Unlevelling the Playing Field: the Investment Value and Capital Market Consequences of Alternative Data

Marcus Painter[a]

[a] *Gatton College of Business and Economics*
University of Kentucky

Abstract

This paper documents the investment value of alternative data and examines how market participants react to the data's dissemination. Using satellite images of parking lots of US retailers, I find a long-short trading strategy based on growth in car count earns an alpha of 1.6% per month. I then show that, after the release of satellite data, hedge fund trades are more sensitive to growth in car count and are more profitable in affected stocks. Conversely, individual investor demand becomes less sensitive to growth in car count and less profitable in affected stocks. Further, the increase in information asymmetry between investors due to the availability of alternative data leads to a decrease in the liquidity of affected firms.

JEL classification: G12, G14, G23

Keywords: Alternative Data, Hedge Funds, Big Data, Information Asymmetry, Retail Investors, Liquidity

Email address: `marcus.painter@uky.edu` (Marcus Painter)

"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."

- Schroders Investment Management[1]

1. Introduction

The proliferation of alternative data has been one of the most striking changes to financial markets in recent years. Alternative data, also commonly referred to as big data, is any non-traditional data that can be used in the investment process.[2] As the introductory quote suggests, proponents of this information revolution believe that alternative data has the potential to uncover fundamental information before the release of more traditional information sources, such as financial statements or macroeconomic announcements. The majority of asset managers appear to share this opinion, with a recent report by JP Morgan estimating asset managers are spending $2-3 billion annually on alternative data.[3] Further, a survey conducted by Standard & Poor's indicates 80% of asset managers plan to increase their investments in big data, with only 6% of asset managers believing big data is not important to their investment process.[4]

The exponential growth of alternative data sources has significant implications for the informational environment of the firm. The growing dissemination of alternative data may only have investment value for the small group of investors who are either able to afford early

[1] "Is Big Data the Key to Bigger Investment Returns?" Morningstar. February 23, 2018.
[2] Examples include social media sentiment analysis (e.g., PsychSignal), crowdsourced investment research (e.g., Seeking Alpha and Estimize), credit card transactions and consumer spending data (e.g., Yodlee and Earnest), and satellite images (e.g., Orbital Insight).
[3] "JP Morgan: Alternative Data Is Altering Investment Landscape" Integrity Research Associates. June 14, 2017.
[4] Morningstar, 2018

access to such information or have a comparative advantage in processing and trading on the information. In fact, survey evidence indicates the two most cited impediments to the use of alternative data are high fixed costs and lack of expertise in managing the data.[5] Therefore, the rise in alternative data creates a potential informational advantage for large sophisticated investors relative to small investors.

The goal of this study is to assess the investment value and capital market consequences of alternative data. To do so, I collect detailed data on parking lot car counts for 163 companies from 2010-2017 from Orbital Insight, a leading provider of data derived from satellite imagery. An important feature of this data is that Orbital Insight began selling the data for these companies at different points in time. Specifically, Orbital Insight began selling data for 54 companies in the summer of 2015, 41 more companies in the summer of 2016, and 33 more companies at the beginning of 2017. Additionally, Orbital Insight collected car count data for 35 companies that were not released in my sample period. This staggered introduction of companies offers a relatively clean setting to explore the causal effects of the dissemination of alternative data on its investment value, its impact on the trading activities of sophisticated and individual investors, and the liquidity of affected firms.

I first demonstrate that a trading strategy based on alternative data reveals new information that predicts stock returns. I construct an investment strategy that each month goes long stocks in the top quintile of year-over-year car count growth and short stocks in the bottom quintile. This strategy generates monthly abnormal returns of 1.6% per month. This result holds in both equal-weighted and value-weighted portfolios and is robust to various risk-adjustments. I also find that an interquartile increase in quarterly car count growth is associated with a 14.47% increase in revenue surprise and a 0.17% increase in price-scaled unexpected earnings. These results suggest that growth in parking lot traffic is able to predict

[5] "This is the Future of Investing, and You Probably can't Afford it" Business Insider. May 28, 2017.

stock prices because it conveys important information about firms' future cash flows.

I next examine how the investment value of satellite data changes following the dissemination of the data to roughly 70 large asset management companies, most of whom are hedge fund managers. Somewhat surprisingly, I find the broader dissemination of the data has virtually no effect on the profitability of the trading strategy. Further, the profitability of the trading strategy persists even for stocks with characteristics that are associated with lower limits to arbitrage.

Given the persistent abnormal returns available from trading on alternative data, it is likely that large sophisticated investors are taking advantage of this opportunity. Hedge funds rank among the most sophisticated investors (Brunnermeier and Nagel (2004); Chen, Kelly, and Wu (2018)) and are able to execute complex trading strategies (Huang (2017); Jame (2017)). Additionally, the majority of initial clients for alternative data providers have been hedge funds.[6] Therefore, I test whether sophisticated investors are taking advantage of alternative data by analyzing the trading behavior and profitability of hedge funds around the dissemination of satellite data. Before Orbital Insight begins selling the data, I find no relation between car count growth and abnormal changes in hedge fund holdings. However, abnormal hedge fund holdings become significantly more responsive to car count growth after the data is released. This change in hedge fund trading behavior increases profitability for the funds, as abnormal hedge fund holdings in firms covered by Orbital Insight are associated with higher abnormal returns once the satellite data is disseminated.

I further investigate the trading behavior of institutional investors in relation to the use of alternative data by examining non-hedge fund asset managers (e.g., mutual funds, banks, and insurance companies). I find no significant change in the trading behavior or profitability of non-hedge fund institutions surrounding the release of the satellite data. This is unsurprising,

[6] "Alternative Data Use Cases Report" Eagle Alpha. April, 2018.

as conversations with industry professionals and survey data both suggest that non-hedge fund institutions will be late adopters of alternative data.[7] Although these funds may be able to afford the high fees associated with alternative data, they may find it too difficult to extract accurate and timely trading signals to justify the expense.

Big data creates a challenge for the subset of investors who are either unaware of the data or unable to take advantage of it. In particular, individual investors lack the resources to obtain and extract profitable signals from alternative data. Further, individuals investors' tendency to be contrarian traders (Grinblatt and Keloharju (2000); Kaniel, Saar, and Titman (2007)) means they are likely the liquidity suppliers for the demand created by alternative data. Consistent with smaller investors being unable to utilize alternative data, I find individual investor demand becomes negatively associated with car count growth after Orbital Insight begins selling the data. Further, individual investor demand predicts negative abnormal announcement returns once the satellite data is released.

My final tests analyze how firms are affected by the release of alternative data. The informational advantage for large sophisticated investors due to alternative data has important implications for firms' liquidity. On the one hand, the availability of alternative data reduces the information asymmetry between firm insiders and outsiders, which could enhance liquidity (Zhu, 2018). On the other hand, big data increases the information asymmetry between sophisticated investors and individual investors. In particular, theory predicts market makers may react to the the risks of dealing with informed traders by increasing bid-ask spreads (Copeland and Galai (1983); Glosten and Milgrom (1985)). Which effect dominates is ultimately an empirical question. Consistent with the rise of information asymmetry between sophisticated and individual investors being the dominant influence, I find that adverse selection arising from the availability of alternative data leads to higher bid-ask spreads and

[7] "Putting Alternative Data to Use in Financial Markets" Greenwich Associates. September 12, 2017.

amihud illiquidity ratios for treated firms.

This paper makes several contributions to the literature. First, it contributes to the growing literature that uses big data to predict stock prices and firm fundamentals. Recent studies show that the use of textual analysis to gauge investor opinion from blogs Seeking Alpha (Chen, De, Hu, and Hwang, 2014) and Twitter (Bartov, Faurel, and Mohanram, 2016) can be used to predict future stock returns and earnings surprises. Huang (2017) uses data on product reviews from Amazon.com to show that consumer opinions contain relevant information for stock pricing and that hedge fund holdings are positively correlated with changes in product reviews. Jame, Johnston, Markov, and Wolfe (2016) find that crowdsourced forecasts from Estimize are incrementally useful in predicting earnings. Da, Engelberg, and Gao (2011) use data on search frequency from Google and find that an increase in search frequency predicts higher short-term stock prices. While prior research using data similar to Orbital Insight's satellite data finds earnings and revenue predictability (Froot, Kang, Ozik, and Sadka (2017); Zhu (2018)), my study is the first to construct a profitable monthly trading strategy based on satellite data. Additionally, this paper is the first to show that the profitability of trading strategies based on alternative data is unrelated to its broader dissemination, suggesting the alpha decay thought to be associated with the implementation of big data investment strategies may be less extreme than once expected.

