, we include the news from both Markit and S3 short interest data, and the R-Square increases slightly relative to Column 2 to 47.8%.

Taken together, we find that about half of the market reaction to FINRA short interest news can be predicted by previously announced daily short interest news from private-sector vendors. We believe this result has important implications. First, it suggests that although the daily private sector data is informative with respect to the short interest, investors do not appear to fully impound this information into stock prices. In other words, the market appears to be underreacting to the news contained in private-sector short interest data. Second, private-sector data (in this case, particularly S3) can reliably predict both FINRA short interest as well as the market reactions to FINRA announcements. Consequently, less costly alternatives to mandating more timely short-interest disclosure through broker-dealers might include: a) increasing investor awareness of

already-provided private-sector data; and/or b) paying private vendors to disseminate it more widely.

4.2.3 Private-sector short interest data and price discovery prior to regulatory disclosure

So far, we have presented evidence that (1) private-sector short interest data are informative and lead to price reaction by investors, but (2) the market still reacts to regulatory short interest disclosure (due to investor under-reaction). The first result implies at least some partial preemption of the market's reaction to regulatory disclosure, while the second confirms this preemption is not complete. In this subsection, we directly examine the extent to which private-sector data helps the market to incorporate the news in regulatory disclosure more quickly.¹⁰ Intuitively, if private-sector data preempts the information content in regulatory short interest disclosure, we expect that the market should react sooner when the private-sector sources contain more information. We use two approaches to measure how much faster the market incorporates regulatory disclosure news.

First, we examine the intraperiod timeliness (*IPT*) of returns around regulatory short interest announcements, following Butler et al.'s (2007) methodology. Intuitively, we use the area under the return curve to gauge how efficient the price responds to news in a certain period. Specifically, we measure IPT_{Treg} over the period from the previous FINRA short interest settlement date to two days past the current announcement date as illustrated in Appendix A. We choose this time horizon in order to match the news that is released by regulatory announcements and by the private-sector data. Since Markit and S3 data are updated and disclosed daily, the news from Markit and S3 that corresponds to the bi-monthly FINRA announcement would start being disclosed following the previous settlement date, which begins approximately 18 days prior to the current announcement date. We choose to extend the horizon up to two days after the FINRA

¹⁰ We note that the previous section also establishes that market reaction to regulatory disclosure is not completely preempted. In this section, we focus on whether any of the reaction is preempted, and to what extent.

announcement date in order to capture all market response associated with the regulatory disclosure. Conceptually, IPT_{Treg} captures the amount of news that is preempted by private-sector data before this news is eventually disclosed by the regulatory announcement.

Second, while intraperiod timeliness is a popular measure capturing the speed of price incorporating news in regulatory short interest disclosure, one potential downside is that the magnitude of IPT_{Treg} is not intuitive to interpret. To address this issue, we borrow an approach from Lee and Zhu (2022) and use an alternative measure capturing the portion of returns in the 20-day window (i.e., from the previous settlement date to two trading days after the current announcement date) realized prior to the current announcement date. Intuitively, when the information content in private-sector short interest data helps the market to preempt the news in regulatory disclosure, we would see the returns *after* the regulatory disclosure become less important relative to the returns prior to the disclosure. Accordingly, we define $Perc_Ret_{Treg}$ as measured as one minus the portion of returns earned during the FINRA announcement window, which in turn is defined as the buy-and-hold return over days $[0, +1]$ around FINRA announcement date divided by buy-and-hold return over the period from 18 days prior to the current announcement date (approximately the date of the previous settlement date) to two days past the current announcement date.

If the private-sector data helps to incorporate the news in short interest into prices in a timelier manner, we would expect that both IPT_{Treg} and $Perc_Ret_{Treg}$ increase with the amount of news that is preemptively disclosed by private-sector data, measured by $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ or $Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]}$, again following the format in Tables 3 and 6. We test these predictions with the following regressions.¹¹

¹¹ We limit the sample to observations that see a greater than 2% absolute buy-and-hold return during this window in order to omit observations without sufficient change in stock prices over this time period. But the results are qualitatively similar without this requirement, and generally become stronger when using 1% as the threshold.

$$\begin{aligned}
IPT_{Treg} \text{ or } Perc_Ret_{Treg} &= \alpha + \beta * Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]} \\
&\text{ or } Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]} + \varepsilon_{Treg}
\end{aligned} \tag{4}$$

Table 7 presents the results of Equation (4). Panel A presents results using IPT_{Treg} and Panel B presents results using $Perc_Ret_{Treg}$. Overall, we find that the short interest news revealed in private-sector sources accelerate the price discovery of news contained in regulatory disclosures. This pattern holds for both Markit and S3, and based on both a continuous variable of the raw increase and an indicator variable of a large increase in short interest. Specifically, we find that IPT_{Treg} increases with $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$, the amount of news that is preemptively disclosed by Markit (Coeff. = 23.6856; t = 9.928) and S3 (Coeff. = 39.7206; t = 9.601) as shown on Columns (1) and (3), respectively. In terms of magnitude, a one standard deviation increase in $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ based on Markit (S3) data is associated with a 0.199 (0.409) increase in IPT_{Treg} , which is roughly equal to 1.9% (3.9%) of the standard deviation of IPT_{Treg} in magnitude. Similarly, we also find that IPT_{Treg} increases with big increase in short interest from Markit (Coeff. = 0.5926; t = 12.285) and S3 (Coeff. = 0.8308; t = 9.877) as shown on Columns (2) and (4), respectively. In terms of magnitude, a large change in short interest based on Markit (S3) data is associated with an increase average in IPT_{Treg} from about 10.60 (10.57) to 11.19 (11.40), or about 5.6% (7.8%) of the standard deviation of IPT_{Treg} .

In Panel B, we find that $Perc_Ret_{Treg}$ similarly increases with $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$, the amount of news that is preemptively disclosed by Markit (Coeff. = 0.4693; t = 4.426) and S3 (Coeff. = 0.7725; t = 4.136) as shown on Columns (1) and (3), respectively. In terms of magnitude, a one standard deviation increase in $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ based on Markit (S3) data is associated with a 0.394% (0.796%) increase in $Perc_Ret$, which is roughly equal to 0.76% (1.53%) of the standard deviation of $Perc_Ret_{Treg}$ in

magnitude. Similarly, we also find that $Perc_Ret_{Treg}$ increases with large increases in short interest from Markit (Coeff. = 0.0074; t = 3.470) and S3 (Coeff. = 0.0130; t = 3.359) as shown on Columns (2) and (4), respectively. In terms of magnitude, a large change in short interest based on Markit (S3) data is associated with an increase average in $Perc_Ret_{Treg}$ from about 91.18% (90.95%) to 91.92% (92.25%), or about 0.8% (1.4%) of its standard deviation, which is quite small.¹²

In summary, we find that the increases in short interest captured by private sector data are associated with higher IPT_{Treg} and a greater proportion of returns earned prior to bi-monthly regulatory announcements. This suggests that private-sector short interest data helps prices reflect the news in regulatory short interest more quickly. However, the economic magnitude of this effect is not large.

