On the Capital Market Consequences of Alternative Data: Evidence from Outer Space

ZSOLT KATONA
U.C. Berkeley
zskatona@berkeley.edu

MARCUS O. PAINTER
Saint Louis University
marc.painter@slu.edu

PANOS N. PATATOUKAS
U.C. Berkeley
panos@haas.berkeley.edu

JEAN ZENG
National University of Singapore
jeanzeng@nus.edu.sg

First Draft: July-30-2018
This Version: May-16-2022

Acknowledgments: We thank RS Metrics and Orbital Insight for providing the satellite imagery data of parking lot traffic, and Markit for providing the securities lending market data. For helpful comments and discussions, we thank Chris Clifford, Mike Farrell, Jill Fisch, Kristine Hankins, Franz Hinzen (discussant), Russell Jame, Thomas Lee, Frank Partnoy, Tamara Nefedova (discussant), the PhD students at Berkeley Haas, former SEC Commissioner Robert Jackson, the DERA of the SEC, and the members of the University of Chicago Applied Math Club. We also thank seminar participants at the 9th Miami Behavioral Finance Conference, the University of Florida, the University of Kentucky, the University of Missouri, the 2019 Future of Financial Information Conference at the Stockholm Business School, the 2019 Annual Meeting of the Financial Management Association, Harvard Business School, Kellogg School of Management, and Tulane University. We gratefully acknowledge financial support from the Fisher Center for Business Analytics at Berkeley Haas. This paper is the product of merging two related papers, one of which had the current title and another that was titled "Un-levelling the playing field: the investment value and capital market consequences of alternative data."

On the Capital Market Consequences of Alternative Data:
Evidence from Outer Space

Abstract

We use the introduction of satellite coverage of major retailers to study the capital market implications of unequal access to alternative data. We find that satellite data allowed sophisticated investors to formulate profitable strategies by targeting the quarterly reports of retailers with bad news. Using a difference-in-differences design, we find that the release of satellite data led to more informed short selling activity, less informed individual buying activity, and lower stock liquidity prior to the quarterly reports of retailers with satellite coverage. Overall, our paper provides evidence that unequal access to alternative data can increase information asymmetry among market participants.

Keywords: Alternative Data; Satellite Imagery; Short Selling; Individual Trading; Liquidity.

JEL Classification: G12, G14, G23.

1. Introduction

Big data is a big deal. From how we connect to our network to how we buy products online or even choose TV shows to watch, big data has been transforming our lives in profound ways. Despite the interest, there is limited evidence on the implications of the rise of big and alternative data for information asymmetry among stock market participants. On one hand, recent advancements in computational power, expanded data storage capacity, and faster interconnection speeds have enabled access to large amounts of alternative data that can inform investment decisions. On the other hand, access to big and alternative data is often only within the reach of sophisticated investors who can afford to incur the substantial costs of acquiring, processing, and integrating such data. These costs lead to unequal access to big and alternative data.

In a groundbreaking study, Zhu (2019) finds that the introduction of alternative data increases long-term price informativeness and disciplines managers' insider trading and investment decisions. While higher price informativeness can lead to a reduction in information asymmetry between firm insiders and sophisticated investors, unequal access to alternative data can lead to an increase in information asymmetry between sophisticated investors and individual investors. In this paper, we study the emergence of satellite imagery data in capital markets to evaluate how alternative data affects information asymmetry among market participants.

Due to the expensive acquisition and processing costs, access to data derived from satellite images is only within the reach of sophisticated investors, with select hedge funds being the typical clients of satellite data vendors. Satellite imagery is one of the most desirable data sources in the alternative data "wish list" of hedge fund managers (Arcadia Data 2018). However, financial analysts and mutual fund managers have not been widespread adopters of alternative data citing various factors, including the cost and difficulty of accessing and quantifying the value of alternative data as well as the challenges of integrating such data with their internal resources (IHS Markit 2019). These factors are significant roadblocks in the path of democratizing access to alternative data.

Standard models of market efficiency under costly information acquisition predict that when information is costly to acquire, process, and integrate, it will not be immediately reflected in prices allowing informed investors to generate a competitive rate of return that is commensurate with the information acquisition costs (e.g., Grossman and Stiglitz 1980). In the context of Blankespoor's et al. (2020) disclosure framework, the substantial costs of acquiring, processing, and integrating satellite data mean that such data can be seen as a form of private information. Within this context, it becomes obvious that the market cannot be perfectly efficient with respect to the information content of satellite data but only "efficiently inefficient" conditional on the data costs (Pedersen 2015).

The concept of unequal access to value-relevant information has existed long before the rise of alternative data. It is well-known that sophisticated investors with greater resources have an advantage in terms of anticipating news before it becomes public (e.g., Hendershott et al. 2015) and that investors exploit opportunities to trade on private information (e.g., Heitzman and Klasa 2020). In fact, even the idea of counting cars to predict retailer performance has existed long before the introduction of satellite data.[1]

A key feature of the satellite data setting is that the data is not truly private, as anyone has the ability to count cars in a parking lot, yet it still creates an asymmetry in anticipating firm news because only a subset of market participants have the access and ability to use the data at scale. This creates a new information environment where information can be public,

[1] Sam Walton, the founder of Walmart, was known for flying over parking lots on a regular basis so he could count cars and monitor store performance (Walton and Huey 1992).

private, or restricted. Advances in computer vision and the increased availability of satellite imagery has enabled daily tracking and processing of retailer parking lots at a large scale and provides an interesting setting to study the capital market implications of unequal access to alternative data.

Our primary data source is RS Metrics, the first vendor to introduce nearly real-time daily data feeds of store-level parking lot traffic signals extracted from satellite imagery analysis in the U.S. market. The store-level dataset includes 4.7 million daily observations across 67,078 unique store locations for 44 major U.S. retailers between 2011:Q1 and 2017:Q4. The daily data feeds include information about parking lot capacity and utilization. From the daily store-level parking lot information, we construct measures of enterprise-level parking lot fill rates. Our major retailer sample accounts for a large fraction of the aggregate sales and market value across U.S. listed companies operating in the same industry.

As a steppingstone for our analysis of short selling and individual trading activity, we first validate the information content of satellite data for anticipating the quarterly performance of covered retailers. We focus on predicting quarterly reports because these are highly anticipated public events with significant price impacts. Consistent with Froot's et al. (2017) and Zhu's (2019) results on the value relevance of alternative data, our evidence shows that growth in parking lot fill rates is incrementally relevant for anticipating retailer performance for the quarter. One implication is that investors with access to satellite data could profit by targeting the quarterly reports of retailers with satellite coverage. Indeed, a trading strategy that takes a long (short) position in the stock of retailers that experience an abnormal increase (decrease) in parking lot fill rates during the quarter generates abnormal returns.

The portfolio returns from targeting retailers with satellite coverage are asymmetric on the long and the short side. We separate retailers that experience an abnormal increase in parking lot fill rates (good news) from retailers that experience an abnormal decrease in parking lot fill rates during the quarter (bad news). The evidence shows that the absolute magnitude of returns is nearly twice as large for the portfolio of bad news retailers relative to the good news portfolio. Around earnings announcements, the good news portfolio outperforms the market by 1.6% while the bad news portfolio underperforms the market by nearly −3%, after accounting for accumulated stock loan fees.

The marked asymmetry in the returns implies that satellite data is especially relevant for short sellers interested in targeting retailers with bad news for the quarter. Using daily stock lending data from Markit, we find evidence of informed short-selling activity leading up to the quarterly reports of retailers with satellite coverage. Focusing on retailers with abnormal decreases in parking lot fill rates, we observe a substantial increase in the lender quantity on loan starting five trading days prior to the quarterly earnings announcement. On the other side, we do not observe pre-earnings announcement changes in short-selling activity for good news retailers with abnormal increases in parking lot fill rates.

Notwithstanding evidence of informed short selling leading to the quarterly reports of bad news retailers, the general public cannot "piggyback" on the buildup of short-selling activity. The reason is that daily short interest data is available only to those who can afford the substantial subscription fees to proprietary data vendors such as Markit, with brokers and hedge funds being the typical clients. In contrast, the general public has access to short interest data only twice per month and only with a significant delay. Indeed, while short sellers are actively targeting the quarterly reports of retailers with bad news for the quarter, we find evidence that individual investors are net buyers of the stock of such retailers.

To estimate the impact of the introduction of satellite data on trading activity, we implement a difference-in-differences (DID) design. The DID compares the group of retailers with satellite coverage to a matched control group of retailers with no satellite coverage in the periods before and after the initiation of satellite coverage. An increase in the informativeness of short-selling activity would imply that short-sellers' ability to anticipate negative news for the quarter increases after the introduction of satellite coverage.

The DID regression results provide evidence that the ability of short sellers to profit from targeting retailers with bad news for the quarter has improved after the introduction of satellite data for the group of retailers with satellite coverage. In contrast, we find that after the introduction of satellite data, individual investors' trades, especially individual investor buys, have become less informative with respect to anticipating retailer news for the quarter. Though the availability of alternative data could decrease information asymmetry between firm insiders and outside investors (Zhu 2019), our evidence suggests that the introduction of satellite data could lead to an increase in information asymmetry among market participants.

Consistent with an increase in information asymmetry among market participants, our DID estimates provide evidence that the introduction of satellite data led to an increase in the bid-ask spread and price impact in the days leading to the quarterly reports of retailers with satellite coverage. In cross-sectional tests, we further find that the effect of satellite data on information asymmetry is more pronounced for smaller, younger, and more volatile retailers for which fundamental uncertainty is likely to be higher.

We do not argue that individual investors are the only market participants on the other side of the alternative data-based informed trades. Rather, we argue that due to the expensive acquisition and processing costs, individual investors are less likely to have access to alternative data. Thus, our evidence of an overall increase in information asymmetry among outside investors does not preclude that unequal access to alternative data may also increase information asymmetry across different groups of sophisticated investors.

The growing importance of alternative data has led to a burgeoning literature studying its capital market effects. Zhu (2019) examines how big data availability influences corporate governance and provides evidence that the introduction of alternative data decreased information asymmetry between corporate insiders and sophisticated investors. We differentiate from this prior work by zeroing in on information asymmetry among outside investors. Together, the two papers offer complementary insights into the capital market implications of alternative data. Satellite data of parking lot traffic has also been used to create measures of local retailer performance. Kang et al. (2021) and Gerken and Painter (2019) study how institutional investors and sell-side analysts, respectively, react to changes in the local performance of retail firms. Different from their focus on the advantage of local institutional investors and analysts when processing information about nearby retailers, our paper provides evidence on the effect of enterprise-level signals extracted from satellite data on informed short selling activity, uninformed individual order flow, and stock liquidity leading to quarterly earnings announcements.

More broadly, our paper relates to growing research on the role of alternative data in capital markets.[2] Concurrent research includes Blankespoor et al. (2022), who use proprietary transaction-level data to construct real-time revenue forecasts and find evidence of dynamic information flow during the quarter, and Dessaint et al. (2021), who provide evidence in the sell-side analyst setting that alternative data are predominantly informative about a firm's short-term prospects but not as informative about a firm's long-

[2] Bollen et al. (2011) find that aggregate Twitter mood has predictive power for aggregate returns; Da et al. (2011) find short-term momentum and long-term reversals for stocks with abnormally high Google search frequency; Jame et al. (2016) find that crowdsourced forecasts are useful in predicting earnings; Froot et al. (2017) use proprietary data of consumer activity and find that managers bias their disclosures when they possess private information motivated in part by insider trading opportunities; Blankespoor et al. (2014a) find that firms can reduce information asymmetry by sending market participants links to press releases through Twitter; Bartov et al. (2018) find that Twitter opinion predicts earnings; Farrell et al (2021) find that stocks with reductions in Seeking Alpha coverage experience a decrease in liquidity.

term prospects. We contribute to this area of research by showing that the introduction of alternative data can lead to an increase in information asymmetry among market participants.

Our paper also relates to research on the impact of technology and data abundance on capital markets. Relevant papers in this area propose that advances in technology and alternative data have nuanced effects on information asymmetry, price discovery, and investor stock market participation (e.g., Blankespoor et al. 2014b; Banerjee et al. 2018; Dugast and Foucault 2018; Mihet 2020).

2. Measuring parking lot traffic from outer space

2.1 Satellite imagery data of parking lot traffic

Our primary source of parking lot data is RS Metrics, the first data vendor to introduce nearly real-time parking lot traffic signals derived from satellite imagery data in the U.S. starting in 2011:Q1. While counting cars to predict retailer performance is not a new idea, the availability of satellite imagery tracking enables daily tracking at a large scale. Appendix 1 in the Online Supplement provides the background on remote sensing technology. The data consists of daily store-level information about parking lot capacity and utilization across major U.S retailers. With respect to the cost of accessing satellite imagery data, data vendors privately negotiate the price depending on the timeliness and extent of access. Our communications indicate that the cost is typically in the range of hundreds of thousands of dollars per year.

