Displaced by Big Data: Evidence from Active Fund Managers

Maxime Bonelli and Thierry Foucault*

February 12, 2024

ABSTRACT

Big data allows active asset managers to find new trading signals but doing so requires new skills. Thus, it can reduce the ability of asset managers lacking these skills to produce superior returns. Consistent with this hypothesis, we find that the release of satellite imagery data tracking firms' parking lots reduces active mutual funds' stock picking abilities in stocks covered by this data. This decline is stronger for funds more likely to rely on traditional sources of expertise, leading them to divest from covered stocks. These results suggest that big data has the potential to displace high-skill workers in finance.

Keywords: Alternative data, active mutual funds, stock-picking skills.

JEL classification: G11, G14, G23.

*Bonelli is affiliated with London Business School and can be reached at mbonelli@london.edu. Foucault is affiliated with HEC Paris can be reached at foucault@hec.fr. We thank Laurent Fresard, Nicholas Hirschey, Augustin Landier, Marcus Painter, Russ Wermers and Tianhao Yao, as well as conference and seminar participants at the NBER Big Data and Securities Markets Fall 2023, the European Winter Finance Conference 2024, Bocconi University, London Business School, Toulouse Business School, UC Irvine, and University of Cologne. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101018214).

"As part of the restructuring, seven of BlackRock's 53 stock pickers are expected to step down from their funds [...] And since last year, BlackRock's dyed-in-the-wool stock pickers have worked in the same division as its quants. These managers [...] might buy (or sell) Wallmart's stock on the basis of a satellite feed that reveals how many cars are in its parking lot as opposed to an insight gleaned from the innards of the retailers' balance sheet.", At BlackRock, Machines are rising over managers to pick stocks, The New-York Times, 2017.

I Introduction

The finance industry is experiencing a significant transformation due to the advent of big data and artificial intelligence (AI). Coping with this evolution requires new skills and could therefore reduce the value of previously sought skills (Acemoglu et al., 2022). A case in point are active asset managers.[1] Indeed, the emergence of alternative data sources, such as social media, point-of-sale data, sensors, and satellite imagery, offers the potential for asset managers to gain more precise insights into stock returns and make better investment decisions.[2] However, effectively leveraging alternative data requires new skills (such as expertise in computer and data science) and significant investments in technology and data sets. Traditional asset managers, who rely on specialized industry knowledge and human judgement, are therefore at risk to be displaced by a new breed of asset managers (the "quants") if the latter makes the skills of the former less valuable. In this paper, we study whether this is the case.

To do so, we study how the availability of new alternative data affects asset managers' stock picking ability. Intuitively, alternative datasets should enable asset managers to obtain more precise signals about stock returns, provided that they buy these datasets and have

[1]See *"Make way for the robot stock pickers"*, Financial Times, June 26, 2016, available at https://www.ft.com/content/84bb5c72-37a9-11e6-9a05-82a9b15a8ee7, *"Fund managers deny AI threatens jobs"*, Financial Times, August 14, 2017, available at https://www.ft.com/content/bd26af40-7dd9-11e7-ab01-a13271d1ee9c, and *"At BlackRock, Machines are rising over managers to pick stocks", The New-York Times, March 28, 2017, available at https://www.nytimes.com/2017/03/28/business/dealbook/blackrock-actively-managed-funds-computer-models.html.*

[2]See JP Morgan (2019) for a catalog of more than 500 alternative data providers.

the skills for exploiting them. As these managers trade on these signals, stock prices become more informative, reducing the return on producing private information for asset managers without these skills (see Section II for a more detailed discussion of this mechanism). If this hypothesis is correct, the stock picking ability of traditional asset managers should decline when new alternative data becomes available for the stocks they have developed specific expertise in.

To test this prediction, we use the staggered introduction of new alternative data for retailers' stock, namely daily store-level car counts from satellite imagery by RS Metrics (an alternative data provider). We use this introduction as a shock to the availability of alternative data for these stocks, and study its effect on the performance ("stock picking ability") of active mutual funds holding these stocks. This shock is well-suited for our tests because satellite imagery is one important type of alternative data, used by asset managers.[3] Moreover, it has predictive power for firms' future cash-flows or stock returns (see Zhu, 2019; Kang et al., 2021; Katona et al., 2023) and it enhances price informativeness (Zhu, 2019). Last, extracting information from the store-level data considered in our tests requires quantitative skills (as the dataset contains several millions of data points).

Our sample comprises "covered" stocks, i.e., stocks for which store level car counts from satellite imagery become available over the period 2009-2017, and control stocks, i.e., stocks for which such data is not available. Moreover, we observe portfolio holdings of about 4,000 different active mutual funds, at the beginning of each quarter, between 2009 and 2021. With this data, in each quarter, we measure a fund's stock picking ability in a given stock by *"Picking"*, the product between (i) the stock's subsequent idiosyncratic return at various horizons and (ii) the deviation of its weight in the fund's portfolio from its weight in the market portfolio.[4] Intuitively, this measure captures a fund's ability to tilt its stock holding

[3] See, for instance, "How satellite imagery is helping hedge funds outperform" available at https://internationalbanker.com/brokerage/how-satellite-imagery-is-helping-hedge-funds-outperform/, "Stock Picks From Space" available at https://www.theatlantic.com/magazine/archive/2019/05/stock-value-satellite-images-investing, and Blackrock's podcast "How geospatial data can inform investment decisions" available at https://www.blackrock.com/us/individual/podcasts/the-bid/geospatial-data.

[4] A similar measure, but a the fund level, is used by Kacperczyk et al. (2014). See also Grinblatt and

in the direction of future return adjusted for systematic risk. Our main tests study whether funds' stock picking ability in treated stocks declines after these stocks become covered using a difference-in-differences specification, with stock fixed effects and fund-quarter fixed effects (which control for changes in variables affecting the overall stock picking ability of a fund).

As predicted, we find that funds' stock picking ability in a given stock drops after it becomes "covered", i.e., after alternative data (from satellite image providers) becomes available for this stock. This drop is statistically and economically significant at all horizons (ranging from 1 to 4 quarters). For instance, at the 1-year horizon, *"Picking"* has a mean and median of zero, with a standard deviation of 0.36 percentage points (p.p.). At this horizon, we find that *"Picking"* drops by 0.11 p.p. for covered stocks in our sample relative to other stocks, that is, about one-third of its standard deviation.

This drop holds after including fund-quarter and fund-stock fixed effects (which ensure that the drop in *"Picking"* is specifically attributed to the same fund holding a stock both before and after coverage initiation), and is robust to various changes in the specification of our main tests, such as, for instance, the definition of a fund's stock picking ability or the set of control stocks. In particular, we consider a narrower set of control stocks matched on industry and market capitalization and find that the results are qualitatively similar. Moreover, we show that the decline in funds' stock picking ability in a covered stock relative to control stocks arises only *after* the stock becomes covered, specifically about one year after coverage initiation.

This negative effect is the average effect across all funds holding covered stocks in our sample, that is, those who don't buy the data considered in our tests and those who do (we do not observe which funds buy the data from RS Metrics). Our hypothesis implies that this effect should be stronger for the former, which are more likely to be funds whose expertise in covered stocks relies on traditional methods for obtaining private signals (see Section II). We test this implication by considering three different measures of expertise: (i) a fund's stock

Titman (1993a), Daniel et al. (1997) and Jiang and Zheng (2018) for related measures of a fund's skill.

picking ability in a given stock prior to its coverage (higher ability signals more expertise), (ii) a fund's focus on industries covered by RS Metrics (Kacperczyk et al., 2005, shows that funds that concentrate their holdings in a given industry are more likely to possess private information), and (iii) a fund's geographical proximity to the main stores or headquarters of covered firms (funds geographically close to a firm's operations are more likely to have private information; see Coval and Moskowitz, 2001). In all cases, we find that the drop in "*Picking*" after coverage initiation is significantly stronger for "experts". For instance, this drop is more than twice larger for funds that have a high stock picking ability in covered stocks before these stocks become covered.

Thus, consistent with our hypothesis, the introduction of alternative data reduces the value created by asset managers who rely on traditional methods to obtain private information ("experts"), creating a risk of displacement. We then study how experts respond to this risk. We first show that they cut their investment in covered stocks (a form of displacement) both at (a) the intensive margin by reducing the weight of covered stocks in their portfolios and (b) the extensive margin by divesting from covered stocks. Specifically, after coverage, the number of funds holding a covered stock drops by about 20%.

We then provide evidence that funds reallocate their portfolios to "peer stocks" in which they can exploit their expertise (based on industry or geography) but which are never covered by the data provider considered in our tests. Specifically, we show that the number of funds holding a stock increases by 10% after one of its geographical or industry peer becomes covered. Overall, these findings suggest that experts respond to the threat of displacement by shifting their investments to stocks where their skills remain valuable.[5] One implication is that one might see greater segmentation in the way information is produced between stocks covered by alternative data (which tends to be large capitalization stocks) and stocks that are not.

[5]These findings echo those regarding securities analysts in Grennan and Michaely (2020). They find that analysts who follow stocks that are the most exposed to the production of information on social medias (measured by social media postings about these stocks) are more likely to reallocate coverage toward less exposed stocks or exit the profession.

According to the mechanism driving our hypothesis, the drop in funds' stock picking ability in covered stocks stems from an increase in the informativeness of the price of these stocks once they become covered (see Section II). Thus, to buttress the interpretation of previous findings, we study whether this is the case, using two (inverse) measures of stock price informativeness: (i) the absolute cumulative abnormal return (Fama et al., 1969; Ball and Brown, 1968) and (ii) the price jump ratio (Weller, 2018) following earning announcements. When these measures are higher, announcements move prices (and therefore investors' beliefs) more, which means that stock prices ahead of these announcements contain less information. With either measure, we find that, as predicted, price informativeness significantly increases for the covered stocks after coverage initiation, which is in line with Zhu (2019).

Our paper contributes to the quickly growing literature on alternative data in financial markets. Several papers show that different types of alternative data (e.g., social media data, geolocation data, employee satisfaction data, or satellite images) contains information on firms' fundamentals and stock returns.[6] Other papers focus on the effects of alternative data on analysts' forecasts informativeness (e.g., Dessaint et al., 2022) and on measures of market quality such as stock price informativeness (Zhu, 2019; Grennan and Michaely, 2021) or liquidity (Katona et al., 2023). To our knowledge, our paper is the first to show that the introduction of new alternative data reduces funds' stock picking ability.[7] This result is in line with Kang (2022), who finds that acquiring private information via participation to syndicated loans for retailers becomes less profitable after retailers become covered by satellite data providers.

Other papers use the introduction of satellite imagery as a shock on the amount of

[6] Dessaint et al. (2022) survey 26 academic papers showing that different types of alternative data has predictive power for future returns. See Table A.1 in their online appendix for the list of these papers and a summary of their main findings.

[7] Zhao (2021) shows that a regulatory change that facilitates the analysis of unstructured corporate information (the adoption of a new format for firms' regulatory filings in the US: the SEC's XBRL mandate) leads to an improvement in measures of the stock picking ability of active funds with more financial analysts relative to other funds and a drop in the the stock picking ability of funds with more IT specialists (which are more likely to be quant funds). In contrast to our analysis, the regulatory shock in Zhao (2021) does not change the amount of data available to investors (it just reduces the cost of data processing).

alternative data to investors (e.g., Zhu, 2019; Kang et al., 2021; Katona et al., 2023; Gerken and Painter, 2022; Liu et al., 2023). In particular, Katona et al. (2023) use data from the same data provider considered in our study. They show that a stock becomes less liquid when it becomes covered and argue that this finding reflects greater informational asymmetries between sophisticated and unsophisticated investors.[8] Instead, we show that the introduction of alternative data can also hurt sophisticated investors, by reducing the ability of funds lacking the skills to use this data to exploit their own private information.

Our paper is also related to the literature on active asset managers' skill, defined by an asset manager's ability to pick stocks (i.e., trade on profitable private information), as in, for instance, Kacperczyk et al. (2014). This literature suggests that the average asset manager has no skill (see Fama and French, 2010). However, there is heterogeneity among asset managers and some display significant ability to pick stocks (Wermers, 2000; Kosowski et al., 2006; Kacperczyk et al., 2014; Jiang and Zheng, 2018). Our results suggest that, like in any other industry, technological innovations (in our case, innovation in the way information is obtained and processed) can displace top performers. In fact, we find that these are the funds with the highest skill before stocks become covered in our sample that experience the largest drop in their picking ability. This implies that one needs to better understand how funds invest in information technologies to understand the source of their skills and its persistence.

Last, our paper is related to the literature on the labor effects of artificial intelligence (see, for instance, Acemoglu and Restrepo, 2018; Acemoglu et al., 2022; Autor et al., 2022). Several papers suggest that higher-skill jobs face greater exposure to AI capabilities (Kanazawa et al., 2022; Brynjolfsson et al., 2023; Eloundou et al., 2023; Agrawal et al., 2023; Hui et al., 2023;

[8]These authors also show how store-level car counts, from satellite imagery, can be used to design profitable trading strategies ahead of earnings announcements and provide evidence of increased informed trading (e.g., short selling activity and decrease in liquidity) after a stock becomes covered. Other researchers study the effects of information in satellite images data on price informativeness (Zhu, 2019) or firms' disclosure (Liu et al., 2023). Gerken and Painter (2022) use store-level car counts data (from satellite imagery) to show that security analysts' forecasts are influenced by information specific to their geographical location. Kang et al. (2021) use a similar approach to show that institutional investors rely on local information to form their portfolios.

Dell'Acqua et al., 2023).[9] Using online job vacancies on Burning Glass from 2010 to 2022, Acemoglu et al. (2022) shows that occupations in the finance/insurance sector are the most impacted by AI and that greater AI exposure has a negative effect on the skills previously sought in posted job vacancies (and a positive effect on new skills related to AI). In the context of active asset management, a pre-requisite for this effect to be at play is that the emergence of alternative data leads to an erosion of traditional asset managers' performance because extracting information from big data requires new skills. Our results are consistent with this possibility.

II Hypotheses Development

Alternative data contains information useful for predicting future cash-flows and returns (see Footnote 6). However, using such data requires a specific forecasting style (based on quantitative data analysis rather than human judgement) and skills (knowledge of techniques from data science). Only asset managers possessing these skills can therefore use new alternative data available for a stock to improve the precision of their signals about this stock. In trading on these signals, they should make the stock price more informative. We predict that, as a result, asset managers lacking the skills required to use alternative data should experience a *decline* in the value of their own private signals.