Second, this paper contributes to the burgeoning literature on how alternative data impacts financial markets. Froot et al. (2017) develop a proxy for real-time corporate sales using data from mobile phones and tablets to show that managers bias earnings forecasts depending on the firm's real-time performance. Zhu (2018) argues the release of alternative datasets reduces information acquisition costs, ultimately leading to greater stock price efficiency, less insider trading, and more efficient investing by the firm's managers. My paper contributes to this literature by being the first to show that sophisticated investors (specifically, hedge funds) benefit from the dissemination of alternative data while individual investors suffer.

Third, this paper adds to the literature on asymmetric information between investors. Looking at a reduction in information asymmetry between investors following Regulation Fair Disclosure, which disallowed the selective disclosure of material information, Chiyachantana, Jiang, Taechapiroontong, and Wood (2004) and Eleswarapu, Thompson, and Venkataraman (2004) find that stock liquidity improves following the regulation. Focusing on the rise of investment research websites, Farrell, Green, Jame, and Markov (2018) show that stocks that have reductions in coverage on the website Seeking Alpha have higher bid-ask spreads and price impact. My paper complements these findings, as I show that information asymmetry and liquidity are also negatively related in the setting of alternative data. My findings are also consistent with studies regarding the rise in information asymmetry due to an exogenous reduction in sell-side analyst coverage. Kelly and Ljungqvist (2012) find that bid-ask spreads and amihud illiquidity ratios both rise after a reduction in analysts and Chen et al. (2018) show that hedge funds trade and profit more after this decreased analyst coverage. Looking specifically at the release of alternative data, Zhu (2018) finds an increase in stock price efficiency, suggesting a decrease in asymmetric information between firm insiders and investors. Rather than studying the relation between insiders and investors, my paper focuses on the information environment between different subgroups of investors, finding that asymmetric information rises in this setting and leads to lower liquidity.

2. Data

2.1. Satellite Image Data

I obtain data on parking lot car counts from Orbital Insight, an image processing company that uses machine learning to convert satellite images into quantitative data. My sample period begins in January 2010 and ends in December 2017.[8] Orbital Insight uses a representative

[8]Note that the data from Orbital Insight starts in 2009, with the first year being used to train their model.

sample of each firm's parking lots to construct a variable for the average number of cars at a firm's retail stores on a given day. Their normalization process accounts for variation in traffic at different times of day as well as factors unique to certain stores. For example, the normalization process considers whether a Wal-Mart is a standard location or a larger Super Wal-Mart location. Notably, data on traffic volume is generally available around 16 hours after an image is taken, so investors can make timely trades based on this information.

Figure 1 shows a sample parking lot image for a Wal-Mart store in Arizona. Orbital Insight draws a "mask" for each parking lot in order to reduce the possibility that cars parked at other stores enter the data. Each circle in the figure represents a car identified by the algorithm. Only circles within the shaded area are counted towards Wal-Mart's car count for that day.

To predict future monthly stock returns, I measure traffic growth as the difference between the log of the average car count for a month minus the log of average car count 12 months prior. This method of measuring growth reduces the possibility that noise from seasonality will affect the predictive power of the measure. Figure 2 shows average daily car counts for firms in my sample, emphasizing the importance of adjusting for seasonality in parking lot traffic. The average car count for all stocks is shown in panel A. The graph shows that average car counts spike during November and December as consumers prepare for the holiday season. Average car counts are lower in January and February, likely due to cold weather dissuading consumers from leaving their homes. Panels B and C show seasonal car counts for Wal-Mart and Home Depot, respectively. Wal-Mart's average car counts are similar to the average covered firms, with more cars in parking lots in December and fewer cars in January. Home Depot follows a unique pattern, as the bulk of the company's sales come in spring. As a result, Home Depot's car counts are higher than average from April to June. These differing patterns in car counts are adjusted for by using the year-over-year car count measure.

To illustrate how micro-level traffic patterns reveal novel information about stock prices, consider Figure 3, which compares cumulative year-over-year growth in car traffic with cumulative stock returns for Bed Bath and Beyond from 2011 to 2017. The two variables trend in similar directions, though the growth in car count variable appears to be a leading indicator of stock returns. The satellite data shows a growth in car count for Bed Bath and Beyond from the beginning of the sample and peaking at the start of 2013. The cumulative stock return follows a similar upward trend but does not peak until the end of 2013. A downward trend in car count begins in February of 2013, which is not reflected in stock prices until November of 2013. A final decline in car counts begins in April 2015, which follows closely with a decline in stock returns. This example represents a systematic pattern across US retailers: traffic patterns reveal novel information about stock prices that can be used to generate abnormal returns.

I merge the satellite data with stock return data from CRSP and firm accounting data from Compustat. I also obtain factor data for market return (MKTRF), size (SMB), value (HML), momentum (UMD), profitability (RMW), and investment (CMA) from Ken French's data library.

Table 1 presents summary statistics for the full sample of covered firms as separated by each release period. The average growth in car count is slightly negative at -0.5% with an interquartile range of 8.08%. Orbital Insight had two main decision criteria for choosing which firms to cover in their initial release. First, the firms had to be large U.S. firms that would therefore be covered by a larger number of investors. Second, the companies must have enough store locations to generate a reliable normalized measure of car counts. Therefore, the size and level of turnover of firms in Release 1 are larger on average than in subsequent releases. Firms in Release 1 also tend to be more value stocks, as evidenced by their lower average book-to-market. I include these characteristics as well as past stock returns in all regressions to control for differences in these characteristics between release periods.

2.2. Institutional Investor Holdings

To identify institutional investors, I use the 13f institutional holdings data from Thomson Reuters. Institutions with over $100 million in assets are required to fill out the quarterly 13f forms for all U.S. equity positions exceeding $200,000 or 10,000 shares. I identify hedge funds in this database following the methodology of Brunnermeier and Nagel (2004) and Griffin and Xu (2009). More specifically, hedge funds are identified by matching the 13f fund names with names from five hedge fund databases: BarclayHedge, HFR, Eureka, Lipper TASS, and Morningstar.[9] I designate all funds not matched through this process as non-hedge funds. Non-hedge funds include other institutional investors like mutual funds, insurance companies, and banks. The final sample for the 2010 to 2017 period includes 659 hedge funds and 4,427 non-hedge funds.

I define abnormal holdings for hedge funds and non-hedge funds using the following equation:

$$Abn\_HF_{i,q} = (\frac{Shares\ Owned^{HF}_{i,q}}{Shares\ Out_{i,q}}) - \frac{\sum_{t=q-4}^{t=q-1}(\frac{Shares\ Owned^{HF}_{i,t}}{Shares\ Out_{i,t}})}{4} \tag{1}$$

where $Shares\ Owned_{i,q}$ is the total number of shares of stock i held by hedge funds in quarter q and $Shares\ Out_{i,q}$ is the total number of shares outstanding for stock i in quarter q. Abnormal holdings for non-hedge funds are defined analogously. Table 1 shows that average abnormal holdings of firms in my sample for hedge funds (non-hedge funds) is -0.07 (0.15) with an interquartile range of 3.37 (5.37).

2.3. Individual Investor Order Imbalances

I use the Trade and Quote (TAQ) dataset to measure individual investor activity. I follow Boehmer, Jones, and Zhang (2017), who provide a clean method to identify individual investor order flow. Boehmer et al. (2017)'s method exploits the tendency for individual investors'

[9] See Cao, Chen, Goetzmann, and Liang (2016) for more detail on the hedge fund identification process.

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 me to identify a clean sample of individual investor initiated transactions in order to measure individual investor demand.

After collecting information on individual investor trading activity, I compute the following order imbalance measures:

$$Indiv\ OIB\ Vol_{i,t} = \frac{Ind\ Buy\ Vol_{i,t} - Ind\ Sell\ Vol_{i,t}}{Ind\ Buy\ Vol_{i,t} + Ind\ Sell\ Vol_{i,t}} \tag{2}$$

$$Indiv\ OIB\ Trade_{i,t} = \frac{Ind\ Buy\ Trade_{i,t} - Ind\ Sell\ Trade_{i,t}}{Ind\ Buy\ Trade_{i,t} + Ind\ Sell\ Trade_{i,t}} \tag{3}$$

where $Ind\ Buy\ Vol_{i,t}$ is the average buy volume by individual investors in stock i during time t and $Ind\ Sell\ Vol_{i,t}$ is the average sell volume in stock i during time t. Similarly, $Ind\ Buy\ Trade_{i,t}$ is the average number of buy trades by individual investors in stock i during time t and $Ind\ Sell\ Trade_{i,t}$ is the average number of sell trades in stock i during time t. The average $Ind\ Buy\ Vol$ in my sample is -1.9% with an interquartile range of 16.8% and the average $Ind\ Buy\ Trade$ is -1.9% with an interquartile range of 15.3%.