The measurement of parking lot traffic is subject to error. First, satellite coverage is available only for a subset of a retailer's stores. One reason for partial coverage is that the cameras have to be pointed in a given direction for a certain store and there is only limited capacity allotted to each satellite user. Another reason for partial coverage is that not all parking lots are visible from outer space, including underground and multi-story lots. In addition, satellite coverage is restricted to domestic store locations. Second, the satellite's orbit is designed in a way that it passes through a given latitude at the same local time of the day at each given location. This time is between late morning and early afternoon for most of the continental U.S., which captures only a snapshot of total parking lot traffic during the day. Third, even though the resolution of satellite imagery has significantly improved over time, it is still hard to measure parking lot traffic due to clouds, haze, trees, shadows, and other environmental factors. RS Metrics processes satellite imagery using a combination of a software for automated counts and human analysts for verifying the counts.

2.2 Measuring parking lot traffic at the individual store level

Our sample starts in 2011:Q1 because this is the first quarter for which RS Metrics started selling satellite imagery data.[3] Our sample ends in 2017:Q4 because this is the last quarter for which we obtained satellite data from RS Metrics per our data service agreement. Our sample includes 4.7 million daily observations across a total of 67,078 unique store locations for the 44 U.S. major retailers with satellite coverage. The key information available from the processed satellite imagery is the daily number of cars parked in an individual store parking lot; denoted by $CARS_{ijd}$, along with the total number of available parking spaces; denoted by $SPACES_{ijd}$, where i indicates the retailer, j indicates the individual store location, and d indicates the day of the satellite imagery.

Table A1 in the Online Supplement reports information about the store count and satellite store coverage for each of the 44 companies in our sample along with the starting date of satellite coverage. The cross-sectional average store count is 2,412 with satellite

[3] While RS Metrics is the first data vendor to sell domestic parking lot signals beginning in 2011:Q1, there are other competing data vendors sourcing imagery from the same satellites with Orbital Insight being the most prominent competitor in the U.S. Orbital Insight, however, started selling parking lot traffic signals to investors beginning in mid-2015; that is, more than four years after investors could subscribe to RS Metrics' data feeds.

coverage available for 58% of the individual store locations. We organize our sample using six-digit Global Industry Classification Standard (GICS) codes. The most represented group in our sample is specialty stores with 21 retailers, including Walmart Inc., Target Corporation, and Bed, Bath & Beyond Inc. We note that the number of retailers with satellite coverage in our sample increased from 10 in 2011 to 30 in 2014 and 44 in 2017. Figure A1 in the Online Supplement maps the geographical coverage of satellite imagery and shows that our store-level data covers 2,571 counties representing 98% of the U.S. population.

2.3 Measuring parking lot traffic at the corporate level

From the daily store-level data, we compile a panel of 650 firm-quarter observations of enterprise-level parking lot fill rates. We start with the daily data for each individual store location j during quarter q and compute the average number of cars parked during the quarter ($CARS_{ijq}$) as well as the average number of parking lot spaces available at each store location during the quarter ($SPACES_{ijq}$). Due to seasonal effects in quarterly data, we focus on year-over-year (YoY) comparisons rather than sequential comparisons; that is, we compare quarter q to quarter $q-4$. To ensure YoY comparability, we restrict our attention to individual store locations with satellite imagery in both quarter q and quarter $q-4$. The same-store comparisons control for YoY growth in parking lot capacity due to acquisitions and the opening of new stores.[4] Figure A2 in the Online Supplement provides an illustrative example of satellite imagery for Target Corporation, the department store company.

Our final sample includes 3.4 million daily observations across 53,647 unique store locations for 44 major U.S. retailers covered from 2011:Q1 to 2017:Q4. For each retailer-

[4] We focus on same-store sales growth as a key indicator of retailer performance from operating activities. For retailers that are growing by opening new stores, same-store comparisons allow investors to differentiate between growth that comes from new stores, and growth from improved operations at existing store locations. Our results are unchanged when we replace growth in same-store parking lot fill rates with overall growth without conditioning on same-store comparisons. We report these results in the Online Supplement (see Table A2). The correlation between the two growth measures is 89%.

quarter, we sum up across individual store locations with YoY satellite coverage to obtain the aggregate parking lot traffic; $CARS_{iq}$, and the aggregate parking lot space; $SPACES_{iq}$. We calculate the enterprise-level parking lot fill rate—our measure of overall parking lot utilization—as the ratio of aggregate parking lot traffic divided by aggregate parking lot space: $FLRT_{iq} = \frac{\sum_{j=1}^{J} CARS_{ijq}}{\sum_{j=1}^{J} SPACE_{ijq}} = \frac{CARS_{iq}}{SPACES_{iq}}$. The key variable of interest is the YoY growth in same-store parking lot fill rates: $\Delta FLRT_{iq} = \frac{FLRT_{iq} - FLRT_{iq-4}}{FLRT_{iq-4}}$. We note that the variability in same-store parking lot fill rates is primarily due to variability in parking lot traffic rather than parking lot capacity. This is because parking lot capacity at the individual store-level is very steady YoY. Indeed, growth in same-store parking lot fill rates is 99% correlated with growth in same-store car traffic.

2.4 Descriptive statistics

Table 1, Panel A, reports the empirical distributions of key variables. Appendix 2 in the Online Supplement provides key variable definitions. The sample includes 650 firm-quarter observations across 44 major U.S. retailers from 2011:Q1 to 2017:Q4. The parking lot fill rate has a mean (median) value of 29.8% (26.8%) with a standard deviation of 9.9% and an interquartile range of 22.9% to 35.3%. The distribution of the YoY growth in parking lot utilization ($\Delta FLRT_{iq}$) is centered at -0.7% and exhibits substantial variation with a standard deviation of 4.9% and an interquartile range of -3.4% to 1.8%. The pairwise correlations in Table 1, Panel B, provide preliminary evidence that $\Delta FLRT_{iq}$ is relevant for anticipating the market reaction to quarterly earnings announcements.

Table 1, Panel C, reports the sample distribution across six-digit GICS industries along with their contribution to the aggregate sales and market value separately for each industry. The evidence highlights that our major retailer sample accounts for a large fraction of the aggregate sales and market value across U.S. listed companies operating in the same industry.

To illustrate, our sample includes 10 multiline retail companies, which account for 85% of the sales and 77% of the market value of the U.S. listed firms in the same GICS code.

3. Empirical Analyses

3.1 Forward-looking content of satellite imagery data

As a steppingstone for our analysis of short selling and individual trading activity, we first validate the relevance of satellite imagery of parking lot fill rates for anticipating retailer performance. The idea is that variation in parking lot utilization should be correlated with shopper conversion across stores. Higher YoY growth in same-store parking lot utilization should indicate higher close rates and, therefore, higher same-store sales growth. We obtain quarterly data on same-store sales (SSS_{iq}) from FactSet Fundamentals. We focus on the domestic portion of sales because satellite imagery covers only individual stores located in the U.S. We measure YoY growth in the domestic portion of same-store sales as $\Delta SSS_{iq} = \left(SSS_{iq} - SSS_{iq-4}\right)/SSS_{iq-4}$. We focus on YoY growth rather than sequential growth due to seasonal effects in retailer performance.

Our first set of tests are based on regression models of the following form:

$$\Delta SSS_{iq} = \alpha + \beta_1 \Delta FLRT_{iq} + \beta_2 \Delta SSS_{iq-1} + \beta_3 QRET_{iq} + C_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \quad (1).$$

The dependent variable is domestic same-store sales growth (ΔSSS_{iq}) and the set of right-hand-side predictors includes same-store growth in parking lot utilization ($\Delta FLRT_{iq}$), lagged same-store sales growth (ΔSSS_{iq-1}), and the stock return from the beginning to the end of quarter q ($QRET_{iq}$). The vector of time-varying firm characteristics (C_{iq}) includes log market capitalization, Tobin's Q, institutional ownership, and indicators for Big-4 auditors, acquisitions, restructurings, asset write-downs and impairments. This vector of time-varying firm characteristics includes previously identified determinants of firm transparency (e.g., Lang et al. 2012). The model includes firm fixed effects (θ_i) to control for firm-specific time-invariant factors and quarter fixed effects (δ_q) to control for aggregate time-varying factors. The coefficient of interest is that on $\Delta FLRT_{iq}$. To ease the interpretation of the estimates, we report regression results using the standardized z-values of the continuous predictors. The standardized regression coefficients measure changes in standard deviation units, which allows us to easily compare the relative importance of each predictor.

Table 2, Panel A, reports regression results for equation (1). The evidence confirms that $\Delta FLRT_{iq}$ is an incrementally relevant predictor of retailer performance after accounting for autocorrelation in retailer growth as well as forward-looking information embedded in the quarterly stock return. Put differently, the predictive content of YoY growth in parking lot utilization contains information that is not subsumed by signals as easily accessible as the quarterly stock return and the past realization of same-store sales growth. A one standard deviation increase in $\Delta FLRT_{iq}$ is associated with a 0.8% increase in ΔSSS_{iq} representing a 61.5% deviation from the mean. Across model specifications, we note that the magnitude of the estimated coefficient on $\Delta FLRT_{iq}$ is not sensitive to the inclusion of time-varying firm characteristics and fixed effects.

Table 2, Panel B, separates negative from positive values of $\Delta FLRT_{iq}$ and provides evidence of asymmetry in the fundamental link of negative and positive changes in parking lot utilization with retailer performance. Across specifications, the magnitude of the estimated coefficient on $\Delta FLRT_{iq}^{-}$ is nearly three times that on $\Delta FLRT_{iq}^{+}$. Our evidence of an asymmetric link between ΔSSS_{iq} and $\Delta FLRT_{iq}$ implies that in-store foot traffic is more sensitive to decreases in parking lot utilization than to increases in parking lot utilization.

Together, the evidence suggests that satellite data is incrementally relevant for anticipating retailer performance, especially for retailers experiencing a decrease in parking lot utilization. If a subset of sophisticated investors are able to profit from trading based on satellite data, the introduction of satellite data could lead to an increase in information asymmetry between investors who can access and process the data and those who cannot. Next, we examine whether investors with access to satellite data could formulate a trading strategy targeting the quarterly reports of retailers with satellite coverage.

3.2 Formulating a trading strategy using satellite imagery data

Table 3, Panel A, reports the cumulative return from a trading strategy that buys (short sells) retailers with same-store parking lot fill rate growth in the top (bottom) quartile of the cross-sectional distribution of $\Delta FLRT_{iq}$. We report raw returns, market-adjusted returns, as well as size and book-to-market factor-adjusted returns cumulated over the three-day window centered on the earnings announcement day. To generate the cutoff values of $\Delta FLRT_{iq}$, we consider retailers with fiscal quarters ending within the last three months. This approach allows us to generate cross-sectional cutoff values on a rolling basis, thereby, allowing for time-series variability in the empirical distribution of $\Delta FLRT_{iq}$. The buy (short sell) portfolio includes retailers in the top (bottom) quartile portfolio of the cross-sectional distribution of $\Delta FLRT_{iq}$.

The evidence shows that around quarterly earnings announcements, the short sell portfolio underperforms the market by -3.10% while the buy portfolio outperforms the market by 1.66%. In terms of factor-adjusted returns, the spread between the buy and sell portfolios is 4.76%, which is statistically significant and economically important. Though we find significant abnormal returns for both portfolios, the absolute magnitude of returns is nearly twice as large for the short sell portfolio relative to the buy portfolio. This is consistent with our findings in Panel B of Table 2 which shows that negative values of $\Delta FLRT_{iq}$ are more informative about same-store sales growth. To evaluate performance net of stock loan fees, we obtain daily data on stock lending market conditions from Markit. Markit aggregates survey information from a consortium of institutional lenders that collectively account for most of the lendable inventory of shares in the U.S. Table 3, Panel B, reports results after adjusting the short sell portfolio for cumulative stock loan fees. We find that the stock loan fees are less than one basis point per day. Therefore, the buy-minus-sell portfolio returns remain intact after accounting for the cumulative cost of short selling.[5]

Figure 1, Panel A, reports the cumulative factor-adjusted returns separately for the top and bottom $\Delta FLRT_{iq}$ portfolios over the $+/-10$ trading-day window centered on the earnings announcement (day zero). The green dashed line presents the performance of the buy portfolio of retailers with abnormal increases in parking lot fill rates. The red dotted (solid) line presents the performance of the short sell portfolio after (before) stock loan fees. The evidence shows that there is only limited pre-announcement activity and the portfolio returns are effectively realized at the time of the earnings announcement. A key implication is that investors with access to satellite data could get ahead of the market and formulate a trading strategy anticipating the market reaction to quarterly earnings announcements. Again, the evidence highlights the asymmetry in the performance of the bad news portfolio relative to the good news portfolio.

Figure 1, Panel B, presents the daily values of abnormal share turnover for the buy and short sell portfolios over the $+/-10$ trading-day window centered on day zero. We measure share turnover as the daily trading volume divided by the number of shares outstanding. To obtain the abnormal values of share turnover, we adjust the daily values of

[5] In additional analysis, we merge store-level parking lot data from RS Metrics and Orbital Insight and find that investors with access to both vendors could formulate even more profitable strategies targeting earnings announcements. The complementarities across data vendors can help investors extract more accurate signals. One question that arises is that if signals based on the satellite data are so valuable why do data vendors like RS Metrics and Orbital Insight choose to sell the data instead of trading on them? Admati and Pfleiderer (1988) study this question and show that while a risk-neutral information owner would want to trade on the information directly, a risk-averse information owner would want to sell the information because that results in better risk-sharing. This analytical result helps explain the state of the market for alternative data.

share turnover during the event window for the daily share turnover prior to the event window. Consistent with the stock price dynamics, abnormal share turnover is close to zero prior to the earnings announcement and the significant spike around day zero suggests that most of the price discovery happens on the announcement day.

