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

ZSOLT KATONA*
zskatona@berkeley.edu

MARCUS O. PAINTER**
marcus.painter@uky.edu

PANOS N. PATATOUKAS*
panos@haas.berkeley.edu

JIEYIN ZENG*
jieyin_zeng@haas.berkeley.edu

Abstract

We study the emergence of satellite imagery of parking lot traffic across major U.S. retailers as a source of alternative data in capital markets. We find that while measures of parking lot traffic from outer space embed timely value-relevant information, such information is not incorporated into stock prices prior to the public disclosure of retailer performance for the quarter. This creates opportunities for sophisticated investors, who can afford to incur the costs of acquiring and processing satellite imagery data, to formulate profitable trading strategies at the expense of individual investors, who tend to be on the other side of the trade. Overall, our evidence suggests that unequal access to alternative data leaves individual investors outside the information loop, thereby, increasing information asymmetry without necessarily facilitating stock price discovery for the general investment community. While so far the focus has been mostly on the bright side of big and alternative data, our paper suggests that there might be a less auspicious side to the rise of such data in capital markets.

Keywords: Alternative Data; Satellite Imagery; Price Discovery; Information Asymmetry.
Data Availability: Data are available from the sources indicated in the text.

---

*Berkeley-Haas; **University of Kentucky. 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, Kristine Hankins, Russell Jame, Tom Lee, Frank Partnoy, the PhD students at Berkeley Haas, and conference and seminar participants at the 9th Miami Behavioral Finance Conference and the University of Kentucky. We gratefully acknowledge financial support from the Fisher Center for Business Analytics and the Center for Financial Reporting and Management 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 "Unlevelling 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 study the emergence of satellite imagery of parking lot traffic across major U.S. retailers as a source of alternative data in capital markets. We find that while measures of parking lot traffic from outer space embed timely value-relevant information, such information is not incorporated into stock prices prior to the public disclosure of retailer performance for the quarter. This creates opportunities for sophisticated investors, who can afford to incur the costs of acquiring and processing satellite imagery data, to formulate profitable trading strategies at the expense of individual investors, who tend to be on the other side of the trade. Overall, our evidence suggests that unequal access to alternative data leaves individual investors outside the information loop, thereby, increasing information asymmetry without necessarily facilitating stock price discovery for the general investment community. While so far the focus has been mostly on the bright side of big and alternative data, our paper suggests that there might be a less auspicious side to the rise of such data in capital markets.

Keywords: Alternative Data; Satellite Imagery; Price Discovery; Information Asymmetry.
Data Availability: Data are available from the sources indicated in the text.

"Far from creating a level playing field, where more readily available information simply leads to greater market efficiency, the impact of the information revolution is the opposite: it is creating hard-to access "realms" for long-term alpha generation for those players with the scale and resources to take advantage of it." (Morningstar 2018)

1. Introduction

Big data is a big deal.[1] From how we connect to our friends to how we buy products online or even choose TV shows, big data has been transforming our lives in profound ways. Despite the hype, however, there is only limited evidence on the implications of the rise of big data in capital markets for individual or "Main Street" investors. 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 such data is often only within the reach of sophisticated investors who can afford to incur the substantial costs of acquiring and processing the data. This generally leads to unequal access to alternative data across Main Street and Wall Street investors.

What are the implications of unequal access to alternative data for capital markets? The conventional view is that access to new data sets should enhance price discovery in the stock market even if access is restricted to sophisticated investors. This view assumes that stock market prices instantaneously aggregate and disseminate value-relevant information embedded in alternative data sets that would otherwise be inaccessible to small investors. This view goes back to Hayek's (1945) idea of the market as a mechanism for aggregating dispersed bits of knowledge in society and prices as a system for communicating all value-relevant information to every individual market participant. The opposite to the conventional view is that unequal access to alternative data leaves small investors outside the "information loop" and creates trading opportunities for sophisticated investors without necessarily enhancing stock price discovery for the general investment community.

[1] The term big data was added to the Oxford English Dictionary in 2013 and the definition reads as follows "*data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges.*"

Bringing all investors, regardless of their size, into the information loop has long been a key challenge for the Securities and Exchange Commission (SEC) in its mission to serve and safeguard the interests of the Main Street investor. This challenge was emphasized in the opening statement of SEC Chair Arthur Levitt nearly eighteen years ago at the open meeting on Regulation Fair Disclosure (Reg FD) on August 10, 2000: "…*Like that neighborhood with gated entrances and tall fences, moving into the information loop is not always an option for many of America's small investors*."[2] While Reg FD addressed the selective disclosure of information by publicly-traded firms so that small investors have access to market-moving information at the same time Wall Street professionals get it, unequal access to alternative data raises the question whether another tall fence has been raised leaving small investors outside the information loop.[3]

In this paper, we study the emergence of high-resolution satellite imagery data in capital markets. Our primary source of store-level parking lot traffic data is RS Metrics, the first data vendor to introduce daily parking lot traffic signals derived from satellite imagery analysis in the U.S.[4] Satellite imagery is within the reach of sophisticated investors, with hedge funds being the typical clients of RS Metrics. The daily data feeds include point-in-time information about parking lot capacity, i.e., the total number of available parking spaces, and utilization, i.e., the number of occupied parking spaces, at a specific time of the day. The raw data from RS Metrics includes 4.8 million daily observations across 67,120 unique store locations for 44 major U.S. retailers over the period from 2011:Q1 to 2017:Q4. The data covers 2,571 counties representing over 98% of the U.S. population. From the daily store-level parking lot information, we compile a panel of firm-quarter observations of enterprise-

[2] The full text of the statement is available from SEC's News Supplement.

[3] Prior studies provide evidence that Reg FD has been effective in decreasing information asymmetry and increasing stock liquidity (e.g., Chiyachantana et al. 2004; Eleswarapu et al. 2004).

[4] While RS Metrics is the first data vendor to sell satellite imagery-based data to investors in the U.S., there are other competing vendors with Orbital Insight being the most prominent one. Orbital Insight began selling parking lot traffic signals using satellite imagery beginning in 2015:Q2. In the Appendix, we merge store-level data of parking lot traffic from the RS Metrics with Orbital Insight and find evidence that investors with access to data from both vendors could formulate even more accurate trading signals.

level parking lot fill rates. The key variable of interest is the year-over-year growth in same-store parking lot fill rates. The year-over-year comparisons control for seasonal effects in quarterly parking lot utilization. The same-store comparisons control for year-over-year growth in parking lot capacity due to acquisitions and the opening of new stores.

Our first set of tests examines whether measuring parking lot traffic from outer space provides timely insights for "nowcasting" quarterly retailer performance.[5] The evidence shows that year-over-year growth in same-store parking lot fill rates is a timely indicator of current quarter growth in same-store sales—a key driver of retailer performance at existing store locations that is widely followed by market participants. The nowcasting content of year-over-year growth in parking lot fill rates is incremental to that embedded in stock price fluctuations during the quarter as well as lagged realizations of same-store sales growth, and it is robust to controlling for firm-specific time-invariant effects as well as aggregate time-varying effects.

After establishing that fluctuations in parking lot fill rates are incrementally relevant for nowcasting retailer performance, we examine whether financial analysts impound this information when projecting current quarter growth in same-store sales. Such information aggregation on the part of financial analysts would allow individual investors to rely on sell-side consensus forecasts for gaining access to the information loop. However, our analysis shows that financial analysts do not fully aggregate information embedded in parking lot traffic signals when revising their expectations for current quarter growth in same-store sales. As a result, financial analysts' forecast errors of same-stores sales growth for the quarter, i.e., the difference between the actual realizations disclosed with the quarterly earnings report minus the prevailing consensus forecast, are predictable based on parking lot traffic signals.

Next, we examine whether satellite imagery of parking lot utilization can be used to anticipate the stock market reaction to quarterly earnings. Our analysis focuses on

---

[5] Nowcasting is a term derived from the contraction of *now* and *forecasting* refers to the prediction of the present, the very near future and the very recent past in economics (e.g., Giannone et al. 2008).

predicting short-window returns centered on the earnings announcement day typically three trading weeks after the end of the quarter. We find that a trading strategy that buys (short sells) retailers that experience an abnormal increase (decrease) in parking lot fill rates during the quarter generates abnormal returns over the three-day window centered on the earnings announcement date. More specifically, over the three-day earnings announcement window, the buy portfolio outperforms the market by 1.63% while the short-sell portfolio underperforms the market by -3.01%. The three-day spread between the buy and sell portfolios is 4.64%, which is statistically significant and economically important.

Short selling entails borrowing the stock for a loan fee. The evidence shows that hedge portfolio returns remain intact after accounting for the cost of shorting. In addition, our analysis of pre-announcement effects yields only limited evidence of price discovery, with most of the portfolio performance being realized on the earnings announcement day. The lack of pre-earnings announcement effects in the stock market does not preclude informed trading activity prior to the quarterly earnings announcements. In fact, using daily stock loan demand data from a proprietary data vendor (Markit), we provide evidence of informed short-selling activity in the securities lending market. Focusing on retailers with abnormal decreases in parking lot fill rates, we document a significant increase in the lender quantity on loan; that is, the quantity of stock from the lendable quantity that has already been lent, starting five trading days prior the earnings announcement for the quarter. Notwithstanding evidence of informed short selling, Main Street investors cannot "piggyback" on the information of short sellers. This is because securities lending operates as an OTC market and the general investment community can observe short interest data only twice per month and only with a significant delay.

When we implement a difference-in-differences (DID) design that compares the treated group of retailers with satellite coverage to a matched control group, we find evidence that short-sellers' ability to anticipate negative news for the quarter has improved after the introduction of satellite imagery data only for the treated group of retailers. Using Boehmer's et al. (2017) measure of individual investor order flow, we also find evidence that while short sellers are targeting retailers with bad news for the quarter, individual investors are net buyers of such retailers. This finding is consistent with prior evidence on individual investors' tendency to be contrarian traders (e.g., Grinblatt and Keloharju 2000; Kaniel et al. 2007). Importantly, our DID design provides evidence that individual investors' contrarian tendency has actually increased after the introduction of satellite imagery data only for the treated group of retailers with satellite coverage, which mirrors our DID evidence of more informed short-selling activity.

In summary, while parking lot traffic signals from outer space provide value-relevant information, our evidence shows that such signals are not impounded into stock prices prior to the public disclosure of retailer performance for the quarter. This creates opportunities for sophisticated investors with access to satellite imagery data to formulate profitable trading strategies at the expense of individual investors, who tend to be on the other side of the trade. What are the implications of unequal access to alternative data for stock liquidity? On one hand, the availability of alternative data could decrease information asymmetry between firm insiders and outsiders, which would increase stock liquidity. On the other hand, the availability of alternative data could increase information asymmetry between sophisticated investors and individual investors, which would decrease stock liquidity. Consistent with the rise of information asymmetry between sophisticated and individual investors being the dominant effect, we find evidence of a significant increase in the bid-ask spreads around quarterly earnings announcements for the stock of retailers with satellite coverage after the introduction of satellite imagery data.

This is the first paper to provide evidence that unequal access to alternative data leaves small investors outside the information loop without necessarily facilitating stock price discovery. Such unequal access increases information asymmetry between sophisticated investors and individual investors, thereby, decreasing stock liquidity. With respect to market efficiency, an increase in stock price informativeness due to the introduction of satellite imagery data would imply that the market impounds more information during the quarter, so that the price reaction to the quarterly earnings announcement is muted. Far from creating a level playing field, our DID design shows that the introduction of satellite imagery data had no detectable effect on the stock price informativeness for retailers with satellite coverage. This null result together with evidence of a significant decrease in liquidity around earnings announcements for the treated group of retailers with satellite coverage provides compelling evidence that market efficiency does not necessarily improve with the introduction of alternative data.

Our paper adds to growing research on the role of alternative data in capital markets. Few relevant papers in this respect are Bollen et al. (2011), who find that aggregate-level Twitter mood has predictive power for stock market returns; Da et al. (2011), who find evidence of short-term momentum and long-term reversals for stocks with abnormally high Google search frequency; Jame et al. (2016), who find that crowdsourced forecasts from Estimize are incrementally useful in predicting earnings; Froot et al. (2017), who use proprietary data of consumer activity and find that managers distort their disclosures in the presence of insider trading opportunities; Farrell et al (2018), who find that stocks with reductions in coverage on the website Seeking Alpha experience a decrease in liquidity; and Zhu (2019), who finds that alternative data availability, such as consumer transactions and satellite imagery, disciplines managers' insider trading and investment decisions by reducing information asymmetry. More broadly, our paper adds to research on the impact of data abundance on price informativeness. Our evidence that the introduction of satellite coverage had no detectable effect on stock price informativeness is broadly consistent with the theoretical predictions of Banerjee et al. (2018) and Dugast and Foucault (2018), whereby improved data availability does not necessarily increase price informativeness.

The paper proceeds as follows. Section 2 provides the background on remote sensing technology and reviews prior applications of satellite imagery data. Section 3 describes how we measure retailer parking lot traffic using satellite imagery data. Section 4 presents the empirical results. Section 5 concludes.

2. Background

2.1 The evolution of 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. Before we review prior applications of satellite imagery data, we provide a brief overview of the characteristics of the satellites that generate the images.

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 much lower altitudes. These remote sensing satellites typically provide full coverage of the earth's surface. Croft (1978) describes the first publicly available data set originating 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 a couple 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.

An important characteristic of remote sensing applications is the type of sensors used. The variety ranges from passive sensing in different spectra of frequencies including, infrared, visible, or ultraviolet light to active sensing such as LiDAR, which uses reflectance from laser pulses emitted by the satellite. LiDAR is similar to radar but instead of microwave signals uses laser emissions and their reflections to generate observations. Its main advantage compared to passive sensing is that it perceives depth and can construct a three-dimensional map of the observed surface.

2.2 Prior applications of satellite imagery data

Early economics research in the area of remote sensing took advantage of nighttime imagery where populated areas become distinctive due to light emissions. Donaldson and Storeygard (2016) provide an extensive overview of remote sensing applications in economics, including the data sources used. In one of the first applications, Welch (1980) combines Landsat and DMSP data to study urban population and energy consumption. In one application, the paper establishes a functional relationship between nighttime lighting intensity and urban population in China. In another, it uncovers a similar relationship between nighttime lights in urban areas in the U.S. and corresponding energy utilization. Sutton et al. (1997) uncovers the links between nighttime lighting data observed by satellites and population density in the U.S. Continuing in this direction, Sutton et al. (2001) provide global population estimates using similar satellite night-lights data.