3. The Investment Value of Satellite Imagery

3.1. Stock Returns

In order to test whether growth in traffic to firms' stores predicts stock returns, I form portfolios based on the growth in car count measure described in section 2.1. At the beginning of each month, I sort stocks into quintile portfolios based on traffic growth and then track their performance over the following month. Quintile 5 contains firms with the highest growth in car count, while quintile 1 contains those with the lowest growth. I then form a high-minus-low (H-L) portfolio which goes long the stocks in quintile 5 and shorts those in quintile 1.

The performance results of the quintile portfolio sorts are reported in Table 2. I analyze returns using raw excess returns as well as CAPM, four-, and six-factor alphas for each portfolio. The four-factor alpha contains the excess market return (MKTRF) as well as the size (SMB), value (HML), and momentum (UMD) factors of Fama and French (1993) and Carhart (1997). The six-factor model adds the profitability (RMW) and investment (CMA) factors of Fama and French (2015). I present equal-weighted returns in Panel A and value-weighted returns in Panel B.

The results in Table 2 show that trading based on signals generated from alternative data can lead to significant outperformance. Across all return specifications, portfolio returns increase monotonically with the prior month's growth in car count. Focusing on excess returns, the highest quintile of stocks earns 1.86% per month while the lowest quintile earns 0.26% resulting in a long-short portfolio of 1.6%. Alphas generated from the CAPM, four-, and six-factor models suggest that factor exposures cannot explain the statistical and economic significance of the long-short portfolio. The H-L portfolios for equal-weighted returns are significant at the 1% level for all factor models and have alphas that range between 1.58% and 1.65% per month. For value-weighted returns, the H-L portfolios earn monthly alphas between 1.56% and 1.67% and are statistically significant at the 1% level.[10]

[10]In the appendix I confirm that growth in car count also predicts firm fundamentals. Specifically, I find an interquartile increase in growth in car count is associated with a 14.47% increase in revenue surprise and a 0.17% increase in price-scaled unexpected earnings. Further, I confirm a substantial portion of the return realization from satellite data stems from earnings announcements, as an interquartile increase in growth in car count is associated with a 71 basis point increase in cumulative abnormal announcement returns.

3.2. Return Predictability Post-Dissemination

In this section, I test the persistence of portfolio returns after the satellite data becomes available to the market. The economically large monthly alpha generated by this trading strategy creates a strong incentive for investors to seek out alternative datasets in order to capture excess profits. If enough investors begin trading on this information then the alpha generated by this strategy will eventually be arbitraged away. However, there are several arguments for why the excess profits from this strategy will persist after the data becomes available. First, the cost of alternative datasets will prohibit smaller investors from accessing them. Second, many larger discretionary funds who can afford alternative datasets do not have the right infrastructure to incorporate the data into their discretionary trading strategies.[11] Finally, the investors who have the resources to purchase and develop trading strategies from alternative data may identify different signals from the same dataset. Therefore, it is possible that the alphas generated from the trading strategy developed in this paper are not immediately arbitraged away.

To test between these two possibilities, I examine the persistence of portfolio returns using a difference-in-differences framework. Specifically, I run the following regression:

$$\begin{aligned} Excess\ Return_{i,t+1} &= \alpha + \beta_1 * Growth\ in\ Car\ Count_{i,t} + \beta_2 * Release \\ &+ \beta_3 * (Growth\ in\ Car\ Count_{i,t} \times Release) + \beta_4 * Firm\ Controls_{i,t} \\ &+ \beta_5 * Firm\ FE + \beta_6 * Year\ Month\ FE + \epsilon_{i,t} \end{aligned} \tag{4}$$

where $Excess\ Return_{i,t+1}$ is the stock return in month $t+1$ for stock i in excess of the risk-free rate, $Growth\ in\ Car\ Count_{i,t}$ is the growth in traffic for firm i in month t, and *Release* is an indicator variable equal to zero for a firm in months prior to Orbital Insight

[11] "Revenge of the Humans: How Discretionary Managers Can Crush Systematics" Leigh Drogen, CEO of Estimize. May 8, 2017.

releasing the data, and equal to one in the months after Orbital Insight begins selling the data. Therefore, $Growth\ in\ Car\ Count_{i,t} \times Release$ captures the marginal change in return predictability of $Growth\ in\ Car\ Count_{i,t}$ for a firm whose data has been disseminated relative to before the data is released. A negative coefficient on the interaction term would indicate that investors trade away the profitability of the trading strategy once the data becomes available.

$Firm\ Controls_{i,t}$ is a set of firm characteristics for firm i in month t. I use the following firm characteristics as controls: the log market value of equity at the end of the prior month, the log of book-to-market at the end of the prior month, the stock return from month $t-12$ to $t-2$, the log of growth in shares outstanding from month $t-36$ to month $t-1$, the change in net working capital minus depreciation in the prior fiscal year, the return on assets in the prior fiscal year, and the log of growth in total assets in the prior fiscal year. Lewellen (2015) shows that these seven characteristics have significant predictive power for stock returns. I include firm and year-month fixed effects and cluster standard errors by firm and year-month.

Table 3 presents the results of the difference-in-differences estimation. I first report results for return predictability for the entire sample in column 1. The coefficient on growth in car count is 0.033, significant at the 1% level. In economic terms, this means that an interquartile range increase in growth in car count is associated with a 26.6 basis point increase in the following month's excess return. This result suggests that traffic growth contains information which cannot be garnered from firm characteristics. Column 2 includes the indicator variable for whether Orbital Insight has released a firm's data as well as its interaction with growth in car count. The coefficient on the interaction term is indistinguishable from zero at standard levels of significance. This result provides evidence that the profitability of the trading strategy is not immediately traded away when the data is disseminated.

3.3. Limits to Arbitrage

A potential explanation for the persistence of the return predictability is arbitrage constraints. The profitability of the trading strategy may be restricted to stocks that limit arbitrageurs by being difficult to analyze or trade in large quantities. I test whether limits to arbitrage explain the persistence of return predictability of alternative data using three proxies to identify stocks more likely to have arbitrage constraints: firm size, bid-ask spread, and the amihud illiquidity ratio. For firm size, I use the market equity of the firm. I calculate bid-ask spread as the difference between the ask and bid prices divided by the average of the ask and bid prices. I follow Amihud (2002) and calculate the illiquidity measure as the average ratio of daily absolute return to dollar trading volume. I test the difference-in-difference regression from equation (4) on subsamples split at the median of these three proxies. If the persistent profitability of the portfolio returns are due to stocks with limits to arbitrage, then the interaction term should be significantly negative for large firms, firms with low bid-ask spreads, and firms with low amihud illiquidity ratios and insignificant for small firms, firms with high bid-ask spreads, and firms with high amihud illiquidity ratios.

Columns 3 through 8 of Table 3 report the results of the subsample splits. I find that, regardless of a firm's arbitrage constraints, the trading strategy based on parking lot traffic growth remains a significant predictor of future excess returns. Specifically, the coefficient on the interaction between growth in car count and release is statistically insignificant in each subsample split. These results suggest that the profitability of the alternative data trading strategy post-dissemination is not due to limits to arbitrage.

4. The Capital Market Consequences of Alternative Data

The investment value contained in alternative data creates an incentive for sophisticated investors to implement the data into their trading strategies. There is evidence that hedge fund managers represent the most sophisticated investors, as they often outperform mutual funds (Ackermann, McEnally, and Ravenscraft, 1999) and are able to develop informed trading strategies based on novel information (Huang, 2017). In this section I investigate how sophisticated investors adapt to the introduction of alternative data by examining the trading behavior of hedge funds around the dissemination of Orbital Insight's satellite data. Additionally, I study the trading behavior of non-hedge fund institutions and individual investors. Though non-hedge funds are likely able to afford the expense of alternative data, they are less likely to be able to turn that data into a profitable trading strategy. Individual investors lack the resources to purchase or develop trading strategies from alternative data. Further, individuals investors' tendency to be contrarian traders (Grinblatt and Keloharju (2000); Kaniel et al. (2007)) means they are likely the liquidity suppliers for the demand created by alternative data. Therefore, I expect a positive relationship between the Orbital Insight trading signal and hedge fund trading, no significant relationship between the trading signal and non-hedge fund institution trading, and a negative relationship between the trading signal and individual investor trading. Further, I expect hedge fund trades to become more profitable after the release of Orbital Insight's data, non-hedge fund trades to have no change in profitability, and individual investor trades to become less profitable.