4.3 Testing RQ3: The implications for retail investors

One key motivation for regulators to consider expanding FINRA disclosure is the concern that smaller retail investors do not have the access to the private-sector data. Such concern is warranted because we show that private-sector data can predict the news content of, and preempt the market reactions to, upcoming regulatory disclosures. In addition, the costs of accessing such private-sector data (i.e., Markit and S3 data of \$100,000 and \$9,000 per year, respectively) are likely too high for retail investors to afford on a regular basis.¹³ These investors in turn may have to wait until short interest becomes more broadly available through regulatory disclosure in order to trade based on this information, which at that point may have become stale.¹⁴ We therefore test

¹² The economic significance is higher when we measure the price discovery using IPT_{Treg} than when we use $Perc_Ret_{Treg}$. One reason is that IPT_{Treg} distinguishes the price discovery in different days before the regulatory announcements, but $Perc_Ret_{Treg}$ pools them together.

¹³ To further support this possibility, our conversation with a few small hedge funds indicate that it is uncommon even for small funds to subscribe to Markit or S3 data.

¹⁴ For instance, a website commonly referred to by retail investors, Yahoo Finance, only displays FINRA short interest. Thus, retail investors may trade on short interest information only once it is updated twice per month on websites such as Yahoo Finance.

whether retail investors are worse off due to the private-sector short interest data and directly examine this concern. Specifically, we study whether the information content in private-sector short interest data enhance or damage retail investors’ trading profitability around regulatory short interest disclosures.

We follow Boehmer et al. (2021) to compute the marketable retail order flow from TAQ data and calculate the firm-day level marketable retail order imbalance based on share volume (i.e., the extent to which buy orders exceed the sell orders).¹⁵ The sample is comprised of daily observations of returns and retail order flow data from 2009 – 2014 due to data availability. As S3 data is not available in this time period, the private-sector data only refers to Markit short interest in this subsection. We calculate $Mroibvol$ as the firm-day level marketable retail order volume imbalance, computed with the following formula from Boehmer et al. (2021):

$$Mroibvol_{(j,t)} = \frac{Mrbvol(j,t) - Mrsvol(j,t)}{Mrbvol(j,t) + Mrsvol(j,t)}$$

We evaluate whether retail investors are hurt due to the information content contained in private-sector short interest data. Our general approach is to investigate whether the profitability of retail trading after regulatory disclosures is affected by the informativeness of private-sector data. Following Boehmer et al. (2021), we interpret the positive association of retail order imbalance and subsequent returns as evidence that retail trading is profitable. We first confirm that

¹⁵ To identify retail order flow, Boehmer et al. (2021) develop a measure exploiting a US regulation (Regulation 606T). Unlike institutional order flow, retail order flow can receive price improvement for market orders, measured as a small fraction of a cent per share. This allows to identify marketable retail price-improved orders from TAQ data because trades that execute at share prices with fractional pennies originate from retail investors. These trades take place off-exchange and are reported to a Trade Reporting Facility (TRF) and are included in the TAQ “consolidated tape” with exchange code “D.” Retail sell transactions tend to be reported at prices that are just above a round penny, whereas retail buy transactions tend to be reported at prices just below a round penny. Institutional trades cannot receive this form of fractional penny price improvement.

retail trading is overall profitable around regulatory disclosure (untabulated), consistent with Boehmer et al. (2021). We then conduct two sets of cross-sectional analyses.

First, we exploit whether the change in short interest ratio calculated using the Markit data is in the same decile as change in short interest ratio calculated using FINRA disclosure. We then create an indicator of $Markit_Match_{Treg}$ equal to one if the deciles of both changes are the same. In other words, $Markit_Match_{Treg}$ captures scenarios where Markit short interest news is a good predictor of regulatory short interest news. If retail investors are disadvantaged due to the information content contained in private-sector data, the profitability of retail trading would be lower when private-sector short interest news can perfectly predict the decile ranks of regulatory short interest news. Table 8 Panel A reports the results. Using both subsample analyses (Columns 1 and 2) and interaction analyses (Column 3), we do not find significant evidence that retail investors' trading profitability is affected by whether private-sector data predict regulatory data perfectly or not. Alternatively, retail investors may be disadvantaged in the days leading up to the FINRA disclosure. As a robustness test, we measure the retail trading profitability prior to the regulatory disclosure, regressing returns from 5 days to 1 day prior to announcement on retail order imbalance 6 days prior to announcement (untabulated). Our inferences that retail profitability appears unaffected holds using this alternative design.

Second, we directly examine the impact of news from private-sector short interest data on retail investors' trading profitability. To do so, we split the firm-day level sample based on whether the daily news in private-sector data is in the top decile or not. If retail investors are hurt due to the information content contained in private-sector data, this impact would be more severe when the Markit daily short interest data contains more news. Table 8 Panel B present the results. Using both subsample (Columns 1 and 2) and interaction analyses (Column 3), we do not find that the

retail trading is less profitable when there is a big increase in Markit short interest. If anything, there is some weak evidence that retail trading profitability is higher when Markit reports large increase in short interest (i.e., bigger news).¹⁶

Taken these analyses together, we do not find evidence that retail investors are hurt by the information content contained in private-sector short interest data. One explanation could be that retail investors trade in a manner that is uncorrelated with the short interest news disclosed by private-sector vendors. Consistent with this possibility, we find that the retail order imbalance is uncorrelated with concurrent news disclosed by Markit both one day before and one day after the FINRA announcement (untabulated). Thus, although it is possible some sophisticated investors are net sellers of firms with large increases in short interest based on private-sector data, it does not appear that retail investors are, on average, net buyers of such firms leading up to FINRA announcements. These results have important regulatory implications. As one key motivation for regulators to expand short interest disclosure is to make sure retail investors do not get hurt due to their lack of access to costly private-sector data, our results suggest that this concern may be unwarranted.

4.4 Additional analyses on short volume data

While our primary focus is on short interest disclosure, we acknowledge this is only one aspect of the debate on short selling transparency. For example, the SEC recently proposed rules involving security lending disclosure (SEC 2021a) and short position reporting by institutional investors (SEC 2022). Closely related to the short interest disclosure, FINRA started publishing

¹⁶ The interaction term in Column 3 is marginally significant (Coeff. = 0.0006; $t = 1.718$). One possibility why retail buy-sell imbalance is associated with higher subsequent returns on days with big increase in short interest is that excessive buying from retail investors could trigger “short squeezes” (i.e., short sellers are forced to cover their short positions, pushing prices even higher and creating a liquidity spiral) for highly-shorted firms. An extreme case of this type would be the GameStop (GME) short squeeze in January 2021 (e.g., Allen et al. 2021; SEC 2021b).

short sale trading volume data on September 30, 2009. According to Wang et al. (2020), even though such data is publicly available in real time, the shorting volume is a significant predictor of negative future stock returns during their sample period of 2010-2015, suggesting that the market does not pay sufficient attention to such disclosure.