In the Internet Appendix (Section I.1), we derive this prediction using a simplified version of Dugast and Foucault (2022), which considers a noisy rational expectations model (similar to Verrecchia, 1982) with two types of active asset managers: "experts" and "data miners". Both types receive informative signals about the future payoff of a stock and trade this stock with noise traders. Experts and data miners use different technologies to produce these signals. The precision of experts' signals is fixed (e.g., by education or industry expertise developed over time).[10] In contrast, data miners can increase the precision of their signals

[9]See also "*Here's what we know about generative AI's impact on white-collar work*", Financial Times, November 10, 2023 available at https://www.ft.com/content/b2928076-5c52-43e9-8872-08fda2aa2fcf

[10]The literature has identified a few fixed characteristics that determine an asset manager's ability to outperform: education (Chevalier and Ellison, 1999), geographical proximity to firms (Coval and Moskowitz, 2001), educational network (Cohen et al., 2008), and industry concentration (Kacperczyk et al., 2005).

when new datasets become available (provided they buy those) because they have the skills to extract information from these datasets (which require manipulating a large amount of data and using, for instance, tools from data science). In practice, data miners represent funds that rely on quantitative analysis of massive amount of data to generate their signals, so-called quant funds. In contrast, experts correspond to discretionary funds since these are more likely to rely on human judgement and expertise to generate their signals.[11]

Let $s(\tau_{ex})$ be the signal of an expert, where τ_{ex} is the precision of her signal, and let $w(s(\tau_{ex}))$ be the weight of the stock in the asset manager's portfolio (the fraction of her AUM invested in this stock). Intuitively, this weight depends on the realization of the signal as a higher signal realization predicts a higher payoff for the asset. The expert's realized excess return on her position in the stock is:

$$R(s(\tau_{ex})) = w(s(\tau_{ex})) \times R^e, \tag{1}$$

where R^e is the excess return on the stock. An informative signal enables the expert to tilt her holdings in the stock in the direction of its future excess return, so that $w(s(\tau_{ex}))$ and R^e co-vary positively. Thus, an expert's expected excess return, that is, $\bar{R}(\tau_{ex}) \equiv \mathrm{E}(R(s(\tau_{ex}))$ is strictly positive if her signal is informative. Following Dugast and Foucault (2022), we show in the Internet Appendix (see eq.(IA.11)) that, in equilibrium, it increases with the precision of the expert's signal and decreases with the informativeness of the stock price. Thus, $\bar{R}(\tau_{ex})$ is a measure of an expert's stock picking ability (everything else equal, the higher is τ_{ex}, the higher the average return of the expert in the stock).

When the precision of data miners' signals increases, price informativeness increases in equilibrium (see eq.(IA.8) in the Internet Appendix) and therefore each expert's expected

[11]For instance, Narang (2013) (p.XV) describes the differences between quant funds and discretionary funds as follows: *"The differences between a quant strategy and a discretionary strategy can be seen in how the strategy is created [...]. By carefully researching their strategies, quants are able to assess their ideas in the same way scientists test their theories. Furthermore, by utilizing a computerized, systematic, implementation, quants eliminate the arbitrariness that pervades so many discretionary trading strategies.*. See also *"The quants run Wall Street now"*, Wall Street Journal, May 21, 2017 available at https://www.wsj.com/articles/the-quants-run-wall-street-now-1495389108.

returns decline. Hence, if as we posit, experts cannot exploit new alternative data, their stock picking ability, $\bar{R}(\tau_{ex})$, should drop when this stock becomes covered by new alternative data. More generally, this prediction should hold for all funds not buying the new data (either because they cannot use them or decide not to buy them) should drop.

Our main goal is to test this prediction. One difficulty for our tests is that we do not directly observe whether a fund buys the new alternative data considered in our tests or not. However, in practice, we expect the fraction of active funds buying the data to be small since the number of quant funds is relatively small and only a fraction of these funds will buy a new alternative dataset.[12] In this case, the average expected excess return *across* all active funds holding a given stock should decrease when new alternative data becomes available for this stock.

To show this in a simple way, we suppose that all asset managers within a given group (e.g., experts) have signals of the same precision.[13] In this case, the average stock picking ability across experts and data miners, $\bar{R}$, is:

$$\bar{R} = (1-\mu)\bar{R}(\tau_{ex}) + \mu\bar{R}(\tau_{dm}), \tag{2}$$

where τ_{dm} is the precision of data miners' signals, $\bar{R}(s(\tau_{dm}))$ is their stock picking ability (defined as for experts but with τ_{dm} instead of τ_{ex}) and μ is the fraction of capital controlled by data miners. The average stock picking ability across funds, $\bar{R}$, is always strictly positive in equilibrium because active asset managers make trading profits at the expense of noise traders. In other words, trading is not a zero sum game among active asset managers because of the presence of noise traders.

Now, consider an increase in the precision of the signal obtained by data miners due, for instance, to the availability of new data and suppose first that all data miners buy the data.

[12] For instance, Abis (2022) (in particular Figure 6) shows that, as of 2018, quant funds account for more less 20% of all active U.S. equity mutual funds and controls less than 10% of assets managed by active mutual funds.

[13] Dugast and Foucault (2022) consider the more general case in which experts and data miners have signals of heterogeneous precisions. This does not affect the conclusions.

This increase reduces experts' stock picking ability, $\bar{R}(\tau_{ex})$, for reasons previously explained, while it increases $\bar{R}(\tau_{dm})$ (as otherwise data miners would not buy the data). However, as shown in the internet appendix, for sufficiently small values of μ, the aggregate value of this drop ($(1-\mu)\bar{R}(\tau_{ex})$) more than offsets the aggregate increase in data miners' average stock picking ability ($\mu\bar{R}(\tau_{dm})$). As a result, the stock picking ability of the average fund, $\bar{R}$, drops.

If only a fraction of data miners experience an increase in the precision of their signal (e.g., because only a few buy the new data), the effect is even stronger (as if μ was smaller) because the data miners who do not buy the data are also negatively affected. Using a different modeling approach, Stambaugh (2020) also considers a shift in the precision of signals for a subset of active managers and obtains a similar conclusion (see Section 4.2.3 and Figure 3 in Stambaugh, 2020), which shows the robustness of this implication.

Hence, we test our main hypothesis in two steps. We first consider the effect of the availability of new alternative data (specifically: store-level car counts based on satellite imagery for U.S retailers) on the average stock picking ability of all actively managed mutual funds (see Sections IV.A and IV.B). Then, in a second step, we test whether effects are stronger (more negative) for funds that are more likely to correspond to experts (Section IV.C).

III Data and Measurements

In this section, we describe the data we use for our tests and how we measure empirically active mutual funds' stock picking ability.

A Data

Our first data set is from the Center for Research on Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database. This database provides comprehensive information about funds, such as their returns and size (total net assets). We focus our analysis on US domestic equity funds from January 2009 to December 2021, for which the holdings data (described below) are most complete and reliable.[14] We exclude index funds, ETFs and money market funds, and we aggregate all share classes of the same fund since they have the same portfolio composition.[15] To address the possibility of incubation bias, we further exclude observations if the year of the observation is prior to the reported fund starting year or if the name of the fund is missing. From this database, we also obtain the contact information of each fund to determine its geographical location.

Our second data set is from the CRSP Mutual Fund Holdings database. This database provides stock holdings of mutual fund portfolios, collected both from mandatory SEC reports by mutual funds and voluntary reports. We use portfolio holdings disclosed by funds at the end of each quarter.[16] We further filter our sample by excluding funds that do not hold more than 80% of equities as well as funds that in the previous month had less than $5 million of assets under management (as in Kacperczyk et al., 2014). Our final sample features 3,962 funds holding 9,781 distinct stocks.

We gather accounting variables of the companies held by mutual funds from Compustat. Specifically, we collect market capitalization, book-to-market ratio, total assets, sales, and total debt of these companies. Moreover, for our tests in Section VI, we collect each company's quarterly earnings announcement date and we source data on analysts' earnings forecasts from the Summary History file in the I/B/E/S database. For each firm in our sample and in each quarter, we compute, across analysts, the average and the standard deviation of their latest available forecasts for the firm's next, second, third and fourth quarterly earnings and the actual realization of these earnings. We do the same for analysts' forecasts of next year annual earnings.[17]

Our last data set comes from RS Metrics (see `https://rsmetrics.com/`). RS Metrics

[14]US equity domestic funds are identified using the CRSP objective code "ED".

[15]We use the index fund and ETF fund flags, and we remove funds which contain any of the following strings in their name: 'index', 'idx', 'indx', 'mkt', 'market', 'composite', 's&p', 'russell','nasdaq', 'dow jones', 'wilshire', 'nyse', 'ishares', 'spdr', 'holdrs', 'ETF', 'Exchange-Traded Fund','Exchange Traded Fund', 'PowerShares', 'StreetTRACKS', '100', '400', '500', '600', '1000', '1500', '2000', '3000', '5000', 'money market', 'money mkt'.

[16]Since 2004, the SEC requires mutual funds to report their holdings at the end of each fiscal quarter.

[17]An analyst can report multiple earnings forecasts for the same horizon in a given quarter. Following the literature, for a given analyst, we always use the latest forecasts at a given horizon in a given quarter.

specializes in selling information from satellite imagery. In particular, it is the first data provider to have released parking lot traffic data for U.S. retail firms based on satellite imagery (see Figure A.1 in the Appendix for examples) starting in 2009. Each store is monitored multiple times in a given month, which enables RS Metrics to provide daily store-level information (using various techniques to analyze images) about parking lot capacity and utilization, across major U.S retailers (e.g., Walmart, Home Depot, Best Buy, Starbucks, Tractor Supply Co., etc.).[18] RS Metrics data track parking lots for more than 65,000 retail store locations around the United States. This information is then sold to subscribers of RS Metrics' services, which, based on our discussions with RS Metrics, include asset management firms.

We obtain from RS Metrics their historical dataset on parking lot traffic, which includes the exact dates at which it starts providing store-level parking lot traffic data for 48 publicly listed U.S. retailers between 2009 and 2017 (all firms covered by RS Metrics during this period). Table B.1 in the Appendix presents the different industries (NAICS sectors) of the stocks covered by RS Metrics. We say that RS metrics "initiates coverage" of a stock when it starts providing parking lot traffic data for this stock issuer and we say that a stock is "covered" if at some point during our sample, RS Metrics initiates coverage of this stock. We use coverage of a stock as a shock to the availability of alternative data about this stock. Figure I shows the number of new coverage initiations in each quarter during our sample period.[19]

Last, there are many other types of alternative data available (credit card data, e.mail

[18]A given store for a given firm is not monitored every day so that there are days in a given month with missing observations for a given store. Moreover, the measurement of parking lot traffic is subject to errors as (i) satellite coverage is available only for a subset of a retailer's stores, (ii) not all parking lots are visible from outer space (e.g., underground lots), and (iii) satellite coverage is restricted to domestic store locations. See Katona et al. (2023) for additional details on the data and RS Metrics' technology.

[19]While RS Metrics is the first data vendor starting to sell parking lot traffic data, another prominent competing data vendor is Orbital Insight. However, Orbital Insight introduced its parking lot traffic product in 2014 only, targeting 20 publicly-traded retailers (see https://www.globenewswire.com/en/Orbital-Insight-Expands-U-S-Retail-Traffic-Product-to-More-Than-100-Retailers.html). In the Internet Appendix, we show that our main findings remain qualitatively similar even when we narrow our sample to the period prior to the launch of Orbital Insight's parking lot traffic product (cf., Internet Appendix Table IA.3).

receipts, geolocation data etc.) and the number of alternative datasets has been steadily growing over time. Thus, what we measure is the marginal effect of a specific alternative dataset. This works against finding any effect if parking lot traffic data contains no or little incremental information.

[Insert Figure I about here]

B Measuring Funds' Stock Picking Ability

We measure the stock picking ability of fund f in stock i at horizon h in quarter t by:

$$Picking^h_{f,i,t} = 100 \times (w^f_{i,t} - w^m_{i,t})(R^i_{t+h} - \beta_{i,t} R^m_{t+h}), \tag{3}$$

where $w^f_{i,t}$ is the fraction of fund f's assets held in stock i at the end of quarter t, $w^m_{i,t}$ is the fraction of total market capitalization in stock i (its weight in the "market portfolio") at the end of quarter t, R^i_{t+h} is the return of stock i over the following h quarters, R^m_{t+h} is the return of the stock market over the following h quarters, and $\beta_{i,t}$ is the beta of stock i with the market (computed using daily returns over the last 252 days). $Picking^h$ is expressed in units of return or percentage points (p.p.) per period of length h-quarter. We consider four different horizons in our empirical analysis, namely $h = 1$, 2, 3 and 4 quarters.

A fund's stock picking ability in stock i, $Picking^h_{f,i,t}$ is a proxy for $R(s) = w(s) \times R^e_i$ in eq.(1), where $w(s)$ is measured by $(w^f_{i,t} - w^m_{i,t})$ and $R^e_i = (R^i_{t+h} - \beta_{i,t} R^m_{t+h})$. It should be positive on average if $(w^f_{i,t} - w^m_{i,t})$ and $(R^i_{t+h} - \beta_{i,t} R^m_{t+h})$ are positively correlated, that if fund f tilts its position in stock i relative to a passive benchmark (the market portfolio) in the direction of its future idiosyncratic return, suggesting that the fund exploits useful signals. This measure is closely related to that used by Kacperczyk et al. (2014) (see also Jiang and Zheng, 2018). The main difference is that we measure picking at the fund-stock level while Kacperczyk et al. (2014) measures it at the fund level (by summing the stock-fund picking ability measure across stocks held by each fund).[20]

[20]Kacperczyk et al. (2014) also introduces another measure ("*Timing*") of a funds' ability to anticipate a

For robustness, we consider two alternative measures of skills. For the first one, we replace $w_{i,t}^{m}$ in eq.(3) by $w_{i,t}^{SP500}$, the weight of stock i in the S&P500 index at the end of quarter t. That is, we change the benchmark index used to measure the extent to which a fund manager deviates from a passive benchmark. For the second measure, called $Trading_{f,i,t}^{h}$, we replace $w_{i,t}^{m}$ in eq.(3) by the past weight of stock i in fund f portfolio, namely

$$Trading_{f,i,t}^{h} = 100 \times (w_{i,t}^{f} - w_{i,t-4}^{f})(R_{t+h}^{i} - \beta_{i,t} R_{t+h}^{m}), \tag{4}$$

where $w_{i,t-4}^{f}$ is the weight of stock i in fund f in quarter $t-4$.[21] This measure captures the ability of fund f to change its holdings in a given stock in the direction of subsequent excess returns. In contrast to the two other measures, it is based on trades (change in funds' holdings in a given stock) rather than on the deviation of the fund's holdings relative to a benchmark portfolio. Our results with these alternative measures are qualitatively similar to those obtained with $Picking_{f,i,t}^{h}$ (reported in Section IV). Thus, we only report them in the Internet Appendix (see Section I.9). For brevity, we do not systematically recall this point when discussing the results regarding $Picking_{f,i,t}^{h}$.