4.1. Institutional Investor Holdings

To test how institutional investor trading behavior changes after the release of alternative data, I run difference-in-differences regressions of abnormal holdings for hedge funds and non-hedge funds on growth in car count around the dissemination of the satellite data. Because institutional holdings are reported quarterly, I measure growth in car count at a quarterly frequency as well. Specifically, in each quarter I examine the change in abnormal holdings for hedge funds and non-hedge funds using the following regression model:

$$\begin{aligned} Abnormal\ Holdings_{i,t} &= \beta_1 * Growth\ in\ Car\ Count_{i,t} + \beta_2 * Release \\ &+ \beta_3 * (Growth\ in\ Car\ Count_{i,t} \times Release) + \beta_4 * Firm\ Controls_{i,t} \\ &+ \beta_5 * Firm\ FE + \beta_6 * Year\ Quarter\ FE + \epsilon_{i,t} \end{aligned} \tag{5}$$

where $Abnormal\ Holdings_{i,t}$ is the level of abnormal holdings for either hedge funds or non-hedge funds in stock i for quarter t. I test the relationship between abnormal holdings and growth in car count in contemporaneous quarters as clients of Orbital Insight are able to access the data with only a 16-hour lag. Therefore, the coefficient on ($Growth\ in\ Car\ Count_{i,t} \times Release$) estimates the sensitivity of fund's holdings to the growth in car count measure in the post-dissemination period relative to the pre-dissemination period. I include stock-level controls for prior quarter returns adjusted for the returns of the CRSP value-weighted index, book-to-market, size, and turnover. I cluster standard errors by firm and year-quarter.

Table 4 presents the regression results. The coefficient on growth in car count is insignificant in all specifications for both hedge funds and non-hedge funds, implying that neither type of institutional investor was able to trade based on parking lot growth prior to the availability of the satellite data. However, the coefficient on the interaction term is significantly positive for abnormal hedge fund holdings and insignificant for non-hedge fund holdings. Specifically, the regression using year-quarter fixed effects (column 1) shows that an interquartile increase in quarterly growth in car count (8.7%) leads to a 50 (8.7*0.057) basis point increase in abnormal hedge fund holdings in firms whose satellite data has been released relative to firms whose data has not been released. A 50 basis point increase in abnormal hedge fund holdings represents 14.2% percent of the interquartile range (3.37%). Looking within-firm, column 2 shows that an interquartile increase in car count growth is associated with a 70.47 basis point increase in abnormal hedge fund holdings for a firm in the post-dissemination period relative to the pre-dissemination period. The coefficient on the interaction term remains positive and significant when both year-quarter and firm fixed effects are included.

The results when looking at non-hedge funds show no effect of satellite data availability on abnormal holdings. The specification including year-quarter and firm fixed effects for abnormal non-hedge fund holdings leads to an insignificant coefficient on the interaction term of -0.014. These findings are consistent with the expectation that sophisticated investors take advantage of the availability of alternative data.

4.2. Institutional Investor Profitability

I next test whether institutional holdings become more profitable after alternative data is available. I hypothesize that hedge funds' better information processing abilities allow them to profit more on trades in stocks that have alternative data available. I estimate the following regression model to examine how abnormal holdings for hedge funds and non-hedge funds perform around the release of Orbital Insight's data:

$$\begin{aligned} Adjusted\ Stock\ Returns_{i,t+1} = {} & \beta_1 * Abnormal\ Holdings_{i,t} + \beta_2 * Release \\ & + \beta_3 * (Abnormal\ Holdings_{i,t} \times Release) + \beta_4 * Firm\ Controls_{i,t} \\ & + \beta_5 * Firm\ FE + \beta_6 * Year\ Quarter\ FE + \epsilon_{i,t} \end{aligned} \tag{6}$$

where all variables are defined the same as in previous regressions. I expect a significantly positive coefficient on ($Abnormal\ Holdings_{i,t} \times Release$) for hedge funds and a coefficient near zero for non-hedge funds.

Results for regressions using equation (6) are reported in Table 5. As expected, abnormal hedge fund holdings become more predictive of future returns once there is alternative data coverage on a stock. In terms of economic magnitude, the regression in column 1 shows that an interquartile increase in abnormal hedge fund holdings is associated with 1.35% (3.734*0.004) increase in the next quarter's adjusted stock returns in stocks whose data has been released relative to unreleased stocks. This 1.35% increase represents 6.36% of the interquartile range for adjusted stock returns. The results when using firm fixed effects show an interquartile increase in abnormal hedge fund holding leads to a 2.02% (3.374*0.006) increase in returns for a stock after its data has been made available relative to the pre-dissemination period. The coefficient on the interaction term remains significant when both time and firm fixed effects are included. Columns 4 through 6 show abnormal non-hedge fund holdings have no change in stock return predictability around the release of the satellite data, as the coefficient on the interaction term is insignificant in all specifications. Collectively, the results in Tables 4 and 5 suggest that hedge funds are better able to take advantage of big data compared to other institutional investors.

4.3. Individual Investor Trading

Given the large expense to access big data, individual investors are not the target client for alternative data providers. Because individual investors are on average contrarian traders (Grinblatt and Keloharju (2000); Kaniel et al. (2007)), it is possible that individuals provide liquidity to meet sophisticated investor demand for stocks with alternative data coverage. Although contrarian trading has historically led to positive excess returns for individual investors (Kaniel et al. (2007)), the asymmetric information environment brought on by alternative data could lead to negative returns on trades made by individual investors. In this section, I test this idea by examining how the sensitivity of individual order demand in relation to growth in car count changes around the release of alternative data. Specifically, I run the following regression:

$$\begin{aligned} Individual\ OIB_{i,t} &= \beta_1 * Growth\ in\ Car\ Count_{i,t} + \beta_2 * Release \\ &+ \beta_3 * (Growth\ in\ Car\ Count_{i,t} \times Release) + \beta_4 * Firm\ Controls_{i,t} \\ &+ \beta_5 * Firm\ FE + \beta_6 * Year\ Month\ FE + \epsilon_{i,t} \end{aligned} \tag{7}$$

where $Individual\ OIB_{i,t}$ refers to either the volume or trade order imbalance for individual investors in stock i in month t. The coefficient on ($Growth\ in\ Car\ Count_{i,t} \times Release$) measures how individual investor demand for a stock relative to a firm's traffic growth changes when alternative data is released.

The first three columns of Table 6 provide results of estimating equation (7) for the order imbalance of individual investor trade volume. Interestingly, the coefficient on growth in car count is significantly positive (thought economically small with a magnitude of 0.0004), suggesting that individual investors are able to trade in the same direction as traffic growth before alternative data is available. This result is consistent with the literature that finds individual investors are informed (Kaniel, Liu, Saar, and Titman (2012); Kelley and Tetlock (2013); Boehmer et al. (2017)). However, individual order flow becomes significantly less sensitive to traffic growth once access to the satellite data becomes available. Looking at column 3, a 1% increase in growth in car count is associated with 6.4% decrease in individual volume order demand in the post-dissemination period relative to the pre-dissemination period. Similar results are found when looking at the order imbalance for the number of individual trades in columns 4 through 6.

4.4. Indiviudal Investor Profitability

I next investigate the profitability of individual investor demand for stocks whose parking lot data is sold by Orbital Insight. Due to the granular nature of TAQ data, I am able to examine how investor demand immediately prior to earning announcements is able to predict announcement returns. The decrease in individual investor demand to traffic growth shown in Table 6 suggests that individual demand will become less informed after alternative data becomes available. I use the following model to examine whether there is a change in individual investor profitability:

$$
\begin{aligned}
Cumulative\ Abnormal\ Returns_{i,t,d\ to\ d+3} = \beta_1 * Individual\ OIB_{i,t,d-3\ to\ d-1} + \beta_2 * Release \\
+ \beta_3 * (Individual\ OIB_{i,t,d-3\ to\ d-1} \times Release) + \beta_4 * Firm\ Controls_{i,t} \\
+ \beta_5 * Firm\ FE + \beta_6 * Year\ Quarter\ FE + \epsilon_{i,t}
\end{aligned}
\tag{8}
$$

where $Cumulative\ Abnormal\ Returns_{i,t,d\ to\ d+3}$ is the abnormal announcement return for stock $i's$ earnings announcement for quarter t from the day of the announcement until three days after and $Individual\ OIB_{i,t,d-3\ to\ d-1}$ is the individual order imbalance for either volume or trades for stock i for the three days prior to the earnings announcement for quarter t. The interaction term will therefore measure how the profitability of individual investor demand changes after the introduction of alternative data to the market.

Table 7 reports the regression results. Consistent with expectations, individual investor demand becomes significantly worse at predicting announcements returns after satellite data for a company becomes available. Looking at column 3, a 1% increase in three day individual order imbalance is associated with a 4.7 basis point decrease in cumulative abnormal announcement returns in the post-dissemination period relative to the pre-dissemination period. This result holds for all specifications shown and for order imbalance measures using both volume of shares and number of shares traded.[12] Together, Tables 6 and 7 provide supporting evidence for the notion that individual investors become relatively less informed when alternative data is introduced.

[12]Results are also robust to using order imbalance measures for a one day or five day period prior to the earning announcement date.

4.5. Liquidity Implications for Affected Firms

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 an increase in stock illiquidity.[13] Therefore, it is possible that the release of satellite data by Orbital Insight leads to a decrease in the liquidity of affected firms' stocks.