It is important to note the major difference between short interest and short volume: while short interest reflects the number of shares that are in outstanding short positions at a *point* of time (like a balance sheet item), short volume reflects the number of shares that are shorted in a *period* of time (like an income statement item). As short sellers usually cover shares rather quickly (e.g., median of eight days according to Diether and Werner (2011)), Blau et al. (2011) show that the correlation between these short interest and monthly-aggregated short volume is only 45%. As such, the availability of short volume data does not alleviate regulators' concern on short interest disclosure (e.g., FINRA 2021; SEC 2014). In Table 9, we re-run Equation (1) and find that the R^2 is only about 1% when we use similarly-constructed short volume news to predict FINRA short interest news, much lower than the roughly 20% for Markit and 60% for S3 as reported in Table 3.¹⁷ As a result, we do not examine short volume disclosure in our main analyses. Nevertheless, in untabulated analyses we confirm that all our inferences regarding private-sector data remain unchanged when we explicitly control for short volume news.

5 Conclusion

Over the past decade, securities regulators have consistently expressed a desire to improve transparency around short selling activities. Recently, in light of the “meme stock” craze and concerns about market manipulation by short sellers, regulators have proposed significant changes to the short selling disclosure environment. For example, regulatory disclosure of the extent to

¹⁷ News is measured as the FINRA short volume for each day. We alternatively measure news as the *change* in FINRA short volume, and find that the explanatory power decreases further.

which a stock is sold “short” is currently mandated twice per month. In June 2021, FINRA requested comments on mandating disclosure of the regulatory short interest on a weekly or even daily basis. In addition, in November 2021, the SEC proposed mandating detailed daily disclosures of securities lending activity, which is a key component of short selling.

Against this regulatory backdrop, daily short interest and lending data from private, third-party vendors has arisen to meet market demand for short selling activity. In this paper, we examine this daily data and find that it strongly predicts bimonthly regulatory short interest. Consequently, the daily private sector data significantly preempts the market reaction to regulatory short interest announcements, but the market still seems to substantially underreact to the information contained in the private-sector short interest data. Although access to this data is relatively expensive, we find no evidence that retail investors are harmed in their trades around short interest disclosures.

These findings are timely and have implications for regulators. First, our evidence suggests that mandating increased frequency of regulatory short interest, which would be quite costly, may not be necessary. Second, despite regulators’ concerns that the security lending data collected by private vendors is incomplete and inaccurate, we find it is incrementally informative to market participants beyond the regulatory short interest.

References

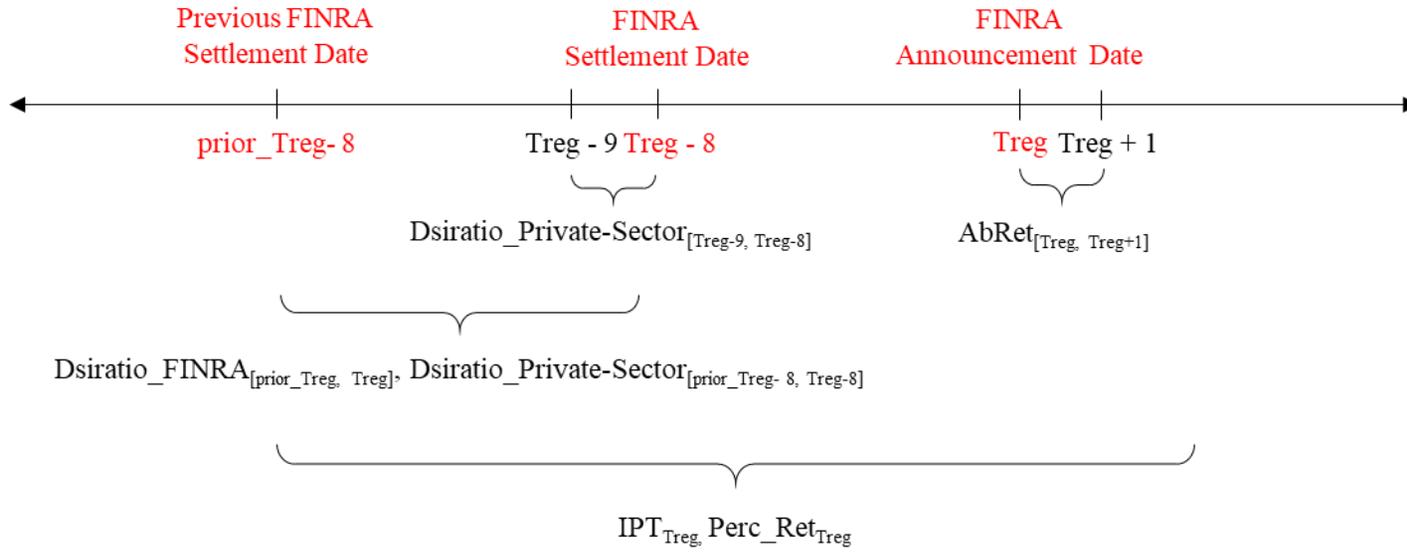
- Allen, F., Haas, M., Nowak, E., Pirovano, M., & Tengulov, A. (2021). Squeezing shorts through social media platforms. *Swiss Finance Institute Research Paper*, (21-31).
- Bao, D., Kim, Y., Mian, G. M., & Su, L. (2019). Do managers disclose or withhold bad news? Evidence from short interest. *The Accounting Review*, 94(3), 1-26.
- Beneish, M. D., Lee, C. M., & Nichols, D. C. (2015). In short supply: Short-sellers and stock returns. *Journal of accounting and economics*, 60(2-3), 33-57.
- Blau, B. M., Van Ness, B. F., & Van Ness, R. A. (2011). Short Interest vs. Short Selling. Working paper.
- Blankespoor, E., Hendricks, B. E., Piotroski, J., & Synn, C. (2022). Real-time revenue and firm disclosure. *Review of Accounting Studies*, 1-38.
- Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2021). Tracking retail investor activity. *The Journal of Finance*, 76(5), 2249-2305.
- Boehmer, E., & Wu, J. (2013). Short selling and the price discovery process. *The Review of Financial Studies*, 26(2), 287-322.
- Bris, A., Goetzmann, W. N., & Zhu, N. (2007). Efficiency and the bear: Short sales and markets around the world. *The Journal of Finance*, 62(3), 1029-1079.
- Brent, A., Morse, D., & Stice, E. K. (1990). Short interest: Explanations and tests. *Journal of Financial and Quantitative Analysis*, 25(2), 273-289.
- Butler, M., Kraft, A., & Weiss, I. S. (2007). The effect of reporting frequency on the timeliness of earnings: The cases of voluntary and mandatory interim reports. *Journal of Accounting and Economics*, 43(2-3), 181-217.
- Dechow, P. M., Hutton, A. P., Meulbroek, L., & Sloan, R. G. (2001). Short-sellers, fundamental analysis, and stock returns. *Journal of financial Economics*, 61(1), 77-106.
- Desai, H., Ramesh, K., Thiagarajan, S. R., & Balachandran, B. V. (2002). An investigation of the informational role of short interest in the Nasdaq market. *The Journal of Finance*, 57(5), 2263-2287.
- Dichev, I. D., & Qian, J. (2022). The benefits of transaction-level data: The case of NielsenIQ scanner data. *Journal of Accounting and Economics*, 101495.
- Diamond, D. W., & Verrecchia, R. E. (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics*, 18(2), 277-311.
- Diether, K. B., & Werner, I. M. (2011). When constraints bind. *Charles A. Dice Center Working Paper*.
- Engelberg, J. E., Reed, A. V., & Ringgenberg, M. C. (2018). Short-selling risk. *The Journal of Finance*, 73(2), 755-786.
- FINRA. (2021). Regulatory Notice 21-19: FINRA Requests Comment on Short Interest Position Reporting Enhancements and Other Changes Related to Short Sale Reporting. <https://www.finra.org/sites/default/files/2021-06/Regulatory-Notice-21-19.pdf>.
- Goldstein, I., Jiang, W., & Karolyi, G. A. (2019). To FinTech and beyond. *The Review of Financial Studies*, 32(5), 1647-1661.
- Grennan, J., & Michaely, R. (2021). Fintechs and the market for financial analysis. *Journal of Financial and Quantitative Analysis*, 56(6), 1877-1907.
- Hu, D. 2019. Does the Public Availability of Market Participants' Trading Data Affect Firm Disclosure? Evidence from Short Sellers. Working Paper. Available at <https://ssrn.com/abstract=3089560>.