Another area of interest is land use. The early work of Skole and Tucker (1993) uses Landsat data to study deforestation in the Amazon. They visually classify the images to ascertain deforestation in the areas under question. Foster and Rosenzweig (2003) combine satellite data with surveys to show that forest growth in India is related to increased demand for products originating in the forest. Muller and Zeller (2002) use manually classified Landsat images to augment meteorological data when examining land use and agricultural output in Vietnam.

Most early work in economics established a relationship between remote sensing data and economic variables. However, the relatively low resolution of traditional nighttime satellite images limits their use as an additional data source. Chen and Nordhaus (2011) compare estimates obtained from DMSP data to traditional output measures and find that satellite data is a valuable proxy for countries with the poorest statistical infrastructure, but high measurement errors in the lighting data limit its use when there is better quality information. Doll et al. (2006) draw attention to potential outliers and possible remedies.

Satellite imagery data has also been used to study growth over time. Henderson et al. (2012) measure real GDP growth from nighttime lights (DMSP) and provide GDP estimates for countries with unreliable economic measures. More recently, a newer technology, Visible Infrared Imaging Radiometer Suite (VIIRS) Lights Data Set has been employed because of its higher accuracy in sensing light intensity provided by onboard radiometric calibration. Li et al. (2013) use VIIRS Lights Data Set as a supplementary source for modeling the regional economy of China. With a less than 1 km × 1 km pixel, wider spectrum, and ability to record dimmer lights, Chen and Nordhaus (2015) also demonstrate the superiority of VIIRS data improving measurements of population and economic output in Africa.

Recent studies using satellite imagery data have further improved detection methods and are able serve a wider variety of applications. For example, Guiteras et al. (2015) study exposure to floods in Bangladesh. They do not rely on nighttime lighting data. Instead, they use Moderate Resolution Imaging Spectroradiometer (MODIS) satellites to record surface reflectance in multiple bands, being able to distinguish between green areas and water covered areas. They find that data is better than other estimates based on precipitation and self-reports of exposure to floods. Another recent example of using high-resolution images is that of Marx et al. (2015), who examine dwelling investments in a Nairobi slum. Algorithmic analysis of these images that are similar to ours in resolution can reveal newly constructed or replaced roofs due to their higher reflectivity than older, rusted ones. Axbard (2016) uses satellite data to construct a monthly measure of local fishing conditions and finds that better income opportunities reduce sea piracy in Indonesia. Easier access to satellite images not only helps academic study of economics, it also has a direct impact on certain industries. Nagaraj (2017) finds that access to Landsat imagery nearly doubled the rate of significant gold discoveries in the mining industry. As we discuss later, availability of satellite images also has an important role in the financial forecasting industry.

2.3 Our application of satellite imagery data

Our application uses high-resolution satellite imagery of parking lot traffic at U.S. retail locations. We obtained satellite imagery data from RS Metrics and Orbital Insight. The data consists of daily store-level information about parking lot capacity and utilization across major U.S retailers. RS Metrics and Orbital insight obtain 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. DigitalGlobe also owns the WorldView constellation and the GeoEye satellite that have similar imaging capabilities.

We note that counting cars from outer space is subject to at least three sources of measurement error. First, satellite coverage is available only for a subset of a retailer's store count. The reason is simply that the cameras have to be pointed in a given direction for a certain store and there is a limited capacity allotted to each satellite user. Relatedly, not all parking lots are visible from outer space. Underground or multi-story lots are obviously hidden from the satellite. 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 US, which captures only a snapshot of total parking lot traffic during the day. Third, even though the resolution of satellite imagery has drastically improved over time, it is still hard to count the cars precisely in a parking lot, due to clouds, haze, trees, shadows and other visual or environmental factors. To mitigate measurement errors, RS Metrics and Orbital Insight process satellite imagery using a combination of automated software and human analysts.

While RS Metrics is the first data vendor to sell satellite imagery-based data to investors in the U.S. starting in 2011:Q1, there are other competing vendors with Orbital Insight being the most prominent one. Orbital Insight began selling parking lot traffic signals using satellite imagery beginning in 2015:Q2. Although RS Metrics and Orbital Insight use the same satellites as their imagery sources, the set of images they use is not the same as they need to decide what images to purchase. Furthermore, data processing and analysis techniques differ between the two vendors. Our main analysis focuses on RS Metrics because it was the first data vendor available to investors. In the appendix (see Table A1 and Figure A1), we expand the RS Metrics data with Orbital Insight data and find evidence that investors with access to data from both data vendors could formulate even more profitable trading strategies at the time of quarterly earnings reports. One implication for investors is that there are complementarities between the two data sets that can help generate more accurate trading signals when combined.

3. Measuring parking lot traffic

3.1 Measuring parking lot traffic at the individual store level

As we explain in Section 2, our primary source of parking lot traffic data is RS Metrics. RS Metrics provides satellite coverage at the individual store level for 44 major U.S. retailers from 2011:Q1 to 2017:Q4. RS Metrics generates the data by first using a proprietary software for automated counts and then analysts for verifying the counts. 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. Our data on parking lot capacity and utilization includes 4.8 million daily observations across a total of 67,120 unique store locations for the 44 U.S. companies with RS Metrics coverage.

Our sample starts in 2011:Q1 because this is the first quarter for which RS Metrics started selling satellite imagery data. We note that RS Metrics was the first data vendor to sell satellite imagery data to investors, with hedge funds being their typical clients. Our sample ends in 2017:Q4 because this is the last quarter for which we obtained satellite imagery data from RS Metrics per our data service agreement.

Table 1 reports information about the store count and satellite store coverage for each of the 44 U.S. companies in our sample along with the starting date of RS Metrics coverage. The cross-sectional average store count is 2,412 with satellite 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 industry group in our sample is specialty stores with 16 retailers, including Walmart Inc., Target Corporation, and Bed, Bath & Beyond Inc. We note that the number of retailers with satellite imagery coverage in our sample increased from 10 in 2011 to 30 in 2014 and 44 in 2017.

Figure 1 presents a heat map to illustrate the geographical coverage of our store-level data at the county level across the U.S. Our store-level data provides coverage for 2,571 counties representing over 98% of the U.S. population. For each individual 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 a 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 some 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).

3.2 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. Specifically, 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; that is, $CARS_{ijq}$, as well as the average number of parking lot spaces available at each store location during the quarter; that is, $SPACES_{ijq}$. Due to seasonal effects in quarterly data, we focus on year-over-year comparisons rather than sequential comparisons; that is, we compare quarter q to quarter $q-4$. To ensure comparability on a year-over-year basis, 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 year-over-year growth in parking lot capacity due to acquisitions and the opening of new stores. Our restricted sample includes 3.4 million daily observations across 53,647 unique store locations for the 44 major U.S. retailers covered from 2011:Q1 to 2017:Q4.

For each retailer-quarter, we sum up across individual store locations with year-over-year satellite coverage to obtain the aggregate parking lot traffic; $CARS_{iq}$, and the aggregate parking lot space; $SPACES_{iq}$. For each retailer-quarter, we calculate the enterprise-level parking lot fill rate—our primary measure of 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 in our analysis is the year-over-year growth in same-store parking lot fill rates measured as:

$$\Delta FLRT_{iq} = \frac{FLRT_{iq} - FLRT_{iq-4}}{FLRT_{iq-4}} \qquad (1).$$

By construction, growth in parking lot fill rates is due to growth in parking lot traffic and growth in parking lot capacity. In our data, most of the variability in same-store parking lot fill rates is due to variability in parking lot traffic rather than parking lot capacity. This is because parking lot capacity at the individual store-level is sticky on a year-over-year basis. Indeed, growth in same-store parking lot fill rates is 99% correlated with growth in same-store car traffic and it is virtually uncorrelated with growth in same-store parking lot capacity. Therefore, our inferences are unchanged when we replace growth in same-store parking lot fill rates with growth in same-store parking lot traffic.

3.3 Illustrative example

Figure 2 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 Richmond, California 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 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%.

4. Empirical Analyses

4.1 Descriptive analysis

Table 2, Panel A, reports the empirical distributions of key variables. Appendix 1 provides the variable definitions. The sample includes 650 firm-quarter observations across 44 major U.S. retailers from 2011:Q1 to 2017:Q4. The mean (median) same-store sales growth is 1.3% (1.6%) with a standard deviation of 5.7% and an interquartile range of -1.2% to 4.3%. The parking lot fill rate level has a mean (median) value of 29.8% (26.8%) with a standard deviation of 9.9% and an interquartile range of 23% to 35%. The distribution of the year-over-year growth in parking lot utilization 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 mean and median cumulative analyst forecast revision is -0.7%, which is consistent with prior evidence of long-term sell-side forecast optimism followed by a walk-down as the forecast horizon shortens. The distribution of earnings announcement stock returns is symmetric around zero and exhibits substantial variability with a standard deviation of 1.9%.

Turning to the pairwise correlations in Table 2, Panel B, we find preliminary evidence that same-store growth in parking lot utilization co-moves with same-store sales growth; the Pearson (Spearman) correlation is 37% (38%). Moreover, the pairwise correlations show that same-store growth in parking lot utilization has predictive power for financial analysts' forecast errors and earnings announcement stock returns. Next, we provide formal empirical tests confirming these preliminary findings.

4.2 Nowcasting same-store sales growth from outer space

Our first objective is to investigate the relevance of satellite imagery of parking lot fill rates for nowcasting current growth in same-store sales. The idea is that seasonally-adjusted changes in parking lot utilization should be correlated with shopper conversion at individual stores. Higher year-over-year growth in same-store parking lot utilization should indicate higher close rates and, therefore, higher same-store sales growth. Our efforts zero in on nowcasting same-store sales growth—a key driver of retailer performance. Indeed, same-store sales are widely reported by publicly-owned retail chains as a crucial element of their operational results. For chains that are growing by opening new stores, same-store sales allows financial analysts to differentiate between sales growth that comes from new stores, and growth from improved operations at existing store locations.

We obtained 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. Year-over-year growth in the domestic component of same-store sales is measured as $\Delta SSS_{iq} = (SSS_{iq} - SSS_{iq-4})/SSS_{iq-4}$. We use year-over-year growth rather than sequential growth due to seasonal effects in quarterly same-store sales data; typically retail sales spike during holiday seasons. Same-store sales growth is mostly auto-correlated at one lag, with a first-order autoregressive coefficient of 0.82, which is consistent with well-documented evidence of mean-reversion in growth rates (e.g., Chan et al. 2003). When nowcasting retailer performance, we are interested in the information content of same-store growth in parking lot utilization that is incremental to that of lagged realizations of same-store sales growth. We note that while same-store sales growth is realized during the quarter, the actual value of realized ΔSSS_{iq} is disclosed only after the end of the quarter at the time of the quarterly earnings report. The typical disclosure lag in our sample, i.e., the lag between the end of the quarter and the quarterly earnings report, is three trading weeks.

Stock price fluctuations within the quarter capture revisions in investors' expectations about corporate value creation. Indeed, a long line of capital markets research provides evidence that stock returns have predictive power for firm performance (e.g., Beaver et al. 1980) and that stock prices lead fundamental performance (e.g., Collins et al. 1987). Given the forward-looking component of stock market prices, we search for the predictive content of growth in same-store parking lot utilization that is incremental to that of contemporaneous stock returns. If the stock market fully incorporates value-relevant information that is correlated with changes in parking lot traffic, one would expect that stock returns cumulated from the beginning to the end of the quarter should subsume the information content of $\Delta FLRT_{iq}$.

Following this discussion, our first set of tests are based on pooled cross-sectional regression models of the following form:

$$\Delta SSS_{iq} = \alpha + \beta_1 \Delta FLRT_{iq} + \beta_2 \Delta SSS_{iq-1} + \beta_3 QRET_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (2).$$

The dependent variable is domestic same-store sales growth and the set of right-hand-side variables includes same-store growth in parking lot utilization ($\Delta FLRT_{iq}$), the lagged value of same-store sales growth (ΔSSS_{iq-1}), and the quarterly stock return including distributions ($QRET_{iq}$). The model also includes firm fixed effects (θ_i) to control for firm-specific time-invariant factors, as well as quarter fixed effects (δ_q) to control for aggregate time-varying factors. The coefficient of interest is that of $\Delta FLRT_{iq}$. A significantly positive β_1 coefficient would imply that same-store growth in parking lot utilization is incrementally relevant for nowcasting current quarter growth in same-store sales.

Throughout the paper, we use two-tailed tests when testing for statistical significance and base statistical inferences on heteroscedasticity-robust standard errors. To ease the interpretation of the estimates and facilitate comparisons across predictors, we report regression results using the standardized z-values of the continuous predictors. The standardized z-values of the predictors are rescaled to have a mean of zero and a standard deviation of one. These standardized regression coefficients measure changes in standard deviation units, which allows us to easily compare the relative importance of each predictor.

Figure 3 illustrates the timeline of our research design. The timeline is organized around the beginning and the end of fiscal quarter q. The figure highlights that the right-hand-side variables in equation (2), including $\Delta FLRT_{iq}$, ΔSSS_{iq-1}, and $QRET_{iq}$, are measured using information available as of the end of quarter q and, therefore, can be used for nowcasting same-store sales growth realizations that will disclosed at the time of quarterly earnings announcement. To be clear, while the quarterly stock return and the lagged realization of same-store sales growth are available to all capital market participants, the $\Delta FLRT_{iq}$ signal is within the reach of sophisticated investors who can afford to incur the substantial costs of acquiring and processing satellite imagery data.