Separating retailers based on changes in parking lot fill rates during the quarter, the evidence highlights that the absolute magnitude of returns is nearly twice as large for the bad news portfolio relative to the good news portfolio. The asymmetry in returns is consistent with our earlier finding that decreases in parking lot traffic are especially relevant for anticipating retailer performance for the quarter. Strategic firm disclosure choices can also be relevant for understanding the asymmetric predictive power of satellite data for negative news. In particular, if managers strategically withhold bad news during the quarter, this can help explain why the market responds more strongly to negative realizations (see, e.g., Blankespoor's et al. 2022).[6] Importantly, the marked asymmetry in the portfolio performance implies that satellite data is especially relevant for short sellers interested in targeting retailers with bad news for the quarter.

3.3 Informed short-selling activity and uninformed individual order flow

The evidence thus far shows that investors with access to satellite data could get ahead of the market and formulate trading strategies targeting quarterly earnings announcements. The strategy works on both the long and the short side, though the returns

[6] Using real-time revenue forecasts constructed using proprietary transaction-level data, Blankespoor et al. (2021) find evidence of dynamic information flow as managers withhold bad news during the quarter which they choose to release only as the public earnings announcement draws near. A key implication of their study is that the probability of bad news disclosures increases as the quarterly earnings announcement approaches. Froot et al. (2017) also use proprietary data to generate real-time revenue forecasts and find evidence that managers strategically choose to withhold part of their private information at the time of the earnings announcement. Different from Blankespoor et al. (2021), however, Froot et al. (2017) conclude that managers bias their disclosures down when in possession of positive private information, which would imply withholding good news rather than bad news

are especially pronounced from short selling retailers with the largest decreases in parking lot fill rates. This asymmetry implies that satellite data is especially relevant for short sellers. In what follows, we use daily data on lender quantity on loan from Markit to provide direct evidence of informed trading in the stock lending market.

Figure 2, Panel A, presents the cumulative change of lender quantity on loan as a percentage of shares outstanding separately for the top and bottom $\Delta FLRT_{iq}$ portfolios.[7] The evidence is consistent with informed short-selling activity prior to the earnings announcement. Focusing on the bottom $\Delta FLRT_{iq}$ portfolio (red solid line), we find evidence of a substantial increase in the lender quantity on loan starting five trading days prior to the earnings announcement. On the other side, we do not find evidence of significant changes in short-selling activity for the top $\Delta FLRT_{iq}$ portfolio (green dashed line). While the evidence is consistent with informed short-selling activity, individual investors cannot "piggyback" on the information content of daily fluctuations in the quantity on loan. The reason is that daily short interest data is available only to those who can afford the substantial subscription fees to Markit's proprietary data feeds, with brokers and hedge funds being the typical Markit clients. In contrast, the general public has access to short interest data only twice per month and only with a significant delay.[8]

[7] Markit records stock lending activity when it becomes known to the market; that is, as of the settlement date. Up until 2017, the settlement date was the trade date plus three trading days. After September 5, 2017, the SEC shortened the standard settlement cycle from three trading days after the trade date to two trading days (Release No. 34-80295). To match stock lending activity to the occurrence of an underlying short sale, we account for the trade settlement period by shifting stock loan transactions back by two or three trading days.

[8] The Financial Industry Regulatory Authority (FINRA) requires member firms to report their short positions as of settlement on the 15th of each month (or the preceding business day if the 15th is not a business day) and as of settlement on the last business day of the month. The short-interest reports must be filed by the 2nd business day after the reporting settlement date. FINRA compiles the short interest data on a stock-by-stock basis across all member firms and provides it for publication on the 8th business day after the reporting settlement date. The current reporting regime became effective in December 2008 (FINRA Rule 4560).

Next, we use Boehmer's et al. (2021) method to measure individual investor buy and sell trades. This method exploits two features of marketable individual order flow. First, most marketable individual order flow is either filled from the broker's own inventory or sold by the broker to wholesalers. Second, marketable individual orders typically receive a small fraction of a penny price improvement over the national best bid or offer, while institutional orders tend to be executed at a round penny or half-penny increments.

To measure individual order flow, we first identify off-exchange trades in the Trade and Quote (TAQ) data (exchange code "D"). Then, we identify trades as individual buys (sells) if the trade took place at a price just below (above) a round penny. We measure individual order imbalance as the individual buys minus the individual sells divided by the total number of shares outstanding. Blankespoor et al. (2020) point out that Boehmer's et al. method has low Type I error rate, since it is unlikely to misclassify institutional trades as individual trades, but high Type II error rate, since it captures only part of the individual order flow.[9]

Figure 2, Panel B, presents the cumulative change in individual order imbalance for the top and bottom $\Delta FLRT_{iq}$ portfolios. The evidence is consistent with uninformed individual trading around earnings announcements. Specifically, the individual order imbalance for retailers in the short sell portfolio increases significantly prior to the earnings announcement while the order imbalance for the buy portfolio remains near zero. Together the evidence shows that the dynamics of short-selling activity and individual trading mirror each other. As short sellers actively target retailers with bad news for the quarter, we find

[9] Institutional trades are almost never sent to wholesalers or internalized but are traded on exchanges or in dark pools, and Regulation National Market System (NMS) prohibits these trades from receiving fractional price improvements. As a result, Boehmer's et al. measure is unlikely to misclassify institutional trades as individual trades. With respect to individual order flow, Boehmer's et al. measure captures only the marketable order flow component, which includes market orders and marketable limit orders, but it does not capture nonmarketable limit orders. This limitation is alleviated by the fact that marketable order flow accounts for the majority of individual order flow.

evidence that individual investors are net buyers of such retailers. Next, we present evidence that the ability of short sellers to profit from targeting quarterly earnings announcements has increased after the introduction of satellite data for the group of retailers with satellite coverage.

3.4 Identifying the effect on short-selling activity

What is the effect of the introduction of satellite data on informed short-selling activity? An increase in the informativeness of short-selling activity would imply that short-sellers' ability to anticipate negative news for the quarter increases after the introduction of satellite coverage. To estimate the effect of satellite coverage on the informativeness of short-selling activity, we implement a difference-in-differences (DID) design using the following model:

$$\begin{aligned} EARET_{iq} = a &+ \beta_1 POST_{iq} \times TREAT_{iq} + \beta_2 \Delta SHORT_{iq} + \beta_3 POST_{iq} \times \Delta SHORT_{iq} \\ &+ \beta_4 TREAT_{iq} \times \Delta SHORT_{iq} + \beta_5 POST_{iq} \times TREAT_{iq} \times \Delta SHORT_{iq} + C_{iq} \\ &+ C_{iq} \times \Delta SHORT_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (2). \end{aligned}$$

The DID design compares the group of covered retailers to a matched control group. For each covered firm, we use a symmetric event window before and after the initiation of satellite coverage. To construct the matched control group, we use the FactSet Revere database to identify all named competitors, including those reported by the target company itself or by the competitors.[10] Then, we zero in on named competitors that operate in the same six-digit GICS industry and do not have satellite coverage. Our choice of a fine industry definition ensures higher comparability across matched pairs. Lastly, we restrict the control

[10] The FactSet Revere database provides the most comprehensive coverage of firm-relationships that is currently available (e.g., Gofman et al. 2020). FactSet analysts monitor the relationships and collect information from firms' annual reports, press releases, investor presentations, and investor relation websites. FactSet Revere's named competitors are widely used for comparable company analysis. In additional analysis, we find similar results when we match target companies to the closest industry peers without conditioning the matched-control group to include named competitors.

group to include the closest named competitors in terms of market capitalization by minimizing the absolute distance across all matched pairs. We note that our matching does not use early treated firms as control firms for late treated ones.[11]

Our procedure identifies an average of 2.4 size-matched competitors per retailer that operate in the same six-digit GICS industry and do not have satellite coverage. Table A3 in the Online Supplement reports the descriptive statistics of the matched pairs. Retailers with satellite coverage are larger, have higher Tobin's Q's, more institutional ownership, and are more likely to have a Big-4 auditor. To control for cross-sectional differences, the vector of time-variant firm characteristics (C_{iq}) in the DID regression model specifications includes log market cap, Tobin's Q, institutional ownership, and the Big-4 auditor indicator, as well as indicators for acquisitions, restructurings, asset write-downs and impairments. We also include interactions of these firm characteristics with $\Delta SHORT_{iq}$ to further control for the potential differences in the relation between $EARET_{iq}$ and short-selling activity across firms.

The dependent variable $EARET_{iq}$ in equation (2) is the cumulative factor-adjusted return from one trading day before to one trading week after the quarterly earnings announcement. Turning to the right-hand-side variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of satellite coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for retailers in the treated group, and $\Delta SHORT_{iq}$ is either the cumulative change in the lender quantity on loan, $\Delta SHORT_{iq}^{Dem}$, or the cumulative change in

[11] Our matching procedure allows us to ensure that our estimates are not impacted by 2×2 DID comparisons of previously treated groups to newly treated groups that typically confound staggered DID applications (Barrios 2021). Goodman-Bacon (2021) discusses the use of group by time fixed effects as one way to address problematic DID comparisons. As an additional test, we repeat our DID tests using matched pair by quarter fixed effects. Table A4 of the Online Supplement reports consistent results, which confirms that our DID estimates are not impacted by comparisons of previously treated groups to newly treated groups.

the available supply of lendable shares, $\Delta SHORT_{iq}^{Sup}$, both measured from the end of the quarter to two days before the earnings announcement.

The coefficient on the triple interaction $POST_{iq} \times TREAT_{iq} \times \Delta SHORT_{iq}^{Dem}$ captures the change in the differential predictive ability of short-selling demand for earnings announcement returns across treated and matched control groups after the introduction of satellite data. An increase in the informativeness of short-selling demand would imply that $\Delta SHORT_{iq}^{Dem}$ is more negatively related to earnings announcement returns in the post period; that is, $\beta_5 < 0$.

Table 4 reports DID regression results for equation (2). The significantly negative coefficient on the triple interaction $POST_{iq} \times TREAT_{iq} \times \Delta SHORT_{iq}^{Dem}$ in Column 1 is consistent with an increase in the informativeness of short-selling demand. More specifically, the evidence shows that the relation between pre-announcement change in lender quantity and earnings announcement returns increases by 1.9 times after the introduction of satellite coverage for the treated group of retailers with satellite data $((\beta_2 + \beta_4 + \beta_3 + \beta_5)/(\beta_2 + \beta_4) = -\,0.055/{-0.029} = 1.9)$. Focusing on the change in the differential predictive ability of short-selling activity for earnings announcement returns, a one standard deviation increase in short-selling demand is associated with a 2.5% lower earnings announcement return for treated retailers relative to the matched control group after the introduction of satellite data. In absolute terms, this magnitude is 6 times the value of the average earnings announcement return.

Turning to Column 2, we observed that the coefficient on the triple interaction $POST_{iq} \times TREAT_{iq} \times \Delta SHORT_{iq}^{Sup}$ is indistinguishable from zero, which implies that there is no detectable change in the informativeness of the supply of lendable shares available from institutions. This null result suggests that evidence of an increase in the informativeness of short demand is not confounded by overlapping changes in the informativeness of short supply.

To be clear, the DID design does not represent a magic bullet that solves the selection problem (e.g., Roberts and Whited 2013). In our setting, the selection problem comes from the fact that data vendors choose which firms to cover and when to begin selling satellite imagery data. As a result, there is no exogenous source of variation in satellite coverage such that firms are randomly assigned as treated or control. Another challenge is that the matched-control firms may be covered by different forms of alternative data. For example, credit card transaction data is available for firms that do not have satellite coverage (e.g., IHS Markit 2019; Zhu 2019).

Zhu (2019) provides related evidence on changes in short seller demand after the introduction of alternative data sources across earnings news partitions. After the introduction of alternative data, Zhu reports that the level of pre-announcement short interest did not change for negative earnings news and it decreased for positive earnings news. While consistent with short demand becoming more sensitive to subsequently disclosed earnings, Zhu's analysis does not provide evidence of increased short interest prior to negative earnings surprises.

Adding to Zhu's study, we find evidence of increased short selling activity leading to the earnings announcement for bad news retailers and we do not find evidence of significant pre-announcement changes in short interest for good news retailers. Furthermore, different from Zhu's analysis, our DID tests zero in on the change in the predictive ability of short-selling activity for earnings announcement returns. The DID estimates provide novel evidence that the introduction of satellite data led to more informed short selling activity leading to the quarterly earnings announcements of retailers with satellite coverage.

3.5 Identifying the effect on individual investor trading

The DID tests so far provide evidence that the ability of short sellers to profit from targeting retailers with bad news for the quarter increased after the introduction of satellite data for the group of retailers with satellite coverage. Next, we attempt to identify the effect of satellite data on individual investor trading.

In the context of Blankespoor's et al. (2020) framework, the substantial costs of acquiring, processing, and integrating satellite data mean that such data can be seen as a form of private information. These costs make it especially difficult for individual investors with limited resources to access the satellite data, making them less informed than sophisticated investors who can afford to incur the substantial costs of acquiring, processing, and integrating such data. The feature of unequal access to satellite data relates to Kyle's (1985) model of the dynamics surrounding insider trading. This model predicts that noise traders will camouflage the trades of insiders who will then profit at the noise traders' expense. In our setting, the satellite data takes the role of inside information for the sophisticated investors with access, and the individual investors without access operate as noise traders. It follows that the introduction of satellite data could lead to individual investor trades becoming less informative.