Last, it is worth stressing that $Picking_{f,i,t}^{h}$ is not fund f's "alpha" since a fund alpha is defined at the fund, not the stock level. One can show however that a fund's abnormal return over a given period (say a quarter) is equal to the sum of $Picking_{f,i,t}^{h}$ across all the stocks held by the fund (see Section I.2 in the Internet Appendix). Thus, the alpha of a fund over a given time period (its average abnormal return) is the average of the sum of $Picking_{f,i,t}^{h}$ across stocks held by the fund. Consequently, a fund alpha is a noisy measure of a fund's stock picking ability in a stock and is therefore less likely to detect the effect of its coverage of by alternative data.

stock's systematic return ($\beta_{i,t} R_{t+h}^{m}$ in our notations). The alternative data considered in our tests are unlikely to affect a fund's "*Timing*" ability since (i) signals about a retailers' sales are likely to be very noisy signals of market returns and (ii) parking lots counts are not obviously related to a retailers' *beta*s. Consistent with this, we show in Section I.4 of the Internet Appendix that the coverage of a stock by RS Metrics has no effect on funds' timing ability.

[21] We use changes in portfolio weights over four quarters to be consistent with the previous literature (e.g., Grinblatt and Titman, 1993b; Daniel et al., 1997; Jiang and Zheng, 2018).

C Summary Statistics

Table I presents summary statistics for our sample. Panel A presents statistics at the fund-quarter level. The average fund in our sample is 17 years old, manages about $1.5 billion of assets, and holds 123 stocks. Panel B presents summary statistics at the stock-quarter level. The average stock in our sample has a beta of one and is held by 57 different mutual funds ("Nb. Funds Holding"). Finally, Panel C presents summary statistics at the fund-stock-quarter, i.e., holding level. The mean value of "*Picking*" (defined in eq.(3)) is close to zero at all four horizons we consider. At the one-year horizon, 95% of observations for Picking are smaller than 0.40% (per year) and 5% are below −0.42%. At all horizons, there is substantial dispersion in Picking.[22] These findings are in line with the literature on active funds' performance, which finds that the average fund has no significant abnormal performance but that there is substantial dispersion in funds' performance. For instance, Barras et al. (2022) (Table II, Panel A) find that the average (gross) annual alpha of active mutual fund (their sample is the entire population of open-end actively managed US equity funds) is 3% with a cross-sectional standard deviation of 4.1% (see also Pastor et al., 2015).

[Insert Table I about here]

IV Empirical Findings

A Stock Picking Ability and Alternative Data

We use the initiation of coverage of a stock by RS Metrics as a shock to the amount of available alternative data for this stock. This shock is well suited to test our displacement hypothesis because (i) satellite traffic data about a firm contain information useful to forecast future sales and stock returns for this firm (see Zhu, 2019; Katona et al., 2023, and Section I.3 in the Internet Appendix for evidence), and (ii) analyzing these data to extract meaningful

[22]This dispersion is due to both dispersion in Picking across funds and across stocks within funds. The dispersion in Picking across stocks for a given fund within a given quarter is substantial: On average, the standard deviation of Picking at the one-year horizon within a given fund-quarter is 0.43%.

signals likely require data science expertise. Notably, the RS Metrics data set tracks parking lots for more than 65,000 retail store locations around the United States, yielding several millions of historical data points to process. Furthermore, there is anecdotal evidence that some funds (most likely quant funds) use these data (see Footnote 3). Thus, as explained in Section II, coverage of a stock by RS Metrics should allow these funds to obtain higher precision signals about this stock and therefore *reduce* other funds' stock picking ability in this stock (in particular, those who rely solely on human expertise to obtain their signals).

To test this prediction, we estimate the following difference-in-differences specification:

$$Picking^h_{f,i,t} = \beta\{Covered \times Post\}_{i,t} + \alpha_i + \gamma_{f\times t} + \epsilon_{f,i,t}, \tag{5}$$

where $Picking^h_{f,i,t}$ measures the picking ability of fund f in stock i in quarter t at the horizon h-quarter as defined in eq.(3), $\{Covered \times Post\}_{i,t}$ is a dummy equal to one after RS Metrics initiates coverage of stock i, α_i are stock fixed effects and $\gamma_{f\times t}$ are fund $\times$quarter fixed effects. Standard errors are double-clustered at the fund and stock levels. Stock fixed effects capture time-invariant determinants of fund picking abilities in each stock. Fund$\times$quarter fixed effects remove any time-varying shocks or fund characteristics that might affect picking ability for all stocks in that fund's portfolio, such as fund size.

The coefficient β measures the extent to which a fund's stock-picking ability in a given stock is affected by the RS metrics' initiation of coverage for this stock. Our displacement hypothesis implies that $\beta < 0$. One concern is that a fund's stock picking ability might vary over time due to unobserved factors, unrelated to alternative data. For example, a fund might decide to beef up its team of analysts at the same time alternative data becomes available. This decision would increase a fund's stock picking ability and bias our estimate of the effect of RS Metrics' coverage initiation on a fund's stock picking ability, β. The inclusion of fund$\times$quarter fixed effect in our specification (cf., eq.(5)) addresses this problem. Indeed, it controls for all unobserved variables that might affect a fund stock picking ability in a given quarter. Given the fund$\times$quarter fixed effects, β captures the change in a fund's picking ability after a stock becomes covered by RS Metrics relative to stocks in the *same* fund's portfolio not covered by RS Metrics.

We also consider a specification including fund×stock fixed effects to make sure that the coefficient β is estimated by comparing picking abilities before versus after coverage initiation for the *same* fund and stock. The inclusion of these additional fixed effects rules out the possibility that change in funds' picking abilities in covered stocks is due to a composition effects (change in the funds holding covered stocks before and after coverage initiation).

[Insert Table II about here]

Table II reports estimates of eq.(5). The horizon at which *Picking* is measured ranges from one quarter (Columns (1) to (3)) to one year (Colums (10) to (12)). In the specifications considered in Columns (1) (one quarter), (4), (7) and (10) (one year), we do not include stock fixed effects and simply control for whether stock i is covered or not at some point during our sample period by RS Metrics. The coefficient on "Covered" is significantly positive. Thus, funds holding covered stocks have significantly higher stock picking ability in these stocks than all funds' stock picking ability in non covered stocks. For instance, at the one year horizon, we observe that, on average, mutual funds display picking skills that are approximately one quarter of a standard deviation higher for covered stocks prior to the release. Thus, funds' performance in covered stocks is not bad to start with.

Consistent with our displacement hypothesis, β is always significantly negative for all horizons in all specifications (even after controlling for stock and fund-quarter fixed effects in columns (2), (5), (8) and (11)). Thus, funds stock picking ability declines after RS Metrics start covering these stocks. The magnitude of this decline is economically significant. For instance, consider the four-quarter horizon where the mean of "Picking" is zero, the standard deviation is 0.361, and only 5% of the realizations exceed 0.4 (cf., Panel C in Table I). Column (11) of Table II shows that, at this horizon, "Picking" for covered stocks experiences a decline of 0.106 after coverage initiation, which is nearly one-third of its standard deviation. In columns (3), (6), (9) and (12), we include fund×stock fixed effects. Results are unchanged with this approach.

To study the dynamics of the impact of coverage initiation on funds' stock picking ability (β), we estimate the following specification:

$$Picking^h_{f,i,t} = \sum_{k=-12,k\neq-1}^{16+} \beta_k\{Covered \times Quarter(k)\}_{i,t} + \alpha_i + \gamma_{f\times t} + \epsilon_{f,i,t}, \qquad (6)$$

where $\{Covered \times Quarter(k)\}_{i,t}$ are dummy variables equal to one if RS Metrics covers stock i's and time t corresponds to k quarters before/after the release (for $k \in \{-12, -11, ..., 15, 16+\}$). Quarter -1 is the quarter just before the first quarter for which RS Metrics initiates coverage of stock i. It serves as the reference point and is therefore omitted in the estimation of eq.(6). Quarters 16+ correspond to quarters more than 4 years after RS Metrics initiates coverage of stock i. Standard errors are double-clustered at the fund and stock levels.

[Insert Figure II about here]

Figure II plots the estimates of the β_ks' in each quarter in eq.(6) when Picking is measured at the yearly horizon (dashed lines correspond to 95% confidence intervals). We observe that these estimates are not significantly different from zero before coverage and become significantly negative about 4 quarters after coverage starts. Thus, a fund's Picking ability is not different for control and covered stocks *before* coverage initiation and the drop in funds' Picking ability for covered stocks after coverage initiation is not the continuation of a "pre-treatment" trend. This observation supports our interpretation that coverage initiation is the cause, and not the consequence, of this drop. Interestingly, the decline in Picking in covered stocks occurs rapidly after coverage begins. Within four quarters following coverage initiation, there is already a significant drop of over 0.1 p.p. in "Picking." This finding suggests that some market participants (e.g., quant funds) swiftly exploit RS Metrics' data after they become available, reducing quickly other funds Picking's ability. Although there is a slight rebound in funds' picking skills for covered stocks after the initial decline, the effect remains negative over time after coverage begins. Moreover, Picking does not revert to pre-coverage level even after four years, indicating a persistent and enduring impact of coverage initiation on Picking.

B Selection of Covered Stocks and Matching

One concern is that stocks covered by RS Metrics are different from stocks that are not. As our results hold when we control for stock or stock-fund fixed effects, this is not an issue if RS Metrics's decision to cover a stock is determined by time-invariant characteristics of covered stocks. The concern arises only if there are time-varying variables that both affect RS Metrics decision to cover a stock and funds' stock picking abilities in covered stocks, differently than for control stocks. To address this issue, we study in more details RS Metrics' coverage decision.

Specifically, we run a series of regressions at the stock-quarter level, where the dependent variable is an indicator equal to 100 in the first quarter of coverage initiation (zero otherwise) and the explanatory variables are a set of potential determinants of the coverage decision (e.g., a stock market capitalization or a stock lagged return). Thus, the data used for this analysis include right-censored stock-quarter observations, i.e., observations up to the coverage initiation for covered stocks and all available observations for control stocks that are never covered by RS Metrics. All regressions include industry and year-quarter fixed effects.

We report the results from this analysis in Table B.2 in the Appendix. In column (1), we explore the relationship between coverage initiation and stock characteristics such as size, assets, and book-to-market ratio. We observe that large stocks are significantly more likely to be selected for coverage. Column (2) introduces variables related to past performance, including stock returns and average earnings from the previous year. These variables are not significantly related to the coverage decision. In column (3), we incorporate the logarithm of the stock's idiosyncratic volatility and the prior year's average analyst forecast error for next quarter's earnings per share, which also are non-significantly related to the coverage decision.

Interestingly, the number of mutual funds holding the stock (a measure of potential demand) and funds' lagged stock picking ability do not predict coverage initiation as well (see Columns (4) and (5)). This observation, with the results from the previous section, mitigates the possibility of reverse causality (RS Metrics initiating coverage in response to a strong demand by funds because of a decline in the quality of their information). Lastly, measures of stock price informativeness and analysts' forecasts errors similarly do not appear to be predictors for coverage initiation (see Column 6).[23] Again, this suggests that RS Metrics' decision to cover a stock is not driven by a decline in the informativeness of a firm's stock price.

In sum, the only robust predictor of coverage for a stock is market capitalization, maybe because data providers expect a stronger demand for large caps, allowing them to better amortize fixed costs of cleaning up and preparing the data.[24] Thus, as a robustness test, we re-estimate our main specifications (those in Table II) with a matched group of control stocks. Specifically, one year prior to the initiation of coverage, we match each covered stock with five non-covered stocks based on industry (NAICS 2-digit sector) and market capitalization. The results with this matching approach are qualitatively similar to those reported in Table II (see Internet Appendix Table IA.4).

C Cross-Sectional Heterogeneity

The tests in the previous section shows that the availability of new alternative data for a stock has a negative effect on active funds' average (across all funds) stock picking ability in this stock ($\bar{R}$ in eq.(2)). However, as explained in Section II, this average effect underestimates the effect of the new data on the stock picking ability of the managers who do not have the skills to exploit the new data (or do not buy them). In this section we test whether this is the case.

As we do not directly observe how managers obtain their signals, we use an indirect approach to identify experts in our data. First, we posit that funds with a high stock picking ability in covered stocks before they become covered are more likely to produce information

[23]Measures of stock price informativeness measures and analyst forecast errors are defined in Section VI.

[24]It is noteworthy that the R^2 of our regressions for predicting coverage is very low. This indicates that predicting coverage with financial variables such as stock returns, volatility, or the number of funds is difficult. One possible reason is that RS Metrics' clients extend beyond the financial industry (e.g., many firms might buy the data to anticipate future dynamics in the retail industry and make production or marketing decisions accordingly).

via judgemental analysis rather than quantitative analysis of large datasets (since they had a high stock picking ability even before new data becomes available). Second, funds that concentrate their holdings in an industry are more likely to rely on industry specific knowledge (see Kacperczyk et al., 2005, for evidence that these funds have a superior stock picking ability) rather than statistical techniques to produce their signals. Third, based on Coval and Moskowitz (2001), we expect funds that are located close to the main stores or to the headquarters of the firms covered by RS Metrics to obtain their information via direct visits to the stores or networking.

Below, we consider each of these measures of expertise in turn and we test whether greater expertise amplifies the negative effect of coverage initiation by expanding our baseline specification (eq.(5)) with interaction terms, between our $Covered \times Post$ dummy and each measure of expertise.