Table 3 reports pooled cross-sectional regression results based on the model specification described in equation (2). Starting with the simple regression results in column (1), the estimated slope coefficient of $\Delta FLRT_{iq}$ is significantly positive and its magnitude implies that a one standard deviation increase in same-store growth in parking lot utilization is expected to result in a 2.2% increase in same-store sales growth. Consistent with prior evidence of mean-reversion in growth rates (e.g. Chan et al. 2003), the regression results in column (2) show that same-store sales growth has a first-order auto-regressive coefficient of 0.82, which is below the benchmark value of one under a random-walk model. The regression results in column (3) also confirm long-standing evidence that stock returns embed forward-looking value-relevant information (e.g., Beaver et al. 1980). The estimated slope coefficient of $QRET_{iq}$ is significantly positive and its magnitude implies that a one standard deviation increase in quarterly stock returns is expected to result in a 1.4% increase in same-store sales growth.

Turning to the multiple regression results in column (4), we find that $\Delta FLRT_{iq}$ is incrementally relevant for nowcasting current quarter same-store sales growth, after controlling for the information content of other predictors. The multiple regressions in the last two columns show that the inferences are not sensitive to controlling for firm-specific time-invariant factors as well as for aggregate time-varying factors, with the complete model specification in column (5) explaining more than 71% of the variation in quarterly same-store sales growth.

4.3 Implications for financial analysts' forecasts

After evaluating the relevance of parking lot traffic data for nowcasting retailer performance, our second research objective is to investigate whether financial analysts incorporate information embedded in parking lot fill rates when projecting same-store sales for the quarter. While financial analysts could help level the playing field across different investor groups with unequal access to data, in practice they may not do so if they themselves do not have access to satellite imagery data or if they do not fully incorporate all value-relevant signals when forecasting same-store sales growth for the quarter.

We obtain sell-side forecasts of domestic same-store sales growth from Factset Estimates and measure the consensus forecast revision from the beginning of quarter q to the most recent forecast prior to the earnings announcement for quarter q. The cumulative forecast revision ($FREVISION_{iq}$) measures the change in financial analysts' expectations over the period that stretches from the beginning of quarter q to the earnings announcement for the quarter, typically three trading weeks after the end of the quarter. Effectively, the measurement window allows several days for financial analysts to incorporate in their forecasts of same-store sales growth for the quarter any value-relevant information available to them as of the end of the quarter.

With this objective in mind, we replace the dependent variable in our baseline model in equation (2) with the cumulative forecast revision ($FREV_{iq}$) and estimate pooled cross-sectional regressions of the following form:

$$FREV_{iq} = \alpha + \beta_1 \Delta FLRT_{iq} + \beta_2 \Delta SSS_{iq-1} + \beta_3 QRET_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (3).$$

With respect to the slope coefficient for $\Delta FLRT_{iq}$, one would expect a positive β_1 estimate if (i) same-store growth in parking lot utilization is relevant for nowcasting current quarter same-store sales growth, and (ii) financial analysts are responsive to the information flow that is correlated with this signal. With respect to the slope coefficients for ΔSSS_{iq-1} and $QRET_{iq}$, one would expect positive estimates for β_2 and β_3. This is because there is long-standing evidence on the auto-correlation properties of sales growth and the forward-looking content of stock returns. Therefore, financial analysts should be responsive to the value-relevant information embedded in these two well-known predictors of firm performance.

This discussion leads to our next set of tests addressing the question whether financial analysts fully incorporate the information flow within the quarter when forecasting same-store sales growth for the quarter. Following our baseline specification, this set of tests is based on pooled cross-sectional regression models of the following form:

$$FERR_{iq} = \alpha + \beta_1 \Delta FLRT_{iq} + \beta_2 \Delta SSS_{iq-1} + \beta_3 QRET_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (4).$$

The left-hand-side variable $FERR_{iq}$ is the consensus forecast error of same-store sales growth for quarter q. We measure the consensus forecast error as the actual value of same-store sales growth released on the earnings announcement day minus the most recent consensus forecast prior to the earnings announcement. If financial analysts fully incorporate information that is relevant for forecasting same-store sales growth, one would expect that their forecast errors for the quarter should be unpredictable based on stale signals measured as of the end of the quarter.

Table 4, Panel A, reports pooled cross-sectional regression results based on the model specification described in equation (3). Starting with the simple regression results, we find a positive association between financial analysts' forecast revisions and same-store parking lot traffic growth. The estimated slope coefficient of $\Delta FLRT_{iq}$ is significantly positive and its magnitude implies that a one standard deviation increase in same-store growth in parking lot utilization is expected to result in a 0.7% increase in financial analysts' consensus forecast of same-store sales growth. Turning to the multiple regression results, we find that $\Delta FLRT_{iq}$ is incrementally relevant for explaining variation in financial analysts' forecast revisions, after controlling for the explanatory power of stock returns and lagged same-store sales growth. The regression results in the last two columns confirm that the inferences are not sensitive to firm- and time-fixed effects, with the complete model specification in column (5) explaining nearly 37% of the variation in financial analysts' forecast revisions from the beginning of the quarter to the most recent forecast date prior to the earnings announcement.

We note that evidence of a positive association between financial analysts' forecast revisions and same-store growth in parking lot traffic comes with two caveats. First, the evidence does not imply that financial analysts directly use the $\Delta FLRT_{iq}$ signal when updating their projections of current quarter growth in same-store sales. Second, it does not necessarily imply that financial analysts fully incorporate all value-relevant information that could have been extracted from $\Delta FLRT_{iq}$. Rather, the evidence merely suggests that financial analysts revise their projections in response to information that is correlated with at least a fraction of the information content embedded in $\Delta FLRT_{iq}$. We note that the same caveats apply for the information embedded in stock returns and lagged same-store sales growth.

Table 4, Panel B, reports pooled cross-sectional regression results based on the model specification described in equation (4) and provides evidence of predictability in financial analysts' forecast errors. The regression results show that current quarter growth in parking lot fill rates, stock returns, as well as lagged realizations of same-store sales growth are incrementally relevant for anticipating variation in financial analysts' forecast errors, after controlling for firm- and time-fixed effects. Evidence of forecast error predictability based on $\Delta FLRT_{iq}$ suggests that financial analysts do not fully incorporate information embedded in parking lot traffic data when projecting current quarter growth in same-store sales. This finding is consistent with the fact that parking lot traffic data is not widely disseminated to capital market participants and can only be accessed for a substantial fee through specialized data vendors. The evidence of forecast error predictability based on $QRET_{iq}$ and ΔSSS_{iq-1}, however, also suggests that financial analysts underreact to public signals that are as easily accessible as stock returns and lagged realizations of same-store sales growth.[6]

To be clear, evidence of predictability in analysts' quarterly forecast errors does not necessarily translate into earnings announcement return predictability. This is because the stock market aggregates into stock prices the beliefs of not just financial analysts but of the general investment community. Next, we test whether the stock market reaction to earnings announcements can be predicted based on satellite imagery of parking lot traffic.

4.4 Implications for capital market participants

4.4.1 Earnings announcement return predictability

If same-store growth in parking lot traffic is incrementally relevant for nowcasting retailer performance, what are the capital market implications of access to parking lot traffic data from outer space?

This is an important question for several reasons. First, satellite imagery of parking lot traffic is *not* available to all investors. While accessible in nearly real-time, the parking lot traffic feeds are only available for a substantial fee. Therefore, the information content of $\Delta FLRT_{iq}$ is available only to sophisticated investors that can afford to incur the substantial data acquisition and processing costs. Second, access to information that is incrementally relevant for anticipating realizations of current quarter same-store sales growth can create opportunities for wealth transfers from small investors, who do not have access to parking lot traffic signals, to sophisticated investors with access to such data prior to the disclosure of retailer performance for the quarter.

[6] Evidence that financial analysts' forecasts do not fully reflect the information in prior price changes can be traced back to Abarbanell (1991), while evidence that analysts under-estimate the serial correlation in accounting data can be traced back to Abarbanell and Bernard (1992).

To address this question, we estimate pooled cross-sectional regressions as follows:

$$EARET_{iq} = \alpha + \beta_1 \Delta FLRT_{iq} + \beta_2 \Delta SSS_{iq-1} + \beta_3 QRET_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (5).$$

The left-hand-side variable $EARET_{iq}$ is the buy-and-hold market-adjusted stock return over the three-day window centered on the earnings announcement for quarter q. We identify earnings announcement dates using the report date of quarterly earnings available from the Compustat Fundamental Quarterly file. We measure market returns based on the CRSP value-weighted index including distributions. The short measurement window $[-1, +1]$ alleviates the issue of risk-adjusting realized returns. This is because the expected value of returns at daily frequencies is close to zero regardless of the asset pricing model. Therefore, the market-adjusted short-window return offer a reasonable measure of earnings announcement surprises for capital market participants. Our results are unchanged to using size and book-to-market factor-adjusted returns in lieu of market-adjusted returns.

Market efficiency with respect to public information would imply that earnings-announcement returns should be unpredictable based on either past stock returns or the lagged realization of same-store sales growth so that the slope coefficient estimates β_2 and β_3 should not be different from zero. This prediction holds regardless of whether or not financial analysts' forecast errors are predictable based on public signals. This is because capital market participants are unlikely to completely fixate on sell-side analysts' forecasts, especially if such forecasts are known to be biased. If capital market participants do not fully incorporate forward-looking information that is correlated with $\Delta FLRT_{iq}$, one would expect that higher (lower) growth in parking lot traffic would translate into positive (negative) earnings announcement surprises so that the slope coefficient β_1 is significantly positive.

To be clear, evidence of earnings announcement return predictability based on $\Delta FLRT_{iq}$ would not be inconsistent with market efficiency under costly information acquisition (e.g., Grossman and Stiglitz 1980). This is because the $\Delta FLRT_{iq}$ signal is not available to the general public and can only be acquired by sophisticated investors for a substantial fee. Importantly, however, such evidence would imply that sophisticated investors with access to satellite imagery of parking lot traffic can formulate a trading strategy that generates abnormal returns at the time of quarterly earnings announcements.

The timeline in Figure 3 visually illustrates the measurement of short-window returns centered on the earnings announcement day along with the measurement of sell-side analysts' forecast revisions from the beginning of the quarter to the most recent consensus prior to the earnings announcement, and forecast error of same-store sales growth for the quarter. The timeline clarifies that the measurement window for analysts' forecast revision does not overlap with the measurement window for earnings announcement returns.

Table 5 reports regression results based on the model specification described in equation (5). Across columns, we find that earnings announcement stock returns are unrelated to lagged stock returns and lagged growth in same-store sales. Evidence that earnings announcement returns are unpredictable based on public information that is as easily accessible as past stock price changes and past realizations of accounting data is consistent with weak-form market efficiency. This finding also offers an interesting contrast to evidence of predictability in financial analysts' forecast errors based on public signals. The key implication here is that the stock market as a whole beats the financial analysts in terms of incorporating public information when forecasting retailer performance for the quarter.

Turning to the predictive power of same-store growth in parking lot fill rates, we *do* find evidence of return predictability around earnings announcements. The significantly positive coefficient for $\Delta FLRT_{iq}$ implies that investors with access to satellite imagery of parking lot traffic can formulate a trading strategy using information as of the end of each quarter that pays off at the time of the earnings announcement for the quarter; that is, typically three trading weeks after the quarter end. The magnitude of the estimated slope coefficient for $\Delta FLRT_{iq}$ implies that a one standard deviation increase in same-store growth in parking lot utilization is expected to result in a 1.2% increase in earnings announcement stock returns. Next, we formulate a long-short strategy based on $\Delta FLRT_{iq}$ and provide direct evidence that sophisticated investors with access to satellite imagery of parking lot fill rates could gain an investment edge and get ahead of the rest of the market.

4.4.2 Formulating a trading strategy from outer space

Table 6, Panel A, reports the buy-and-hold stock 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 $\Delta FLRT_{iq}$. The return measurement window is from one trading day before to one trading day after the earnings announcement day for the quarter (day 0). We report raw returns, market-adjusted returns, as well as size and book-to-market factor-adjusted returns. We use the value-weighted CRSP index including distributions when calculating market-adjusted returns. We use the portfolio data from Kenneth French's website when calculating 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. This approach allows 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 at the time of the quarterly earnings announcements the bottom portfolio underperforms the market by -3.01% while the top portfolio outperforms the market by 1.63%. The spread between the top and bottom portfolios in terms of market-adjusted returns is 4.64%, which is statistically significant and economically important. The buy-sell spread is 4.76% in terms of factor-adjusted returns. While we find significant abnormal returns for both the buy portfolio and the short-sell portfolio, the absolute magnitude of abnormal returns is nearly twice as large for the short-sell portfolio. In practice, short selling entails borrowing the stock for a loan fee (see, e.g., Reed 2013 for an overview of the short selling). To evaluate portfolio performance net of short-selling cost, we obtained daily data on stock loan fees from Markit (formerly known as Data Explorers). The Markit data aggregates survey information from a consortium of more than 100 institutional lenders that collectively account for most of the lendable inventory of shares in the U.S. Table 6, Panel B, reports results after adjusting the short-sell portfolio for stock loan fees. The evidence shows that stock loan fees are less than one basis point per day. Therefore, the hedge portfolio returns remain intact after accounting for the cost of short selling.

To shed light on pre-announcement effects leading to the earnings announcement day, Figure 4, Panel A, reports the cumulative market-adjusted returns for the buy and short-sell portfolios over the two trading weeks (i.e., ten trading days) before and after the earnings announcement day. The green (red) solid line presents the performance of the top (bottom) portfolio of retailers with abnormal increases (decreases) in parking lot fill rates. The red dashed line presents the performance of the short-sell portfolio net of stock loan fees. The plot shows that there are only limited pre-announcement effects with most of the price discovery happening on the earnings announcement day. Indeed, while $\Delta FLRT_{iq}$ positively predicts earnings announcement returns, the correlation between $\Delta FLRT_{iq}$ and pre-earnings announcement returns is not significantly different from zero.

Figure 4, Panel B, presents the hedge portfolio returns measured as the spread of the buy portfolio minus the short-sell portfolio. The black solid (dashed) line presents the hedge portfolio returns as the spread in the buy portfolio minus the short-sell portfolio before (after) stock loan fees. The hedge portfolio performance provides consistent evidence of limited pre-announcement stock price effects with most of the hedge portfolio return accruing on the earnings announcement day. In fact, the hedge portfolio return accumulated from day -10 to day -2 is less than 0.50%, which is not reliably different from zero.