- Kang, J. K., Stice-Lawrence, L., & Wong, Y. T. F. (2021). The firm next door: Using satellite images to study local information advantage. *Journal of Accounting Research*, 59(2), 713-750.
- Kahraman, B. (2021). Publicizing arbitrage. *Journal of Financial and Quantitative Analysis*, 56(3), 789-820.
- Khanna, N., & Mathews, R. D. (2012). Doing battle with short sellers: The conflicted role of blockholders in bear raids. *Journal of Financial Economics*, 106(2), 229-246.
- Lee, C., & Zhu, C. (2022). Active funds and bundled news. *The Accounting Review*, 97(1), 315-339.
- Langley, K. & Banerji, G. (2022). Bearish Bets Against Markets Are Surging. *Wall Street Journal*. February 22, 2022. Available at <https://www.wsj.com/articles/bearish-bets-against-markets-are-surging-11645525800>.
- Michaels, D., & Lim, D. (2021). GameStop Frenzy Prompts SEC to Weigh More Short Sale Transparency. *Wall Street Journal*. February 21, 2021. Available at <https://www.wsj.com/articles/gamestop-frenzy-prompts-sec-to-weigh-more-short-sale-transparency-11613593827>.
- Noh, S., So, E. C., & Zhu, C. (2021). Financial Reporting and Consumer Behavior. Available at SSRN 3932590.
- Pownall, G., & Simko, P. J. (2005). The information intermediary role of short sellers. *The Accounting Review*, 80(3), 941-966.
- Saffi, P. A., & Sigurdsson, K. (2011). Price efficiency and short selling. *The Review of Financial Studies*, 24(3), 821-852.
- SEC. (2008). SEC Halts Short Selling of Financial Stocks to Protect Investors and Markets. available at <https://www.sec.gov/news/press/2008/2008-211.htm>.
- SEC. (2014). Short Sale Position and Transaction Reporting. <https://www.sec.gov/dera/reportspubs/special-studies/short-sale-position-and-transaction-reportingpdf.html>.
- SEC. (2021a). Proposed Rule: Reporting of Security loans. Available at: <https://www.sec.gov/rules/proposed/2021/34-93613.pdf>.
- SEC. (2021b). Staff Report on Equity and Options Market Structure Conditions in Early 2021. Available at: <https://www.sec.gov/files/staff-report-equity-options-market-struction-conditions-early-2021.pdf>.
- SEC. (2022). Proposed Rule: Short Position and Short Activity Reporting by Institutional Investment Managers. Available at <https://www.sec.gov/rules/proposed/2022/34-94313.pdf>.
- Senchack, A. J., & Starks, L. T. (1993). Short-sale restrictions and market reaction to short-interest announcements. *Journal of Financial and quantitative analysis*, 28(2), 177-194.
- Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, 289-315.
- Sloan, R. (2009). Don't blame the shorts: why short sellers are always blamed for market crashes and how history is repeating itself. McGraw Hill Professional.
- Soros, G. (2009). The game changer. *Financial Times*. January 29, 2009.
- Terazono, E. (2022). Short sellers tuck into Beyond Meat. *Financial Times*. January 12, 2022. Available at <https://www.ft.com/content/6e97ac14-5ccb-4643-b3d6-c3592b241032>.
- Wang, X., Yan, X. S., & Zheng, L. (2020). Shorting flows, public disclosure, and market efficiency. *Journal of Financial Economics*, 135(1), 191-212.

- Wienberg, C. (2022). Hedge Funds Made \$1.25 Billion Profit on Europe's Top Short Ambu. *Bloomberg*. May 12, 2022. Available at <https://www.bloomberg.com/news/articles/2022-05-12/hedge-funds-made-1-25-billion-profit-on-europe-s-top-short-ambu>.
- Zhu, C. (2019). Big data as a governance mechanism. *The Review of Financial Studies*, 32(5), 2021-2061.

Appendix A Timing of FINRA Short Interest Disclosures

This appendix shows the typical timeline of FINRA short interest settlement (i.e., the date on which short interest is determined) and announcement dates. The timing of key variable construction is presented.



Appendix B Variable Definitions

Variable definitions are presented.

Variable	Definition
$Dsiratio_FINRA_{[prior_Treg, Treg]}$	Change in short interest ratio (measured as number of shares held short divided by shares outstanding) from the previous FINRA announcement to the current FINRA announcement.
$Big_Inc_FINRA_{[prior_Treg, Treg]}$	Indicator variable capturing the largest decile of increase in short interest ratio from the previous FINRA announcement to the current FINRA announcement.
$Dsiratio_Private-Sector_{[Treg-x-1, Treg-x]}$	Change in daily short interest ratios using Markit or S3 data between Treg-x-1 and Treg-x, where Treg is the mandatory announcement date. This variable is also referred to as Dsiratio_Markit or Dsiratio_S3 when necessary.
$Dsiratio_Private-Sector_{[t-1, t]}$	Change in daily short interest ratios using Markit or S3 data between t-1 and t, where t indexes a firm-day observation.
$Big_Inc_Private-Sector_{[Treg-x-1, Treg-x]}$	Indicator variable capturing the largest decile of increase in short interest from day Treg-x-1 to Treg-x using Markit or S3 data, where Treg is the mandatory announcement date. This variable is also referred to as Big_Inc_Markit or Big_Inc_S3 when necessary.
$Big_Inc_Private-Sector_{[t-1, t]}$	Indicator variable capturing the largest decile of increase in short interest from t-1 to t, where t indexes a firm-day observation.
$Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$	Change in short interest ratios using Markit or S3 data between the current and previous mandatory settlement dates, where Treg is the mandatory announcement date. This variable is also referred to as Dsiratio_Markit or Dsiratio_S3 when necessary.
$AbRet_t$	Daily abnormal return adjusted for the weighted average daily returns of firms in the same size decile.
$CAR_{[Treg, Treg+1]}$	FINRA announcement window return, measured as cumulative abnormal daily return adjusted for the weighted average daily returns of firms in the same size decile.
IPT_{Treg}	Intraperiod timeliness following Butler et al. (2007), measured over the period from 18 days prior to the current announcement date (approximately the date of the previous settlement date) to two days past the current announcement date. The sample is limited to observations with buy-and-hold returns of greater than 1% over this time period.
$Perc_Ret_{Treg}$	One minus the percentage of returns earned during the FINRA announcement window, measured as the buy-and-hold return over days [0, +1] around FINRA announcement date divided by buy-and-hold return over the period from 18 days prior to the current announcement date (approximately the date of the previous settlement date) to two days past the current announcement date.
$Pred_CAR_{[Treg, Treg+1]}$	Fitted value of regressing $CAR_{[Treg, Treg+1]}$ on $Dsiratio_FINRA$ for a subset of sample with increases in FINRA short interest.