Table 8 reports evidence consistent with this conjecture. The table reports regression results of stock liquidity measures on an indicator variable equal to one if a firm's satellite data has become available to investors. I use the bid-ask spread and the amihud illiquidity ratio as measures of stock liquidity. Looking at columns 2 and 5, which analyze within-firm changes in liquidity by including firm-level fixed effects, I find a significant increase in both liquidity measures. Specifically, the release of Orbital Insight's satellite data is associated with an increase in the amihud illiquidity ratio of 0.003 which represent a 30.00% increase from the mean amihud of 0.01. Similarly, the bid-ask ratio increases by 0.04, representing a 30.08% increase from the mean.

I account for omitted factors that could introduce a time trend by including year-month fixed effects in columns 3 and 6. Release is identified in this setting due to the staggered introduction of the satellite data by Orbital Insight. Though the magnitudes on the coefficients are reduced, I continue to find positive coefficients on the release indicator variable for both liquidity measures. Under this specification, the amihud illiquidity ratio for firms whose data has been disseminated is 20% higher than the average firm. Likewise, the bid-ask spread is 3.8% higher (though the coefficient is insignificant at conventional levels). These results are

[13]Some examples of theoretical papers in this line of research include Copeland and Galai (1983), Glosten and Milgrom (1985), Kyle (1985), and Easley and O'hara (1987).

consistent with the literature that finds that stock liquidity decreases in environments with increasing informational symmetry.

5. Conclusion

I contribute to the understanding of alternative data's impact on financial markets by documenting the investment value of a leading satellite imagery provider's data and examining how the dissemination of this data affects investor trading behavior. Using a measure of the growth in the number of cars in US retail firms' parking lots, I show that a long-short trading strategy based on alternative data earns monthly alphas of 1.6%. These abnormal returns persist even after the data is made available to market participants.

Using the staggered introduction of the satellite data as a natural experiment, I find market participants react to the dissemination of this data. In particular, hedge funds begin to trade in the direction of the growth in car count measure after the data is released, while individual investor demand trends the opposite way. Further, hedge fund trades become more profitable and individual investor demand becomes less profitable. Finally, the increase in information asymmetry between market participants leads to lower liquidity for covered firms' stocks.

Alternative data has become an essential resource for investors looking for an informational advantage. This type of data will continue to impact financial markets as more datasets are introduced. Collectively, my findings illustrate the power of alternative data both in its ability to generate value and in its influence on the functioning of capital markets.

References

Ackermann, Carl, Richard McEnally, David Ravenscraft. 1999. The performance of hedge funds: Risk, return, and incentives. *The Journal of Finance*, **54**(3) 833–874.

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

Bartov, Eli, Lucile Faurel, Partha S Mohanram. 2016. Can twitter help predict firm-level earnings and stock returns? *Rotman School of Management Working Paper No. 2631421. Available at SSRN: https://ssrn.com/abstract=2782236 or http://dx.doi.org/10.2139/ssrn.2782236*.

Berkman, Henk, Cameron Truong. 2009. Event day 0? after-hours earnings announcements. *Journal of Accounting Research*, **47**(1) 71–103.

Boehmer, Ekkehart, Charles Jones, Xiaoyan Zhang. 2017. Tracking retail investor activity. *Working Paper*.

Brunnermeier, Markus K., Stefan Nagel. 2004. Hedge funds and the technology bubble. *The Journal of Finance*, **59**(5) 2013–2040.

Cao, Charles, Yong Chen, William N Goetzmann, Bing Liang. 2016. The role of hedge funds in the security price formation process. *Working Paper*.

Carhart, Mark M. 1997. On persistence in mutual fund performance. *The Journal of Finance*, **52**(1) 57–82.

Chen, Hailiang, Prabuddha De, Yu Hu, Byoung-Hyoun Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, **27**(5) 1367–1403.

Chen, Yong, Bryan Kelly, Wei Wu. 2018. Sophisticated investors and market efficiency: Evidence from a natural experiment. Tech. rep., National Bureau of Economic Research.

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. *Financial Review*, **39**(4) 549–577.

Copeland, Thomas E, Dan Galai. 1983. Information effects on the bid-ask spread. *The Journal of Finance*, **38**(5) 1457–1469.

Da, Zhi, Joseph Engelberg, Pengjie Gao. 2011. In search of attention. *The Journal of Finance*, **66**(5) 1461–1499.

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.

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

Fama, Eugene F., Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics*, **116**(1) 1–22.

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*.

Froot, Kenneth, Namho Kang, Gideon Ozik, Ronnie Sadka. 2017. What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics*.

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.

Griffin, John M, Jin Xu. 2009. How smart are the smart guys? a unique view from hedge fund stock holdings. *The Review of Financial Studies*, **22**(7) 2531–2570.

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.

Huang, Jiekun. 2017. The customer knows best: The investment value of consumer opinions. *Journal of Financial Economics (JFE), Forthcoming*.

Jame, Russell. 2017. Liquidity provision and the cross section of hedge fund returns. *Management Science*.

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

Jegadeesh, Narasimhan, Joshua Livnat. 2006. Revenue surprises and stock returns. *Journal of Accounting and Economics*, **41**(1) 147–171.

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.

Kelly, Bryan, Alexander Ljungqvist. 2012. Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, **25**(5) 1366–1413.

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

Lewellen, Jonathan. 2015. The cross section of expected stock returns. *Critical Finance Review* 4: 1–44.

Zhu, Christina. 2018. Big data as a governance mechanism. *Working Paper*.

Table 1
Summary Statistics

This table reports descriptive statistics for firms with satellite data from January 2010 through December 2017. Growth in car count is the natural log of car count minus the natural log of car count from 12 months prior. Car count is the monthly average of cars observed each day in a firm's parking lots. Excess return is the monthly stock return in excess of the risk free rate. Abn_HF is the current quarter aggregate hedge fund holdings for a stock minus the average aggregate hedge fund holdings over the prior four quarters. Abn_nonHF is calculated analogously for all non-hedge fund institutions. Indiv OIB Vol is the monthly volume of trades initiated as buys by individual investors minus the volume of trades initiated as sells by individual investors divided by the sum of both buy and sell individual-initiated orders. Indiv OIB Trade is calculated analogously for the number of individual trades. Bid-Ask is the monthly average of the daily ask price of a stock minus the bid price of a stock divided by the average of the ask and bid price. Amihud is the absolute return of a stock divided by that stock's dollar volume and multiplied by 1000000. ln(Size) is the log of the market capitalization of the firm, calculated as the share price multiplied by the number of shares outstanding. ln(B/M) is the log book value of equity minus the log market value of equity. ln(Turnover) is the log of the number of shares traded by shares outstanding over the prior 12 months. Return data comes from CRSP, accounting data from Compustat, institutional investor data from 13f filings, individual investor data from TAQ, and satellite data from Orbital Insight.

	Mean	Std Dev	25th	Median	75th
Growth in Car Count	-0.495	7.201	-4.580	-0.702	3.495
Excess Return	0.642	18.425	-10.774	-0.508	10.939
Abn_HF	-0.067	3.538	-1.776	-0.063	1.598
Abn_nonHF	0.152	5.921	-2.417	0.097	2.951
Indiv OIB Vol	-0.019	0.167	-0.100	-0.009	0.068
Indiv OIB Trade	-0.019	0.146	-0.091	-0.01	0.062
Bid-Ask	0.133	0.213	0.028	0.053	0.119
Amihud	0.010	0.0280	0.0001	0.0006	0.004

	Full Sample		Release=1		Release=2		Release=3		Beta-Mode	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ln(Size)	7.559	1.70	8.373	1.82	6.846	1.24	7.382	1.69	7.059	1.29
ln(B/M)	-0.898	0.89	-0.990	0.77	-0.852	1.14	-0.851	0.76	-0.810	0.79
ln(Turnover)	-1.553	0.72	-1.375	0.58	-1.686	0.73	-1.530	0.88	-1.787	0.69