In the appendix, we merge store-level parking lot traffic data from RS Metrics with Orbital Insight and find evidence that investors with access to data from both vendors could formulate even more profitable trading strategies at the time of quarterly earnings reports. The combined data includes 4.9 million daily observations across 75,992 unique store locations for the sample 44 major U.S. retailers over the period from 2011:Q1 to 2017:Q4. The evidence in Table A1 and Figure A1 shows that the trading strategy based on parking lot traffic signals is even more profitable when using the combined data. Focusing on factor-adjusted returns after accounting for short-selling cost, the sell portfolio underperforms the market by 3.14% and the buy portfolio outperforms the market by 1.81%. The buy minus sell portfolio earns a factor-adjusted return of 4.95%, which is 22 basis points higher than using the RS Metrics data alone (compare Table A1, Panel B, to Table 6, Panel B).

Overall, the portfolio analysis provides evidence that stock prices do not aggregate information embedded in parking lot fill rates prior to the earnings announcement, with most of the price discovery happening in the short-window centered on the earnings-announcement day. One key implication is that sophisticated investors who can afford to incur the substantial costs of acquiring and processing satellite imagery data could get ahead of the market and formulate a long-short strategy that generates abnormal returns. Small investors with no access to satellite imagery data are left out of the information loop and do not have the opportunity to formulate a similar trading strategy.

4.4.3 Informed short-selling and uninformed individual trading

The evidence thus far shows that investors with access to parking lot traffic signals have the opportunity to formulate a profitable trading strategy. The strategy works on both the long and the short side, though the returns are especially pronounced from short selling retailers with abnormal decreases in parking lot fill rates. In what follows, we use daily data on stock loan demand from Markit and provide direct evidence of informed trading in the securities lending market. Markit provides the daily lender quantity on loan; that is, the quantity of stock from the lendable quantity that has already been lent. Markit aggregates survey information from a consortium of institutional lenders that collectively account for the vast majority of stock lending. Therefore, Markit's daily measure of lender quantity on loan offers a representative measure of the overall short interest.

Figure 5, 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. The evidence is consistent with informed short-selling activity prior to the earnings announcement day. Focusing on the bottom $\Delta FLRT_{iq}$ portfolio (red solid line), we find evidence of a significant 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 solid line). While the evidence is consistent with informed short-selling activity, individual investors cannot piggyback on the information content of daily fluctuations in the lendable quantity on loan. This is because daily short interest data is available only to those who can afford the substantial subscription fees to Markit's data feeds, with brokers and hedge funds being the typical clients of Markit. In contrast, the general investment community has access to short interest data only twice per month and only with a significant delay.[7]

Next, we use the Trade and Quote (TAQ) dataset to measure individual investor flow following Boehmer's et al. (2017) method. Boehmer's et al. (2017) method exploits the tendency for individual investors' order flow to be internalized or sent to wholesalers. Individual orders typically receive price improvements of a fraction of a penny to compensate for this internalization or whole-selling. Therefore, individual initiated buy orders will have transaction prices slightly below the round penny and sell orders slightly above the round penny. Further, these transactions typically happen off the exchange and are therefore labeled in TAQ with exchange code "D". These institutional features allow us to identify a clean sample of individual investor-initiated transactions in order to measure individual investor trades. It is important to note that Boehmer's et al. (2017) method identifies only a subset of individual investor trades, as limit orders made by individual investors do not have price improvements of a fraction of a penny. Notwithstanding this limitation, the misclassification of an institutional trade as an individual trade would likely bias against our results. After identifying trades initiated by individual investors, we measure individual order imbalance as the total individual investor-initiated buys minus the total individual-investor initiated sells scaled by the total number of shares outstanding.

Figure 5, Panel B, presents the cumulative change in individual investor order imbalance for the top and bottom $\Delta FLRT_{iq}$ portfolios. The evidence is consistent with uninformed trading by individual investors around the quarterly earnings announcements of retailers covered by satellite data. Specifically, the individual order imbalance for retailers in the sell portfolio increases significantly prior to the earnings announcement while the

---

[7] Under the current reporting regime, 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. This means that general investment community can observe short interest only twice per month and only with a significant delay.

order imbalance for the buy portfolio remains near zero. Comparing across Panels A and B of Figure 5, we observe that the dynamics of short-selling activity and individual trading are similar, as both series experience a sharp increase around five days prior to the quarterly earnings announcement.

The message is clear. As short sellers target retailers with bad news for the quarter, individual investors are net buyers of such retailers. The key implication is that short sellers with access to satellite imagery data have the opportunity to benefit at the expense of individual investors, who tend to take the other side of the trade. One potential policy implication is that more timely and detailed information about short selling could help level the playing field for individual investors and facilitate access to the information loop prior to the earnings announcement day. However, as pointed out in SEC's (2014) report on short-sale position and transaction reporting prepared as part of the Dodd-Frank Wall Street Reform and Consumer Protection Act, a more frequent short-selling reporting regime "…*could also facilitate copycat and anticipation strategies that could discourage both the fundamental analysis that is vital to price efficiency and hedging that facilitates capital formation*."[8]

In section 4.5, we provide additional evidence that the ability of short sellers to benefit at the expense of contrarian individual investors has increased after the introduction of satellite imagery data only for the group of retailers with satellite coverage.

4.4.4 A note on insider trading activity

In general, insiders are privy to much more information regarding store performance compared to what outsiders can access through alternative data sources, including satellite imagery of parking lot traffic. While insiders are "on the inside" by definition, trading on material nonpublic information is illegal and the SEC treats the detection and prosecution of insider trading violations as one of its enforcement priorities. To mitigate their own legal risk, firms have adopted blackout policies that restrict their insiders from trading their stock.

[8] The report on short-sale position and transaction reporting was prepared by the Staff of the Division of Economic and Risk analysis of the SEC and it is available from the SEC website.

These blackout policies allow insiders to trade only within a short window after the quarterly earnings announcement and outside this window, insiders are restricted from trading without pre-approval from their firms (e.g., Bettis et al. 2000; Lee et al. 2014). In addition, pursuant to Sections 16(a) and 23(a) of the Securities Exchange Act of 1934, and Sections 30(h) and 38 of the Investment Company Act of 1940, insiders are required to disclose their trading activity in a Form-4 SEC filing within two business days following the execution of the trade. The Form-4 filings are publicly available via the SEC's EDGAR system and are widely tracked by popular websites (see, e.g., Finviz; GuruFocus).

In additional analysis, we examine whether insider-trading activity during the quarter is incrementally relevant for nowcasting current quarter same-store sales growth as well as earnings announcement returns. Using data from Thomson Financial, we measure insider-trading activity during the quarter as the number of shares purchased less the number of shares sold by officers and directors scaled by the number of shares outstanding. There are two takeaways from this additional analysis. First, we do not find evidence that insider-trading activity during the quarter is incrementally relevant for predicting current quarter performance. Second, we do not find evidence of return predictability around earnings announcements based on insider trading activity during the quarter. These findings imply that in our setting insiders do not use their access to material nonpublic information for personal trading gains. It follows that Main Street investors cannot piggyback on insider trading activity to gain access to the information loop.[9]

4.5 Difference-in-differences analyses

The evidence so far shows that stock prices do not fully incorporate value-relevant information embedded in parking lot traffic data prior to the disclosure of quarterly retailer

[9] Consistent with the widespread adoption of blackout policies restricting insiders from trading their stock during periods when financial statements are being prepared but have not yet been publicly disclosed, we do not find evidence of insider buying or selling activity from the first day following the end of a quarter to two days after the quarterly earnings announcement. In line with prior research, we find that a disproportionate fraction of insider-trading activity occurs during the $(+3, +12)$ window after the quarterly earnings announcement (e.g., Bettis et al. 2000).

performance. Next, we investigate whether or not the introduction of satellite imagery data had an impact on stock price informativeness and informed trading.

4.5.1 The effect on stock price informativeness

An increase in stock price informativeness due to the introduction of satellite imagery data would imply that the stock prices of retailers with satellite coverage impound more information pre-earnings announcement, so that the price reaction to the earnings announcement is muted.[10] To identify the effect of satellite coverage on stock price informativeness, we use a difference-in-differences (DID) approach that compares the treated group of retailers with satellite coverage initiated at different points in time by RS Metrics to a matched control group of retailers without satellite coverage throughout the sample period. To construct the matched control group, we match each retailer in the treated group to a retailer without satellite coverage throughout the sample period that operates in the same six-digit GICS industry and has similar quarterly same-store sales growth. For each retailer in the treated group, we use a symmetric event window before and after the initiation of RS Metrics coverage.

To search for changes in the forward-looking content of stock returns for retailer performance, we implement the DID approach using the following regression model:

$$\begin{aligned} Y_{iq} = \alpha + \beta_1 POST_{iq} + \beta_2 TREAT_{iq} + \beta_3 POST_{iq} \times TREAT_{iq} + \beta_4 QRET_{iq} \\ + \beta_5 POST_{iq} \times QRET_{iq} + \beta_6 TREAT_{iq} \times QRET_{iq} + \beta_7 POST_{iq} \\ \times TREAT_{iq} \times QRET_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (6). \end{aligned}$$

We consider two dependent variables Y_{iq}. First, $\Delta SSS_{iq}$is the year-over-year growth in domestic same-store growth. Second, $FERR_{iq}$ is financial analysts' consensus forecast error of same-store sales growth for quarter q cumulated from the beginning of quarter q to the day prior to the earnings announcement date. Turning to the right-hand-side variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of satellite

---

[10] This argument is consistent with rational expectations models showing that earnings announcement is decreasing in the amount of pre-announcement information (e.g., Kim and Verrecchia 1991).

coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for retailers in the treated group, and $QRET_{iq}$ is the buy-and-hold stock return cumulated from the beginning to the end of quarter q. The β_7 coefficient on the triple interaction term $POST_{iq} \times TREAT_{iq} \times QRET_{iq}$ captures the change in the differential predictive ability of stock returns for either same-store sales growth or financial analysts' forecast errors of growth across the treated and matched control groups after the introduction of satellite coverage by RS Metrics. A greater amount of pre-announcement information for retailers with satellite coverage would imply that returns embed more forward-looking content in the post period; that is, $\beta_7 > 0$.

Table 7 reports regression results based on equation (6). The significantly positive coefficient on $QRET_{iq}$ is consistent with long-standing evidence that stock returns embed forward-looking content for fundamental firm performance. The insignificant coefficients on the two-way interactions shows that there are no detectable differences across the treated and the control groups. Importantly, the insignificant coefficient on the triple-interaction term $POST_{iq} \times TREAT_{iq} \times QRET_{iq}$ implies that the introduction of satellite coverage had no detectable effect on the amount of pre-announcement information embedded in returns. Next, we provide consistent evidence that the introduction of satellite imagery data had no detectable effect on the stock price reaction to the public disclosure of retailer performance for the quarter.

To search for changes in the stock price reaction to news about retailer performance, we implement the DID approach using the following regression model:

$$\begin{aligned} EARET_{iq} = \alpha + \beta_1 POST_{iq} + \beta_2 TREAT_{iq} + \beta_3 POST_{iq} \times TREAT_{iq} + \beta_4 FERRR_{iq} \\ + \beta_5 POST_{iq} \times FERR_{iq} + \beta_6 TREAT_{iq} \times FERR_{iq} + \beta_7 POST_{iq} \\ \times TREAT_{iq} \times FERR_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (7). \end{aligned}$$

The dependent variable $EARET_{iq}$ is the buy-and-hold market-adjusted stock return over the three-day window centered on the earnings announcement for quarter q. We focus on changes in the stock price reaction to the consensus forecast error of same-store sales growth for the quarter. The β_7 coefficient on the triple interaction term $POST_{iq} \times TREAT_{iq} \times FERR_{iq}$ captures the change in the differential sensitivity of stock prices to news about same-store sales growth across the treated and matched control groups after the introduction of satellite coverage by RS Metrics. A greater amount of pre-announcement information for retailers with satellite coverage would imply that the market reaction should be muted in the post period; that is, $\beta_7 < 0$.

Table 8 reports regression results based on equation (7). The significantly positive coefficient on $FERR_{iq}$ is consistent with a positive (negative) market reaction to good (bad) news about growth. The insignificant coefficients on the two-way interactions shows that there are no detectable differences across the treated and the control groups. Importantly, the insignificant coefficient on the triple-interaction term $POST_{iq} \times TREAT_{iq} \times FERR_{iq}$ implies that the introduction of satellite coverage had no detectable effect on the stock price reaction to news about retailer performance.

Taken together, the evidence shows that the introduction of satellite coverage had no detectable effect on stock price informativeness. The evidence is broadly consistent with models where improved data availability does not necessarily increase stock price informativeness (e.g., Banerjee et al. 2018; Dugast and Foucault 2018).[11]

4.5.2 The effect on informed short-selling and uninformed individual trading

Our next set of DID analyses zero in on the dynamics of informed short-selling activity. Specifically, we estimate the following regression model:

$$\begin{aligned} EARET_{iq} = \alpha &+ \beta_1 POST_{iq} + \beta_2 TREAT_{iq} + \beta_3 POST_{iq} \times TREAT_{iq} + \beta_4 \Delta SHORT_{iq} \\ &+ \beta_5 POST_{iq} \times \Delta SHORT_{iq} + \beta_6 TREAT_{iq} \times \Delta SHORT_{iq} + \beta_7 POST_{iq} \\ &\times TREAT_{iq} \times \Delta SHORT_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (8). \end{aligned}$$

---

[11] We note that our evidence is different from Zhu (2019), who follows a different DID research design and finds that the introduction of alternative data increases price informativeness. The following key differences in the research design could help explain the different results. First, Zhu's (2019) definition of treated firms mixes firms that are truly covered by alternative data vendors with matched firms in the same industry that are not covered. Second, Zhu's (2019) definition of control firms identifies for each treated firm a matched firm operating in a supplier-customer industry rather than a matched firm operating in the same industry. Third, Zhu's (2019) definition of the $POST$ indicator assumes that alternative data coverage started only in 2014:Q3, which effectively ignores the fact that RS Metrics started selling satellite imagery data as early as 2011:Q1.