C.1 Pre-coverage Picking Abilities

We first estimate the following specification:

$$\begin{aligned} Picking^{h}_{f,i,t} = {} & \beta\{Covered \times Post\}_{i,t} \\ & + \beta_{high}\{Covered \times Post\}_{i,t} \times HighPickingPre_{f,i} \\ & + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t}, \end{aligned} \tag{7}$$

where $HighPickingPre_{f,i}$ is a dummy variable equal to 1 if Picking for fund f in stock i is above the median value of Picking of all funds in this stock, in the quarters *before* coverage of this stock begins. Thus, β measures the average effect of coverage across funds identified as inferior stock pickers on that stock prior to its coverage while $(\beta + \beta_{high})$ measures this effect across funds identified as superior stock pickers. Thus, β_{high} measures the differential effect of coverage between superior and inferior stock pickers. Other variables are the same as in eq. (5) and standard errors are double-clustered at the fund and stock levels. In estimating eq. (7), we exclude observations for funds that began holding covered stocks after coverage initiation since, by definition, $HighPickingPre_{f,i}$ cannot be observed for these funds.

We emphasize that we include fund×stock fixed effects, $\alpha_{f\times i}$, to make sure that the coefficients β and β_{high} are estimated by comparing picking abilities before versus after coverage initiation for the *same* fund and stock. The inclusion of such fixed effects rules out the possibility that the change in picking skills we observe is driven by a shift in the composition of the funds holding covered stocks before and after coverage initiation. Because $HighPickingPre_{f,i}$ is constant over time for a given fund-stock, the interaction term $Covered_i \times HighPickingPre_{f,i}$ is absorbed by the fund-stock fixed effects.

[Insert Table III about here]

Table III presents the estimates of eq.(7). For each horizon (ranging from one quarter to one year), we find that the effect of coverage initiation on the stock-picking ability of the funds with high expertise (those for which $HighPickingPre_{f,i} = 1$) is negative and strongly significant (t-stat above 3). The magnitude of this effect is also more than twice as large as the effect of coverage initiation in our baseline specification (the estimate of β in Table II). In contrast, the effect of coverage initiation on the stock-picking ability of funds with low expertise (those for which $HighPickingPre_{f,i} = 0$) is not significant. Interestingly, this differential effect is also consistent with the mechanism described in Section II. Indeed, as shown by eq.(IA.12) in the Internet Appendix, the negative effect of new alternative data on an expert's stock picking ability should be stronger for experts with a higher signal precision (and a fund's stock picking ability increases with the precision of its signal).

To describe the dynamics of the impact of coverage initiation on funds with high stock picking ability before coverage initiation, we estimate the following specification:

$$
\begin{aligned}
Picking^{h}_{f,i,t} = & \sum_{k=-12,k\neq-1}^{16+} \beta_k \{Covered \times Quarter(k)\}_{i,t} \\
& + \sum_{k=-12,k\neq-1}^{16+} \beta_{high,k} \{Covered \times Quarter(k)\}_{i,t} \times HighPickingPre_{f,i} \\
& + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t},
\end{aligned}
\tag{8}
$$

where $\{Covered \times Quarter(k)\}_{i,t}$ is defined as in eq.(6). The coefficients β_k's capture the dynamics of the effect on the funds identified as inferior stock pickers on the covered stock before coverage starts, while the coefficients $\beta_{high,k}$'s capture the dynamics of the additional effect on the funds identified as superior stock pickers. Standard errors are double-clustered at the fund and stock levels.

[Insert Figure III about here]

Panel A and B of Figure III plot the estimates of respectively the β_ks' and the $\beta_{high,k}$s' for each quarter in eq.(8) when Picking is measured at the yearly horizon. In both panels, the estimates are not significantly different from zero before coverage and remain so, after coverage initiation for funds with low stock picking ability (Panel A). In contrast, we observe a significant drop in Picking after coverage initiation for funds with high stock picking ability. The effect remains negative over time, around -0.2 p.p, and does not revert to pre-coverage level even after four years.

A potential concern is that the above results may be due to reversals in mutual funds' stock picking abilities.[25] Even though Panels A and B of Figure III depict no discernible pre-trend in picking abilities before coverage initiation, funds with abnormally high picking ability levels might experience a subsequent relative decline post-coverage due to *Picking* reverting to its mean level. To address this concern, we re-estimate equations (5) and (7) with controls for the lagged value of "Picking". If our previous results are just mechanically due to reversals in funds' performance, the lagged value of picking should subsume the effect of coverage initiation on funds with high picking abilities pre-coverage.

The estimation results are presented in Appendix Table B.3. Across all specifications, we observe a negative coefficient on the lagged value of Picking ("Past Picking"), between -0.065 (for Picking 1-Q) and -0.207 (for Picking 4-Q). Thus, there is relative reversal in picking abilities at the *fund-stock* level. For instance, assuming lagged Picking 1-Q is 0.19 (its 95th

[25] For example, Kacperczyk et al. (2014) find that stock picking estimated at the fund level does not exhibit much persistence.

percentile), this translates to a reduction of current Picking 1-Q by -0.065 × 0.19 = -0.01. More important for our purposes, the coefficients on the interaction term "Covered × Post" and on the triple interaction "Covered × Post × High Picking Pre" remain significantly negative, and their magnitudes align with the estimates in Tables II and III. Thus our findings are robust, even after accounting for reversals in funds' stock picking abilities.

C.2 Industry Expertise

In the previous section, we identify experts by the level of their stock picking ability prior coverage. In this section and the next, we relate expertise to traditional methods of obtaining private signals, namely (i) industry expertise and (ii) geographical location.

To study the role of industry expertise, we proceed as follows. We define a fund as specialized in the industry of a stock covered by RS Metrics if (i) the fund is classified by CRSP as a "Sector Fund" and invest in the industry of this stock (sector funds are funds that invest primarily in one type of industry or sector) or (ii) the fund is an "Industry Specialist" in the sense that it invests, on average, more than 75% of its assets in stocks that belong to industries covered by RS Metrics, that is, the NAICS sectors in which RS Metrics covers at least one company. These sectors include Manufacturing, Wholesale Trade, Retail Trade, Real Estate and Rental and Leasing, Accommodation and Food Services and Other Services (cf., Table B.1 in the Appendix).[26]

We then estimate the following specification:

$$\begin{aligned} Picking^h_{f,i,t} = \beta\{Covered \times Post\}_{i,t} \\ + \beta_{expert}\{Covered \times Post\}_{i,t} \times IndustryExpert_{f,i} \\ + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t}, \end{aligned} \tag{9}$$

where $IndustryExpert_{f,i}$ is a dummy variable equal to 1 if fund f is a"Sector Fund" or an "Industry Specialist" for the industry of stock i. Other variables are the same as in eq.(7)

[26] Most of the covered stocks are in the retail trade industry but there is substantial variation across funds in industry specialization. 4% of funds in our sample are classified as "Industry Specialist" in industries covered by RS Metrics, and 9% are "Sector Fund".

and standard errors are double-clustered at the fund and stock levels. The coefficient of interest is β_{expert}, which measures the effect of coverage initiation of a given stock on industry experts' stock picking ability, above and beyond the baseline effect (measured by β) for all funds holding this stock. Again, we include fund×stock fixed effects to make sure that the coefficients β and β_{expert} are estimated by comparing picking abilities before versus after coverage initiation for the *same* fund and stock.

[Insert Table IV About Here]

Table IV reports estimates of eq.(9). Panel A includes the dummy "Industry Specialist", while Panel B includes the dummy "Sector Fund". In both panels, in columns (1), (3), (5), and (7), we estimate the equation without fund-stock fixed effects but we include stock fixed effects and the interaction variables $Covered_{i,t} \times IndustrySpecialist_{f,i}$ (Panel A) or $Covered_{i,t} \times SectorFund_{f,i}$ (Panel B), to measure the difference in picking skills between industry experts and non-experts funds for covered stocks prior to coverage initiation. We observe that this difference is significantly positive, which confirms that mutual funds focusing their investments in the industries of covered companies have a higher stock picking ability.

Moreover, as predicted, we find that, after coverage initiation, funds with specific industry expertise experience a significant decline in their stock picking ability, 10 to 20 times larger than that for other funds (depending on which specification is considered). The asymmetry of the effect between experts and non experts provides additional evidence that funds with high stock selection ability, in this case resulting from their industry specialization, are the most affected by the availability of new alternative data for a stock.

C.3 Geographical Location

Our second measure of expertise builds on the idea that a fund's location can be a source of private information and therefore expertise (see Coval and Moskowitz, 2001; Bae et al., 2008; Cicero et al., 2023). In particular, funds that are located in proximity to the main stores or headquarters of firms covered by RS Metrics are better able to directly monitor sales by these firms or benefit from networking opportunities with managers of these firms. As for industry expertise, this source of information is fixed and unaffected by the introduction of the alternative data considered in our tests, as we assume for experts in the mechanism described in Section II.

To measure location-based expertise, we identify, for each covered stock, the metropolitan statistical areas (MSAs) corresponding to (i) the firm's headquarters (sourced from Compustat) and (ii) the location where the highest number of the firm's parking lots are located. Specifically, using RS Metrics data, we identify the firm's primary MSA: the one with the highest number of the firm's parking lots. Then, we categorize a fund as "Local" for a specific stock if it is located in the same MSA as either (i) the firm's headquarters or (ii) the stock's primary MSA based on parking lots.[27] We do not exclusively rely on firms' headquarters for identifying local funds because, among the retailers covered by RS Metrics in our sample, very few share the MSA of their headquarters with funds. Specifically, only 0.10% of observations in our sample correspond to funds categorized as "Local" based on the MSA of a covered firm's headquarters. In contrast, this percentage increases to 6% when "Local" is defined using both the firm's headquarters and the primary location of its parking lots.[28]

We then estimate the following specification:

$$\begin{aligned} Picking^{h}_{f,i,t} = {} & \beta\{Covered \times Post\}_{i,t} \\ & + \beta_{local}\{Covered \times Post\}_{i,t} \times Local_{f,i} \\ & + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t}, \end{aligned} \tag{10}$$

where $Local_{f,i}$ is a dummy variable indicating whether fund f is classified as "Local" for stock i. Other variables are the same as in eq.(7) and standard errors are double-clustered at

[27] To ensure that the primary MSA provides valuable information, we only consider a fund as "Local" based on parking lots if at least 5% of the firm's parking lots are situated within its primary MSA. Among covered stocks, the average (median) proportion of parking lots located in the primary MSA is 7% (5%), with the 10th percentile at 3% and the 90th percentile at 12%.

[28] When "Local" is defined based only on firm's headquarters, our results presented below are qualitatively similar but not statistically significant.

the fund and stock levels.

[Insert Table V about here]

Table V reports estimates of eq.(10). In columns (1), (3), (5), and (7), we estimate this equation without fund-stock fixed effects but we include stock fixed effects and the interaction variable $Covered_{i,t} \times Local_{f,i}$ equal to 1 if fund f is classified as "Local." The coefficient on this variable is positive and significant (see Table V). Thus, mutual funds located in areas where covered companies have the highest number of their stores or their headquarters have a higher stock picking ability prior to coverage initiation. More important for our purpose, in all specifications, the coefficient, β_{local}, is negative and significant. Thus, following coverage initiation, local funds experience a more pronounced decline in their stock picking ability than other funds holding covered stocks (for which we also find a drop in stock picking ability; β is significantly negative in all specifications). The magnitude of the coefficients suggests that the decline in the stock picking ability of local funds for covered stocks is four to five times larger than for those of other funds.

C.4 Robustness Tests including Stock-Quarter Fixed Effects

We conduct supplementary robustness tests of our cross-sectional heterogeneity results. Specifically, we re-estimate regressions (7), (9) and (10) adding stock-quarter fixed effects. These fixed effects capture time-varying fluctuations in picking abilities that are common across all funds holding a specific stock. As such, they absorb the interaction term $Covered \times Post$. In particular, stock-quarter fixed effects control for any time-varying variables at the stock level that may affect both RS Metrics's decision to cover a stock and funds' picking abilities (e.g., changes in the stock's factor exposures). However, this specification still enables us to estimate the coefficient on the interaction terms between $Covered \times Post$ and each measure of expertise. Therefore, this additional test serves to confirm that the observed heterogeneous changes in Picking across funds persist even when accounting for unobserved variations at the stock level. The estimation results are presented in Table IA.5 (Internet Appendix) and remain consistent with the results discussed earlier.

C.5 Is the Effect Weaker for Quant Funds?

Our findings so far are consistent with our hypothesis: Alternative data reduces the performance of funds relying on traditional methods to obtain information (e.g., industry-specific expertise or geographical location). Our interpretation is that these funds lack the expertise required to exploit alternative data. In contrast, we expect those who buy the alternative data considered in our tests to experience an increase in their stock picking ability ($\bar{R}_i(s_i(\tau_{dm}))$ in eq.(2)). Accordingly, if we zoom in on the funds that are more likely to buy these data in our sample, we should observe a weaker negative effect of RS Metrics coverage on their stock picking ability (the effect is not necessarily positive because we do not perfectly observe the funds buying the data).

As explained previously, we conjecture that quant funds are more likely to buy alternative data as they have the skills required to use them. Identifying these funds in our sample is not straightforward as we do not directly observe a fund's type. To overcome this issue, we build proxies for quant funds in our sample, exploiting the text of mutual fund prospectuses, as explained in Section I.8 of the Internet Appendix. We search for specific keywords, such as "quantitative stock selection", in the strategy section of fund prospectuses which provides information regarding funds' investment process.[29] We then estimate a specification similar to eq.(9), interacting "Covered $\times$ Post" with various proxies for quant funds. Internet Appendix Table IA.6 presents estimates of these tests, and shows that the coefficients on these triple interaction terms are positive and statistically significant in most specifications. Thus, as predicted, coverage initiation has a weaker negative effect on funds that are more likely to be able to use the data considered in our tests. Our inference however is limited by our inability to directly observe funds that buy the data and by the fact we only focus on mutual funds. Indeed, we expect sophisticated investors out of our sample, such as hedge funds, to also use alternative data, and therefore to benefit after coverage initiation.

[29] Abis (2022) uses machine learning to categorize US active equity mutual funds as quants or discretionaries. Here, we use a more direct and simpler methodology.

V Evidence of Displacement

In line with our hypothesis, the availability of new alternative data for a stock reduces the value of traditional expertise in this stock. In this section, we study how asset managers respond. We do not expect them to immediately switch to other occupations (an extreme form of displacement). However, they could reduce the capital allocated to the stocks in which their expertise is less valuable (a form of displacement) and shift it to stocks in which this expertise remains valuable. We examine whether this is the case in Sections V.A and V.B, respectively.