Retail Order Imbalance _[Treg, Treg+1]	Difference between retail purchase volume and retail sell volume, scaled by total retail volume following Boehmer et al. (2020), measured over days [0, +1] around the FINRA announcement date.
Retail Order Imbalance _t	Difference between retail purchase volume and retail sell volume, scaled by total retail volume following Boehmer et al. (2020), measured on day t.
Markit_Match _{Treg}	Indicator variable that captures instances in which change in short interest ratio calculated using the Markit data is in the same decile as change in short interest ratio calculated using FINRA disclosure
CAR _[Treg+2, Treg+5]	Cumulative abnormal return adjusted for the weighted average daily returns of firms in the same size decile, over days [+2, +5] around the FINRA announcement date.
CAR _[t+1, t+5]	Cumulative abnormal return adjusted for the weighted average daily returns of firms in the same size decile, from [+1, +5] relative to each daily observation.
Short_Volume _[Treg-8-x]	The daily FINRA short volume (scaled by shares outstanding) on day Treg-8-x, where Treg-8 is the settlement date for mandatory short interest disclosures.

Figure 1
Cumulative Returns Around FINRA Short Interest Announcements

This figure presents the response to FINRA short interest announcements. We examine the cumulative abnormal returns (adjusted by weighted average daily returns of firms in the same size decile) for the [-5, +5] window around each short interest announcement. The sample begins with all FINRA short interest announcements from 2009 - 2020. We keep only the short interest announcements in the top decile of short interest change (i.e., highest increase in short interest). We then graph the average cumulative abnormal returns around announcement dates.

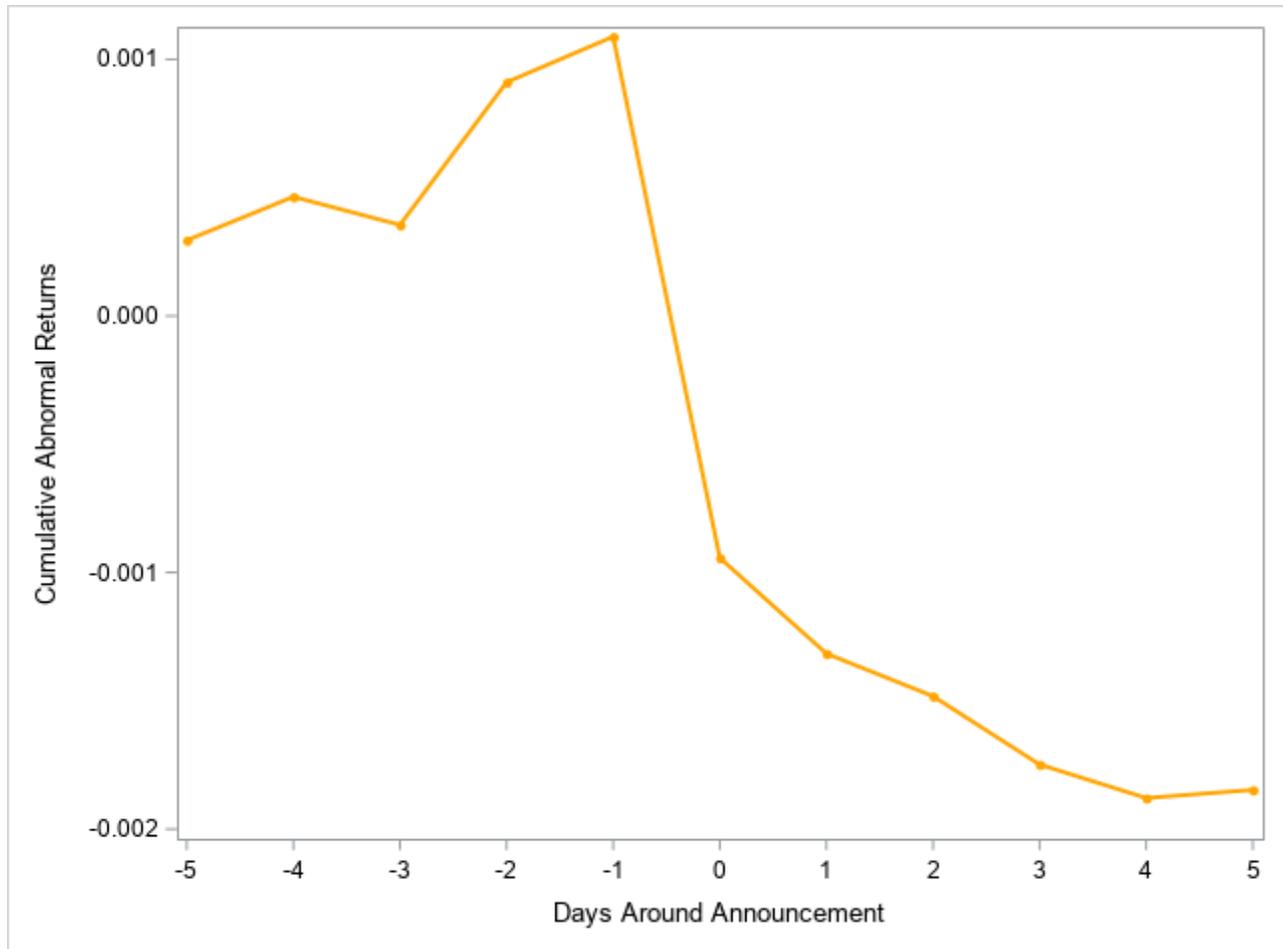


Table 1 Panel A
Sample Selection Summary

This table presents the detailed steps in the sample selection process.

Compustat Short Interest Announcements from 2009 to 2020	2,344,696
Less duplicate GVKEY - settlement date observations	207,785
Less observations without a matching PERMNO	571,069
Less observations missing shares outstanding	614
Less observations missing lags to calculate change in short interest	3,749
Less observations that are not common stock	499,009
Less observations missing announcement window returns	4,718
Total FINRA Short Interest Announcements	1,057,752

Samples:

Announcements with matching Markit Data from 2009 to 2020 (Tables 3 and 5)	985,050
Announcements with greater than 2% absolute buy-and-hold returns (Table 7)	382,644
Announcements with matching S3 Data from 2016 to 2020 (Tables 3 and 5)	319,498
Announcements with greater than 2% absolute buy-and-hold returns (Table 7)	123,936
Announcements with matching retail trading data from 2009 to 2014 (Table 8 Panel A)	512,791

Table 1 Panel B
Sample Distribution by Year

This table presents the sample of FINRA Short Interest Announcements by year.