Table 2
Growth in Car Count Portfolio Sorts - Quintiles

This table reports abnormal return estimates for a trading strategy that sorts stocks into quintiles based on car counts from satellite parking lot data. Stocks are sorted at the beginning of every calendar month based on the growth in car count in the prior month. In Panel A (B) stocks are equal (value) weighted within a given portfolio. Portfolios are rebalanced monthly. The "H-L" portfolio represents the difference in returns between stocks with the highest growth in car count (Q5) and stocks with the lowest car count (Q1). Abnormal returns are reported as the return in excess of the market, CAPM alpha, four-factor alpha, and six-factor alpha. The four-factor model includes factors for the excess market return (MKTRF) as well as size (SMB), value (HML), and momentum (UMD) factors of Fama and French (1993) and Carhart (1997). The six-factor model adds the profitability (RMW) and investment (CMA) factors of Fama and French (2015). t-statistics are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Panel A: Equal Weights	Q1(Low)	Q2	Q3	Q4	Q5(High)	**H-L**
Excess Returns	0.26	0.69	0.98*	1.07**	1.86***	**1.60*** **
	(0.46)	(1.27)	(1.76)	(2.06)	(3.13)	**(5.57)**
CAPM Alpha	-1.03***	-0.55	-0.24	-0.13	0.54	**1.58*** **
	(-2.77)	(-1.40)	(-0.60)	(-0.37)	(1.31)	**(5.49)**
Four-Factor Alpha	-0.96***	-0.50	-0.16	-0.08	0.69*	**1.65*** **
	(-2.67)	(-1.28)	(-0.42)	(-0.26)	(1.94)	**(5.70)**
Six-Factor Alpha	-1.15***	-0.72**	-0.41	-0.30	0.45	**1.60*** **
	(-3.38)	(-2.04)	(-1.20)	(1.09)	(1.47)	**(5.41)**

Panel B: Value Weights	Q1(Low)	Q2	Q3	Q4	Q5(High)	**H-L**
Excess Returns	0.74	0.97**	1.23***	1.59***	2.41***	**1.67*** **
	(1.41)	(2.05)	(3.00)	(3.61)	(4.79)	**(4.43)**
CAPM Alpha	-0.38	-0.03	0.32	0.71*	1.27***	**1.65*** **
	(-0.98)	(-0.09)	(0.97)	(1.93)	(3.68)	**(4.19)**
Four-Factor Alpha	-0.47	-0.11	0.17	0.59*	1.15***	**1.62*** **
	(-1.20)	(-0.29)	(0.55)	(1.71)	(3.41)	**(4.41)**
Six-Factor Alpha	-0.57	-0.3	-0.03	0.42	0.99***	**1.56*** **
	(-1.46)	(-0.91)	(-0.10)	(1.35)	(3.07)	**(4.24)**

Table 3
Alternative Data Return Predictability Post-Dissemination
This table reports regressions of the following form:

$$Excess\ Return_{i,t+1} = \alpha + \beta_1 * Growth\ in\ Car\ Count_{i,t} + \beta_2 * Release + \beta_3 * (Growth\ in\ Car\ Count_{i,t} \times Release) + \beta_4 * Cont_{i,t} + \epsilon_{i,t}.$$

$Excess\ Return_{i,t+1}$ is the return on stock i in month t in excess of the risk-free rate. $Growth\ in\ Car\ Count_{i,t}$ is the natural log of car count for firm i in month t minus the natural log of car count from 12 months prior. *Release* is an indicator variable equal to one in months when a stock's satellite data is available to market participants and zero otherwise. $Cont_{i,t}$ is a vector of the following controls: the log market value of equity at the end of the prior month, the log of book-to-market at the end of the prior month, the stock return from month $t-12$ to $t-2$, the log of growth in shares outstanding from month $t-36$ to month $t-1$, the change in net working capital minus depreciation in the prior fiscal year, the return on assets in the prior fiscal year, and the log of growth in total assets in the prior fiscal year. Columns 3 through 8 split the sample at the median of market equity, bid-ask spread, and amihud illiquidity ratio (defined in Table 1). t-statistics, calculated using standard errors clustered by firm and year-month, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Dependent Variable: Excess Stock Return					
	Full Sample	Full Sample	Large Firms	Low Spread	Low Amihud	Small Firms	High Spread	High Amihud
Growth in Car Count	**0.033*****	**0.029*****	**0.049*****	**0.052*****	**0.050*****	**0.018***	**0.013**	**0.017***
	(3.860)	**(3.312)**	**(4.000)**	**(4.148)**	**(3.746)**	**(1.985)**	**(1.364)**	**(1.913)**
Growth in Car Count × Release		**0.028**	**0.012**	**0.002**	**0.018**	**0.026**	**0.042**	**0.023**
		(0.870)	**(0.221)**	**(0.037)**	**(0.317)**	**(0.698)**	**(1.095)**	**(0.608)**
Release		0.035	1.136**	0.368	0.785	-1.555	-0.333	-1.586
		(0.067)	(2.152)	(0.753)	(1.583)	(-1.350)	(-0.269)	(-1.353)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YM FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,679	10,679	5,388	5,458	5,534	5,290	5,162	5,086
R-squared	0.195	0.195	0.259	0.265	0.250	0.199	0.202	0.203

Table 4
Growth in Car Count and Institutional Holdings
This table reports difference-in-differences regressions of the following form:

$$Abnormal\ Holdings_{i,t} = \beta_1 * Growth\ in\ Car\ Count_{i,t} + \beta_2 * Release + \ \beta_3 * (Growth\ in\ Car\ Count_{i,t} \times Release) + \beta_4 * Cont_{i,t} + \epsilon_{i,t}$$

$Abnormal\ Holdings_{i,t}$ is the level of abnormal holdings for either hedge funds or non-hedge funds in stock i for quarter t. $Cont_{i,t}$ is a vector of the following controls: prior quarter returns adjusted for the returns of the CRSP value-weighted index, book-to-market, size, and turnover. All other variables are defined in previous tables. t-statistics, calculated using standard errors clustered by firm and year-quarter, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Abnormal HF Holdings			Abnormal non-HF Holdings		
Growth in Car Count	-0.003	-0.008	-0.003	-0.017	0.003	-0.008
	(-0.351)	(-1.036)	(-0.305)	(-0.702)	(0.145)	(-0.324)
Growth in Car Count × Release	**0.057****	**0.081****	**0.059***	**0.006**	**-0.005**	**-0.014**
	(2.037)	**(2.423)**	**(1.882)**	**(0.074)**	**(-0.066)**	**(-0.192)**
Release	0.505	0.185	0.285	-0.422	-0.662	-0.215
	(1.682)	(0.682)	(0.820)	(-0.702)	(-1.252)	(-0.293)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
YQ FE	Yes	No	Yes	Yes	No	Yes
Observations	3,327	3,327	3,327	3,327	3,327	3,327
R-squared	0.028	0.108	0.124	0.026	0.132	0.147

Table 5

The Effect of Alternative Data on Institutional Investor Profitability

This table reports difference-in-differences regressions of the following form:

$$Adjusted\ Stock\ Returns_{i,t+1} = \beta_1 * Abnormal\ Holdings_{i,t} + \beta_2 * Release + \ \beta_3 * (Abnormal\ Holdings_{i,t} \times Release) + \beta_4 * Cont_{i,t} + \epsilon_{i,t}$$

$Cont_{i,t}$ is a vector of the following controls: prior quarter returns adjusted for the returns of the CRSP value-weighted index, book-to-market, size, and turnover. All other variables are defined in previous tables. t-statistics, calculated using standard errors clustered by firm and year-quarter, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Adjusted Stock Returns					
Abnormal HF	-0.001	-0.003**	-0.002*			
	(-1.492)	(-2.485)	(-1.812)			
Abnormal HF × Release	**0.004***	**0.006****	**0.006****			
	(1.847)	**(2.034)**	**(2.095)**			
Abnormal non-HF				-0.000	-0.000	-0.000
				(-0.747)	(-0.610)	(-0.610)
Abnormal non-HF × Release				**0.000**	**0.001**	**0.001**
				(0.086)	**(0.470)**	**(0.470)**
Release	0.008	-0.064**	0.000	0.007	0.000	0.000
	(0.577)	(-2.217)	(0.021)	(0.552)	(0.029)	(0.029)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
YQ FE	Yes	No	Yes	Yes	No	Yes
Observations	3,279	3,279	3,279	3,279	3,279	3,279
R-squared	0.202	0.078	0.252	0.201	0.251	0.251

Table 6
Growth in Car Count and Individual Investor Order Demand

This table reports difference-in-differences regressions of the following form:

$$Individual\ OIB_{i,t} = \beta_1 * Growth\ in\ Car\ Count_{i,t} + \beta_2 * Release + \ \beta_3 * (Growth\ in\ Car\ Count_{i,t} \times Release) + \beta_4 * Cont_{i,t} + \epsilon_{i,t}$$

$Individual\ OIB_{i,t}$ refers to either the volume or trade order imbalance for individual investors in stock i in month t. $Cont_{i,t}$ is a vector of the following controls: the log market value of equity at the end of the prior month, the log of book-to-market at the end of the prior month, the stock return from month $t-12$ to $t-2$, and the log of stock turnover in the prior month. All other variables are defined in previous tables. t-statistics, calculated using standard errors clustered by firm and year-month, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:		Individual OIB Volume			Individual OIB Trades	
Growth in Car Count	0.0004***	0.0004**	0.0004***	0.0005***	0.0004***	0.0005***
	(2.840)	(2.609)	(2.899)	(4.366)	(3.411)	(3.739)
Growth in Car Count x Release	**-0.056***	**-0.053**	**-0.064***	**-0.073****	**-0.064***	**-0.072****
	(-1.741)	**(-1.456)**	**(-1.786)**	**(-2.195)**	**(-1.953)**	**(-2.207)**
Release	0.002	0.010**	0.003	0.003	0.003	0.001
	(0.4634)	(2.579)	(0.583)	(0.585)	(0.588)	(0.164)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
YM FE	Yes	No	Yes	Yes	No	Yes
Observations	10,593	10,593	10,593	10,593	10,593	10,593
R-squared	0.031	0.035	0.055	0.054	0.065	0.097