To estimate the effect of satellite coverage on the informativeness of individual trading activity, we use the following DID regression model:

$$
\begin{aligned}
EARET_{iq} = \alpha &+ \beta_1 POST_{iq} \times TREAT_{iq} + \beta_2 IndOIB_{iq} + \beta_3 POST_{iq} \times IndOIB_{iq} \\
&+ \beta_4 TREAT_{iq} \times IndOIB_{iq} + \beta_5 POST_{iq} \times TREAT_{iq} \times IndOIB_{iq} + C_{iq} \\
&+ C_{iq} \times IndOIB_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (3).
\end{aligned}
$$

The independent variable $IndOIB_{iq}$ measures abnormal individual order imbalance as the average individual order imbalance over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement adjusted for the average individual order imbalance during the quarter. We use Boehmer's et al. (2021) method to identify individual trades and measure the daily individual order imbalance as individual buys minus individual sells divided by the total number of shares outstanding.

The coefficient on the triple interaction $POST_{iq} \times TREAT_{iq} \times IndOIB_{iq}$ captures the change in the informativeness of individual trading activity across the treated and matched control groups after the introduction of satellite coverage. A decrease in the informativeness of individual trading activity in the stock of retail companies with satellite coverage would imply that individual order flow is more negatively related to earnings announcement returns in the post period; that is, $\beta_5 < 0$.

Table 5 reports the DID regression results for equation (3). Column 1 reports that the coefficient on the triple interaction $POST_{iq} \times TREAT_{iq} \times IndOIB_{iq}$ is significantly negative, which is consistent with a decrease in the informativeness of individual order flow for the treated group of retailers with satellite coverage after the introduction of satellite data. In contrast, the significantly positive coefficient on the interaction $POST_{iq} \times IndOIB_{iq}$ implies that the informativeness of individual order flow has actually improved over time for the control group of retailers with no satellite coverage. Put differently, while there is a general positive trend in the informativeness of individual trading activity over time (e.g., Kelley et al. 2013), this trend is overturned for retailers with satellite coverage after the introduction of satellite data.

To gain more insights into the dynamics of individual order flow, we separately examine the effect of the introduction of satellite data on the informativeness of individual buying activity ($IndBUYS_{iq}$) and individual selling activity ($IndBUYS_{iq}$). Columns 2 and 3 provide evidence that the overall decrease in the informativeness of individual order flow is primarily due to changes in the informativeness of individual buying activity. The β_5 coefficient on the triple interaction term is significantly negative for $IndBUYS_{iq}$ and it is indistinguishable from zero for $IndSELLS_{iq}$. Focusing on the change in the differential predictive ability of individual trading for earnings announcement returns, a one standard deviation increase in individual buying activity is associated with a 2.1% lower earnings announcement return for treated retailers relative to the matched control group after the introduction of satellite data. In absolute terms, this magnitude is 5 times the value of the average earnings announcement return.

Together, the evidence suggests that the dynamics of short-selling activity and individual order flow mirror each other. The introduction of satellite data led to an increase in the informativeness of short-selling activity and a decrease in the informativeness of individual buying activity. To be clear, our results do not imply a one-to-one transfer of wealth from individual investors to short sellers, but simply show that short-selling activity is becoming more informed while individual investor activity is becoming less informed with respect to anticipating retailer news for the quarter.

3.6 Identifying the effect on stock liquidity

So far, the DID results provide evidence that the ability of short sellers to preempt bad quarterly news for retailers with satellite coverage has improved after the introduction of satellite data. In contrast, individual investors' trades have become less informative with respect to anticipating retailer news for the quarter. Together, the evidence suggests that the introduction of satellite data could lead to an increase in information asymmetry among market participants.

Prior research concludes that greater information asymmetry among market participants leads to lower stock liquidity (e.g., Copeland and Galai 1983; Glosten and Milgrom 1985; Kyle 1985; Easley and O'Hara 1987). In these models, information asymmetry can increase in either the proportion of informed traders or the precision of their information. Given that the information asymmetry between sophisticated and unsophisticated investors is likely to increase before the earnings announcement (Lee, Mucklow, and Ready, 1993), we expect a decrease in stock liquidity for treated firms in the trading days leading to earnings announcements.

We use the following DID regression model to estimate the effect of satellite coverage on stock liquidity:

$$Y_{iq} = a + \beta_1 POST_{iq} \times TREAT_{iq} + C_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \quad (4).$$

We estimate equation (4) using two complementary measures of stock liquidity. First, we compute the effective spread of a firm's trades using intraday trading data from the TAQ database. We measure the effective spread as the daily firm average of $2 \times (|P_k - M_k|)$, where P_k the price on the observed trade and M_k is the midpoint of the National Best Bid and Offer (NBBO) quotes for that trade ($Spread_{iq}$). Relative to the quoted bid-ask spread, the effective spread offers a more accurate measure of stock liquidity since trades are often executed within the quoted spread (e.g., Petersen and Fialkowski 1994). We then compute the average effective spread from the day after the end of the quarter to two days before the earnings announcement. This pre-announcement measurement window is the same used for the short selling and individual investor order flow tests and allows us to examine the window over which sophisticated investors are likely to have the greatest information advantage. Second, we examine whether the price impact on a trade, the permanent component of the effective spread, is affected by the release of satellite data. We follow Holden and Jacobsen (2014) and measure price impact as the daily average of $2 \times D_k(M_{k+5} - M_k)$, where M_{k+5} is the midpoint five minutes after M_k, and D_k is equal to +1 if we identify the trade as buyer initiated and -1 if it is seller initiated.

Table 6 reports the DID regression results for equation (4). Starting with Column 1, the significantly positive coefficient on the interaction $POST_{iq} \times TREAT_{iq}$ is consistent with an increase in the effective spread and, therefore, a decrease in stock liquidity for the treated group of retailers with satellite coverage after the introduction of satellite data. The inclusion of quarter fixed effects alleviates concerns that our result is due to an aggregate time-trend in the effective spread. Turning to Column 2, we find consistent evidence of an increase in price impact and, therefore, a decrease in stock liquidity for the treated group of retailers with satellite coverage after the introduction of satellite data. Given an average effective spread (price impact) of 0.031 (0.023), the increase in effective spread (price impact) after the release of satellite data represents a shift of 19% (29%) from the unconditional mean.

The effect of satellite coverage on information asymmetry is likely to vary with the information environment of each retailer. In particular, we expect that the introduction of satellite data provides sophisticated investors with a greater edge when uncertainty about firm fundamentals is higher. We use firm size, firm age, and volatility to proxy for a firm's fundamental uncertainty. We conjecture that the effect of satellite data on information asymmetry will be more pronounced for smaller, younger, and high volatility retailers for which fundamental uncertainty is likely to be higher. We test this conjecture using the following DID regression model:

$$Y_{iq} = \alpha + \beta_1 F_{iq} + \beta_2 POST_{iq} \times TREAT_{iq} + \beta_3 POST_{iq} \times F_{iq} + \beta_4 TREAT_{iq} \times F_{iq} + \beta_5 POST_{iq} \times TREAT_{iq} \times F_{iq} + C_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (5).$$

Equation (5) interacts the baseline model in equation (4) with F_{iq}, which denotes the alternative indicators of a retailer's information environment. Focusing on the treated group of retailers, the $I(SMALL_{iq})$ indicator identifies retailers with below median market capitalization, the $I(YOUNG_{iq})$ indicator identifies retailers with below median firm age, and the $I(HVLT_{iq})$ indicator identifies retailers with above median stock return volatility. For the dependent variable, Y_{iq}, we continue to use the effective spread and price impact as our measures of stock liquidity. A significantly positive value for β_5 would suggest that the introduction of satellite data led to a larger increase in information asymmetry for firms with higher fundamental uncertainty.

Table 7 presents the cross-sectional tests and provides evidence of heterogenous treatment effects on stock liquidity. The estimated coefficient on the triple interaction term $POST_{iq} \times TREAT_{iq} \times F_{iq}$ is significantly positive for all three indicators of fundamental uncertainty. The evidence is consistent with the conjecture that the effect of satellite data on information asymmetry is concentrated in smaller, younger, and high volatility retailers for which fundamental uncertainty is likely to be higher. These results are consistent with the notion that introduction of satellite data led to a larger increase in information asymmetry for firms with higher fundamental uncertainty.

To summarize, the stock liquidity results provide evidence consistent with a significant increase in information asymmetry around earnings announcements for retailers with satellite coverage after the introduction of such data. These effects are concentrated in firms with poor information environments. As we also point out in Section 1, individual investors may not be the only group of market participants on the other side of the alternative data-based informed trades. Instead, we argue that individual investors are less likely to have access to alternative data due to the substantial subscription fees to proprietary data vendors and expensive processing costs. Indeed, our communication with RS Metrics highlights that access to data derived from satellite images is only within the reach of sophisticated investors, with select hedge funds being their typical clients. Within this context, our evidence of an overall increase in information asymmetry among outside investors does not preclude that unequal access to alternative data may also increase information asymmetry across groups of sophisticated investors.

3.7 Parallel-trends assumption

A key assumption of the DID estimation is that in the absence of treatment, the average change in outcomes would have been the same for both the treatment and control groups. While we cannot formally test for the parallel-trends assumption, we can evaluate whether pre-treatment trends in outcomes of interest are the same for covered and control firms. We report the parallel-trends analysis in the Table A5 of the Online Supplement. Our analysis shows that covered and control firms are indistinguishable from each other in terms of pre-treatment trends in short-selling and individual trading informativeness as well as in terms of stock liquidity around earnings announcements. [12]

We acknowledge that while evidence of similar pre-treatment trends is comforting, it is not a sufficient condition to ensure that the DID coefficient estimates are valid (e.g., Roberts and Whited 2013). Though we cannot rule out alternative explanations, our tests provide suggestive evidence that the observed effects are more likely to be a result of the introduction of satellite data as opposed to an alternative force.

4. Conclusion

We study the introduction of satellite coverage of major U.S. retailers as a source of alternative data in capital markets. We present evidence that satellite data enabled sophisticated investors with access to such data to formulate profitable trading strategies, especially by targeting the quarterly reports of retailers with bad news for the quarter. Using a DID design, we find that the introduction of the satellite data led to more informed short selling activity, less informed individual buying activity, and lower stock liquidity leading to the quarterly reports of retailers with satellite coverage.

---

[12] As an additional test, we find evidence that the effect of satellite data on short selling activity, individual order flow, and stock liquidity does not kick in immediately. This is consistent with the idea that it takes time for sophisticated investors to procure access to the data. We report these results in Table A6 of the Online Supplement .

Overall, our paper adds to research on the role of alternative data in capital markets by providing evidence that unequal access to satellite data sources can increase information asymmetry among market participants. We acknowledge that the consequences of alternative data may extend beyond the capital markets. Our paper does not explore the social welfare implications of alternative data. Mihet (2020) proposes a theory whereby innovations in financial technology can exacerbate capital income inequality between more sophisticated investors, who can afford to acquire costly private information, and less sophisticated investors, who have little access to private information.

Over time, the value of publicized signals should decay (e.g., McLean and Pontiff 2016). Still, data hunters will continue to scour data from anywhere there is a digital footprint and sophisticated investors will continue to invest in data in their quest to gain an edge. As the market's most sophisticated players come to rely on sources of data that are ever more out of reach for the general public, regulators and policy makers may need to grapple with the question of what the social welfare implications of unequal access to alternative data are.

References

Admati, A., R., and P. Pfleiderer. 1988. "Selling and Trading on Information in Financial Markets", *American Economic Review* 78(2): 96-103

Arcadia Data. 2018. "Alternative Data Strategy: How and Why." Available Online.

Amihud, Y. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets* (5): 31-56.

Banerjee, S., J. Davis, and N. Gondhi. 2018. "When Transparency Improves, Must Prices Reflect Fundamentals Better?". *Review of Financial Studies* 31(6): 2377-2414.

Barrios, J.M., 2021. "Staggeringly Problematic: A Primer on Staggered DiD for Accounting Researchers." Available at SSRN 3794859.

Bartov, E., L. Faurel, and P.S. Mohanram. 2018. "Can Twitter Help Predict Firm-Level Earnings and Stock Returns?" *The Accounting Review* 93(3): 25-57.

Blankespoor, E., G.S. Miller, and H.D. White, 2014a. "The Role of Dissemination in Market Liquidity: Evidence from Firms' Use of Twitter." *The Accounting Review* 89(1): 79-112.

Blankespoor, E., G.S. Miller, and H.D. White, 2014b. "Initial Evidence on the Market Impact of the XBRL Mandate. *Review of Accounting Studies* 19(4): 1468-1503.

Blankespoor, E., deHaan, E. and Marinovic, I., 2020. "Disclosure Processing Costs, Investors' Information Choice, and Equity Market Outcomes: A Review." *Journal of Accounting and Economics*, 70(2-3).

Blankespoor, E., Hendricks, B.E., Piotroski, J.D. and Synn, C., 2022. "Real-Time Revenue and Firm Disclosure." *Review of Accounting Studies*, Forthcoming

Boehmer, E., C. Jones, X. Zhang, and X. Zhang. 2021. "Tracking Retail Investor Activity." *Journal of Finance* 75(5): 2249-2305.