The independent variable $\Delta SHORT_{iq}$ is the cumulative changes in the lender quantity on loan from the end of quarter to two days before earnings announcement. The β_7 coefficient on the triple interaction term $POST_{iq} \times TREAT_{iq} \times \Delta SHORT_{iq}$ captures the change in the predictive ability of the predictive ability of short-selling activity for earnings announcement returns across the treated and matched control groups after the introduction of satellite coverage by RS Metrics. An increase in informed short-selling activity for retailers with satellite coverage would imply that $\Delta SHORT_{iq}$ is more negatively related to earnings announcement returns in the post period; that is, $\beta_7 < 0$.

Table 9 reports regression results based on equation (8). The significantly negative coefficient on the triple-interaction term $POST_{iq} \times TREAT_{iq} \times \Delta SHORT_{iq}$ is consistent with an increase in informed short-selling activity after the introduction of satellite imagery data. Specifically, the evidence shows that the change in lender quantity on loan cumulated in the pre-earnings announcement period becomes a stronger predictor of earnings announcement returns after the introduction of RS Metrics coverage for retailers with satellite coverage.

We next investigate the profitability of individual trading after the introduction of satellite imagery data. To do so, we estimate the following regression model:

$$\begin{aligned} EARET_{iq} = \alpha + \beta_1 POST_{iq} + \beta_2 TREAT_{iq} + \beta_3 POST_{iq} \times TREAT_{iq} + \beta_4 IndOI_{iq} \\ + \beta_5 POST_{iq} \times IndOI_{iq} + \beta_6 TREAT_{iq} \times IndOIB_{iq} + \beta_7 POST_{iq} \\ \times TREAT_{iq} \times IndOI_{iq} + \theta_i + \delta_q + \varepsilon_{iq} \qquad (9). \end{aligned}$$

The independent variable $IndOI_{iq}$ is Boehmer's et al. (2017) measure of individual investor order imbalance as described in section 4.4.3. The slope on the triple interaction term of $POST_{iq} \times TREAT_{iq} \times IndOIB_{iq}$ measures the change in the relative informativeness of individual order flow between the treated and control groups after the introduction of satellite imagery data. Table 10 reports the regression results based on equation (9). First, we observe that individual order flow negatively predicts announcement returns, which is consistent with prior evidence on individual investors' tendency to be contrarian traders (e.g., Grinblatt and Keloharju 2000; Kaniel et al. 2007). Second, we observe that while in general individual investors' trades are becoming more informative over time (e.g., Kaniel et al. 2012; Kelley and Tetlock 2013; Boehmer et al. 2017), this is not true for the individual order flow in the stock of retailers with satellite coverage after the introduction of satellite imagery data. In fact, the significantly negative coefficient on the triple interaction term $POST_{iq} \times TREAT_{iq} \times IndOIB_{iq}$ implies that individual investors' contrarian tendency has actually increased for the treated group of retailers with satellite coverage after the introduction of satellite imagery data.

Overall, the evidence suggests that the introduction of satellite imagery data has created profitable trading opportunities for sophisticated investors, who can use parking lot traffic signals to target retailers with bad news for the quarter, at the expense of individual investors, who tend to be on the other side of the trade.

4.5.3 The effect on stock liquidity

The introduction of alternative data to the market represents a change in the informational environment between market participants. As large, sophisticated investors are the only group that can feasibly make use of alternative data, these investors gain an informational advantage over other investors in the market. Theoretical research generally concludes that an increase in information asymmetry between market participants leads to a decrease in stock liquidity and an increase in bid-ask spreads (e.g., Copeland and Galai 1983; Glosten and Milgrom 1985; Kyle 1985; Easley and O'Hara 1987). Therefore, it is possible that the release of satellite data leads to a decrease in the liquidity of affected firms' stocks. Our final regressions test this implication. Specifically, we conduct a DID analysis with the average bid-ask spread scaled by the midpoint from the end of the quarter to two days before the earnings announcement as the dependent variable. A higher bid-ask spread for stocks covered by satellite data relative to the matched control sample would suggest that the rise in information asymmetry due to the dissemination of alternative data leads to lower liquidity for covered firms.

Table 11 reports evidence that liquidity does decrease for firms covered by satellite data. Specifically, the coefficient on $POST_{iq} \times TREAT_{iq}$ is positive and statistically significant, meaning that the bid-ask spread rises more for treated firms in the post period relative to the matched control sample. The inclusion of firm and quarter fixed effects alleviates concern that the results are driven by any time-invariant firm characteristics or by an overall time-trend in the bid-ask spread. Our evidence of a decrease in liquidity after the introduction of satellite imagery for the treated group of retailers with satellite coverage is consistent with prior theory and evidence that stock liquidity decreases when information asymmetry across investors increases.

Viewed as a whole, our DID analyses provide additional support for the notion that the introduction of satellite imagery data created trading opportunities for large sophisticated investors, who can afford to access and process such data, without necessarily enhancing stock price discovery for the general investment community. In fact, consistent with an increase in information asymmetry across sophisticated investors and individual investors, our evidence shows a significant increase in the bid-ask spread around quarterly earnings announcements for the stock of retailers with satellite coverage after the introduction of satellite imagery data.

5. Conclusion

We study the emergence of satellite coverage of parking lot traffic across major U.S. retailers as a source of alternative data in capital markets. We document that counting cars from outer space provides timely and incrementally relevant information for predicting current quarter performance. Such material information, however, is not impounded into stock prices prior to the public disclosure of retailer performance for the quarter. This creates opportunities for sophisticated investors with access to satellite imagery data to formulate profitable trading strategies at the expense of individual investors, who tend to be on the other side of the trade. Far from creating a level playing field, we find that the introduction of satellite imagery data has led to a decrease in stock liquidity around quarterly earnings announcements due to an increase in information asymmetry across sophisticated investors and individual investors.

Overall, the evidence is consistent with the view that unequal access to big and alternative data leaves small investors outside the information loop without necessarily enhancing stock price informativeness. Looking ahead, technological advancements will continue to stimulate growth in the volume, velocity, and variety of alternative data sets. Data hunters will continue to scour alternative data from anywhere there is a digital footprint. Sophisticated investors will continue to invest in new data sets in their quest to gain an edge, especially as the value of publicized signals extracted from existing data sets should decay over time (e.g., McLean and Pontiff 2016). As sophisticated investors increasingly rely on big and alternative data that are ever more out of reach for individual investors, the fence separating sophisticated investors from individual investors might be getting taller. In a market setting where the line separating public from material nonpublic information might be getting more blurry, the question that regulators need to answer is what is their role in terms of leveling the playing field for individual investors?

References

Abarbanell, J. S. 1991. "Do Analysts' Earnings Forecasts Incorporate Information in Prior Stock Price Changes?" *Journal of Accounting and Economics* 14 (2): 147–165.

Abarbanell, J. S., and V. L. Bernard. 1992. "Tests of Analysts Overreaction Underreaction To Earnings Information As an Explanation for Anomalous Stock-Price Behavior." *Journal of Finance* 47 (3): 1181–1207.

Axbard, Sebastian. 2016. "Income Opportunities and Sea Piracy in Indonesia: Evidence from Satellite Data." *American Economic Journal: Applied Economics* 8 (2): 154–94.

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

Beaver, W., R. Lambert, and D. Morse. 1980. "The Information Content of Security Prices." *Journal of Accounting and Economics* 2 (1): 3–28.

Bettis, J. C., J. L. Coles, and M. L. Lemmon. 2000. "Corporate Policies Restricting Trading by Insiders." *Journal of Financial Economics* 57 (2): 191-220.

Boehmer, Ekkehart, Charles Jones, Xiaoyan Zhang. 2017. "Tracking retail investor activity." Working Paper. Available at SSRN.

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

Chan, L.K., J. Karceski, and J. Lakonishok. 2003. "The Level and Persistence of Growth Rates" *Journal of Finance* 58 (2): 643-684.

Chen, X., and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *PNAS* 108 (21): 8589–94.

Chen, X., and William D. Nordhaus. 2015. "A Test of the New VIIRS Lights Data Set: Population and Economic Output in Africa." *Remote Sensing* 7 (4): 4937–4947.

Chiyachantana, Chiraphol N, Christine X Jiang, Nareerat Taechapiroontong, Robert A Wood. 2004. "The impact of regulation fair disclosure on information asymmetry and trading: An intraday analysis." *The Financial Review* 39(4): 549-577.

Collins, D.W., S.P. Kothari, and J.D. Rayburn. 1987. "Firm Size and the Information Content of Prices with Respect to Earnings." *Journal of Accounting and Economics* 9(2): 111-138.

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

Croft, T.A. 1978. "Nighttime Images of the Earth from Space." *Scientific American* 239: 68-79.

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

Doll, Christopher N. H., J.-P. Muller, and Jeremy G. Morley. 2006. "Mapping Regional Economic Activity from Night-time Light Satellite Imagery." *Ecological Economics* 57 (1): 75–92.

Donaldson, D. and Storeygard A. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171-198.

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

Easley, David, Maureen O'Hara. 1987. "Price, trade size, and information in securities markets." *Journal of Financial Economics*, 19(1): 69-90.

Eleswarapu, Venkat R, Rex Thompson, Kumar Venkataraman. 2004. "The impact of regulation fair disclosure: Trading costs and information asymmetry." *Journal of Financial and Quantitative Analysis*, 39(2): 209-225.

Farrell, Mike, T Clifton Green, Russell Jame, Stan Markov. 2018. "The democratization of investment research: Implications for retail investor profitability and firm liquidity." Working Paper. Available at SSRN.

Foster, A. D. and Rosenzweig, M.R. 2003. "Economic Growth and the Rise of Forests." *Quarterly Journal of Economics* 118 (2): 601–637.

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.

Giannone, D., L. Reichlin, and D. Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55 (4): 665-676.

Glosten, Lawrence R, Paul R Milgrom. 1985. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *Journal of Financial Economics*, 14(1): 71-100.

Grinblatt, Mark, Matti Keloharju. 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.

Guiteras, Raymond, Amir Jina, and A. Mushfiq Mobarak. 2015. "Satellites, Self-Reports, and Submersion: Exposure to Floods in Bangladesh." *The American Economic Review* 105 (5): 232–236.

Hayek, F. A. 1945. "The Use of Knowledge in Society." *The American Economic Review* 35 (4): 519–530.

Henderson, J.V., A. Storeygard, and D. N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994-1028.

Jame, Russell, Rick Johnston, Stanimir Markov, Michael C Wolfe. 2016. "The value of crowdsourced earnings forecasts." *Journal of Accounting Research*, 54(4): 1077-1110.

Kaniel, Ron, Shuming Liu, Gideon Saar, Sheridan Titman. 2012. "Individual investor trading and return patterns around earnings announcements." *The Journal of Finance*, 67(2): 639-680.

Kaniel, Ron, Gideon Saar, Sheridan Titman. 2007. "Individual investor trading and stock returns." *The Journal of Finance*, 62(3): 1139-1168.

Kelley, Eric K, Paul C Tetlock. 2013. "How wise are crowds? insights from retail orders and stock returns." *The Journal of Finance*, 68(3): 1229-1265.

Kim, O., and R.E. Verrecchia. 1991. "Trading Volume and Price Reactions to Public Announcements." *Journal of Accounting Research* 29 (2): 302-321.

Kyle, Albert S. 1985. "Continuous auctions and insider trading." *Econometrica: Journal of the Econometric Society* 1315-1335.

Lee, I., M. L. Lemmon, Y. Li, and J. M. Sequeira.2014. "Do Voluntary Corporate Restrictions on Insider Trading Eliminate Informed Insider Trading?" *Journal of Corporate Finance* 29 (Dec.): 158-178.

Li Xi, Huimin Xu, Xiaoling Chen and Chang Li. 2013. "Potential of NPP-VIIRS Nighttime Light Imagery for Modeling the Regional Economy of China." *Remote Sensing* 5 (6): 3057-3081.

Marx, Benjamin, Thomas M. Stoker, and Tavneet Suri. 2017. "There is No Free House: Ethnic Patronage in a Kenyan Slum." Forthcoming in *American Economic Journal: Applied Economics*.

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

Morningstar. 2018 "Is Big Data the Key to Bigger Investment Returns?" Available online at Morningstar Fund Research & Insights.

Muller D. and Zeller M. 2002. "Land Use Dynamics in the Central Highlands of Vietnam: A Spatial Model Combining Village Survey Data with Satellite Imagery Interpretation." *Agricultural Economics* 27: 333-354.

Nagaraj, Abhisek. 2017. "The Private Impact of Public Maps: Landsat Satellite Imagery and Gold Exploration." Working Paper. Available at SSRN.

Reed, A. V. 2013. "Short Selling." *Annual Review of Financial Economics* 5 (1): 245-258.

Skole D. and Tucker C. 1993. "Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978 to 1988." *Science* 260: 1905-09.

Sutton P., Roberts D., Elvidge C., and Baugh K. 2001. "Census from Heaven: An estimate of the global human population using night-time satellite imagery." *International Journal of Remote Sensing* 22 (16): 3061-3076

Sutton P., Roberts D., Elvidge C., and Meij H. 1997. "A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States." *Photogrammetric Engineering & Remote Sensing* 63 (11): 1303-1313

Welch R. 1980. "Monitoring Urban Population and Energy Utilization Patterns from Satellite Data." *Remote Sensing of Environment* 9: 1-9.

Zhu C. 2019. "Big Data as Governance Mechanism." Forthcoming in *Review of Financial Studies*.

Appendix 1
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 stock return (including distributions) cumulated over the three-month window from the beginning to the end of quarter q. We obtained daily stock return data from CRSP.
$FREV_{iq}$	Financial analysts' consensus forecast revision of same-store sales growth for quarter q cumulated from the beginning of quarter q to the day prior to the earnings announcement date. We obtained sell-side consensus forecasts from Factset Estimates.
$FERR_{iq}$	Financial analysts' consensus forecast error of same-store sales growth for quarter q measured as realized same-store sales growth for quarter q minus the prevailing consensus forecast as of the day prior to the earnings announcement date.
$EARET_{iq}$	Buy-and-hold market-adjusted stock return cumulated over the three-day window centered on the earnings announcement for quarter q. We obtained daily stock return data from CRSP. We measure market returns based on the CRSP value-weighted index including distributions.
$\Delta SHORT_{iq}$	The cumulative change in the lender quantity on loan from the end of quarter to two days before earnings announcement. We obtained daily data on the lender quantity on loan from Markit.
$IndOIB_{iq}$	The cumulative order imbalance of individual investors (net buy volume) from the end of quarter to two days before earnings announcement. We use the TAQ dataset to measure individual investor activity following Boehmer's et al. (2017) method to identifying individual investor order flow.
$Spread_{iq}$	The average bid-ask spread scaled by the midpoint from the end of quarter to two days before the quarterly earnings announcement.