A Divestment

We first study whether coverage leads funds to reduce the capital invested in covered stocks. To do so, we begin by analyzing the rank of covered stocks in mutual fund portfolios. A stock's rank is determined by its weight (in terms of dollar invested as a fraction of total net assets) in a fund's portfolio. For example, a rank of 1 means that the stock is the largest investment made by a fund in dollar value, a rank of 2 represents the second largest investment, and so on. Therefore, the rank of a stock is a relative measure that takes the variation in fund size into account.

We then estimate specification (5) using as dependent variable the (negative) natural logarithm of the stock rank in the fund portfolio (so that a higher value of the dependent variable corresponds to a greater investment in a stock).[30] Results are presented in columns (1) and (2) of Table VI. After coverage initiation, covered stocks experience a 9% decrease in their rank, that is, a fall of 11 places (since the average fund in our sample holds 123 stocks). We emphasize that the regression in column (2) incorporates fund-stock fixed effects, ensuring that the coefficient on "Covered × Post" is estimated by comparing a given fund's holdings in covered stocks after and before coverage, relative to the fund's variation in holdings of

[30]The relationship between the negative of the logarithm of rank and Picking is positive and significant (about 0.05 when controlling for fund-quarter fixed effects). This is consistent with Anton et al. (2021) who find empirically that mutual funds managers overweight the stocks identified as their "best ideas".

uncovered stocks.

[Insert Table VI about Here]

To test whether divestment is stronger for funds that have more expertise, we estimate the same specifications including triple interactions as in equations (7), (9) and (10), using the negative of the logarithm of the stock rank as the dependent variable. Table VII presents the results. Consistent with our previous results, we find that funds classified as superior stock pickers before coverage initiation, industry specialists, and "local" funds divest to a larger extent from covered stocks after coverage initiation.

[Insert Table VII about Here]

The previous results indicate a decline in funds' holdings of covered stocks at the intensive margin. To further explore the extensive margin, we now study how the number of funds holding covered stocks varies around coverage initiation. Specifically, we estimate a difference-in-differences regression with stock-quarter observations, using the logarithm of the number of funds holding the stock in a given quarter as dependent variable. Columns (3) and (4) of Table VI present the estimation results. Both regressions incorporate quarter and stock fixed effects. Column (3) encompasses all stocks in our sample, while column (4) focuses solely on stocks within the covered industries. In both cases, we observe a reduction of approximately 20% in the number of funds holding a covered stock after coverage of this stock by RS Metrics begins, that is, about 11 funds per covered stocks (on average, a stock is held by 57 distinct funds).

[Insert Figure IV about Here]

Figure IV describes the dynamics of the effect of a stock coverage on its rank in a fund portfolio (Panel A) and on the number of funds holding this stock (Panel B). Specifically, it plots the quarterly coefficients of an event study regression that includes interactions between the variable "Covered" and a set of dummy variables indicating quarters before and after coverage starts (similar to the specification used in eq.(6)). Prior to coverage, the evolution of the rank of a stock in a fund's portfolio and the number of funds holding this stock is similar for covered and control stocks. However, within eight quarters after RS Metrics begins covering a stock, we observe a significant decline of about 10% in the rank of this stock in funds' portfolios and in the number of funds holding this stock. Exits do not happen instantaneously as it takes time for a fund to learn that the value of its signal for a covered stock has declined. Importantly, the negative trend in the number of funds holding covered stocks persists even after four years, indicating a lasting and substantial impact of the availability of alternative data on funds' investment intensity in covered stocks and the number of funds holding these stocks.

B Reallocation of Assets

We now examine whether, after coverage, funds redirect their investments towards uncovered stocks in which the value of their expertise is valuable. To test this conjecture, we study, around coverage initiations, the evolution of the number of funds holding non-covered stocks in the same industry or geographical area as covered stocks (we refer to these uncovered stocks as "peer stocks").

More specifically, we estimate regressions at the stock-quarter level, exclusively using data from control stocks non-covered by RS Metrics. Our dependent variable is the logarithm of the number of funds holding a specific stock in a given quarter. We define two independent variables: one denoted as "Industry Peer Covered $\times$ Post", which is equal to one when RS Metrics begins covering a stock in the same industry (defined by the 2-digit NAICS sector) as the focal stock. The second, denoted "Local Peer Covered $\times$ Post" is a dummy equal to one when RS Metrics starts covering a stock whose highest number of parking lots are in the same Metropolitan Statistical Area (MSA) as the focal stock's headquarters.[31]

[31]We leverage headquarters location as a proxy for geographical activity areas of non-covered stocks as we cannot directly observe their point-of-sale locations.

[Insert Table VIII about Here]

Table VIII reports the results. As expected, we observe that after a stock becomes covered, there is a significant increase in the number of funds holding non-covered peers of the covered stock. In fact, the estimates suggest a roughly 10% increase in funds investing in non-covered peer stocks following the coverage initiation of their peers. This result is consistent with industry experts and local funds reorienting their portfolio towards stocks where their expertise retains value.

This last finding suggests that coverage of a stock might, indirectly, affect its peers even though they are not covered. To rule out the possibility that such spillover effects from covered stocks to non-covered stocks contaminate our main results, we re-estimate our main specifications regarding the effect of coverage on funds' stock picking (those reported in Table II), omitting non-covered peer stocks from the set of control stocks. Our findings with this approach are unchanged (see Table IA.12 in the online appendix).

VI Mechanism: The Role of Price Informativeness

As shown in Section IV.A, the availability of new alternative data for a stock reduces the average funds' stock picking ability in that stock. According to the mechanism described in Section II, this effect is due to the fact that the availability of new alternative data makes the price of covered stocks more informative. In this section, we test whether this is the case.

A Measuring Price Informativeness

We use two distinct measures to capture price informativeness. The first measure is the absolute cumulative abnormal return (ACAR) (Fama et al., 1969; Ball and Brown, 1968) following earning announcements. More specifically, we compute two ACAR measures for each stock-quarter: $ACAR[0, 2]$ and $ACAR[-1, 2]$, where $ACAR[-m, n]$ represents the absolute cumulative abnormal return from the m^{th} day prior to earnings announcements to the n^{th} day after earnings announcements. We estimate abnormal returns relative to a Fama and French (1992) three-factor model, estimated using daily returns over a 252-day window ending 90 days before the earnings announcement. We exclude observations with estimation windows comprising fewer than 63 (preceding) trading days, i.e., one calendar quarter. Lower ACARs are indicative of information being incorporated in stock prices before the announcement and thus greater price informativeness.

Our second measure of price informativeness is the price jump ratio (Weller, 2018), which is the ratio of post-announcement returns to total returns before and including the earnings announcement. A lower jump ratio indicates that stock prices are more informative prior to the announcement since their variation post announcement account for a smaller fraction of total return over a window comprising the announcement. Following Weller (2018), we employ a 21-day pre-announcement window. For each stock-quarter, we calculate two jump ratio measures: $CAR[0,2]/CAR[-21,2]$ and $CAR[-1,2]/CAR[-21,2]$, where $CAR[-m,n]$ represents the cumulative abnormal return from the m^{th} day prior to earnings announcements to the n^{th} day after earnings announcements, estimated as in the ACAR measure described above. Following Weller (2018), we require the announcement period returns to be substantially larger compared to scaled daily volatility, indicating substantial earnings announcement information. Specifically, we require $ACAR[-21,2] > \sqrt{24}\sigma$, where σ denotes the daily volatility of the stock during the month preceding the earnings announcement window.

B Empirical Findings

To test whether the availability of satellite imagery data enhances price informativeness, we estimate the following specification at the stock-quarter level:

$$Informativeness_{i,t} = \beta\{Covered \times Post\}_{i,t} + \delta X_{i,t} + \alpha_i + \gamma_t + \epsilon_{f,i,t}, \tag{11}$$

where $Informativeness_{i,t}$ denotes either an ACAR or a jump ratio measure, reflecting the price informativeness of stock i in quarter t, $\{Covered \times Post\}_{i,t}$ is a dummy equal to one after RS Metrics initiates coverage of stock i, α_i are stock fixed effects and γ_t are quarter fixed effects. The vector $X_{i,t}$ comprises the following control variables lagged by one quarter: the natural logarithm of the stock's market capitalization, book-to-market ratio, total assets, and sales, as well as leverage (ratio of debt to total assets). Following Weller (2018), we also include the logarithm of the daily volatility during the month preceding the earnings announcement window. Standard errors are clustered at the stock level. To ensure comparability among stocks, our analysis focuses solely on stocks within covered industries, that is, NAICS sectors where RS Metrics provides satellite imagery data for at least one company (cf., Appendix Table B.1).

[Insert Table IX about here]

Table IX reports estimates of eq.(11). In Columns (1) and (2), we use the jump ratio measure as a measure of stock price informativeness, while in columns (3) and (4) we use the ACAR measure. In all cases, we observe that coverage initiation is associated with a significant increase in price informativeness prior to earnings announcements (that is, a decrease in the jump ratio or ACAR). Specifically, columns (1) and (2) indicate a significant decline of approximately 9% in the jump ratio for covered stocks following coverage initiation. Similarly, columns (3) and (4) show a reduction of approximately 0.5% in the ACAR.

Thus, consistent with the mechanism described in Section II, the availability of new alternative data for a stock is associated with an increase in stock price informativeness. A similar finding (in the case of satellite imagery as well) is obtained in Zhu (2018). In contrast, Katona et al. (2023) find no statistically significant effect of the introduction of satellite data on stock price informativeness. They conjecture that this result could be due to the fact the effect of satellite data on price informativeness takes time to materialize and provide evidence consistent with this conjecture. The difference between their findings and ours may reflect the fact that our sample is not exactly the same (in particular, our sample ends in 2021 while their ends in 2017) and that our measure of stock price informativeness is not defined exactly in the same way. In any case, the novelty of our paper is not to study the effect of alternative data on price informativeness but its effect on funds' stock picking ability.

C Better Private Signals or Better Public Signals?

As expected, we find an improvement in price informativeness for covered stocks. This finding is consistent with the mechanism described in Section II, according to which the availability of new alternative data enables some funds (the quants) to improve the precision of their private signals. Another, non exclusive, possibility is that the availability of new alternative data leads to more accurate public signals (the "public information channel"). In either case, price informativeness increases and therefore the effect on funds relying on traditional source of information is the same: Their performance will be reduced (as we find in Section IV). However, Katona et al. (2023) finds that the stocks covered by RS Metrics data become less liquid after coverage begins. This evolution is consistent with a scenario in which this data enables some market participants to obtain more precise private signals, as assumed in the mechanism described in Section II. Indeed, more precise private signals make the order flow more informative, which increases illiquidity (we show that this is the case in the model used to develop our testable hypotheses; see Section I.1 in the Internet Appendix). More precise public signals should have the opposite effect.

Nevertheless, to study whether the public information channel plays a role, we study whether sell-side equity analysts' earnings forecasts become more accurate after a stock becomes covered since equity analysts are an important source of public information in stock markets (see, for instance, Lee and So, 2017; Chen et al., 2020). For this, we use data from I/B/E/S to calculate standardized analysts' earning forecast errors for each stock in industries covered by RS Metrics. For every quarter, we compute these errors by taking the absolute difference between the actual earnings and the average of the latest analyst forecasts, divided by the standard deviation of those forecasts. We compute such forecasting errors for the subsequent quarter, for the second, third, and fourth quarter ahead, as well as for the forthcoming year. We then estimate eq. (11) using the standardized analysts' forecasting errors at various horizons (1 quarter to 4 quarters and next year) as dependent variables. All other variables are the same as in eq. (11).

The results, presented in Appendix Table B.4, reveal no significant coefficients on the interaction term between Covered and Post. Thus, there is no significant reduction in analysts' forecasts errors after a stock becomes covered.[32] Thus, the improvement in price informativeness for covered stocks is unlikely to be due to an improvement in the quality of analysts' forecasts for these stocks. Of course, the availability of new alternative data could enhance the accuracy of public signals for investors via other sources than sell-side analysts' forecasts (e.g., managerial disclosure).

VII Conclusion

The availability of new alternative data enables fund managers with the skills required to use this data to improve the precision of their signals. In trading on these improved signals, these investors improve price informativeness and thereby exert a negative externality on fund managers who rely on more traditional methods (industry specific knowledge, networking and location, human judgement and economic analysis, etc.) to obtain private information. In this paper, we provide evidence of this mechanism.

Specifically, we show that the availability of satellite imagery data tracking retailer firms' parking lots reduces the average stock picking ability of active mutual fund managers in stocks covered by this data. This decline is particularly pronounced for funds that heavily rely on traditional sources of expertise, such as industry expertise and geographical location. We also find that the initiation of coverage is associated with a drop in active mutual funds' holdings on average or even an exit of some funds from covered stocks. These findings are consistent with the idea that the rise of alternative data could gradually "crowd out" asset managers lacking the skills to exploit this data from the stocks covered by such data.

We obtain these results by considering the introduction of one specific type of alternative

[32]This, maybe, is not surprising given that only few analysts seem to rely on data sourced from satellite imagery for forming their forecasts. Chi et al. (2022) study the frequency with which financial analysts mention the use of alternative data in their written reports. They show that, while analysts frequently rely on alternative data to form their forecasts, geospatial and satellite imagery data are among the least popular categories of alternative data used by analysts. Specifically, as shown in Table 2 of Chi et al. (2022), only 3% of analysts' reports in their data mention the use of data based on satellite imagery as a source of information.

data (based on satellite imagery) for only a limited subset of stocks. However, subject to data availability, one could use the same approach to check the robustness of our conclusions when other types of alternative data is introduced. Another interesting question, beyond the scope of this paper, is whether there are cases in which traditional asset managers' skills are less at risk of displacement. For instance, alternative data might be less useful for forecasting long-term firms' earnings than short-term earnings (Dessaint et al., 2022). Thus, traditional source of expertise (e.g., economic reasoning) might especially be valuable for forecasting earnings of firms with long horizon projects (high duration stocks). If so, at least for some stocks, a combination of quantitative and a discretionary approaches might be optimal.

References

Abis, S. (2022). Man vs machine: Quantitative and discretionary equity management. Working paper, Columbia University.

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40:293–340.