Bollen, J., H. Mao, and X. Zeng, 2011. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2 (1): 1-8.

Chakrabarty, B., Li, B., Nguyen, V. and Van Ness, R.A., 2007. "Trade Classification Algorithms for Electronic Communications Network Trades." *Journal of Banking & Finance* 31(12): 3806-3821.

Copeland, T. E., and D. Galai. 1983. "Information Effects on the Bid-Ask Spread." *The Journal of Finance* 38(5): 1457-1469.

Da, Z., J. Engelberg, and P. Gao. 2011. "In Search of Attention." *The Journal of Finance* 66 (5): 1461–1499.

Dessaint, O., Foucault, T. and Frésard, L., 2021. "Does Alternative Data Improve Financial Forecasting? The Horizon Effect." Available at SSRN 3702411.

Dugast, J., and T. Foucault. 2018. "Data Abundance and Asset Price Informativeness." *Journal of Financial Economics* 130 (2): 367-391.

Easley, D., and M. O'Hara. 1987. "Price, Trade Size, and Information in Securities Markets." *Journal of Financial Economics* 19(1): 69-90.

Farrell, M., T. C. Green, R. Jame, and S. Markov. 2021. "The Democratization of Investment Research: Implications for Retail Investor Profitability and Firm Liquidity." *Journal of Financial Economics* , Forthcoming.

Froot, K., N. Kang, G. Ozik, and R. Sadka. 2017. "What Do Measures of Real-Time Corporate Sales Say about Earnings Surprises and Post-Announcement Returns?" *Journal of Financial Economics* 125 (1): 143–162.

Gao, Gao, X., and Ritter, J.R., 2010. "The Marketing of Seasoned Equity Offerings." *Journal of Financial Economics* 97(1): 33-52.

Gerken, W.C. and Painter, M., 2019. "The Value of Differing Points of View: Evidence from Financial Analysts' Geographic Diversity." Available at SSRN 3479352.

Glosten, L. R., and P. R. Milgrom. 1985. "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *Journal of Financial Economics* 14(1): 71-100.

Gofman, M., Segal, G. and Wu, Y., 2020. "Production Networks and Stock Returns: The Role of Vertical Creative Destruction." *Review of Financial Studies* 33(12): 5856-5905.

Grinblatt, M. and Keloharju, M., 2000. "The Investment Behavior and Performance of Various Investor Types: A study of Finland's Unique Data Set." *Journal of Financial Economics*, 55(1): 43-67.

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

Goodman-Bacon, A., 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254-277.

Heitzman, S. and Klasa, S., 2021. "Informed Trading Reactions to New Private Information: Evidence from nonpublic merger negotiations." *Management Science*, 67(4): 2630-2656.

Hendershott, T., Livdan, D. and Schürhoff, N., 2015. "Are Institutions Informed About News?" *Journal of Financial Economics*, 117(2): 249-287.

Holden, C.W. and Jacobsen, S., 2014. "Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions." *The Journal of Finance* 69(4): 1747-1785.

IHS Markit. 2019. "Demystifying Alternative Data: Can Alternative Data Really Enhance Your Investment Strategy?" Available at the IHS Markit website.

Jame, R., R. Johnston, S. Markov, and M. C. Wolfe. 2016. "The Value of Crowdsourced Earnings Forecasts." *Journal of Accounting Research* 54 (4): 1077-1110.

Kang, J.K., Stice-Lawrence, L. and 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.

Kelley, E. K., and P. C. Tetlock. 2013. "How Wise are Crowds? Insights from Retail orders and Stock Returns." *The Journal of Finance* 68 (3): 1229-1265.

Kyle, A. S. 1985. "Continuous Auctions and Insider Trading." *Econometrica: Journal of the Econometric Society* 53 (6): 1315-1335.

Lang, M., K.V. Lins, and M. Maffett. 2012. "Transparency, Liquidity, and Valuation: International Evidence on When Transparency Matters Most." *Journal of Accounting Research* 50 (3): 729-774.

Lee, C.M., Mucklow, B. and Ready, M.J., 1993. "Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis." *Review of Financial Studies* 6 (2): 345-374.

McLean, R. D., and J. Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance* 71 (1): 5-32.

Mihet, R., 2020. "Financial Innovation and the Inequality Gap." Working Paper. Available at SSRN 3474720

Pedersen, L.H., 2015. "Efficiently inefficient." Princeton University Press.

Petersen, M.A. and Fialkowski, D., 1994. "Posted Versus Effective Spreads: Good Prices or Bad Quotes?" *Journal of Financial Economics* 35(3): 269-292.

Walton, S. and Huey, J., 1992. Made in America (Vol. 216). New York: Doubleday.

Zhu C. 2019. "Big Data as a Governance Mechanism." *Review of Financial Studies* 32 (5): 2021-2061.

Table 1
Descriptive analysis

Panel A: Empirical distributions.

	Mean	Std. Dev.	Min	p25	p50	p75	Max
ΔSSS_{iq}	0.013	0.057	-0.317	-0.012	0.016	0.043	0.240
$FLRT_{iq}$	0.298	0.099	0.131	0.229	0.268	0.353	0.604
$\Delta FLRT_{iq}$	-0.007	0.049	-0.295	-0.034	-0.007	0.018	0.415
$QRET_{iq}$	-0.026	0.163	-0.624	-0.120	-0.016	0.066	0.773
$EARET_{iq}$	-0.004	0.097	-0.403	-0.054	-0.002	0.048	0.472

Panel B: Pairwise correlations.

	(1)	(2)	(3)	(4)
(1) ΔSSS_{iq}		0.371^{***}	0.336^{***}	0.203^{***}
(2) $\Delta FLRT_{iq}$	0.383^{***}		0.047	0.130^{***}
(3) $QRET_{iq}$	0.280^{***}	0.032		0.021
(4) $EARET_{iq}$	0.174^{***}	0.123^{***}	-0.017	

Panel C: Industry breakdown.

Industry Name	GICS code	# of firms	% Sales	% Mkt Cap
Multiline Retail	255030	10	85%	77%
Specialty Retail	255040	21	53%	70%
Food & Staples Retailing	301010	6	57%	52%
Hotels, Restaurants & Leisure	253010	6	9%	12%
Chemicals	151010	1	4%	4%

This table presents descriptive statistics. Panel A reports the empirical distributions of key variables. Panel B reports Pearson (Spearman) pairwise correlations below (above) the main diagonal. Panel C reports the distribution of the 44 U.S. companies in our sample across their six-digit GICS industries and their aggregate contribution to the sales and market value separately for each industry. *** indicates statistical significance at the 1% level using two-tailed tests. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4.

Table 2
Forward-looking content of satellite imagery data

Panel A: Baseline specification.

	Dependent Variable $= \Delta SSS_{iq}$		
	(1)	(2)	(3)
$\Delta FLRT_{iq}$	0.008**	0.008***	0.008***
	(3.46)	(3.75)	(4.41)
ΔSSS_{iq-1}	0.044***	0.043***	0.039***
	(15.39)	(16.53)	(12.59)
$QRET_{iq}$	.	0.007***	0.007***
		(3.87)	(7.64)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adjusted R^2	69.1%	70.6%	71.9%
OBS.	650	650	649

Panel B: Alternative specification.

	Dependent Variable $= \Delta SSS_{iq}$		
	(1)	(2)	(3)
$\Delta FLRT_{iq}^{-}$	0.017***	0.016***	0.015**
	(5.57)	(4.95)	(3.39)
$\Delta FLRT_{iq}^{+}$	0.006*	0.006**	0.005**
	(2.33)	(2.79)	(2.55)
ΔSSS_{iq-1}	0.043***	0.042***	0.038***
	(18.63)	(19.52)	(14.02)
$QRET_{iq}$		0.007***	0.006***
		(3.79)	(7.92)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adj. R^2	70.0%	71.3%	72.3%
OBS.	650	650	649

This table provides evidence that growth in same-store parking lot fill rates predicts growth in same-store retailer sales. Panel A reports results from the baseline linear regression model specification. Panel B reports results from the alternative specification that allows for a different coefficient on negative and positive values of growth in same-store parking lot fill rates. We report regression results using the standardized z-values of the continuous predictors. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests.

Table 3
Formulating a trading strategy using satellite imagery data

Panel A: Portfolio performance before stock loan fees.

	Portfolio returns before stock loan fees		
	Raw Returns	Market Adjusted	Factor Adjusted
Sell Portfolio	-2.82%***	-3.01%***	-3.10%***
	(-2.90)	(-3.13)	(-3.25)
Buy Portfolio	1.78%**	1.63%**	1.66%**
	(2.38)	(2.17)	(2.22)
Buy-minus-sell	4.60%***	4.64%***	4.76%***
	(3.75)	(3.80)	(3.93)

Panel B: Portfolio performance after stock loan fees.

	Portfolio returns after stock loan fees		
	Raw Returns	Market Adjusted	Factor Adjusted
Sell Portfolio	-2.79%***	-2.98%***	-3.07%***
	(-2.87)	(-3.10)	(-3.22)
Buy Portfolio	1.78%**	1.63%**	1.66%**
	(2.38)	(2.17)	(2.22)
Buy-minus-sell	4.57%***	4.62%***	4.73%***
	(3.73)	(3.78)	(3.90)

This table reports returns from a trading strategy that buys (short sells) retailers with same-store parking lot fill rate growth in the top (bottom) quartile of the cross-sectional distribution of $\Delta FLRT_{iq}$. We report raw, market-adjusted, and factor-adjusted returns over the three-day window centered on quarterly earnings announcements. We use the value-weighted CRSP index including distributions when calculating market-adjusted returns. We use the portfolio data from Kenneth French's website to calculate factor-adjusted returns. To generate the cross-sectional quartile cutoff values of $\Delta FLRT_{iq}$, we consider retailers with fiscal quarters ending within the last three months. Panel A (Panel B) report results before (after) adjusting the short-sell portfolio returns for accumulated stock loan fees. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests.

Table 4
Difference-in-differences: The effect on short-selling activity

	$Dependent\ Variable = EARET_{iq}$	
$Z_{iq} =$	$\Delta SHORT_{iq}^{Dem}$	$\Delta SHORT_{iq}^{Sup}$
$POST_{iq} \times TREAT_{iq}$	0.001	-0.000
	(0.08)	(-0.02)
Z_{iq}	-0.036*	-0.015
	(-1.95)	(-0.64)
$POST_{iq} \times Z_{iq}$	-0.001	-0.001
	(-0.12)	(-0.18)
$TREAT_{iq} \times Z_{iq}$	0.007*	-0.013*
	(1.93)	(-1.82)
$POST_{iq} \times TREAT_{iq} \times Z_{iq}$	-0.025***	0.009
	(-3.16)	(0.77)
Characteristics (C_{iq})	Yes	Yes
$C_{iq} \times Z_{iq}$	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes
Adj. R^2	6.4%	6.3%
OBS.	4,139	4,152

This table provides evidence that the informativeness of short-selling demand increased after the introduction of satellite data. The dependent variable $EARET_{iq}$ is the cumulative factor-adjusted return from one trading day before to one trading week after the quarterly earnings announcement. Turning to the independent variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of satellite coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage, $\Delta SHORT_{iq}^{Dem}$ is the cumulative change in the lender quantity on loan from the end of the quarter to two days before the quarterly announcement, and $\Delta SHORT_{iq}^{Sup}$ is the cumulative change in the available supply of lendable share from the end of the quarter to two days before the quarterly announcement. We report regression results using the standardized z-values of the continuous predictors. For each retailer in the treated group, we use a symmetric event window before and after the initiation of satellite coverage. The treated group of retailers includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. The matched control group includes an average of 2.4 size-matched competitors per retailer that operate in the same six-digit GICS industry and do not have satellite coverage. We obtain information about named competitors from FactSet Revere. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 5
Difference-in-differences: The effect on individual order flow

	$Dependent\ Variable = EARET_{iq}$		
$Z_{iq} =$	$IndOIB_{iq}$	$IndBUYS_{iq}$	$IndSELLS_{iq}$
$POST_{iq} \times TREAT_{iq}$	0.001	0.002	0.002
	(0.06)	(0.17)	(0.18)
Z_{iq}	-0.025	-0.038*	-0.026*
	(-1.64)	(-2.15)	(-1.87)
$POST_{iq} \times Z_{iq}$	0.015***	0.013*	0.002
	(3.17)	(1.81)	(0.31)
$TREAT_{iq} \times Z_{iq}$	0.017***	0.016**	0.008
	(3.21)	(2.88)	(1.12)
$POST_{iq} \times TREAT_{iq} \times Z_{iq}$	-0.025***	-0.021**	-0.010
	(-3.09)	(-2.94)	(-1.24)
Characteristics (C_{iq})	Yes	Yes	Yes
$C_{iq} \times Z_{iq}$	Yes	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes	Yes
Adj. R^2	7.3%	6.9%	6.4%
OBS.	3,877	3,877	3,877