Table 1
Satellite coverage

	Company Name	GICS Industry	Average Store Count	Average Satellite Coverage	Starting Date of Coverage
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%	2011:Q3
38	Target Corporation (TGT)	Specialty Retail	1,819	71%	2014:Q4
39	The Container Store Group, Inc. (TCS)	Specialty Retail	80	66%	2016:Q1
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 store count and satellite store coverage for each of the 44 U.S. companies in our sample along with the starting date of RS Metrics coverage.

Table 2
Descriptive statistics

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
$QRET_{iq}$	-0.013	0.154	-0.553	-0.106	-0.013	0.080	0.624
$FREV_{iq}$	-0.007	0.023	-0.205	-0.013	-0.003	0.003	0.154
$FERR_{iq}$	-0.001	0.008	-0.138	-0.002	0.000	0.000	0.063
$EARET_{iq}$	-0.001	0.019	-0.091	-0.010	0.000	0.008	0.068

Panel B: Pairwise correlations.

	(1)	(2)	(3)	(4)	(5)	(6)
(1) ΔSSS_{iq}		0.371	0.305	0.672	0.548	0.201
(2) $\Delta FLRT_{iq}$	0.383		0.066	0.263	0.254	0.127
(3) $QRET_{iq}$	0.246	0.045		0.383	0.210	0.018
(4) $FREV_{iq}$	0.728	0.260	0.335		0.280	0.050
(5) $FERR_{iq}$	0.614	0.257	0.172	0.270		0.380
(6) $EARET_{iq}$	0.172	0.121	-0.007	0.058	0.338	

This table presents descriptive statistics. Panel A reports the empirical distributions of key variables. Panel B reports Pearson (Spearman) pairwise correlations above (below) the main diagonal. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4.

Table 3
Nowcasting same-store sales growth from outer space

	Dependent Variable = ΔSSS_q				
	(1)	(2)	(3)	(4)	(5)
$\Delta FLRT_{iq}$	0.022***	.	.	0.008***	0.008***
	(5.35)			(4.44)	(4.13)
ΔSSS_{iq-1}	.	0.047***	.	0.043***	0.039***
		(20.62)		(19.33)	(12.39)
$QRET_{iq}$	.	.	0.014***	0.006***	0.005***
			(4.81)	(3.93)	(3.11)
Firm Fixed Effects	No	No	No	No	Yes
Quarter Fixed Effects	No	No	No	No	Yes
Adjusted R^2	14.5%	67.5%	5.9%	70.3%	71.4%

This table reports pooled cross-sectional regression results and provides evidence that growth in same-store parking lot fill rates for quarter q is incrementally relevant for forecasting current quarter growth in same-store sales. 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. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 4
Implications for financial analysts' forecasts

Panel A: Financial analysts' forecast revisions.

	$Dependent\ Variable = FREV_{iq}$				
	(1)	(2)	(3)	(4)	(5)
$\Delta FLRT_{iq}$	0.007***	.	.	0.003***	0.003**
	(4.24)			(2.92)	(2.35)
ΔSSS_{iq-1}	.	0.014***	.	0.012***	0.013***
		(10.73)		(9.72)	(6.64)
$QRET_{iq}$	.	.	0.009***	0.007***	0.006***
			(5.75)	(5.12)	(4.89)
Firm Fixed Effects	No	No	No	No	Yes
Quarter Fixed Effects	No	No	No	No	Yes
Adjusted R^2	6.6%	28.3%	11.1%	35.3%	36.9%

Panel B: Financial analysts' forecast errors.

	$Dependent\ Variable = FERR_{q}$				
	(1)	(2)	(3)	(4)	(5)
$\Delta FLRT_{iq}$	0.005***	.	.	0.003***	0.004***
	(4.73)			(4.01)	(3.73)
ΔSSS_{iq-1}	.	0.006***	.	0.005***	0.003**
		(6.54)		(5.12)	(2.44)
$QRET_{iq}$	.	.	0.003***	0.002***	0.002**
			(3.96)	(3.02)	(2.46)
Firm Fixed Effects	No	No	No	No	Yes
Quarter Fixed Effects	No	No	No	No	Yes
Adjusted R^2	6.5%	10.6%	2.8%	14.4%	20.4%

Panel A of this table provides evidence that growth in same-store parking lot fill rates for quarter q is incrementally relevant for explaining financial analysts' consensus forecast revisions from the beginning of the quarter to the most recent date prior to the earnings report for the quarter. Panel B provides evidence that growth in same-store parking lot fill rates for quarter q predicts financial analysts' consensus forecast error of growth in same-store sales for the quarter. 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. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 5
Are earnings announcement returns predictable from outer space?

	$Dependent\ Variable = EARET_q$				
	(1)	(2)	(3)	(4)	(5)
$\Delta FLRT_{iq}$	0.012***	.	.	0.011**	0.012**
	(2.77)			(2.56)	(2.04)
ΔSSS_{iq-1}	.	0.006	.	0.002	0.002
		(1.38)		(0.57)	(0.29)
$QRET_{iq}$	.	.	-0.001	-0.002	-0.004
			(-0.12)	(-0.28)	(-0.81)
Firm Fixed Effects	No	No	No	No	Yes
Quarter Fixed Effects	No	No	No	No	Yes
Adjusted R^2	1.3%	0.2%	-0.2%	1.1%	1.5%

This table reports pooled cross-sectional regression results and provides evidence that growth in same-store parking lot fill rates for quarter q predicts three-day stock returns centered on the earnings announcement for the quarter. 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. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 6
Formulating a trading strategy from outer space

Panel A: Before short-selling cost.

	Earnings announcement returns		
	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)
$Buy - Sell$	4.60%***	4.64%***	4.76%***
	(3.75)	(3.80)	(3.93)

Panel B: After short-selling cost.

	Earnings announcement returns		
	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)
$Buy - Sell$	4.57%***	4.62%***	4.73%***
	(3.73)	(3.78)	(3.90)

This table reports the buy-and-hold 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 returns, market-adjusted returns, as well as size and book-to-market 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. The return measurement window is from one trading day before to one trading day after the earnings announcement day for the quarter (day 0). 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. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 7
Difference-in-differences: The effect on stock price informativeness

	Dependent Variable =			
	ΔSSS_{iq}		$FERR_{iq}$	
	(1)	(2)	(1)	(2)
$POST_{iq}$	-0.005*	-0.017***	-0.002*	-0.001
	(-1.82)	(-5.22)	(-1.65)	(-0.79)
$TREAT_{iq}$	0.001	-0.004	0.000	-0.001
	(0.27)	(-1.03)	(0.26)	(-0.52)
$POST_{iq} \times TREAT_{iq}$	-0.003	0.002	-0.003	-0.002
	(-0.57)	(0.49)	(-1.62)	(-1.28)
$QRET_{iq}$	0.007***	0.006***	0.004***	0.004***
	(2.65)	(3.36)	(3.77)	(3.98)
$POST_{iq} \times QRET_{iq}$	0.003	-0.003	-0.001	-0.001
	(1.00)	(-1.29)	(-0.63)	(-0.85)
$TREAT_{iq} \times QRET_{iq}$	0.002	0.003	0.000	0.000
	(0.62)	(1.14)	(0.11)	(0.08)
$POST_{iq} \times TREAT_{iq} \times QRET_{iq}$	0.005	0.007	0.001	0.001
	(0.82)	(1.53)	(0.41)	(0.60)
Firm Fixed Effects	No	Yes	No	Yes
Quarter Fixed Effects	No	Yes	No	Yes
Adjusted R^2	4.9%	39.9%	4.6%	19.9%

This table provides evidence that the introduction of satellite imagery data had no detectable effect on the amount of pre-announcement information embedded in stock returns. We consider two dependent variables. ΔSSS_{iq} is the year-over-year growth in domestic same-store growth. $FERR_{iq}$ is financial analysts' consensus forecast error of same-store sales growth measured as realized same-store sales growth minus the prevailing consensus forecast as of the day prior to the earnings announcement. Turning to the independent variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of RS Metrics coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage, and $QRET_{iq}$ is the buy-and-hold stock return (including distributions) cumulated over the three-month window from the beginning to the end of quarter. 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 match each retailer in the treated group to a matched control group of retailers without satellite coverage that operates in the same six-digit GICS industry and has similar quarterly same-store sales growth. For each retailer in the treated group, we use a symmetric event window before and after the initiation of RS Metrics coverage. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 8
Difference-in-differences: The effect on stock price reaction to news

	$Dependent\ Variable = EARET_{iq}$	
	(1)	(2)
$POST_{iq}$	-0.005	0.005
	(-0.82)	(0.62)
$TREAT_{iq}$	-0.005	-0.002
	(-0.91)	(-0.21)
$POST_{iq} \times TREAT_{iq}$	0.005	-0.001
	(0.63)	(-0.11)
$FERR_{iq}$	0.019***	0.027***
	(4.32)	(4.98)
$POST_{iq} \times FERR_{iq}$	0.008	-0.002
	(1.19)	(-0.21)
$TREAT_{iq} \times FERR_{iq}$	0.01	0.003
	(1.59)	(0.34)
$POST_{iq} \times TREAT_{iq} \times FERR_{iq}$	0.001	0.009
	(0.11)	(0.85)
Firm Fixed Effects	No	Yes
Quarter Fixed Effects	No	Yes
Adjusted R^2	8.2%	15.4%

This table provides evidence that the introduction of satellite imagery data had no detectable effect on the stock price reaction to news about retailer performance. The dependent variable $EARET_{iq}$ is the buy-and-hold market-adjusted stock return over the three-day window centered on the earnings announcement for quarter q. Turning to the independent variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of RS Metrics coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage, $FERR_{iq}$ is the consensus forecast error of same-store sales growth and $QRET_{iq}$ is the buy-and-hold stock market return (including distributions) cumulated over the three-month window from the beginning to the end of quarter. 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 match each retailer in the treated group to a matched control group of retailers without satellite coverage that operates in the same six-digit GICS industry and has similar quarterly same-store sales growth. For each retailer in the treated group, we use a symmetric event window before and after the initiation of RS Metrics coverage. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 9
Difference-in-differences: The effect on informed short-selling activity

	$Dependent\ Variable = EARET_{iq}$	
	(1)	(2)
$POST_{iq}$	-0.005	0.004
	(-0.86)	(0.48)
$TREAT_{iq}$	-0.002	-0.002
	(-0.35)	(-0.26)
$POST_{iq} \times TREAT_{iq}$	-0.005	-0.008
	(-0.55)	(-0.92)
$\Delta SHORT_{iq}$	-0.010**	-0.009*
	(-2.12)	(-1.93)
$POST_{iq} \times \Delta SHORT_{iq}$	-0.001	0.006
	(-0.17)	(0.76)
$TREAT_{iq} \times \Delta SHORT_{iq}$	0.016**	0.017**
	(2.30)	(2.27)
$POST_{iq} \times TREAT_{iq} \times \Delta SHORT_{iq}$	-0.019*	-0.026**
	(-1.80)	(-2.41)
Firm Fixed Effects	No	Yes
Quarter Fixed Effects	No	Yes
Adjusted R^2	1.2%	7.8%

This table provides evidence that the informativeness of short-selling activity increased after the introduction of satellite imagery data. The dependent variable $EARET_{iq}$ is the buy-and-hold market-adjusted stock return over the three-day window centered on the earnings announcement for quarter q. Turning to the independent variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of RS Metrics coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage, $\Delta SHORT_{iq}$ is the cumulative change in the lender quantity on loan from the end of quarter to two days before the quarterly earnings announcement. 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 match each retailer in the treated group to a matched control group of retailers without satellite coverage that operates in the same six-digit GICS industry and has similar quarterly same-store sales growth. For each retailer in the treated group, we use a symmetric event window before and after the initiation of RS Metrics coverage. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 10
Difference-in-differences: The effect on uninformed individual trading

	$Dependent\ Variable = EARET_{iq}$	
	(1)	(2)
$POST_{iq}$	-0.006	0.004
	(-0.92)	(0.46)
$TREAT_{iq}$	-0.003	-0.003
	(-0.60)	(-0.33)
$POST_{iq} \times TREAT_{iq}$	-0.003	-0.011
	(-0.37)	(-1.15)
$IndOIB_{iq}$	-0.007**	-0.007*
	(-2.26)	(-1.78)
$POST_{iq} \times IndOIB_{iq}$	0.022***	0.021***
	(3.24)	(3.10)
$TREAT_{iq} \times IndOIB_{iq}$	0.006	0.009
	(1.10)	(1.49)
$POST_{iq} \times TREAT_{iq} \times IndOIB_{iq}$	-0.016*	-0.018*
	(-1.73)	(-1.87)
Firm Fixed Effects	No	Yes
Quarter Fixed Effects	No	Yes
Adjusted R^2	0.7%	10.2%