Acemoglu, D. and Restrepo, P. (2018). Artificial intelligence, automation and work. Working Paper 24196, National Bureau of Economic Research.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2023). Do we want less automation? *Science*, 381(6654):155–158.

Anton, M., Cohen, R. B., and Polk, C. (2021). Best ideas. *Available at SSRN 1364827*.

Autor, D., Chin, C., Salomons, A. M., and Seegmiller, B. (2022). New frontiers: The origins and content of new work, 1940–2018. Technical report, National Bureau of Economic Research.

Bae, K.-H., Stulz, R. M., and Tan, H. (2008). Do local analysts know more? a cross-country study of the performance of local analysts and foreign analysts. *Journal of financial economics*, 88(3):581–606.

Ball, R. and Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of accounting research*, pages 159–178.

Barras, L., Gagliardini, P., and Scaillet, O. (2022). Skill, scale, and value creation in the mutual fund industry. *The Journal of Finance*, 77(1):601–638.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative ai at work. Technical report, National Bureau of Economic Research.

Chen, Y., Kelly, B., and Wu, W. (2020). Sophisticated investors and market efficiency: Evidence from a natural experiment. *Journal of Financial Economics*, 138(2):316–341.

Chevalier, J. and Ellison, G. (1999). Are some mutual fund managers better than others? cross-sectional patterns in behavior and performance. *The Journal of Finance*, 54(3):875–899.

Chi, F., Hwang, B.-H., and Zheng, Y. (2022). The use and usefulness of big data in finance. Working paper, Cornell University.

Cicero, D. C., Puckett, A., Wang, A. Y., and Zhang, S. (2023). Taxi! do mutual funds pursue and exploit information on local companies? Forthcoming Journal of Financial and Quantitative Analysis, Auburn University.

Cohen, L., Frazzini, A., and Malloy, C. (2008). The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy*, 116(5):951–979.

Coval, J. D. and Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, 109(4):811–841.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, 52(3):1035–1058.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Dessaint, O., Foucault, T., and Fresard, L. (2022). Does alternative data improve financial forecasting? the horizon effect. Journal of Finance, Forthcoming.

Dugast, J. and Foucault, T. (2022). Equilibrium data mining and data abundance. Forthcoming Journal of Finance.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

Fama, E. and French, K. (1992). The cross section of expected stock returns. *Journal of Finance*, 47(2):427–465.

Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, 10(1):1–21.

Fama, E. F. and French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, 65(5):1915–1947.

Gerken, W. C. and Painter, M. O. (2022). The Value of Differing Points of View: Evidence from Financial Analysts' Geographic Diversity. *The Review of Financial Studies*, 36(2):409–449.

Grennan, J. and Michaely, R. (2020). Artificial intelligence and high-skilled work: Evidence from analysts. Working paper, Swiss Finance Institute.

Grennan, J. and Michaely, R. (2021). Fintechs and the market for financial analysis. *Journal of Financial and Quantitative Analysis*, 56:1877–1907.

Grinblatt, M. and Titman, S. (1993a). Performance measurement without benchmark: An examination of mutual fund returns. *Journal of Business*, 66:47–68.

Grinblatt, M. and Titman, S. (1993b). Performance measurement without benchmarks: An examination of mutual fund returns. *Journal of business*, pages 47–68.

Hui, X., Reshef, O., and Zhou, L. (2023). The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Available at SSRN 4527336*.

Jiang, H. and Zheng, L. (2018). Active fundamental performance. *The Review of Financial Studies*, 31(12):4688–4719.

Kacperczyk, M., Sialm, C., and Zheng, L. (2005). On the industry concentration of actively managed equity mutual funds. *The Journal of Finance*, 60(4):1983–2011.

Kacperczyk, M., van Nieuwerburgh, S., and Veldkamp, L. (2014). Time-varying fund manager skills. *Journal of Finance*, 69:1455–1483.

Kanazawa, K., Kawaguchi, D., Shigeoka, H., and Watanabe, Y. (2022). Ai, skill, and productivity: The case of taxi drivers. Technical report, National Bureau of Economic Research.

Kang, J. K. (2022). Gone with the big data: Institutional lender demand for private information. Working paper, Harvard Business School.

Kang, J. K., Stice-Lawrence, L., and Wong, Y. T. F. (2021). The firm next door: Using satellite images to study local information advantage. *Journal of Accounting Research*, 59(2):713–750.

Katona, Z., Painter, M., Patatoukas, P. N., and Zeng, J. (2023). On the capital market consequences of alternative data: Evidence from outer space. Forthcoming Journal of Financial and Quantitative Analysis.

Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2006). Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis. *The Journal of finance*, 61(6):2551–2595.

Lee, C. M. and So, E. C. (2017). Uncovering expected returns: Information in analyst coverage proxies. *Journal of Financial Economics*, 124(2):331–348.

Liu, C., Qiu, Y., Wang, S., and Yeung, E. (2023). Watching from the sky: Business observability and voluntary disclosure. Working paper, Columbia University.

Narang, R. (2013). *Inside the Black Box: A simple guide to quantitative and high-frequency trading*. Wiley, New-York.

Pastor, L., Stambaugh, R., and Taylor, L. (2015). Scale and skill in active management. *The Journal of Financial Economics*, 116:23–45.

Stambaugh, R. (2020). Skill and profits in active management. Working paper, University of Pennsylvania.

Verrecchia, R. (1982). Information acquisition in a noisy rational expectations economy. *Econometrica*, pages 1415–1430.

Weller, B. M. (2018). Does algorithmic trading reduce information acquisition? *The Review of Financial Studies*, 31(6):2184–2226.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance*, 55(4):1655–1695.

Zhao, J. (2021). Who benefits (more) from the growth of data? Working paper, Bayes Business School.

Zhu, C. (2019). Big data as a governance mechanism. *Review of Financial Studies*, 32(5):2021–2061.

Zhu, M. (2018). Informative fund size, managerial skill, and investor rationality. *Journal of Financial Economics*, 130.

Tables

Table I: Sample Descriptive Statistics

This table presents descriptive statistics for the main variables employed in our main analysis. The sample includes 3,962 funds and 9,781 distinct stocks between 2009 and 2021. In Panel B, Market Cap., Book to Market, Assets, Leverage, price informativeness measures ($ACAR$ and CAR measures), and analysts' earning forecast measures ("Std. FE") are reported for the sample of stocks used in our regressions studying price informativeness in Section VI, i.e., stocks in the industries covered by RS Metrics (cf. Appendix Table B.1).

Panel A: Fund-Quarter level

	Obs	Mean	Sd	5%	25%	50%	75%	95%
TNA (mm)	108,439	1,475.469	5,957.176	13.300	76.700	281.000	982.100	5,654.900
Nb. Stocks	108,439	123.188	211.065	29.000	47.000	74.000	115.000	361.000
Age (Years)	108,439	16.960	12.544	2.333	8.833	15.250	21.583	36.667
Return	108,439	0.006	0.050	-0.086	-0.015	0.010	0.032	0.076
Flow	107,744	0.280	30.791	-0.053	-0.015	-0.006	0.004	0.057

Panel B: Stock-Quarter level

	Obs	Mean	Sd	5%	25%	50%	75%	95%
Beta	229,263	1.007	0.559	0.141	0.674	0.995	1.323	1.916
Return	229,263	0.012	0.152	-0.190	-0.049	0.008	0.064	0.212
Nb. Funds Holding	229,263	57.434	78.070	1.000	6.000	30.000	79.000	202.000
Market Cap. (bn)	58,174	7.599	35.307	0.037	0.222	0.850	3.224	30.000
Book to Market	58,174	0.580	0.606	0.082	0.247	0.438	0.736	1.488
Assets (bn)	58,174	5.612	21.555	0.037	0.190	0.706	2.871	21.965
Leverage	58,174	0.030	0.062	0.000	0.000	0.007	0.031	0.141
$ACAR[0,2]$	58,174	0.069	0.076	0.004	0.021	0.048	0.092	0.204
$ACAR[-1,2]$	58,174	0.072	0.079	0.004	0.022	0.049	0.095	0.209
$CAR[0,2]/CAR[-21,2]$	19,160	0.479	0.461	-0.235	0.185	0.478	0.768	1.205
$CAR[-1,2]/CAR[-21,2]$	19,160	0.503	0.460	-0.216	0.212	0.507	0.792	1.223
Std. FE 1-Q	44,197	2.487	2.686	0.000	0.680	1.600	3.250	8.500
Std. FE 2-Q	44,541	2.723	2.942	0.061	0.758	1.800	3.500	9.000
Std. FE 3-Q	42,355	3.190	3.660	0.111	0.897	2.000	4.000	11.000
Std. FE 4-Q	39,352	3.502	4.103	0.143	1.000	2.000	4.400	12.400
Std. FE 1-Y	11,021	2.441	2.872	0.000	0.600	1.500	3.000	9.000

Panel C: Holding (Fund-Stock-Quarter) level

	Obs	Mean	Sd	5%	25%	50%	75%	95%
Picking 1-Q	1.28e+07	0.000	0.162	-0.188	-0.024	-0.000	0.021	0.187
Picking 2-Q	1.24e+07	-0.001	0.242	-0.280	-0.037	-0.000	0.029	0.271
Picking 3-Q	1.20e+07	-0.003	0.310	-0.356	-0.049	-0.000	0.035	0.340
Picking 4-Q	1.16e+07	-0.005	0.361	-0.424	-0.061	-0.001	0.039	0.400
Trading 1-Q	1.28e+07	0.001	0.082	-0.060	-0.004	0.000	0.004	0.060
Trading 2-Q	1.24e+07	0.000	0.123	-0.089	-0.006	0.000	0.006	0.087
Trading 3-Q	1.20e+07	0.000	0.156	-0.115	-0.008	0.000	0.007	0.109
Trading 4-Q	1.16e+07	-0.000	0.182	-0.136	-0.009	0.000	0.008	0.127

Table II: Alternative Data and Stock Picking Skills

This table presents our main results on the effect of the release of alternative data on fund picking abilities. Regressions are estimated at the fund-stock-quarter level. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (3). *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. *Covered* $\times$ *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. Standard errors are double-clustered at the fund and stock levels.

	Picking 1-Q			Picking 2-Q			Picking 3-Q			Picking 4-Q		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Covered $\times$ Post	-0.017***	-0.021***	-0.021**	-0.032**	-0.042***	-0.038**	-0.052**	-0.067***	-0.058*	-0.080***	-0.106***	-0.098**
	(0.006)	(0.007)	(0.008)	(0.013)	(0.016)	(0.019)	(0.021)	(0.025)	(0.032)	(0.030)	(0.036)	(0.047)
Covered	0.024***			0.048***			0.075***			0.107***		
	(0.005)			(0.010)			(0.016)			(0.023)		
Fund $\times$ Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Fund $\times$ Stock FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	1.28e+07	1.28e+07	1.28e+07	1.24e+07	1.24e+07	1.24e+07	1.20e+07	1.20e+07	1.20e+07	1.16e+07	1.16e+07	1.16e+07
R^2	0.08	0.10	0.20	0.08	0.12	0.28	0.09	0.14	0.34	0.08	0.15	0.39

Table III: Heterogeneous Effects based on Skills before Coverage Initiation

The table displays the results of our study on how the impact of alternative data varies depending on a fund's ability to pick stocks before the release of satellite data imagery by RS Metrics. Regressions are estimated at the fund-stock-quarter level. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (3). *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. *Covered* $\times$ *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. The table presents estimation results of specifications that include interactions with a dummy variable, "High Picking Pre", which equals one if the stock is covered by RS Metrics and the fund has a picking ability above the median for that stock before the release of satellite data imagery. The regressions in the table do not include the picking skills for stocks covered by RS Metrics for funds that start holding the stock after the release of satellite data imagery. In other words, we only analyze the effect of alternative data on funds that had a certain level of stock-picking ability before the satellite data imagery was released. All picking skills for uncovered stocks are included. Standard errors are double-clustered at the fund and stock levels.

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	0.002	0.012	0.024	0.018
	(0.005)	(0.010)	(0.015)	(0.019)
Covered × Post × High Picking Pre	-0.053***	-0.113***	-0.183***	-0.253***
	(0.013)	(0.030)	(0.053)	(0.080)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.27e+07	1.23e+07	1.19e+07	1.15e+07
R^2	0.20	0.28	0.34	0.39

Table IV: Heterogeneous Effect based on Industry Expertise

The table presents the results of our study on the differential impact of alternative data on funds with industry expertise. Regressions are estimated at the fund-stock-quarter level. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (3). *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. *Covered* × *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. Panel A presents estimation results of specifications that include interactions with a dummy variable, "Industry Specialist", which equals one if the fund has on average more than 75% of its assets invested in stocks that belong to covered industries. Covered industries are NAICS sectors in which RS Metrics covers at least one company (cf., Appendix Table B.1). Panel B presents estimation results of specifications that include interactions with a dummy variable, "Sector Fund", which equals one if the fund is classified as a sector fund by CRSP, i.e., invest primarily in a given sector. Standard errors are double-clustered at the fund and stock levels.

Panel A: Industry Specialists

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.018***	-0.015**	-0.033***	-0.021*	-0.053***	-0.031	-0.084***	-0.056**
	(0.005)	(0.006)	(0.011)	(0.012)	(0.018)	(0.019)	(0.026)	(0.027)
Covered × Post × Industry Specialist	-0.140**	-0.161***	-0.347**	-0.419***	-0.576**	-0.711***	-0.866**	-1.086***
	(0.065)	(0.054)	(0.165)	(0.142)	(0.275)	(0.235)	(0.406)	(0.358)
Covered × Industry Specialist	0.105***		0.250***		0.399***		0.568***	
	(0.035)		(0.085)		(0.132)		(0.179)	
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	No	Yes	No	Yes	No	Yes	No
Fund × Stock FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1.28e+07	1.28e+07	1.24e+07	1.24e+07	1.20e+07	1.20e+07	1.16e+07	1.16e+07
R^2	0.10	0.20	0.12	0.28	0.14	0.34	0.15	0.39

Panel B: Sector Funds

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.018***	-0.015**	-0.033***	-0.022*	-0.054***	-0.031	-0.086***	-0.056**
	(0.005)	(0.006)	(0.012)	(0.012)	(0.019)	(0.019)	(0.026)	(0.027)
Covered × Post × Sector Fund	-0.103	-0.132**	-0.256	-0.342**	-0.428	-0.583**	-0.635*	-0.888**
	(0.064)	(0.053)	(0.159)	(0.142)	(0.261)	(0.236)	(0.385)	(0.361)
Covered × Sector Fund	0.089**		0.213**		0.342**		0.484***	
	(0.036)		(0.085)		(0.133)		(0.183)	
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	No	Yes	No	Yes	No	Yes	No
Fund × Stock FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1.28e+07	1.28e+07	1.24e+07	1.24e+07	1.20e+07	1.20e+07	1.16e+07	1.16e+07
R^2	0.10	0.20	0.12	0.28	0.14	0.34	0.15	0.39

Table V: Heterogeneous Effect based on Geographical Location

The table presents the results of our study on the differential impact of alternative data on fund picking abilities depending on fund location. Regressions are estimated at the fund-stock-quarter level. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (3). *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. *Covered* × *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. The table presents estimation results of specifications that include interactions with a dummy variable, "Local", which equals one if the fund is located in the same MSA as either (i) the firm's headquarters or (ii) the stock's primary MSA based on parking lots, as identified through satellite imagery data (the MSA where the highest number of the firm's parking lots are located). The regressions in the table do not include the picking skills for funds for which we are unable to obtain the official address from the CRSP database. Standard errors are double-clustered at the fund and stock levels.