This table provides evidence that the informativeness of individual order flow decreased after the introduction of satellite data. The dependent variable $EARET_{iq}$ is the cumulative factor-adjusted return from one trading day before to one trading week after the quarterly earnings announcement. Turning to the independent variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of satellite coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage, $IndOIB_{iq}$ is the abnormal individual order imbalance, $IndBUYS_{iq}$ is the abnormal individual buying activity, and $IndSELLS_{iq}$ is the abnormal individual selling activity over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement. We use Boehmer's et al. (2021) method to identify the individual order flow in the TAQ data. We report regression results using the standardized z-values of the continuous predictors. The treated group of retailers includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. We report regression results using the standardized z-values of the continuous predictors. For each retailer in the treated group, we use a symmetric event window before and after the initiation of satellite coverage. The treated group of retailers includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. The matched control group includes an average of 2.4 size-matched competitors per retailer that operate in the same six-digit GICS industry and do not have satellite coverage. We obtain information about named competitors from FactSet Revere. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 6
Difference-in-differences: The effect on stock liquidity

	$Dependent\ Variable =$	
	$Spread_{iq}$	$Price\ Impact_{iq}$
$POST_{iq} \times TREAT_{iq}$	0.00584***	0.00667***
	(3.19)	(3.89)
Characteristics (C_{iq})	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes
Adj. R^2	79.7%	74.8%
OBS.	3,726	3,726

This table provides evidence that stock liquidity decreased after the introduction of satellite data. The dependent variable in column 1 is the effective spread, measured as two times the absolute difference between the price on a trade and the midpoint of the National Best Bid and Offer (NBBO) quotes for that trade. The dependent variable in column 2 is price impact, which measures the permanent component of the effective spread by comparing the midpoint five minutes after the trade to the midpoint at the time of the trade. Both measures are computed from the day after the end of the quarter to two days before the earnings announcement. Turning to the independent variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of satellite coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage. For each retailer in the treated group, we use a symmetric event window before and after the initiation of satellite coverage. The treated group of retailers includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. The matched control group includes an average of 2.4 size-matched competitors per retailer that operate in the same six-digit GICS industry and do not have satellite coverage. We obtain information about named competitors from FactSet Revere. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

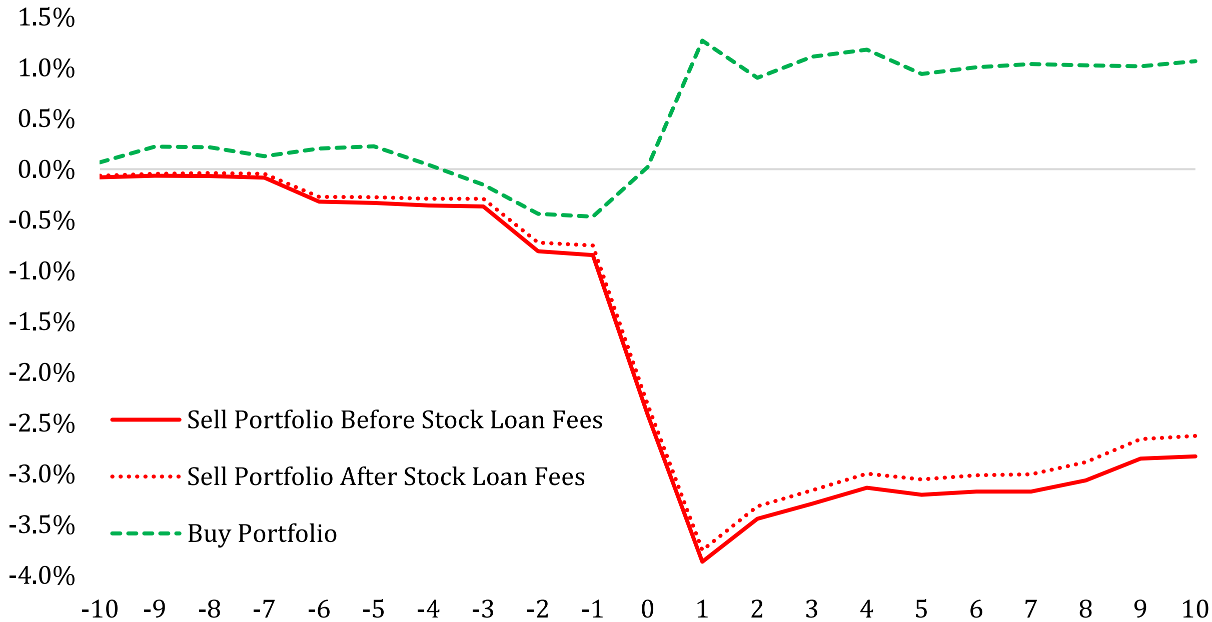
Table 7
Difference-in-differences: Heterogeneous effect on stock liquidity

	$Dependent\ Variable = Spread_{iq}$			$Dependent\ Variable = Price\ Impact_{iq}$		
$F_{iq} =$	$I(SMALL_{iq})$	$I(YOUNG_{iq})$	$I(HVLT_{iq})$	$I(SMALL_{iq})$	$I(YOUNG_{iq})$	$I(HVLT_{iq})$
F_{iq}	-0.00257**	-0.00191	-0.00058	-0.00183**	-0.00174**	-0.00049
	(-2.42)	(-1.54)	(-0.51)	(-2.36)	(-2.64)	(-0.64)
$POST_{iq} \times TREAT_{iq}$	0.00200	-0.00072	0.00234	0.00375*	0.00199	0.00416**
	(0.91)	(-0.32)	(0.86)	(1.82)	(1.09)	(2.20)
$POST_{iq} \times TREAT_{iq} \times F_{iq}$	0.00798***	0.01691***	0.00712**	0.00607***	0.01207***	0.00536**
	(4.42)	(4.04)	(2.27)	(4.70)	(3.79)	(2.75)
Characteristics (C_{iq})	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	79.9%	80.2%	80.7%	74.9%	75.3%	74.9%
OBS.	3,726	3,726	3,717	3,726	3,726	3,717

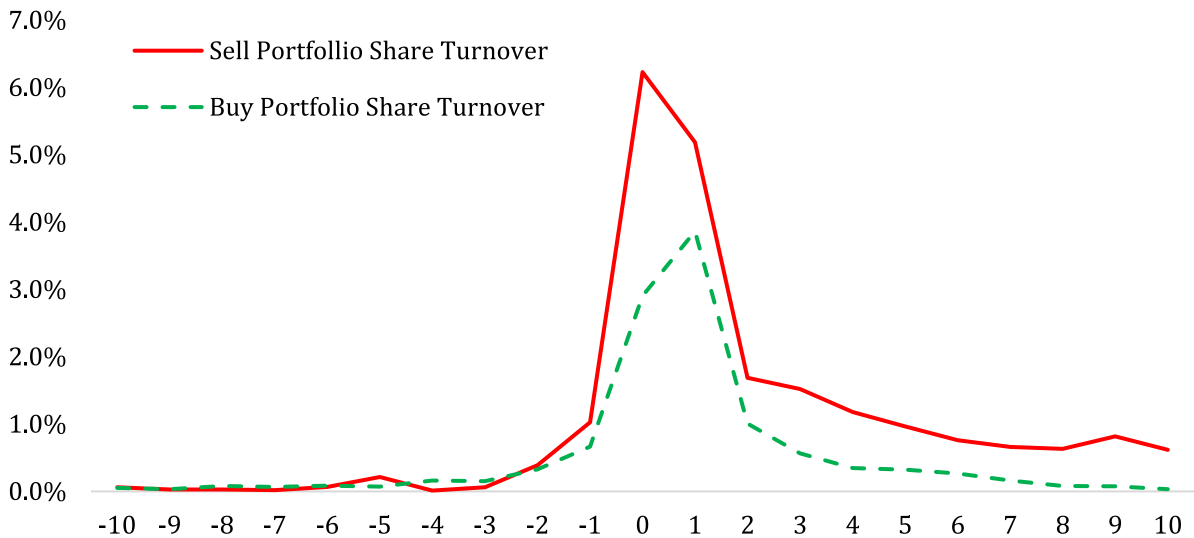
This table provides evidence that effect of satellite data on stock liquidity is more pronounced for smaller, younger, and high volatility retailers. We use two measures of stock liquidity: the effective spread and price impact. as The effective spread measured as two times the absolute difference between the price on a trade and the midpoint of the NBBO quotes for that trade. Price impact, which measures the permanent component of the effective spread by comparing the midpoint five minutes after the trade to the midpoint at the time of the trade. Both measures are computed from the day after the end of the quarter to two days before the earnings announcement. $POST_{iq}$ is an indicator variable that takes the value one after the initiation of satellite coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage. F_{iq} denotes our indicators of a retailer's information environment. Focusing on the treated group of retailers, the $I(SMALL_{iq})$ indicator identifies retailers with below median market capitalization, the $I(YOUNG_{iq})$ indicator identifies retailers with below median firm age, and the $I(HVLT_{iq})$ indicator identifies retailers with above median stock return volatility. We measure age relative to the firm's founding year using data available from Jay Ritter's website. We measure volatility as the standard deviation of the past twelve months prior to the quarter. For each retailer in the treated group, we use a symmetric event window before and after the initiation of satellite coverage. The treated group of retailers includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. The matched control group includes an average of 2.4 size-matched competitors per retailer that operate in the same six-digit GICS industry and do not have satellite coverage. We obtain information about named competitors from FactSet Revere. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Figure 1
Formulating a trading strategy using satellite imagery data

Panel A: Buy and sell portfolio returns.



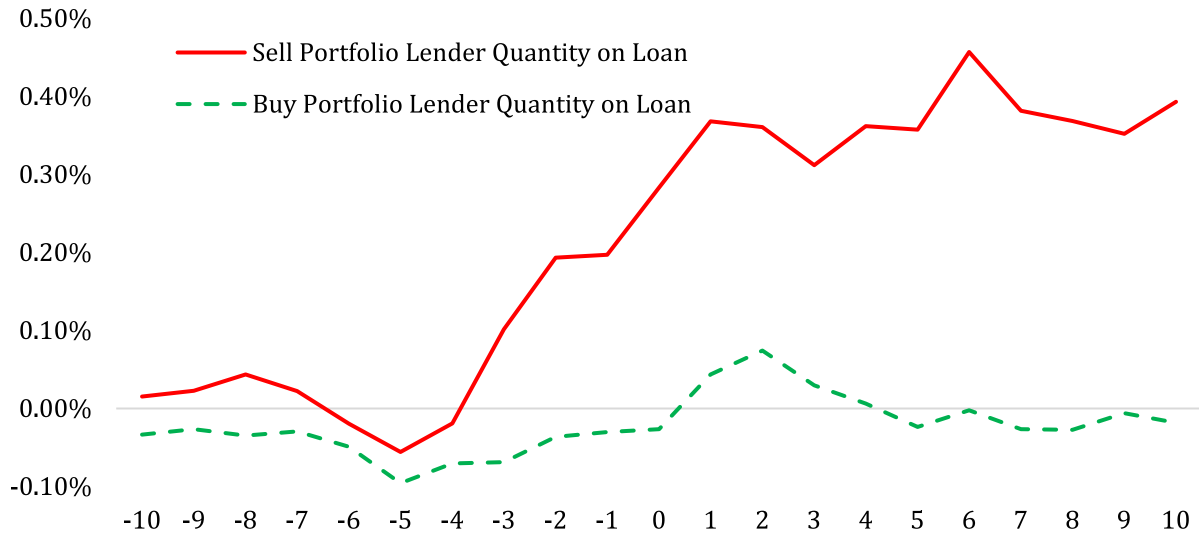
Panel B: Abnormal share turnover.



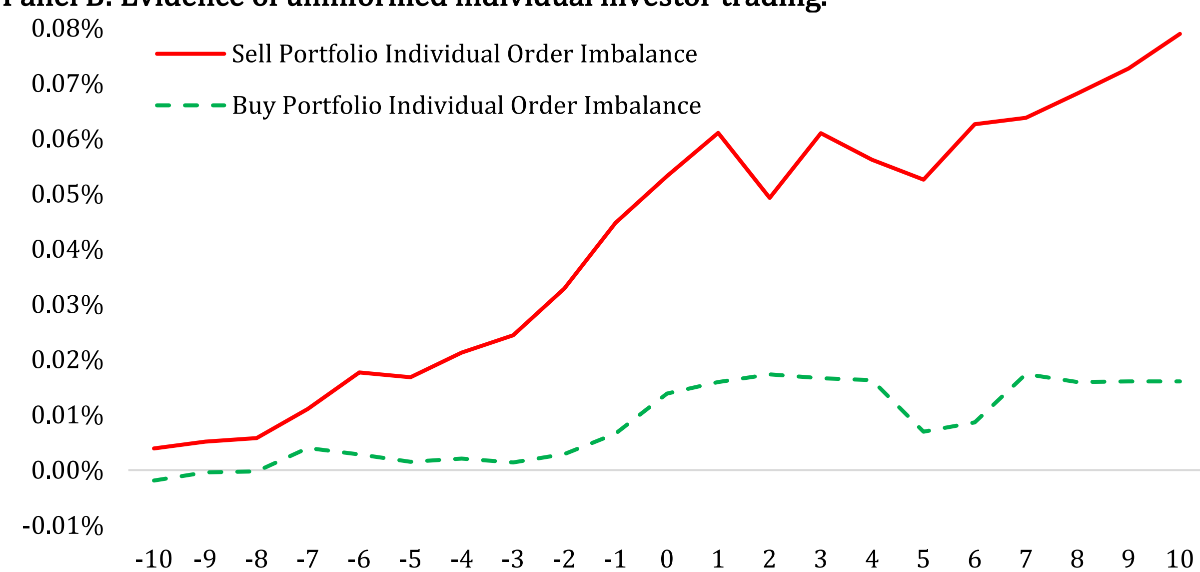
This figure presents the cumulative factor-adjusted return from a strategy that buys (short sells) retailers with same-store parking lot fill rate growth in the top (bottom) quartile of the cross-sectional distribution of $\Delta FLRT_{iq}$. The measurement window is from ten trading days before to ten trading days after the quarterly earnings announcement (day zero). In Panel A, the green dashed (red solid) line presents the performance of the portfolio that buys (short sells) retailers in the top (bottom) quartile portfolio of $\Delta FLRT_{iq}$, while the red dotted line presents the performance of the short sell portfolio net of stock loan fees. In Panel B, we present the daily abnormal share turnover for the buy and short-sell portfolios. We measure share turnover as the daily trading volume divided by the number of shares outstanding. To obtain the abnormal values of share turnover, we adjust the daily values of share turnover during the event window for the daily share turnover prior to the event window. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4.