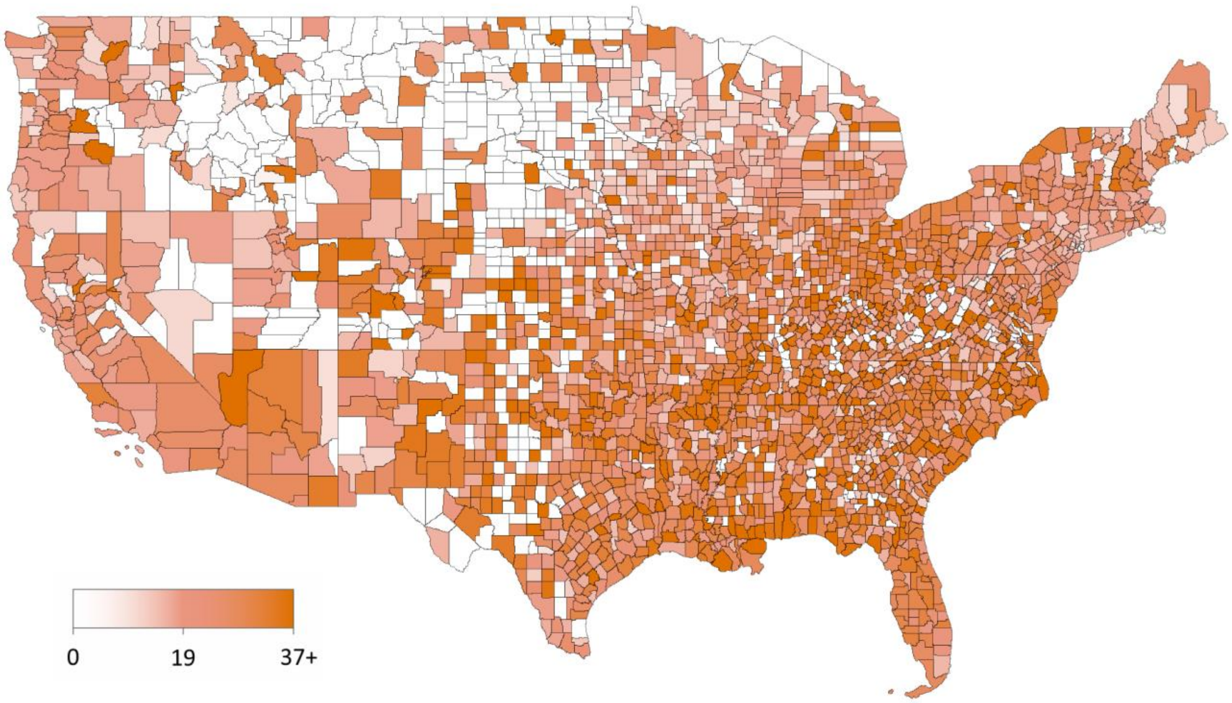
This table provides evidence that the informativeness of individual order flow decreased after the introduction of satellite imagery data. The dependent variable $EARET_{iq}$ is the buy-and-hold market-adjusted stock return over the three-day window centered on the earnings announcement for quarter q. Turning to the independent variables, $POST_{iq}$ is an indicator variable that takes the value one after the initiation of RS Metrics 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 cumulative order imbalance of individual investors (net buy volume) from the end of quarter to two days before the quarterly earnings announcement. 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 match each retailer in the treated group to a matched control group of retailers without satellite coverage that operates in the same six-digit GICS industry and has similar quarterly same-store sales growth. For each retailer in the treated group, we use a symmetric event window before and after the initiation of RS Metrics coverage. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Table 11
Difference-in-differences: The effect on stock liquidity

	$Dependent\ Variable = \ Spread_{iq}$	
	(1)	(2)
$POST_{iq}$	-0.0017***	-0.0003
	(-2.85)	(-1.28)
$TREAT_{iq}$	-0.0027***	-0.0001
	(-4.69)	(-0.58)
$POST_{iq} \times TREAT_{iq}$	0.0015**	0.0007***
	(2.48)	(2.81)
Firm Fixed Effects	No	Yes
Quarter Fixed Effects	No	Yes
Adjusted R^2	2.9%	71.3%

This table provides evidence that stock liquidity decreased after the introduction of satellite imagery data. The dependent variable is the average value of the bid-ask spread at close scaled by the midpoint measured over the window from one day before to one day 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 RS Metrics coverage, $TREAT_{iq}$ is an indicator variable that takes the value one for the treated group of retailers with satellite coverage. The treated group of retailers includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. We match each retailer in the treated group to a matched control group of retailers without satellite coverage that operates in the same six-digit GICS industry and has similar quarterly same-store sales growth. For each retailer in the treated group, we use a symmetric event window before and after the initiation of RS Metrics coverage. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Figure 1
Geographical coverage of satellite imagery

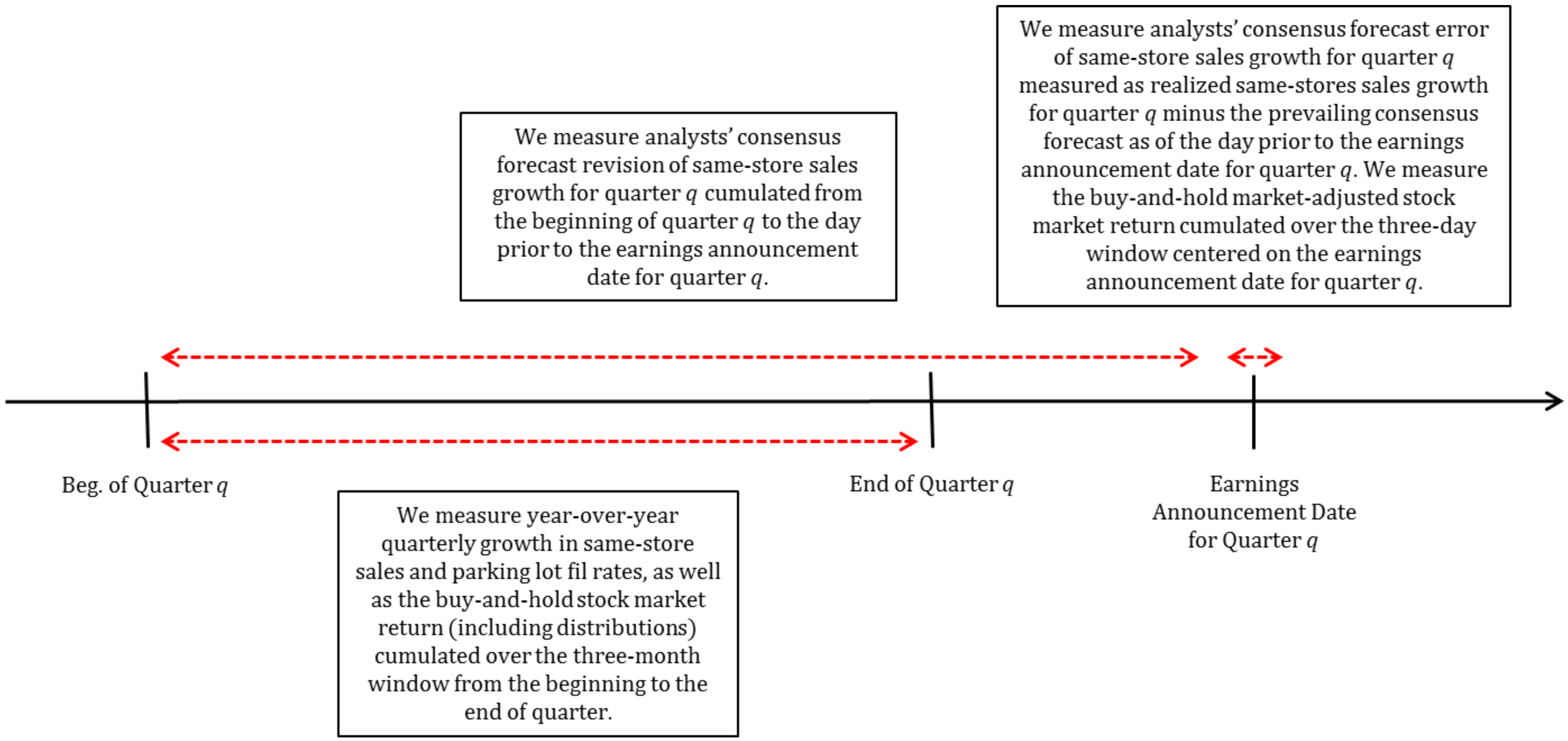


This figure presents the number of individual store locations with satellite coverage per 100,000 residents across counties in the U.S. The underlying data covers 67,210 individual store locations across 2,571 counties covering 98% of the U.S. population. The color spectrum across counties is proportionately 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 19 stores with interquartile range from 10 to 27.

Figure 2
Illustrative example of satellite imagery

This figure presents the parking lot satellite image of the Target store located at 4500 Macdonald Ave, Richmond CA 94805. The image was captured by RS Metrics on September 19, 2016 at 11:03am. The red line outlines the boundary of the parking lot associated with Target and the red dots indicate the occupied parking lot spaces. For this case, RS Metrics identifies 540 parking lot spaces with 146 of them filled.

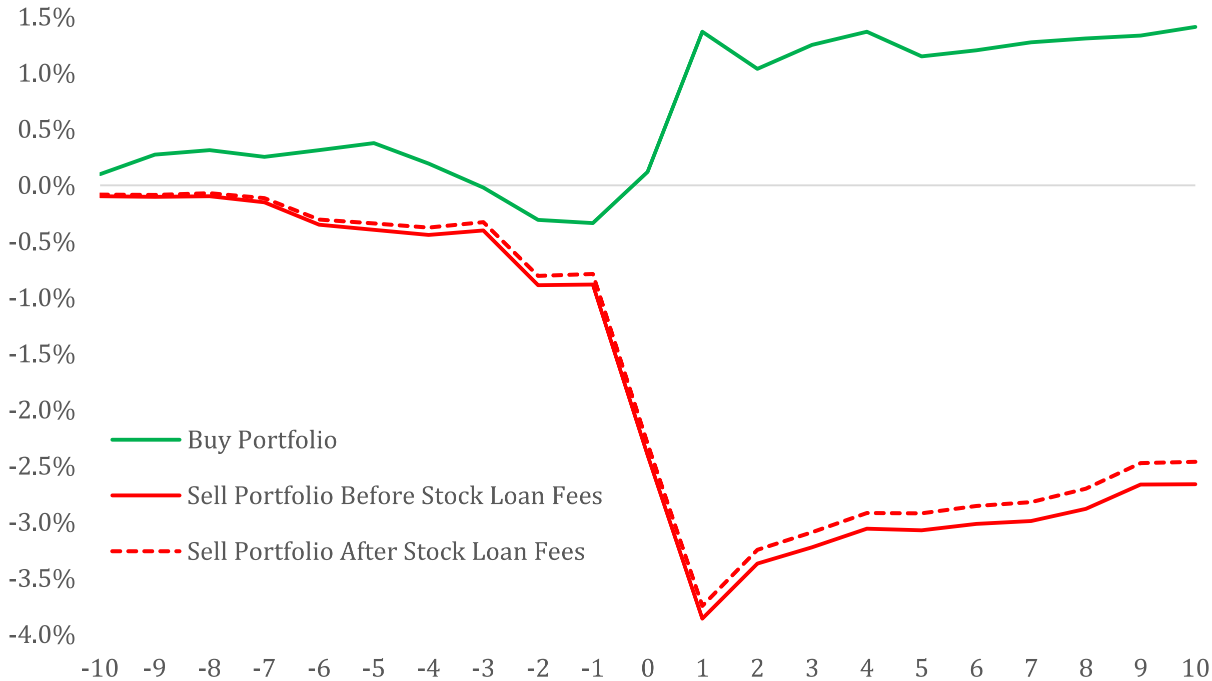
Figure 3
Timeline of research design



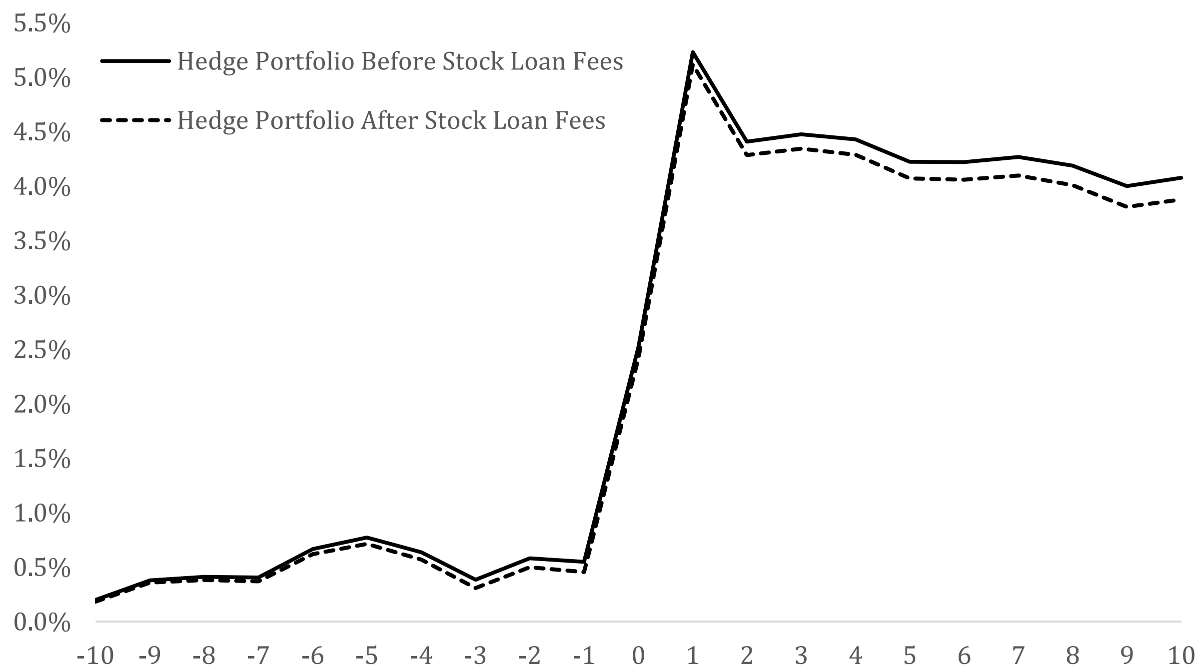
This figure presents the timeline of our research design along with the measurement of key variables.

Figure 4
Formulating a trading strategy from outer space

Panel A: Buy and sell portfolio returns.