Table 7
The Effect of Alternative Data on Individual Investor Profitability
This table reports difference-in-differences regressions of the following form:

$$CAR_{i,t,d\ to\ d+3} = \beta_1 * Individual\ OIB_{i,t,d-3\ to\ d-1} + \beta_2 * Release + \ \beta_3 * (Individual\ OIB_{i,t,d-3\ to\ d-1} \times Release) + \beta_4 * Cont_{i,t} + \epsilon_{i,t} \quad (9)$$

$CAR_{i,t,d\ to\ d+3}$ is the abnormal announcement return for stock $i's$ earnings announcement for quarter t from the day of the announcement until three days after. $Individual\ OIB_{i,t,d-3\ to\ d-1}$ is the individual order imbalance for either volume or trades for stock i for the three days prior to the earnings announcement for quarter t. $Cont_{i,t}$ is a vector of the following controls: the standardized unexpected earnings for the quarter, prior quarter returns adjusted for the returns of the CRSP value-weighted index, book-to-market, size, and turnover. All other variables are defined in previous tables. t-statistics, calculated using standard errors clustered by firm and year-quarter, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Cumulative Abnormal Returns					
Individual OIB Vol	0.007	0.004	0.002			
	(1.024)	(0.613)	(0.382)			
Individual OIB Vol x Release	**-0.053****	**-0.047***	**-0.047***			
	(-2.105)	**(-1.725)**	**(-1.662)**			
Individual OIB Trad				0.017*	0.010	0.008
				(1.793)	(1.024)	(0.807)
Individual OIB Trad x Release				**-0.112****	**-0.115***	**-0.110***
				(-2.282)	**(-1.959)**	**(-1.847)**
Release	0.009**	0.006	0.010*	0.011**	0.007*	0.012**
	(2.170)	(1.310)	(1.920)	(2.375)	(1.849)	(2.301)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	No	Yes	Yes
YQ FE	Yes	No	Yes	Yes	No	Yes
Observations	3,537	3,537	3,537	3,537	3,537	3,537
R-squared	0.023	0.078	0.089	0.024	0.080	0.090

Table 8

The Effect of Alternative Data on Firm Liquidity

This table reports regressions of the following form:

$$Liquidity_{i,t} = \beta_1 * Release + \ \beta_2 * Cont_{i,t} + \epsilon_{i,t}$$

$Liquidity_{i,t}$ is measured using the log of the amihud illiquidity ratio and the bid-ask spread. $Cont_{i,t}$ is a vector of the following controls: the log market value of equity at the end of the prior month, the log of book-to-market at the end of the prior month, the stock return from month $t-12$ to $t-2$, and the log of stock turnover in the prior month. All other variables are defined in previous tables. t-statistics, calculated using standard errors clustered by firm and year-month, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)
		ln(Amihud)			Bid-ask Spread	
Release	**0.004*** **	**0.003** **	**0.002*** **	**0.047*** **	**0.040*** **	**0.005**
	(7.002)	**(2.524)**	**(2.952)**	**(11.657)**	**(4.077)**	**(1.268)**
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	No	Yes	No
YM FE	No	No	Yes	No	No	Yes
Observations	12,628	12,628	12,628	12,628	12,628	12,628
R-squared	0.380	0.835	0.385	0.445	0.835	0.459

Fig. 1
Sample Satellite Image
This figure shows an example of how satellite images of parking lots are converted into car counts. This example is of a Wal-Mart store in Arizona. The circles identify each car. Only cars inside the shaded region are counted towards Wal-Mart's total car count. Source: Orbital Insight.

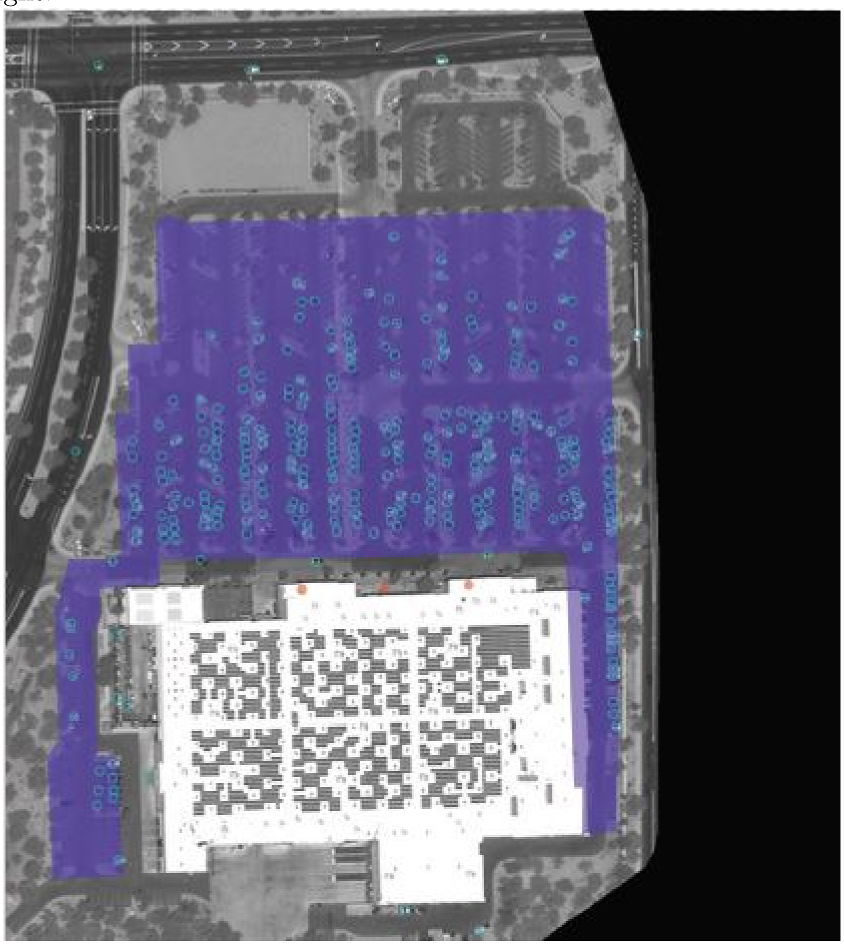
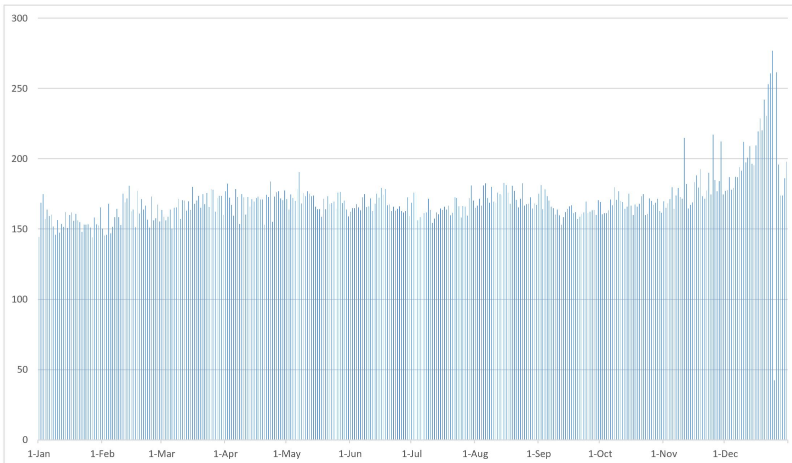


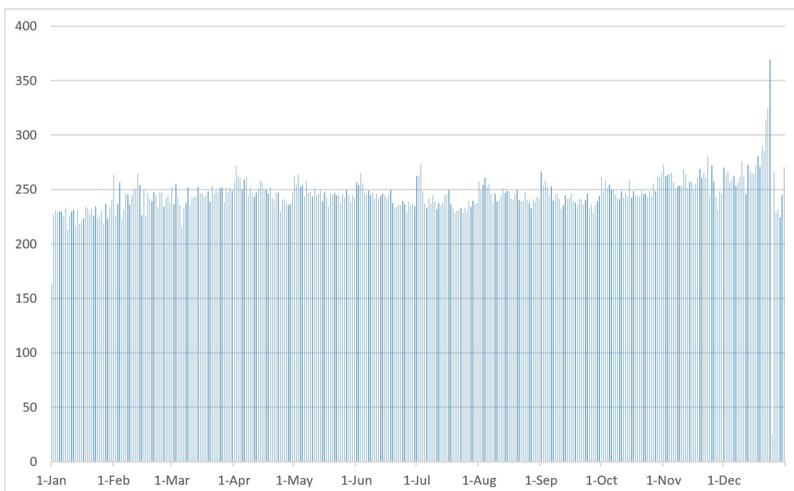
Fig. 2
Daily Calendar Car Count

This figure shows the average parking lot car counts for each calendar day. Panel A reports averages for all stocks in the sample, Panel B reports Wal-Mart's car counts, and Panel C reports Home Depot's car counts.