Figure 2
Informed short-selling activity and uninformed individual investor trading

Panel A: Evidence of informed short-selling activity.



Panel B: Evidence of uninformed individual investor trading.



Panel A of this figure presents the cumulative changes in the lender quantity on loan for the portfolio that buys (short sells) retailers in the top (bottom) quartile portfolio of $\Delta FLRT_{iq}$. To generate cross-sectional quartile cutoff values of $\Delta FLRT_{iq}$, we consider retailers with fiscal quarters ending within the last three months. The measurement window is from ten trading days before to ten trading days after the quarterly earnings announcement day (day zero). The green dashed (red solid) line presents the cumulative change in lender quantity on loan as a percentage of the number of shares outstanding for the portfolio that buys (short sells) retailers in the top (bottom) quartile portfolio of $\Delta FLRT_{iq}$. Panel B of this figure presents the cumulative daily order imbalance of individual investors in the top (bottom) quartile portfolio of $\Delta FLRT_{iq}$. We measure individual order imbalance as the total individual investor-initiated buys minus the total individual initiated sells divided by the total number of shares outstanding. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4.

ONLINE SUPPLEMENT

On the Capital Market Consequences of Alternative Data:

Evidence from Outer Space

Appendix 1: Background on remote sensing technology.

Appendix 2: Key variable definitions.

Supplemental Tables

- **Table A1:** Satellite coverage by company.
- **Table A2:** Forward-looking content of satellite imagery data.
- **Table A3:** Comparison of covered and control companies.
- **Table A4:** Matched pairs by quarter fixed effect analysis.
- **Table A5:** Parallel-trends analysis.
- **Table A6:** Post-breakdown analysis.

Supplemental Figures

- **Figure A1:** Geographical coverage of satellite imagery.
- **Figure A2:** Illustrative example of satellite imagery.

Appendix 1
Background on remote sensing technology

Mounting cameras to take pictures of the surface of the earth was the driving force behind early satellite launches. While the original purpose was oriented towards military applications and weather forecasting, it was not long before the first applications in economics research. Unlike most communication satellites that follow a geostationary orbit (at about 36,000km altitude) and remain in a fixed point above the equator relative to the surface of the earth, the satellites of interest to us orbit the earth at lower altitudes.

These remote sensing satellites typically provide full coverage of the earth's surface. The first publicly available data set originated from the U.S. Air Force's Defense Meteorological Satellite Program (DMSP) and NASA's Landsat system. The spacecraft in this program orbit the earth at altitudes around 700km and take advantage of the smaller distance and the different orbital characteristics to produce higher resolution images. Because of the lower orbit, theses satellites move fast above the surface and orbit the earth about every 99 minutes or over 14 times a day. The near polar orbits are set up such that they miss the poles only by few degrees and move mostly in a northerly/southerly direction taking images of the surface in "vertical" strips. Furthermore, the orbits are designed to be "sun-synchronous" such that the satellite passes a given latitude the same time of the day, every day. Since the earth rotates under the orbit, the cameras record a different strip of the surface on each revolution. Combining the different strips results in a full coverage of the surface, where each point is covered at least once a day, at the same time of the day. With multiple satellites in a system the frequency can be increased.

RS Metrics sources raw satellite imagery from companies such as DigitalGlobe Inc., a division of Maxar Technologies and Airbus Defense and Space, formerly known as the European Aeronautic Defense and Space Company (EADS). These companies provide raw satellite images using their low orbit satellite constellations. For example, EADS launched the Pleiades 1A & 1B satellites as part of a new constellation in December 2011 and December 2012, respectively. Both satellites share the same sun-synchronous orbit, 180 degrees apart at an altitude of 694 kilometers with an orbital period of 99 minutes. The orbits are designed that at least one of the satellites crosses over a given latitude/longitude at roughly the same local time every day. Each satellite photographs a north-south oriented swath of the surface of the Earth, with each swath shifting in the direction opposite to the rotation of the earth. Given the wide viewing angle and the resulting over 1 million square kilometers per day coverage capacity, the constellation provides daily revisit of each point at around the same local time. The satellites have very high-resolution cameras that provide a 0.5m resolution panchromatic and pan-sharpened multispectral images that capture a large part of the electromagnetic spectrum. This level of resolution makes it possible to measure parking lot traffic at the individual store-level.

Appendix 2
Key variable definitions

Variable	Definition
ΔSSS_{iq}	Year-over-year growth in domestic same-store sales. We obtained information on quarterly realizations of same-store sales from FactSet Fundamentals.
$\Delta FLRT_{iq}$	Year-over-year growth in parking lot fill rates. We obtained daily store-level information on parking lot traffic and capacity from RS Metrics.
$QRET_{iq}$	Buy-and-hold size and book-to-market adjusted stock return cumulated from the beginning to the end of quarter q. We obtained stock return data from CRSP.
$EARET_{iq}$	Buy-and-hold size and book-to-market adjusted stock return centered on the earnings announcement. We obtained stock return data from CRSP. We use the portfolio data from Kenneth French's website to calculate factor-adjusted returns.
$\Delta SHORT_{iq}^{Dem}$	The cumulative change in lender quantity on loan from the end of the quarter to two days before the earnings announcement. We obtained daily data on lender quantity on loan from Markit.
$\Delta SHORT_{iq}^{Sup}$	The cumulative change in active supply of lendable shares from the end of the quarter to two days before the earnings announcement. We obtained daily data on active lendable quantity from Markit.
$I(HVLT_{iq})$	Indicator variable equal to one if a firm's stock return volatility is above the median. We measure volatility as the standard deviation of stock returns for the past twelve months prior to the quarter.
$I(SMALL_{iq})$	Indicator variable equal to one if a firm's market cap is below the median market cap of all firms in that quarter.
$I(YOUNG_{iq})$	Indicator variable equal to one if a firm's age is below the median. We measure age relative to the firm's founding year using data available from Jay Ritter's website.
$IndOIB_{iq}$	The average individual order imbalance over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement adjusted for the average individual order imbalance during the quarter. We use Boehmer's et al. (2021) method to identify individual trades and measure the daily individual order imbalance as buys minus sells divided by the total number of shares outstanding.

$IndBUYS_{iq}$	The average individual buying activity over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement adjusted for the average individual buying activity during the quarter. We use Boehmer's et al. (2021) method to identify individual trades and measure the daily individual buying activity as individual buys divided by the total number of shares outstanding.
$IndSELLS_{iq}$	The average individual selling activity over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement adjusted for the average individual selling activity during the quarter. We use Boehmer's et al. (2021) method to identify individual trades and measure the daily individual selling activity as individual sells divided by the total number of shares outstanding.
$Spread_{iq}$	Effective spread defined as $2 \times (\|P_k - M_k\|)$, where P_k the price on a trade and M_k is the midpoint of the National Best Bid and Offer (NBBO) quotes for that trade, then averaged from the end of the quarter to two days before the earnings announcement. Intraday trading data comes from the TAQ database.
$Price\ Impact_{iq}$	Price impact defined as $2 \times D_k(M_{k+5} - M_k)$, where M_k is the midpoint of the NBBO quotes on trade k, M_{k+5} is the midpoint five minutes after M_k, and D_k is equal to $+1$ if we identify the trade as buyer initiated and -1 if it is seller initiated. We follow Chakrabarty et al. (2006) to identify buyer versus seller-initiated trades. This measure is averaged from one trading day before to one trading week after the earnings announcement.

Table A1
Satellite coverage by company

	Company Name	GICS Industry	Store Count	Coverage	Starting Date
1	Bed Bath & Beyond Inc. (BBBY)	Specialty Retail	1,468	46%	2011:Q3
2	Best Buy Co., Inc. (BBY)	Specialty Retail	2,403	38%	2011:Q3
3	Big 5 Sporting Goods Corporation (BGFV)	Specialty Retail	433	79%	2013:Q4
4	Big Lots, Inc. (BIG)	Multiline Retail	1,491	69%	2012:Q4
5	BJ's Restaurants, Inc. (BJRI)	Hotels, Restaurants & Leisure	170	77%	2013:Q4
6	Buffalo Wild Wings, Inc. (BWLD)	Hotels, Restaurants & Leisure	1,077	60%	2012:Q2
7	Burlington Stores, Inc. (BURL)	Specialty Retail	587	67%	2016:Q1
8	Cabela's Incorporated (CAB)	Specialty Retail	66	64%	2013:Q1
9	CarMax, Inc. (KMX)	Specialty Retail	177	81%	2016:Q4
10	Chipotle Mexican Grill, Inc. (CMG)	Hotels, Restaurants & Leisure	1,830	56%	2012:Q2
11	Conn's, Inc. (CONN)	Specialty Retail	108	79%	2015:Q2
12	Costco Wholesale Corporation (COST)	Food & Staples Retailing	741	40%	2017:Q4
13	Dick's Sporting Goods, Inc. (DKS)	Specialty Retail	769	61%	2015:Q2
14	Dillard's, Inc. (DDS)	Multiline Retail	293	66%	2016:Q4
15	Dollar General Corporation (DG)	Multiline Retail	12,246	33%	2013:Q2
16	Dollar Tree, Inc. (DLTR)	Multiline Retail	11,448	46%	2014:Q2
17	El Pollo Loco Holdings, Inc. (LOCO)	Hotels, Restaurants & Leisure	460	86%	2016:Q2
18	Home Depot, Inc. (HD)	Specialty Retail	2,264	61%	2011:Q1
19	J. C. Penney Company, Inc. (JCP)	Multiline Retail	1,062	66%	2011:Q4
20	Kohl's Corporation (KSS)	Multiline Retail	1,158	69%	2011:Q4
21	Kroger Co. (KR)	Food & Staples Retailing	3,892	51%	2016:Q1
22	Lowe's Companies, Inc. (LOW)	Specialty Retail	1,862	68%	2011:Q1

23	Lumber Liquidators Holdings, Inc. (LL)	Specialty Retail	361	72%	2013:Q3
24	Macy's Inc. (M)	Multiline Retail	855	30%	2013:Q1
25	Monro Inc. (MNRO)	Specialty Retail	1,085	42%	2012:Q3
26	Nordstrom, Inc. (JWN)	Multiline Retail	352	42%	2016:Q4
27	Panera Bread Company (PNRA)	Hotels, Restaurants & Leisure	1,821	59%	2011:Q4
28	Party City Holdco, Inc. (PRTY)	Specialty Retail	929	71%	2016:Q2
29	PetSmart, Inc. (PETM)	Specialty Retail	1,320	63%	2012:Q2
30	Pier 1 Imports, Inc. (PIR)	Specialty Retail	1,035	72%	2014:Q4
31	Ross Stores, Inc. (ROST)	Specialty Retail	1,491	63%	2015:Q1
32	Safeway Inc. (SWY)	Food & Staples Retailing	1,371	63%	2013:Q2
33	Sears Holdings Corporation (SHLDQ)	Multiline Retail	1,693	74%	2014:Q2
34	Sherwin-Williams Company (SHW)	Chemicals	4,202	48%	2012:Q3
35	Smart & Final Stores, Inc. (SFS)	Food & Staples Retailing	311	77%	2016:Q4
36	Staples, Inc. (SPLS)	Specialty Retail	2,067	43%	2011:Q4
37	Starbucks Corporation (SBUX)	Hotels, Restaurants & Leisure	21,942	8%	2012:Q2
40	TJX Companies Inc. (TJX)	Multiline Retail	3,785	38%	2016:Q1
38	Target Corporation (TGT)	Specialty Retail	1,819	71%	2011:Q3
39	The Container Store Group, Inc. (TCS)	Specialty Retail	80	66%	2014:Q4
41	Tractor Supply Company (TSCO)	Specialty Retail	1,412	46%	2012:Q1
42	Ulta Beauty Inc. (ULTA)	Specialty Retail	781	64%	2012:Q3
43	Walmart Inc. (WMT)	Food & Staples Retailing	10,957	18%	2011:Q1
44	Whole Foods Market, Inc. (WFM)	Food & Staples Retailing	444	61%	2015:Q2

This table reports information about the average store count and satellite coverage for each of the 44 U.S. companies in our sample along with the starting date of RS Metrics coverage.

Table A2
Additional evidence for overall parking lot fill rates

Panel A: Baseline specification.