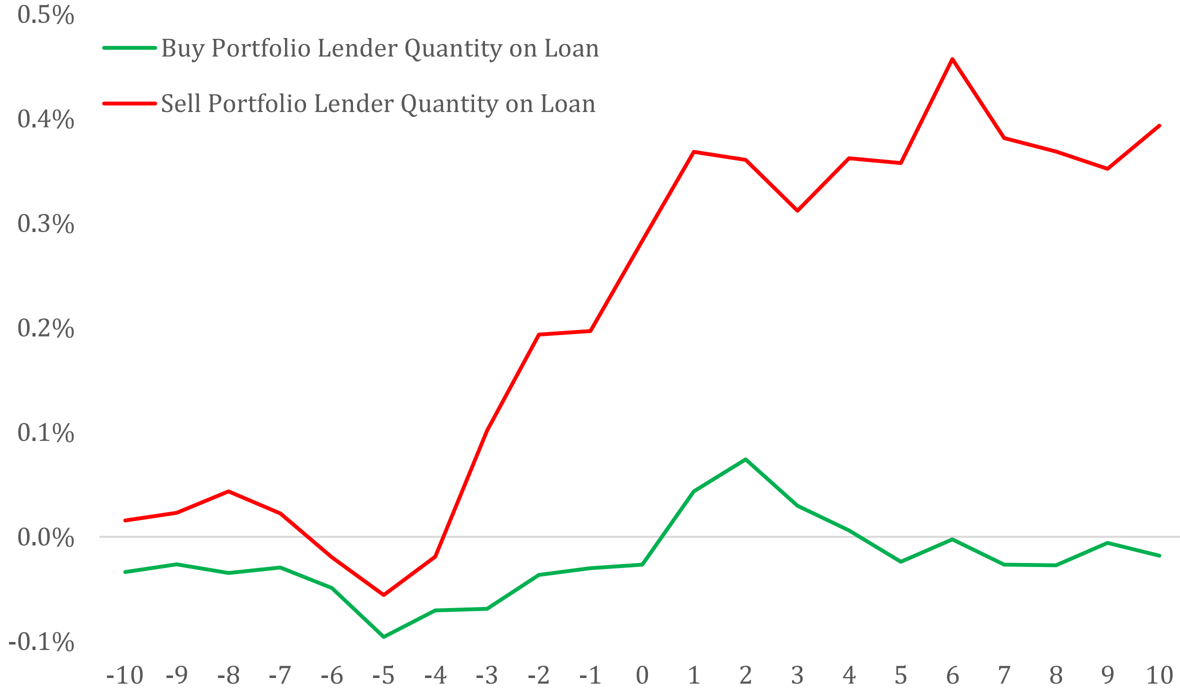
Panel B: Hedge portfolio returns.



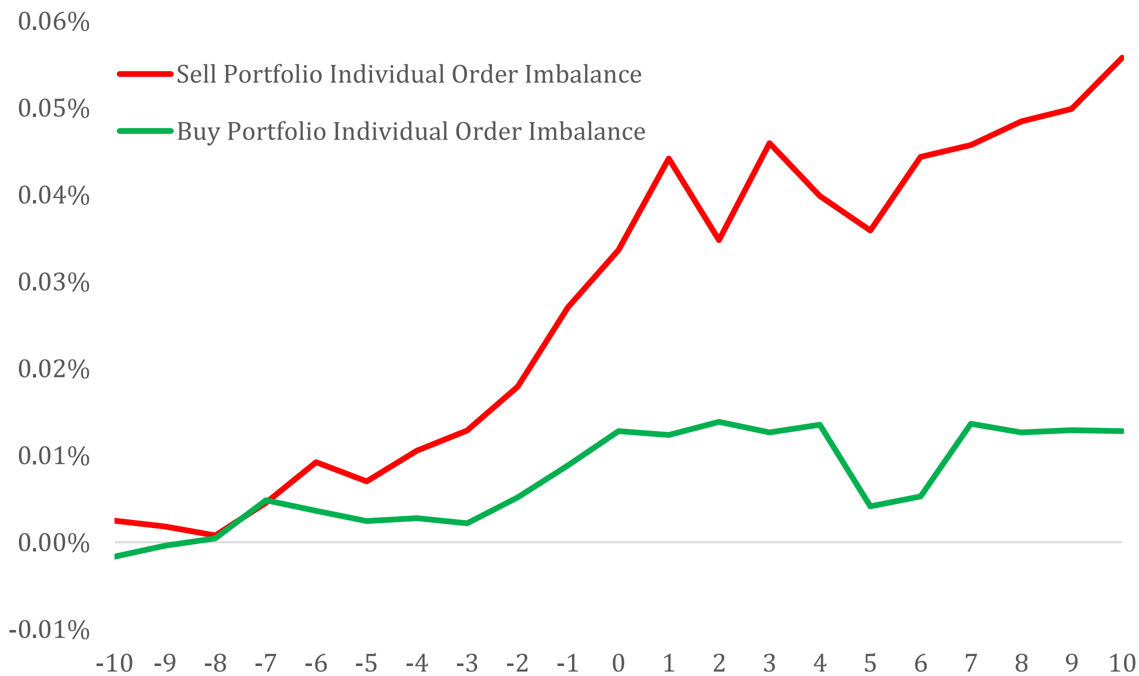
This figure presents the buy-and-hold market-adjusted return from a trading strategy that buys (short sells) retailers with same-store parking lot fill rate growth in the top (bottom) quartile of the cross-sectional distribution of $\Delta FLRT_{iq}$. To generate the cross-sectional quartile cutoff values 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 earnings announcement day for the quarter (day 0). In Panel A, the green (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 dashed line presents the performance of the short-sell portfolio net of cumulated stock loan fees. In Panel B, the black solid (dashed) line presents the hedge portfolio returns as the spread in the buy portfolio minus the short-sell portfolio before (after) stock loan fees. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4.

Figure 5
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 the 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 earnings announcement day for the quarter (day 0). The green (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 scaled by the total number of shares outstanding. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4.

Table A1
Formulating a trading strategy with RS Metrics and Orbital Insight data

Panel A: Before short-selling cost.

	Earnings announcement returns		
Portfolios	*Raw Returns*	*Market adjusted*	*Factor adjusted*
Sell portfolio	-3.03%***	-3.12%***	-3.17%***
	(-3.35)	(-3.49)	(-3.59)
Buy portfolio	1.93%**	1.82%**	1.81%**
	(2.57)	(2.43)	(2.41)
Buy − Sell	4.96%***	4.94%***	4.98%***
	(4.22)	(4.24)	(4.30)

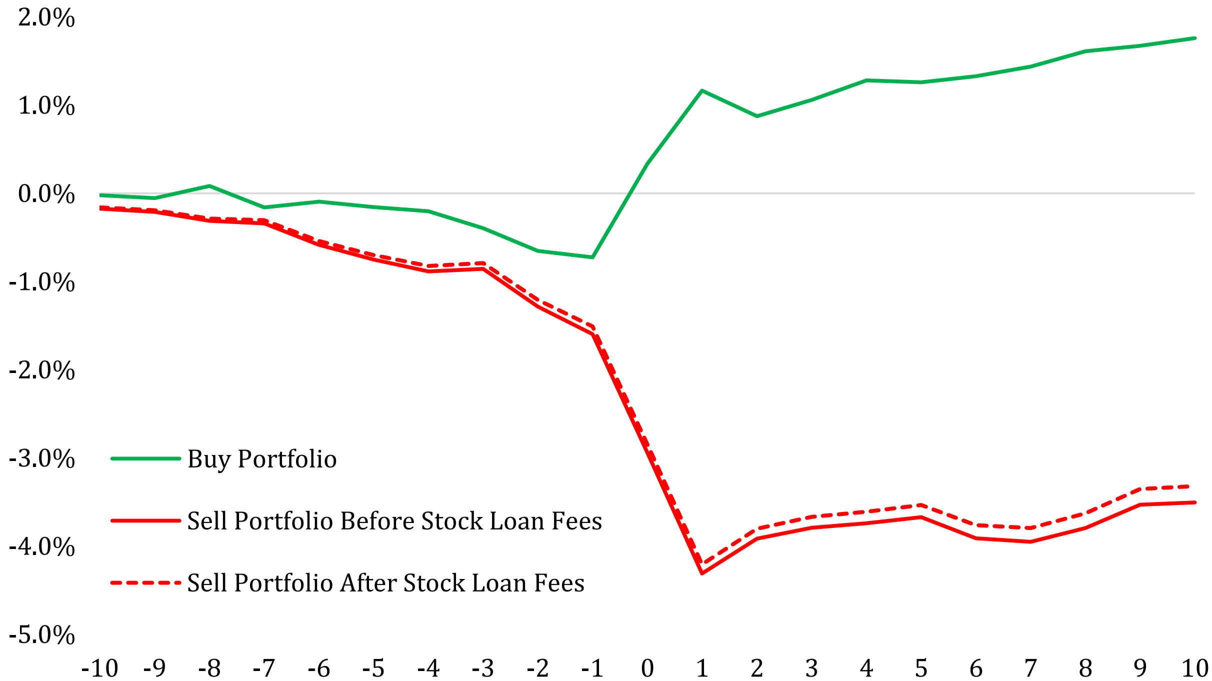
Panel B: After short-selling cost.

	Earnings announcement returns		
Portfolios	*Raw Returns*	*Market adjusted*	*Factor adjusted*
Sell portfolio	-3.00%***	-3.09%***	-3.14%***
	(-3.33)	(-3.47)	(-3.57)
Buy portfolio	1.93%**	1.82%**	1.81%**
	(2.57)	(2.43)	(2.41)
Buy − Sell	4.93%***	4.92%***	4.95%***
	(4.20)	(4.22)	(4.28)

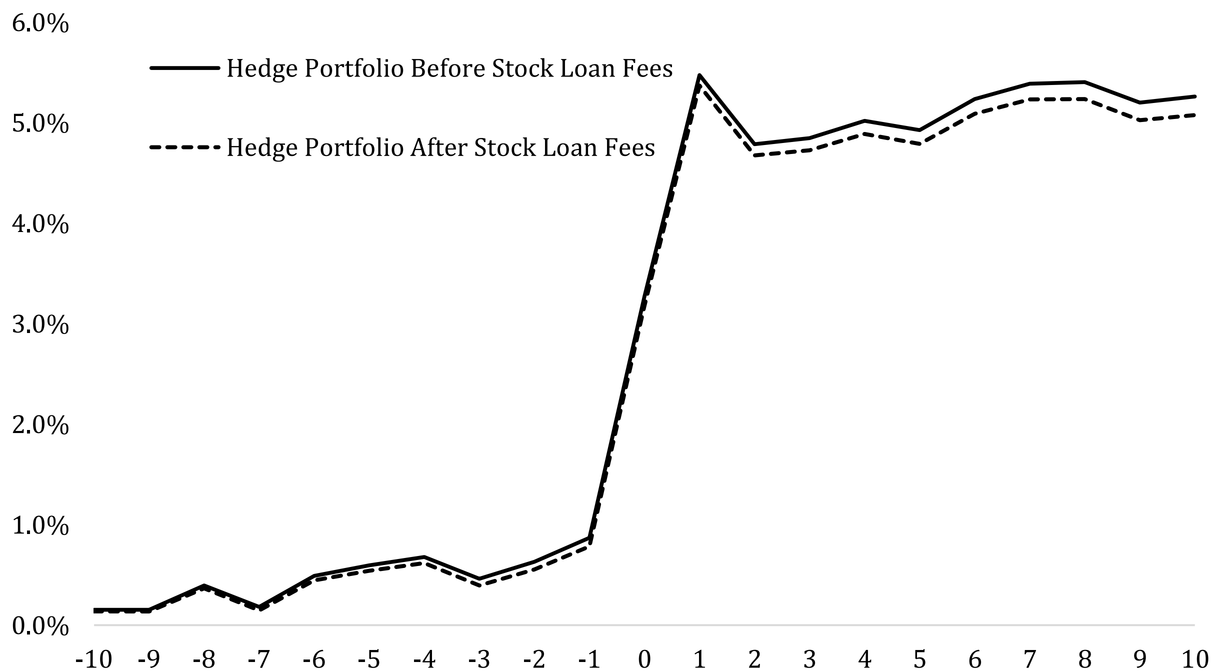
This table reports the buy-and-hold returns from a trading strategy that buys (short sells) retailers with same-store parking lot fill rate growth in the top (bottom) quartile of the cross-sectional distribution of parking lot traffic growth for the combined RS Metrics and Orbital Insight data. We expand the RS Metrics data with Orbital Insight data for the 44 retailers. We combine daily store-level observations for store locations with year-over-year satellite coverage. The combined data includes 4.9 million daily observations across 75,992 unique store locations for 44 major U.S. retailers over the period from 2011:Q1 to 2017:Q4. From the combined data, we compute the year-over-year same-store growth in parking lot traffic for our sample of 650 firm-quarters. We report raw returns, market-adjusted returns, as well as factor-adjusted returns. To generate the cross-sectional quartile cutoff values of parking lot traffic growth, we consider retailers with fiscal quarters ending within the last three months. The return measurement window is from one trading day before to one trading day after the earnings announcement day for the quarter (day 0). Panel A (Panel B) report results before (after) adjusting the short-sell portfolio returns for accumulated stock loan fees. The t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

Figure A1
Formulating a trading strategy with RS Metrics and Orbital Insight data

Panel A: Buy and sell portfolio returns.



Panel B: Hedge portfolio returns.



This table reports the buy-and-hold 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 parking lot traffic growth for the combined RS Metrics and Orbital Insight data. We expand the RS Metrics data with Orbital Insight data for the 44 retailers. We combine daily store-level observations for store locations with year-over-year satellite coverage. The combined data includes 4.9 million daily observations across 75,992 unique store locations for 44 major U.S. retailers over the period from 2011:Q1 to 2017:Q4. From the combined data, we compute the year-over-year same-store growth in parking lot traffic for our sample of 650 firm-quarters. We report raw returns, market-adjusted returns, as well as factor-adjusted returns. To generate the cross-sectional quartile cutoff values of parking lot traffic growth, we consider retailers with fiscal quarters ending within the last three months. The return measurement window is from ten trading days before to ten trading days after the earnings announcement day for the quarter (day 0). In Panel A, the green (red) solid line presents the performance of the portfolio that buys (short-sells) retailers in the top (bottom) quartile portfolio of parking lot traffic growth, while the red dashed line presents the performance of the short-sell portfolio net of cumulated stock loan fees. In Panel B, the black solid (dashed) line presents the hedge portfolio returns as the spread in the buy portfolio minus the short-sell portfolio before (after) stock loan fees.