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.019***	-0.017**	-0.038***	-0.031*	-0.062***	-0.048*	-0.097***	-0.081**
	(0.007)	(0.007)	(0.014)	(0.016)	(0.023)	(0.026)	(0.032)	(0.038)
Covered × Post × Local	-0.038**	-0.057***	-0.077**	-0.116**	-0.116**	-0.181**	-0.167*	-0.271**
	(0.015)	(0.019)	(0.037)	(0.055)	(0.059)	(0.087)	(0.085)	(0.126)
Covered × Local	0.027**		0.056**		0.082**		0.115**	
	(0.011)		(0.024)		(0.038)		(0.054)	
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	No	Yes	No	Yes	No	Yes	No
Fund × Stock FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	9972770	9972770	9681321	9681321	9383392	9383392	9083207	9083207
R^2	0.10	0.21	0.12	0.29	0.14	0.34	0.15	0.39

Table VI: Divestment from Covered Stocks

The table presents the results of our study on the impact of alternative data on fund holdings. Regressions in columns (1) and (2) are estimated at the fund-stock-quarter level and the dependent variable is the natural logarithm of the stock rank in the fund portfolio. To facilitate interpretation, we use the negative of the logarithm of rank in the regression, whereby larger values correspond to the largest investments. Regressions in columns (3) and (4) are estimated at the stock-quarter level and the dependent variable is the logarithm of the number of funds holding the stock in a given quarter. Column (3) encompasses all stocks in our sample, while column (4) focuses solely on stocks within the covered industries. Covered industries are NAICS sectors in which RS Metrics covers at least one company (cf., Appendix Table B.1). *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. *Covered* × *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. Standard errors are clustered at the fund and stock levels in columns (1) and (2), and at the stock level in columns (3) and (4).

	Stock Rank		Nb. Funds Holding the Stock	
	(1)	(2)	(3)	(4)
Covered × Post	-0.084**	-0.086**	-0.167***	-0.200***
	(0.037)	(0.038)	(0.064)	(0.065)
Fund × Year-Quarter FE	Yes	Yes	No	No
Year-Quarter FE	No	No	Yes	Yes
Stock FE	Yes	No	Yes	Yes
Fund × Stock FE	No	Yes	No	No
Only Stocks in Covered Industries	No	No	No	Yes
Observations	1.28e+07	1.28e+07	229,263	101,784
R^2	0.72	0.87	0.89	0.87

Table VII: Divestment of Experts from Covered Stocks

The table presents the results of our study on the impact of alternative data on fund holdings depending on fund expertise. We investigate this by examining the investment-size rank of covered stocks in portfolios of funds managed by experts. Regressions are estimated at the fund-stock-quarter level and the dependent variable is the natural logarithm of the stock rank in the fund portfolio. To facilitate interpretation, we use the negative of the logarithm of rank in the regression, whereby larger values correspond to the largest investments. *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. $Covered \times Post$ is a dummy equal to one after RS Metrics initiates coverage of the stock. In column (1), "High Picking Pre" is a dummy variable that equals one if the stock is covered by RS Metrics and the fund has a picking ability above the median for that stock before the coverage starts. In column (2), "Industry Specialist" is a dummy variable that equals one if the fund has on average more than 75% of its assets invested in stocks that belong to covered industries. Covered industries are NAICS sectors in which RS Metrics covers at least one company (cf., Appendix Table B.1). In column (3), "Sector Fund" is a dummy variable that equals one if the fund is classified as a sector fund by CRSP, i.e., invest primarily in a given sector. In column (4), "Local" is a dummy variable that equals one if the fund is located in the same MSA as either (i) the firm's headquarters or (ii) the stock's primary MSA based on parking lots, as identified through satellite imagery data (the MSA where the highest number of the firm's parking lots are located). Standard errors are double-clustered at the fund and stock levels.

	Stock Rank			
	(1)	(2)	(3)	(4)
Covered × Post	0.005	-0.072**	-0.071**	-0.068*
	(0.050)	(0.035)	(0.035)	(0.035)
Covered × Post × High Picking Pre	-0.198***			
	(0.054)			
Covered × Post × Industry Specialist		-0.378**		
		(0.176)		
Covered × Post × Sector Fund			-0.331**	
			(0.154)	
Covered × Post × Local				-0.150
				(0.128)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes
Observations	1.27e+07	1.28e+07	1.28e+07	9972770
R^2	0.87	0.87	0.87	0.87

Table VIII: Investment in Peers of Covered Stocks

The table presents the results of our study on the impact of alternative data on fund holdings of peer stocks. We investigate this by examining the number of funds holding non-covered stocks that are in the same industry or same geographical area as covered stocks. Regressions are estimated at the stock-quarter level and include only control (non-covered) stocks. The dependent variable is the logarithm of the number of funds holding the stock in a given quarter. "Industry Peer Covered × Post" is a dummy equal to one if RS Metrics initiates coverage of a stock in the same industry (2-digit NAICS sector) as the focal stock. "Local Peer Covered × Post" is a dummy equal to one if RS Metrics initiates coverage of a stock whose highest number of parking lots are in the same Metropolitan Statistical Area (MSA) as the headquarter of the focal stock. Standard errors are clustered at the stock level.

	Nb. Funds Holding the Stock		
	(1)	(2)	(3)
Industry Peer Covered × Post	0.112***		0.109***
	(0.030)		(0.030)
Local Peer Covered × Post		0.086***	0.084***
		(0.027)	(0.027)
Year-Quarter FE	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
Observations	227,295	227,295	227,295
R^2	0.89	0.89	0.89

Table IX: Alternative Data and Price Informativeness

The table presents the results of our study on the impact of alternative data on stock price informativeness. Regressions are estimated at the stock-quarter level. In columns (1) and (2), the dependent variable is the jump ratio. $CAR[-m, n]$ represents the cumulative abnormal return from the mth day prior to earnings announcements to the nth day after earnings announcements. We estimate abnormal returns relative to a Fama and French (1992) three-factor model, using daily returns over a 252-day window ending 90 days before the earnings announcement. Consistent with Weller (2018), we require $ACAR[-21, 2] > \sqrt{24}\sigma$, where σ denotes the daily volatility of the stock during the month preceding the earnings announcement period. In columns (3) and (4), the dependent variable is the absolute cumulative abnormal return. $ACAR[-m, n]$ represents the absolute cumulative abnormal return from the mth day prior to earnings announcements to the nth day after earnings announcements. *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. *Covered* $\times$ *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. Control variables include the one-quarter lagged log of market capitalization, log of the book-to-market ratio, leverage (debt over assets), and log of assets, as well as the log of the daily volatility of the stock during the month preceding the earnings announcement period. All regressions focuses solely on stocks within the covered industries. Covered industries are NAICS sectors in which RS Metrics covers at least one company (cf., Appendix Table B.1). Standard errors are clustered at the stock level.

	$\frac{CAR[-1,2]}{CAR[-21,2]}$	$\frac{CAR[0,2]}{CAR[-21,2]}$	$ACAR[-1, 2]$	$ACAR[0, 2]$
	(1)	(2)	(3)	(4)
Covered × Post	-0.089**	-0.089**	-0.005**	-0.004*
	(0.044)	(0.036)	(0.003)	(0.002)
Control Variables	Yes	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Observations	19,160	19,160	58,174	58,174
R^2	0.22	0.23	0.36	0.37

Figures

Figure I: Number of Satellite Data Releases over Time

This figure reports the number of stocks for which RS Metrics initiates coverage in each quarter. The sample includes U.S. retail firms whose satellite imagery data of parking lot traffic are released by RS Metrics from 2009 to 2017.

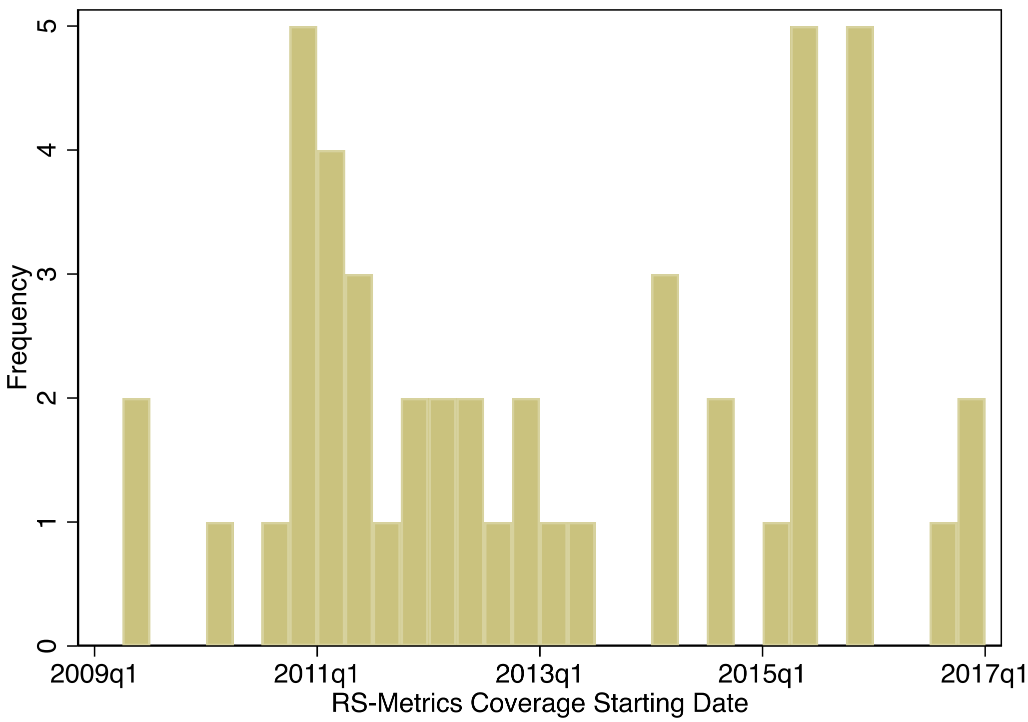


Figure II: Alternative Data and Fund Picking Abilities

This figure reports the dynamic effect of the release of alternative data on fund picking abilities. The specification corresponds to equation (6). Each circle corresponds to the coefficient on the interaction of "Covered" and a specific quarter dummy. Dashed lines are 95% confidence intervals. The dependent variable is Picking 4-Q, defined in equation (3). Standard errors are clustered at the fund and stock levels.

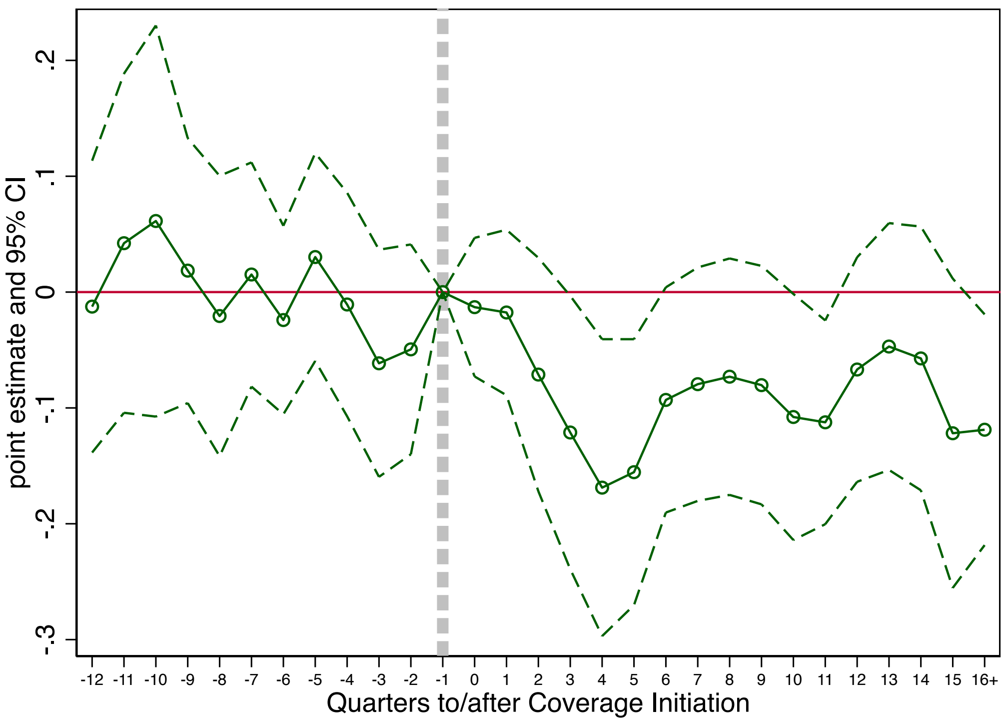
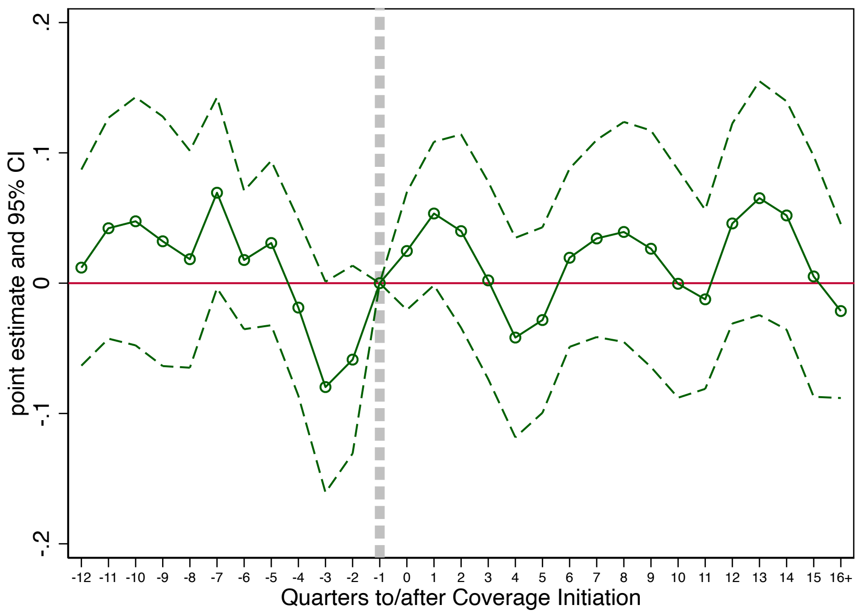


Figure III: Fund Picking Abilities depending on Skills before Coverage Initiation

This figure reports the dynamic effect of the release of alternative data on fund picking abilities. The specification corresponds to equation (8). In Panel A, each circle corresponds to the coefficient on the interaction between "Covered" and a specific quarter dummy. In Panel B, each circle corresponds to the coefficient on the triple interaction between "Covered", "High Picking Pre" and a specific quarter dummy. Dashed lines are 95% confidence intervals. The dependent variable is Picking 4-Q, defined in equation (3). Standard errors are clustered at the fund and stock levels.