2009	98,357
2010	94,596
2011	91,087
2012	87,936
2013	85,995
2014	88,029
2015	89,025
2016	86,509
2017	84,830
2018	84,824
2019	84,972
2020	81,592
Total	1,057,752

Table 2
Descriptive Statistics

This table reports descriptive statistics of variables used in this paper.

Variable	N	Mean	SD	P10	P25	P50	P75	P90
<i>Calculated Around FINRA Disclosures (2009 - 2020):</i>								
Dsiratio_FINRA _[Treg_Prior, Treg]	985,050	0.0001	0.0064	-0.0053	-0.0016	0.0000	0.0015	0.0055
CAR _[Treg, Treg+1]	985,050	-0.0002	0.0527	-0.0391	-0.0166	-0.0011	0.0144	0.0372
IPT _{Treg}	789,012	10.6636	10.4383	-0.2467	5.8007	10.6555	15.4227	21.5957
Perc_Ret _{Treg}	789,012	0.9126	0.5199	0.3766	0.7087	0.9314	1.1230	1.4194
Pred_CAR _[Treg, Treg+1]	985,050	-0.0004	0.0010	-0.0012	-0.0006	-0.0004	-0.0001	0.0004
<i>Calculated Using Markit (2009 - 2020):</i>								
Dsiratio_Private-Sector _[Treg-9, Treg-8]	985,050	0.0000	0.0032	-0.0023	-0.0006	0.0000	0.0006	0.0024
Dsiratio_Private-Sector _[Prior_Treg-8, Treg-8]	985,050	0.0000	0.0084	-0.0074	-0.0021	0.0000	0.0020	0.0075
<i>Calculated Using S3 (2016 - 2020):</i>								
Dsiratio_Private-Sector _[Treg-9, Treg-8]	319,498	0.0000	0.0012	-0.0009	-0.0003	0.0000	0.0003	0.0010
Dsiratio_Private-Sector _[Prior_Treg-8, Treg-8]	319,498	0.0000	0.0103	-0.0079	-0.0026	-0.0001	0.0024	0.0081
<i>Variables For Retail Analysis (2009 - 2014):</i>								
CAR _[Treg+2, Treg+5]	512,791	0.0001	0.0762	-0.0560	-0.0246	-0.0018	0.0215	0.0547
Retail Order Imbalance _[Treg, Treg+1]	512,791	-0.0384	0.4147	-0.5568	-0.2527	-0.0280	0.1732	0.4474
Markit_Match _{Treg}	512,791	0.2280	0.4195	0.0000	0.0000	0.0000	0.0000	1.0000
CAR _[t+1, t+5]	5,286,115	0.0005	0.0811	-0.0628	-0.0276	-0.0019	0.0244	0.0622
Retail Order Imbalance _t	5,286,115	-0.0375	0.4767	-0.6888	-0.3118	-0.0277	0.2283	0.5765

Table 3

Using Private-Sector Data to Predict FINRA Short Interest Disclosures

This table reports the ability of private-sector data vendors, Markit and S3, to predict FINRA short interest disclosures. The sample includes FINRA short interest announcements with matching data from Markit and S3 for the corresponding sample periods. The dependent variable, $Dsirat_{i,t}^{FINRA} [Prior_Treg, Treg]$, captures the change in short interest ratio (measured as number of shares held short divided by shares outstanding) from the previous FINRA announcement to the current FINRA announcement. The independent variable, $Dsirat_{i,t}^{Private-Sector} [Treg-9, Treg-8]$, measures the change in daily short interest using Markit or S3 data between $Treg-9$ and $Treg-8$, where $Treg-8$ is the settlement date for mandatory short interest disclosures. All following independent variables are measured analogously, with corresponding time periods in brackets. All variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement date. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	$Dsirat_{i,t}^{FINRA} [Prior_Treg, Treg]$			
	Markit		S3	
Private-Sector Data Source	2009-2020		2016-2020	
Sample Period	(1)	(2)	(3)	(4)
$Dsirat_{i,t}^{Private-Sector} [Treg-9, Treg-8]$	0.3174*** (20.556)		0.7637*** (20.430)	
$Dsirat_{i,t}^{Private-Sector} [Treg-10, Treg-9]$	0.3742*** (23.148)		0.7492*** (10.562)	
$Dsirat_{i,t}^{Private-Sector} [Treg-11, Treg-10]$	0.4042*** (23.376)		0.6307*** (14.814)	
$Dsirat_{i,t}^{Private-Sector} [Treg-12, Treg-11]$	0.4071*** (33.364)		0.7849*** (20.222)	
$Dsirat_{i,t}^{Private-Sector} [Treg-13, Treg-12]$	0.4237*** (46.879)		0.6202*** (12.444)	
$Dsirat_{i,t}^{Private-Sector} [Treg-14, Treg-13]$	0.4061*** (50.691)		0.6958*** (11.226)	
$Dsirat_{i,t}^{Private-Sector} [Treg-15, Treg-14]$	0.4125*** (44.518)		0.7772*** (20.419)	
$Dsirat_{i,t}^{Private-Sector} [Treg-16, Treg-15]$	0.3867*** (38.790)		0.4453*** (6.762)	
$Dsirat_{i,t}^{Private-Sector} [Treg-17, Treg-16]$	0.3454*** (24.042)		0.3210*** (7.858)	
$Dsirat_{i,t}^{Private-Sector} [Treg-18, Treg-17]$	0.2968*** (19.035)		0.1175*** (7.635)	
$Dsirat_{i,t}^{Private-Sector} [Treg-19, Treg-18]$	0.1737*** (17.499)		0.0528*** (2.850)	
$Dsirat_{i,t}^{Private-Sector} [Prior_Treg-8, Treg-8]$		0.3571*** (58.904)		0.4867*** (58.310)
Constant	0.0000 (1.055)	0.0000 (0.962)	-0.0000 (-0.974)	0.0001 (1.238)

Observations	985,050	985,050	319,498	319,498
R-squared	0.217	0.218	0.672	0.560

Table 4
Informativeness of Private-Sector Short Interest Announcements

This table presents the market response to private sector short interest announcements. The full sample includes all Markit and S3 short interest announcements, and Columns (1) and (3) include only Markit or S3 announcements with increases in short interest. The main dependent variable, $AbRet_t$, is the cumulative abnormal return adjusted for the weighted average daily returns of firms in the same size decile for day t . $Dsiration_Private-Sector_{[t-1, t]}$ captures the change in short interest ratio from the previous private-sector short interest announcement to the current announcement. $Big_Inc_Private-Sector_{[t-1, t]}$ is an indicator variable capturing the largest decile of increase in short interest ratio from the previous private-sector announcement to the current private-sector announcement. All continuous independent variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	AbRet _t			
	2009 - 2020		2016 - 2020	
Sample Period	Markit		S3	
Private-Sector Data Source	Markit		S3	
Sampling Procedure	Increases in	Full Sample	Increases in	Full Sample
	Short Interest		Short Interest	
	(1)	(2)	(3)	(4)
$Dsiration_Private-Sector_{[t-1, t]}$	-0.1377*** (-4.447)		-0.1777** (-2.416)	
$Big_Inc_Private-Sector_{[t-1, t]}$		-0.0005*** (-5.504)		-0.0003* (-1.948)
Constant	0.0005*** (11.022)	0.0001* (1.876)	0.0002** (2.104)	-0.0000 (-0.080)
Observations	5,303,936	10,758,425	1,914,682	3,441,257
R-squared	0.0001	0.0000	0.0000	0.0000