	Dependent Variable = ΔSSS_{iq}		
	(1)	(2)	(3)
$\Delta FLRT_{iq}$	0.007**	0.007**	0.008***
	(3.35)	(3.59)	(4.82)
ΔSSS_{iq-1}	0.045***	0.044***	0.039***
	(16.46)	(18.23)	(13.23)
$QRET_{iq}$	.	0.007***	0.007***
		(3.77)	(8.47)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adjusted R^2	68.7%	70.2%	71.7%
OBS.	650	650	649

Panel B: Alternative specification.

	Dependent Variable = ΔSSS_{iq}		
	(1)	(2)	(3)
$\Delta FLRT_{iq}^{Q1}$	0.012**	0.012**	0.012*
	(2.89)	(2.99)	(2.31)
$\Delta FLRT_{iq}^{Q4}$	0.006**	0.006**	0.006**
	(2.57)	(3.31)	(2.98)
ΔSSS_{iq-1}	0.044***	0.043***	0.038***
	(17.11)	(18.89)	(12.83)
$QRET_{iq}$	.	0.007**	0.007***
		(3.65)	(8.11)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adj. R^2	69.2%	70.6%	71.8%
OBS.	650	650	649

This table provides evidence that our results are unchanged when we replace growth in same-store parking lot fill rates with overall growth without conditioning on same-store comparisons. Panel A reports results from the baseline linear regression model specification. Panel B reports results from the alternative specification that allows for a different coefficient on negative and positive values of growth in same-store parking lot fill rates. We report regression results using the standardized z-values of the continuous predictors. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests.

Table A3
Comparison of covered and control companies

	Covered Firms		Control Firms		Dif. of Means	t-stat
	Mean	Median	Mean	Median		
ln(Market Cap)	8.82	8.90	7.23	7.36	1.60	5.53***
Tobin's Q	3.51	2.38	2.64	2.04	0.87	1.77*
Institutional Ownership	85.9%	87.8%	77.8%	79.6%	8.10%	2.41**
Big-4 Auditor	97.7%	100.0%	82.0%	100.0%	15.70%	3.13***
Acquisition	2.27%	0.00%	5.70%	0.00%	-3.43%	-1.01
Restructuring	0.00%	0.00%	0.88%	0.00%	-0.88%	-1.08
Write-Down	2.27%	0.00%	6.14%	0.00%	-3.87%	-1.09

This table presents descriptive statistics, measured at the last quarter before the satellite coverage for the treated and control groups: ln(Market Cap) is the natural log of market cap; Tobin's Q is the ratio of the sum of market value of equity plus the book value of long-term and short-term debt, divided by the sum of book value of equity plus the book value of long-term and short-term debt; Institutional Ownership is the percentage of stock owned by institutional investors; Big-4 Auditor is an indicator variable that is equal to one if the firm is audited by is audited by Deloitte, Ernest & Young, KPMG, or PwC; Acquisition is an indicator variable that is equal to one if the firm reported acquisition costs; Restructuring is an indicator variable that is equal to one if the firm reported restructuring costs; Write-Down is an indicator variable that is equal to one if the firm reported goodwill impairments or other asset write-offs. ***, **, * denote significance at the 1%, 5%, and 10% level for a two-tailed test, respectively.

Table A4
Matched pairs by quarter fixed effects

Panel A: Short-selling activity and individual investor trading.

	$Dependent\ Variable = EARET_{iq}$	
$Z_{iq} =$	$\Delta SHORT_{iq}^{Dem}$	$IndOIB_{iq}$
$POST_{iq} \times TREAT_{iq} \times Z_{iq}$	-0.019**	-0.025**
	(-2.52)	(-2.11)
Characteristics (C_{iq})	Yes	Yes
$C_{iq} \times Z_{iq}$	Yes	Yes
Matched Pairs by Quarter Fixed Effects ($\xi_i \times \delta_q$)	Yes	Yes
Adj. R^2	4.7%	5.9%
OBS.	4,139	3,877

Panel B: Stock Liquidity.

	$Dependent\ Variable =$	
	$Spread_{iq}$	$Price\ Impact_{iq}$
$POST_{iq} \times TREAT_{iq}$	0.01062**	0.00905***
	(2.72)	(3.22)
Characteristics (C_{iq})	Yes	Yes
Matched Pairs by Quarter Fixed Effects ($\xi_i \times \delta_q$)	Yes	Yes
Adj. R^2	26.6%	25.0%
OBS.	3,726	3,726

This table provides more evidence on the effect of satellite data on short selling activity, individual order flow, and stock liquidity. Panel A estimates the main DID regression models in Tables 4 and 5 in the manuscript using matched pairs by quarter fixed effects. Similarly, Panel B estimates the DID regression models in Table 6 in the manuscript using matched pair by quarter fixed effects. We report t-statistics in parentheses based on clustered standard errors by time.

Table A5
Parallel-trends analysis

Panel A: Short-selling activity and individual investor trading.

	Dependent Variable = $EARET_{iq}$	
Z_{iq} =	$\Delta SHORT_{iq}^{Dem}$	$IndOIB_{iq}$
$TREND_{iq} \times TREAT_{iq} \times Z_{iq}$	-0.000	-0.000
	(-0.55)	(-0.01)
Characteristics (C_{iq})	Yes	Yes
$C_{iq} \times Z_{iq}$	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes
Adj. R^2	5.6%	3.8%
OBS.	2,099	2,143

Panel B: Stock Liquidity.

	Dependent Variable =	
	$Spread_{iq}$	$Price\ Impact_{iq}$
$TREND_{iq} \times TREAT_{iq}$	-0.00020	0.00004
	(-1.60)	(0.44)
Characteristics (C_{iq})	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes
Adj. R^2	80.1%	73.8%
OBS.	1,873	1,873

This table provides evidence that the group of covered firms and control firms are indistinguishable from each other in terms of pre-treatment trends in short-selling activity and individual trading informativeness (Panel A) as well as in terms of stock liquidity leading to quarterly earnings announcements (Panel B). Panel A estimates the main DID regression models in Tables 4 and 5 in the manuscript using $Post = 0$ observations and replacing the $POST$ indicator with a time-trend variable that is increasing by one for each calendar year-quarter, denoted as $Trend$. Similarly, Panel B estimates the DID regression models in Table 6 in the manuscript. We report t-statistics in parentheses based on clustered standard errors by time.

Table A6
Post-breakdown analysis

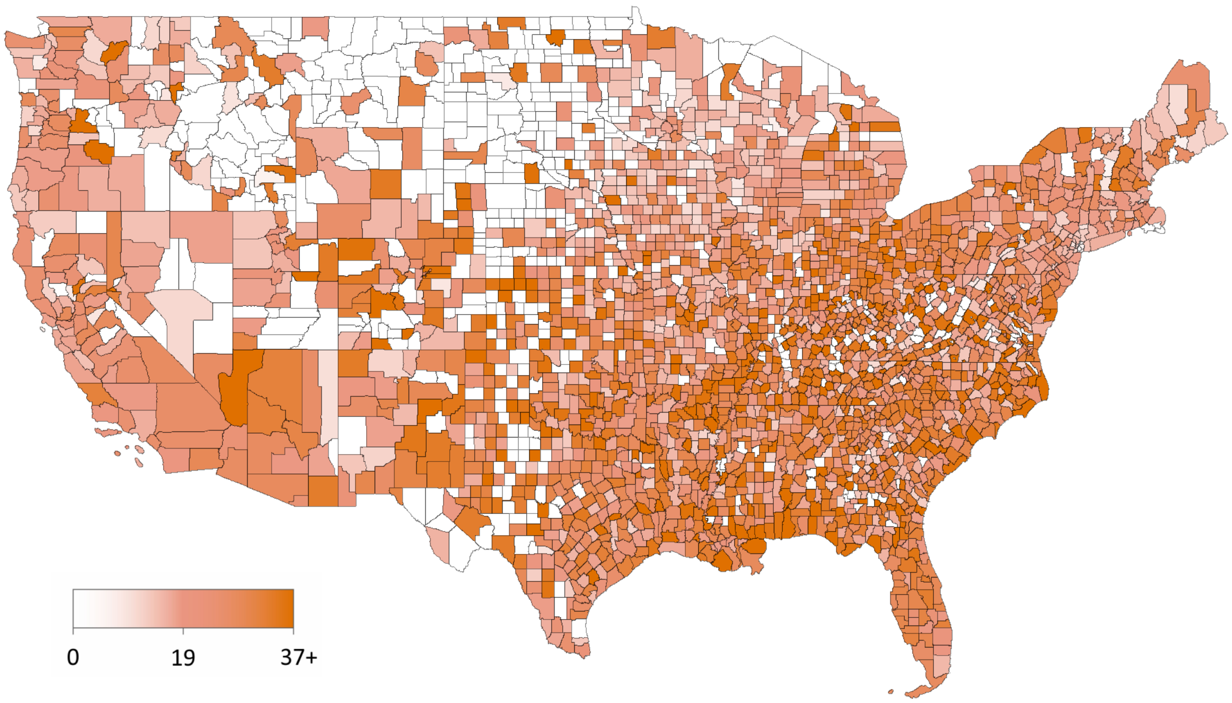
$Z_{iq} =$	Dependent Variable = $EARET_{iq}$	
	$\Delta SHORT_{iq}^{Dem}$	$IndOIB_{iq}$
$POST\_Q1_{iq} \times TREAT_{iq} \times Z_{iq}$	-0.030	-0.009
	(-1.15)	(-0.68)
$POST\_Q2_{iq} \times TREAT_{iq} \times Z_{iq}$	-0.025***	-0.027***
	(-3.18)	(-3.06)
Characteristics (C_{iq})	Yes	Yes
$C_{iq} \times Z_{iq}$	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes
Adj. R^2	6.3%	7.3%
OBS.	4,139	3,877

Panel B: Stock Liquidity.

	Dependent Variable =	
	$Spread_{iq}$	$Price\ Impact_{iq}$
$POST\_Q1_{iq} \times TREAT_{iq}$	0.00309*	0.00320
	(1.81)	(1.66)
$POST\_Q2_{iq} \times TREAT_{iq}$	0.00613***	0.00701***
	(3.24)	(4.07)
Characteristics (C_{iq})	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes
Adj. R^2	79.7%	74.8%
OBS.	3,726	3,726

This table provides evidence that the effect of satellite data on short selling activity, individual order flow, and stock liquidity does not kick in immediately. Panel A estimates the main DID regression models in Tables 4 and 5 in the manuscript after decomposing the $POST$ indicator into the $POST\_Q1_{iq}$ an indicator variable that takes the value one in the first quarter after the initiation of satellite coverage and the $POST\_Q2_{iq}$ an indicator variable that takes the value one after the first quarter following the initiation of satellite coverage. Similarly, Panel B estimates the DID regression models in Table 6 in the manuscript. We report t-statistics in parentheses based on clustered standard errors by time.

Figure A1
Geographical coverage of satellite imagery



This figure presents the number of individual store locations with satellite coverage per 100,000 residents across counties in the U.S. The underlying data covers 67,078 individual store locations across 2,571 counties covering 98% of the U.S. population. For each county, we compute the number of individual store locations with satellite coverage per 100,000 residents. Across counties, the mean (median) store count per 100,000 residents is 18.11 (18.61) stores, with standard deviation of 12.48 and interquartile range from 9.55 to 26.55. The heat map shows that satellite coverage is extensive not only in densely populated areas, but also in more rural counties with the exception of the most sparsely populated ones. In fact, the mean (median) population of counties with no coverage in our data is 7,537 (5,705), while that of counties with coverage is 117,725 (35,767). The color spectrum across counties is proportionately dark to the number of store coverage per capita ranging from white to dark orange. Across counties, the median store count per 100,000 residents is 18.6 stores with interquartile range from 9.6 to 26.6.

Figure A2
Illustrative example of satellite imagery

This figure illustrates the measurement of key variables using satellite imagery data for Target Corporation, the department store company. The satellite image is for the Target store located at 4500 Macdonald Ave, Richmond CA 94805. The image was captured on September 19, 2016, at 11:03am. The processed image indicates the number of cars present within a fixed area of parking lot spaces that RS Metrics assigns to each store. The yellow line outlines the boundary of the parking lot associated with Target and the red dots indicate the occupied parking lot spaces. The parking lot spaces assigned to each store do not change over time unless the company renovates the parking lot. At the time of the satellite image, RS Metrics reports 540 parking lot spaces with 146 of them filled. The parking lot spaces on the bottom right of this Target store are excluded because they may represent employee parking. As a general rule for any individual store location, RS Metrics defines the "most likely parking area" for customers and keeps that parking lot boundary relatively fixed over time so that the variability in the data comes from the number of cars parked at any time. Starting with the granular parking lot data for Target Corporation in 2016:Q3, we identify 1,210 individual store locations across the U.S. with year-over-year satellite coverage, i.e., coverage in both 2016:Q3 and 2015:Q3. We calculate the average parking lot size and parking lot traffic per Target store during the quarter, and we sum across stores to obtain the enterprise-level information. For 2016:Q3 across the 1,210 Target store locations with year-over-year satellite coverage, the aggregate parking lot traffic is 156,977 while the aggregate parking lot space is 595,340. It follows that the parking lot fill rate for Target Corporation in 2016:Q3 is 26.37%. Repeating the steps for 2015:Q3, we find a fill rate of 26.94%. Hence, the year-over-year growth rate in the fill rate is −2.14%.