(a) All Stocks



(b) Wal-Mart



(c) Home Depot

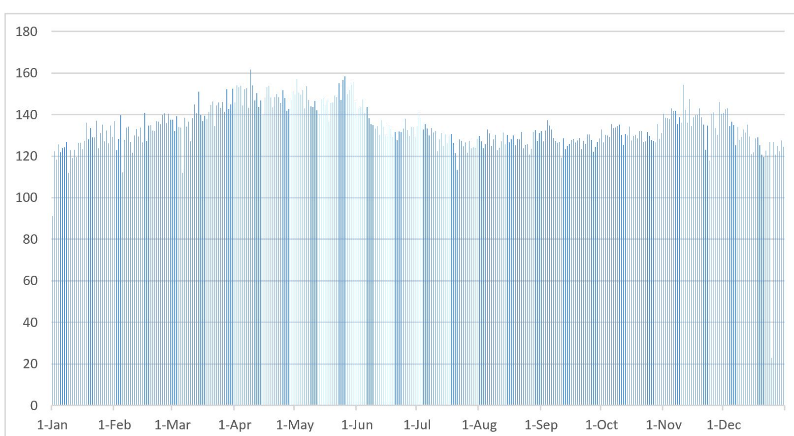
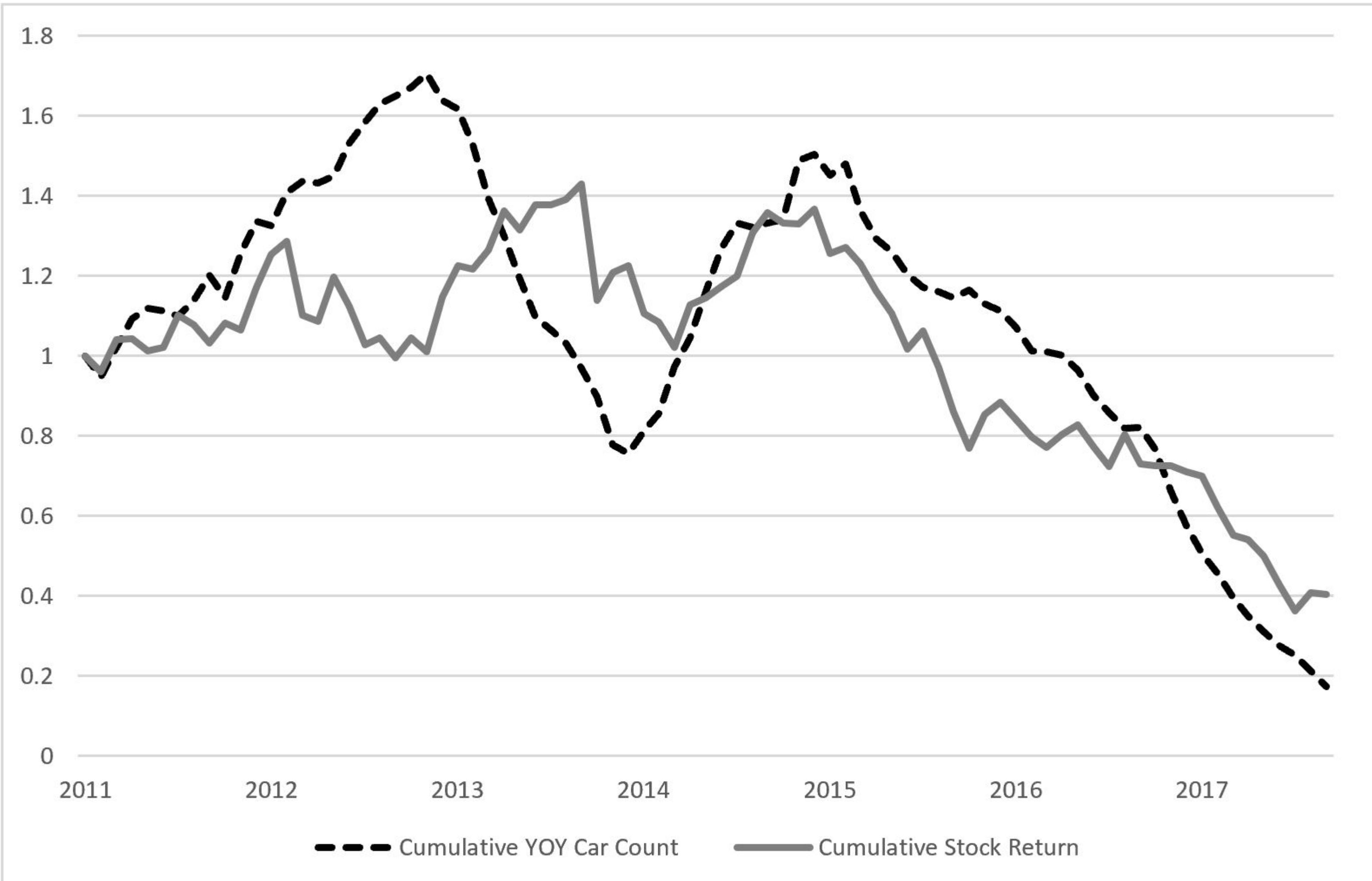


Fig. 3

Growth in Car Count vs. Stock Return Example: Bed Bath and Beyond

This figure compares cumulative year-over-year growth in car traffic with cumulative stock returns for Bed Bath and Beyond from 2011 to 2017. Car count data is from Orbital Insight and stock return data is from CRSP.



Appendix

A.1. Predicting Firm Fundamentals with Satellite Data

I follow the literature to develop measures for cash flow surprises. I follow Jegadeesh and Livnat (2006) and Froot, Kang, Ozik, and Sadka (2017) to construct the standardized unexpected revenue (SUR), which assumes revenue follows a seasonal random walk with a drift. Specifically, SUR for firm i in quarter q is defined using the following equation:

$$SUR_{i,q} = \frac{(Rev_{i,q} - Rev_{i,q-4}) - r_{i,q}}{\sigma_{i,q}}, \tag{10}$$

where $\mathrm{r}_{i,q}$ and $\sigma_{i,q}$ are the average and standard deviation, respectively, of $(Rev_{i,q} - Rev_{i,q-4})$ for the prior eight quarters. The second measure I use for cash flow is standardized unexpected earnings (SUE), defined for firm i in quarter q as follows:

$$SUE_{i,q} = \frac{A(EPS_{i,q}) - F(EPS_{i,q})}{P_{i,q}}, \tag{11}$$

where $A(EPS_{i,q})$ is the actual earnings per share on the announcement date, $F(EPS_{i,q})$ is the average analyst forecasted earnings per share, and $P_{i,q}$ is the stock price at the end of the quarter. Finally, I calculate the cumulative abnormal return (CAR) for the earnings announcement as the stock return in excess over the market from one day before the earnings announcement date until three days after the announcement date. Because earnings are often reported after trading hours (Berkman and Truong, 2009), I use a five day window announcement in order to ensure that the market's complete reaction to earnings is measured. Results are robust to using a three day or four day window.

Table A.1
Firm Fundamentals and Announcement Returns

This table presents regressions of SUE, SUR, and CAR on growth in car count as well as control variables. Growth in Car Count$_Q$ is calculated as the log of average car count in the current quarter minus the log of average car count over the prior four quarters. SUR is calculated as $[(Rev_{i,t} - Rev_{i,t-4}) - r_{i,t}]/\sigma_{i,t}$ where $r_{i,t}$ and $\sigma_{i,t}$ are the average and standard deviation, respectively, of $(Rev_{i,t} - Rev_{i,t-4})$ over the prior eight quarters and $Rev_{i,t}$ is the revenue for stock i in quarter t. The control variables are the log of market equity in *t-1*, the log of book-to-market in *t-1*, the cumulative stock return from 30 days to 3 days before the earnings announcement date, and the log of turnover over the prior 12 months, and the standard deviation of analysts' EPS forecasts. t-statistics, calculated using standard errors clustered by firm and year-quarter, are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

	(1) SUR	(2) SUE	(3) CAR
Growth in Car Count	1.793***	0.021**	0.088***
	(4.814)	(2.228)	(3.515)
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
YQ FE	Yes	Yes	Yes
Observations	3,419	3,426	3,424
R-squared	0.153	0.217	0.094