Table 5
Informativeness of FINRA Short Interest Announcements

This table presents the market response to FINRA short interest announcements. The full sample includes all FINRA short interest announcements meeting the criteria presented in Table 1, and Columns (1) and (3) include only FINRA announcements with increases in short interest. The main dependent variable, $CAR_{[Treg, Treg+1]}$, is the cumulative abnormal return adjusted for the weighted average daily returns of firms in the same size decile, over days [0, +1] around the FINRA announcement date. $Dsiratio_FINRA_{[prior_Treg, Treg]}$ captures the change in short interest ratio from the previous FINRA announcement to the current FINRA announcement. $Big_Inc_FINRA_{[prior_Treg, Treg]}$ is an indicator variable capturing the largest decile of increase in short interest ratio from the previous FINRA announcement to the current FINRA announcement. All continuous independent variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable Sample Period	CAR _[Treg, Treg+1]			
	2009 - 2020		2016 - 2020	
Sampling Procedure	Increases in Short Interest	Full Sample	Increases in Short Interest	Full Sample
	(1)	(2)	(3)	(4)
$Dsiratio_FINRA_{[prior_Treg, Treg]}$	-0.1500*** (-4.808)		-0.1099* (-1.798)	
$Big_Inc_FINRA_{[prior_Treg, Treg]}$		-0.0025*** (-8.242)		-0.0023*** (-4.176)
Constant	-0.0004 (-1.423)	0.0000 (0.100)	-0.0008 (-1.555)	-0.0004 (-0.852)
Observations	475,833	985,050	153,678	319,498
R-squared	0.0003	0.0002	0.0001	0.0001

Table 6 Panel A

Predictive Ability of Private Sector Data for Market Response to FINRA Disclosures

This table presents the ability of private-sector data to predict the market response to FINRA short interest announcements, 8 days in advance. The full sample includes all FINRA short interest announcements meeting the criteria presented in Table 1, and Columns (1) and (3) include only increases in short interest, measured with Markit or S3, respectively. The main dependent variable, $CAR_{[Treg, Treg+1]}$, is the cumulative abnormal return adjusted for the weighted average daily returns of firms in the same size decile, over days [0, +1] around the FINRA announcement date. $Dsratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ captures the change in short interest ratio from the previous mandatory settlement date to the current mandatory settlement date using Markit or S3 data. $Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]}$ is an indicator variable capturing the largest decile of increase in short interest ratio from the previous mandatory settlement date to the current mandatory settlement date using Markit or S3 data. All continuous independent variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	CAR _[Treg, Treg+1]			
	Markit		S3	
Private-Sector Data Source	2009 - 2020		2016 - 2020	
Sample Period				
Sampling Procedure	Increases in	Full Sample	Increases in	Full Sample
	Short Interest		Short Interest	
	(1)	(2)	(3)	(4)
$Dsratio_Private-Sector_{[prior_Treg-8, Treg-8]}$	-0.1028*** (-3.361)		-0.0877* (-1.812)	
$Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]}$		-0.0018*** (-4.844)		-0.0026*** (-3.798)
Constant	-0.0001 (-0.242)	-0.0001 (-0.252)	-0.0005 (-0.935)	-0.0004 (-0.793)
Observations	481,596	985,050	151,558	319,498
R-squared	0.0002	0.0001	0.0002	0.0002

Table 6 Panel B**Proportion of Market Response to FINRA Disclosure Predicted by Private-Sector Data**

This table quantifies the proportion of the market's response to FINRA disclosures that can be predicted by using private-sector data. The sample includes FINRA short interest announcement observations with increases in FINRA short interest, within the respective sample periods. The dependent variable, $Pred_CAR$, is the fitted value of regressing $CAR_{[Treg, Treg+1]}$ on $Dsiratio_FINRA_{[prior_Treg, Treg]}$ for a subset of sample with increases in FINRA short interest. $Dsiratio_Markit(S3)_{[prior_Treg-8, Treg-8]}$ captures the change in short interest ratio from the previous mandatory settlement date to the current mandatory settlement date using Markit (S3) data. All continuous independent variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable	Pred_CAR _[Treg, Treg+1]		
	Markit	S3	Markit & S3
Private-Sector Data Source			
Sample Period	2009 - 2020	2016 - 2020	2016 - 2020
Sample Selection	Increases in Short Interest	Increases in Short Interest	Increases in Short Interest
	(1)	(2)	(3)
$Dsiratio_Markit_{[prior_Treg-8, Treg-8]}$	-0.0404*** (-38.705)		-0.0106*** (-11.321)
$Dsiratio_S3_{[prior_Treg-8, Treg-8]}$		-0.0446*** (-38.632)	-0.0407*** (-35.651)
Constant	-0.0008*** (-156.799)	-0.0010*** (-167.425)	-0.0010*** (-189.821)
Observations	475,833	153,678	151,493
R-squared	0.1618	0.4563	0.4775

Table 7 Panel A

Intraperiod Timeliness of Returns Around FINRA Short Interest Announcements

This table examines whether the news from private-sector data vendors increases the speed of price discovery. The full sample includes all FINRA short interest announcement observations within the respective sample periods, and Columns (1) and (3) include only FINRA announcements with increases in short interest. The dependent variable, IPT_{Treg} , is the intraperiod timeliness measured following Butler et al. (2007). It is measured over the period from 18 days prior to the current announcement date (approximately the date of the previous settlement date) to two days past the current announcement date. The sample is limited to observations with buy-and-hold returns of greater than 2% over this time period to reduce noise in estimates. $Dsratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ captures the change in short interest ratio from the previous mandatory settlement date to the current mandatory settlement date using Markit or S3 data. $Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]}$ is an indicator variable capturing the largest decile of increase in short interest ratio from the previous mandatory settlement date to the current mandatory settlement date using Markit or S3 data. All continuous variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable Private-Sector Data Source Sample Period	IPT_{Treg}			
	Markit 2009 - 2020		S3 2016 - 2020	
Sample Selection	Increases in Short Interest	Full Sample	Increases in Short Interest	Full Sample
	(1)	(2)	(3)	(4)
$Dsratio_Private-Sector_{[prior_Treg-8, Treg-8]}$	23.6856*** (9.928)		39.7206*** (9.601)	
$Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]}$		0.5926*** (12.285)		0.8308*** (9.877)
Constant	10.7795*** (169.385)	10.6013*** (180.027)	10.5728*** (88.797)	10.5708*** (97.224)
Observations	382,644	789,012	123,936	255,917
R-squared	0.0004	0.0003	0.0014	0.0006

Table 7 Panel B
Percentage of Returns Earned Around FINRA Short Interest Announcements

This table examines whether the news from private-sector data vendors increases the speed of price discovery. The full sample includes all FINRA short interest announcement observations within the respective sample periods, and Columns (1) and (3) include only FINRA announcements with increases in short interest. $Perc_Ret_{Treg}$ is one minus the percentage of returns earned during the FINRA announcement window, and is measured as the buy-and-hold return over days [0, +1] around FINRA announcement date divided by buy-and-hold return over the period from 18 days prior to the current announcement date (approximately the date of the previous settlement date) to two days past the current announcement date. The sample is limited to observations with buy-and-hold returns of greater than 2% over this 20-day period to reduce noise in estimates. $Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$ captures the change in short interest ratio from the previous mandatory settlement date to the current mandatory settlement date using Markit or S3 data. $Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]}$ is an indicator variable capturing the largest decile of increase in short interest ratio from the previous mandatory settlement date to the current mandatory settlement date using Markit or S3 data. All continuous variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable Private-Sector Data Source Sample Period	Markit 2009 - 2020		Perc_Ret _{Treg} S3 2016 - 2020	
	Increases in Short Interest (1)	Full Sample (2)	Increases in Short Interest (3)	Full Sample (4)
$Dsiratio_Private-Sector_{[prior_Treg-8, Treg-8]}$	0.4693*** (4.426)		0.7725*** (4.136)	
$Big_Inc_Private-Sector_{[prior_Treg-8, Treg-8]}$		0.0074*** (3.470)		0.0130*** (3.359)
Constant	0.9127*** (302.447)	0.9118*** (325.198)	0.9069*** (159.869)	0.9095*** (168.666)
Observations	382,644	789,012	123,936	255,917
R-squared	0.0001	0.0000	0.0002	0.0001

Table 8 Panel A
Are Retail Investors Harmed?
Using Predictive Ability of Markit Data

This table examines whether the return-predictive ability of retail trading declines in cases in which retail investors may be at an informational disadvantage. The sample includes FINRA short interest announcements for which retail trading can be measured from 2009-2014 (limited due to data availability). $CAR_{[T_{reg}+2, T_{reg}+5]}$ is the cumulative abnormal return adjusted for the weighted average daily returns of firms in the same size decile, over days [+2, +5] around the FINRA announcement date. Retail Order Imbalance is the difference between retail purchase volume and retail sell volume, scaled by total retail volume, measured over days [0, +1] around the FINRA announcement date. $Markit_Match_{T_{reg}}$ is an indicator variable that captures instances in which change in short interest ratio calculated using the Markit data is in the same decile as change in short interest ratio calculated using FINRA disclosure. All continuous independent variables are winsorized at 1% and 99%. Firm fixed effects are included to control for time-invariant firm characteristics. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable Sample Period	CAR _[T_{reg}+2, T_{reg}+5] 2009-2014		
	Markit_Match = 1	Markit_Match = 0	Full Sample
Sampling Procedure	(1)	(2)	(3)
Retail Order Imbalance _[T_{reg}, T_{reg}+1]	0.0015** (2.238)	0.0013*** (3.104)	0.0013*** (3.098)
Markit_Match _{T_{reg}}			-0.0002 (-0.898)
Retail Order Imbalance _[T_{reg}, T_{reg}+1] * Markit_Match _{T_{reg}}			0.0002 (0.257)
Constant	0.0000 (0.087)	0.0002 (0.489)	0.0002 (0.541)
Firm FE	Yes	Yes	Yes
Observations	116,755	395,866	512,791
R-squared	0.0668	0.0414	0.0327

Table 8 Panel B
Are Retail Investors Harmed?

Examining Informativeness of Markit Data

This table examines whether the return-predictive ability of retail trading declines in cases in which retail investors may be at an informational disadvantage. The sample consists of *daily* observations for which Markit short interest and retail trading can be measured. $CAR_{[t+1, t+5]}$ is the cumulative abnormal return adjusted for the weighted average daily returns of firms in the same size decile, from [+1, +5] relative to each daily observation. Retail Order Imbalance is the difference between retail purchase volume and retail sell volume, scaled by total retail volume, measured for each day t . $Big_Inc_Markit_{[t-1, t]}$ is an indicator variable capturing the largest decile of increase in Markit short interest ratio from the previous date. All continuous independent variables are winsorized at 1% and 99%. Firm fixed effects are included to control for time-invariant firm characteristics. Standard errors are clustered by each mandatory announcement window in order to address market-wide correlation in returns. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable Sample Period	CAR _[t+1, t+5] 2009-2014		
	Big_Inc_Markit = 1	Big_Inc_Markit = 0	Full Sample
Sampling Procedure	(1)	(2)	(3)
Retail Order Imbalance _t	0.0021*** (5.786)	0.0015*** (15.163)	0.0015*** (15.085)
Big_Inc_Markit _[t-1, t]			-0.0014*** (-7.056)
Retail Order Imbalance _t * Big_Inc_Markit _[t-1, t]			0.0006* (1.718)
Constant	-0.0011*** (-3.214)	0.0007*** (4.967)	0.0007*** (4.437)
Firm FE	Yes	Yes	Yes
Observations	528,835	4,757,125	5,286,115
R-squared	0.0245	0.0170	0.0151

Table 9**Using Daily FINRA Short Volume Disclosures to Predict FINRA Short Interest Disclosures**

This table reports the ability of daily FINRA short volume disclosures to predict FINRA short interest disclosures. The sample includes FINRA short interest announcements with matching data from FINRA daily short volume disclosures. The dependent variable, $Dsratio_FINRA_{[prior_Treg, Treg]}$, captures the change in short interest ratio (measured as number of shares held short divided by shares outstanding) from the previous FINRA announcement to the current FINRA announcement. The independent variable, $Short_Volume_{Treg-8}$, measures the daily short volume (scaled by shares outstanding) on day $Treg-8$, where $Treg-8$ is the settlement date for mandatory short interest disclosures. All following independent variables are measured analogously, with corresponding time periods in brackets. All variables are winsorized at 1% and 99%. Standard errors are clustered by each mandatory announcement date. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable Sample Period	$Dsratio_FINRA_{[prior_Treg, Treg]}$ 2009-2020	
	(1)	(2)
$Short_Volume_{Treg-8}$	-0.0024 (-0.204)	
$Short_Volume_{Treg-9}$	-0.0136 (-1.037)	
$Short_Volume_{Treg-10}$	0.0719*** (4.370)	
$Short_Volume_{Treg-11}$	0.0688*** (2.749)	
$Short_Volume_{Treg-12}$	0.0458* (1.684)	
$Short_Volume_{Treg-13}$	0.0713** (1.985)	
$Short_Volume_{Treg-14}$	0.0205 (1.359)	
$Short_Volume_{Treg-15}$	0.0176* (1.778)	
$Short_Volume_{Treg-16}$	0.0101 (1.047)	
$Short_Volume_{Treg-17}$	-0.0020 (-0.158)	
$Short_Volume_{Treg-18}$	-0.0418** (-2.096)	
$Short_Volume_{[prior_Treg-8, Treg-8]}$		0.0175*** (9.189)
Constant	-0.0003*** (-4.329)	-0.0002*** (-3.370)
Observations	849,249	849,249
R-squared	0.013	0.011