Panel A: Inferior Stock Pickers before Coverage Initiation



Panel B: Superior Stock Pickers before Coverage Initiation

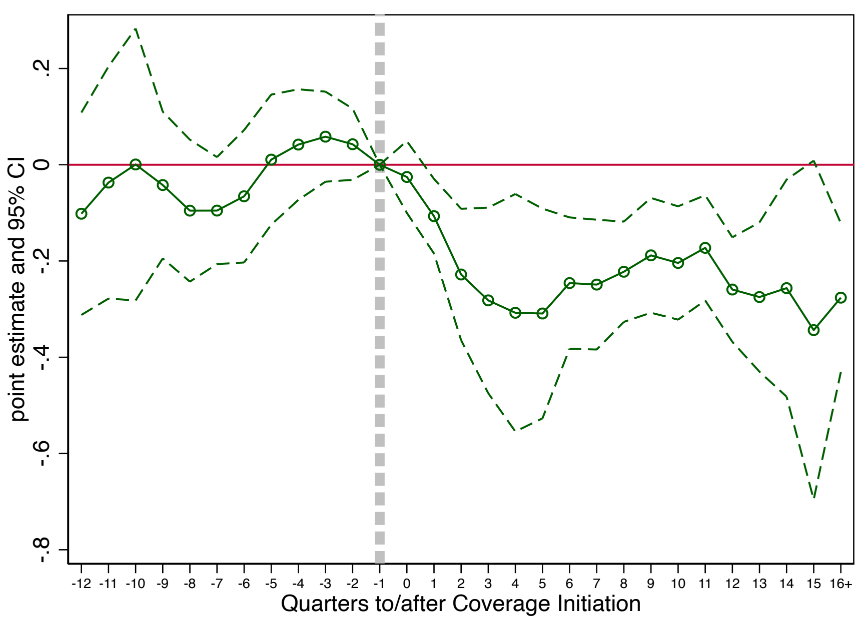
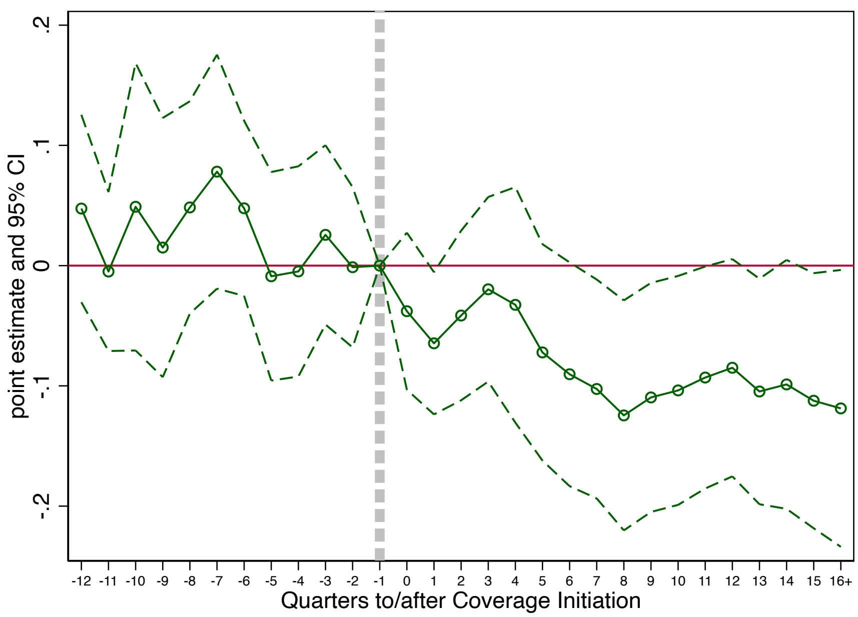


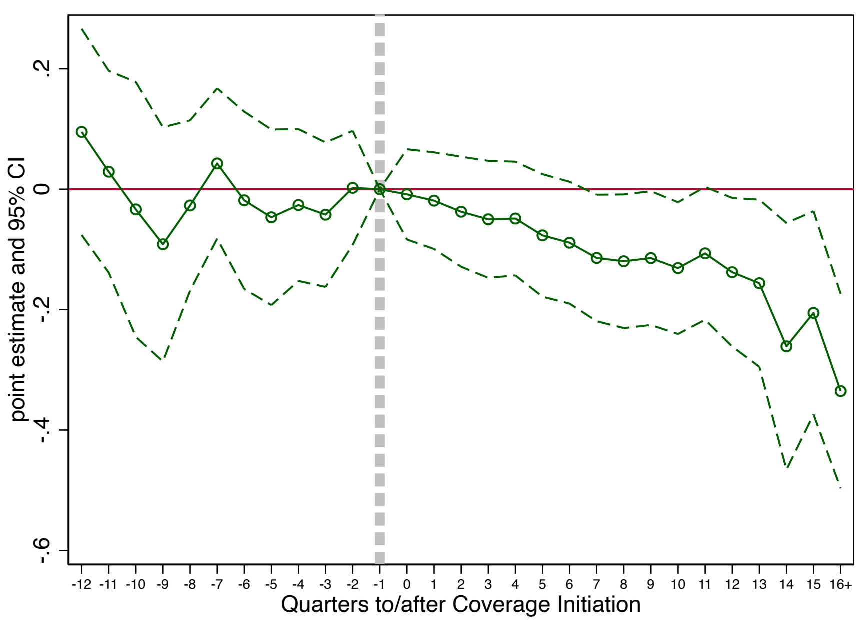
Figure IV: Alternative Data and Fund Divestment

This figure reports the dynamic effect of the release of alternative data on the stock rank in fund portfolio and the number of funds holding the stock. In Panel A, the specification is an event study estimated at the fund-stock-quarter level. The dependent variable is the natural logarithm of the stock rank in the fund portfolio. To facilitate interpretation, we use the negative of the logarithm of rank in the regression, whereby larger values correspond to the largest investments. In Panel B, the specification is an event study estimated at the stock-quarter level. The dependent variable is the logarithm of the number of funds holding the stock in a given quarter. In both panels, each circle corresponds to the coefficient on the interaction between "Covered" and a specific quarter dummy. Dashed lines are 95% confidence intervals. Standard errors are clustered at the fund and stock levels in Panel A, and at the stock level in Panel B.

Panel A: Stock Rank in Fund Portfolio



Panel B: Number of Funds Holding the Stock

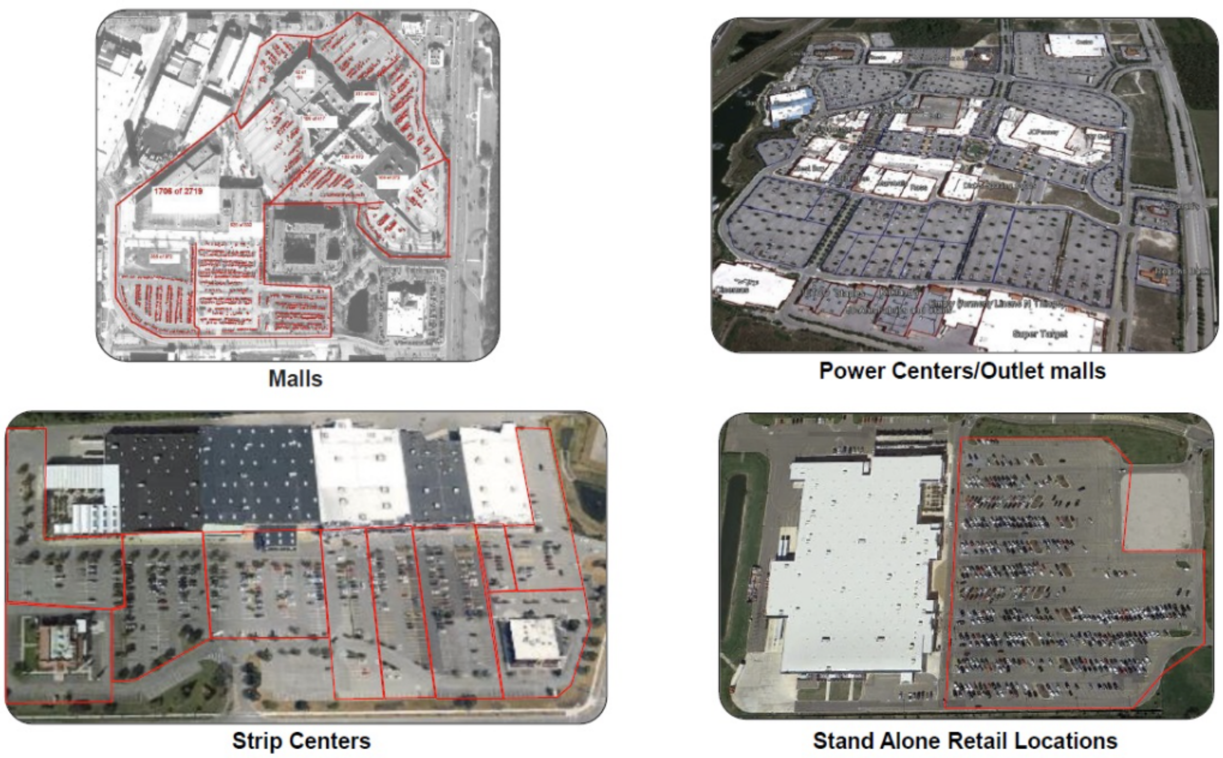


Appendix

A Figures

Figure A.1: Examples of Satellite Images Exploited by RS Metrics

This figure reports four examples of satellite images processed by RS Metrics to determine vehicle counts at parking lots and the actual parking lot size of retail stores. Each location is monitored with a multiple times a month frequency. Source: https://learn.rsmetrics.com/trafficsignals/retail/monitoring.



B Tables

Table B.1: Covered Stocks in each Industry

This table presents the industries of stocks covered by RS Metrics. Each line corresponds to a 2-digit NAICS sector and indicates the number of stocks eventually covered.

NAICS Sector	Description	Nb. Covered Stocks
11	Agriculture, Forestry, Fishing and Hunting	0
21	Mining, Quarrying, and Oil and Gas Extraction	0
22	Utilities	0
23	Construction	0
31-33	Manufacturing	1
42	Wholesale Trade	1
44-45	Retail Trade	38
48-49	Transportation and Warehousing	0
51	Information	0
52	Finance and Insurance	0
53	Real Estate and Rental and Leasing	2
54	Professional, Scientific, and Technical Services	0
55	Management of Companies and Enterprises	0
56	Administrative and Support and Waste Management and Remediation Services	0
61	Educational Services	0
62	Health Care and Social Assistance	0
71	Arts, Entertainment, and Recreation	0
72	Accommodation and Food Services	5
81	Other Services (except Public Administration)	1
92	Public Administration (not covered in economic census)	0

Table B.2: Stock-level Characteristics and Coverage Initiation

This table reports the association of different stock-level characteristics with coverage initiation by RS Metrics. The data used for this analysis include right-censored stock-quarter observations, i.e., observations up to the coverage initiation for covered stocks and all available observations for control stocks that are never covered by RS Metrics. The dependent variable is an indicator equal to 100 in the first quarter of coverage initiation, zero otherwise. "Idiosyncratic Vol." is the volatility of the stock return adjusted for market beta compused using daily returns over the last 252 days. "Analyst FE 1-Q" is the average over the prior year of the absolute value of next quarter's actual earnings minus the average of the most recent analyst forecasts, divided by the standard deviation of those forecasts. $ACAR[-m, n]$ represents the absolute cumulative abnormal return from the mth day prior to earnings announcements to the nth day after earnings announcements. $CAR[-m, n]$ represents the cumulative abnormal return from the mth day prior to earnings announcements to the nth day after earnings announcements. "Industry FE" correspond to 2-digit NAICS sector fixed effects. Standard errors are clustered at stock level.

	Coverage Initiation × 100					
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Market Cap.)	0.041***	0.045***	0.047***	0.047***	0.048***	0.049***
	(0.013)	(0.014)	(0.015)	(0.015)	(0.015)	(0.016)
Log(Assets)	-0.017	-0.017	-0.017	-0.017	-0.019	-0.019
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Log(Book-to-Market)	0.005	0.006	0.006	0.006	0.006	0.006
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Stock Return (Year -1)		0.021	0.021	0.021	0.027	0.028
		(0.029)	(0.029)	(0.030)	(0.031)	(0.031)
Earnings (Year -1)		-0.013	-0.012	-0.012	-0.012	-0.012
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Log(Idiosyncratic Vol.)			0.009	0.009	0.009	-0.002
			(0.022)	(0.022)	(0.022)	(0.020)
Analyst FE 1-Q (Year -1)			0.002	0.002	0.001	0.002
			(0.003)	(0.003)	(0.003)	(0.003)
Log(Nb. Funds Holding)				0.001	0.001	-0.001
				(0.014)	(0.014)	(0.015)
Picking 1-Q (Year -1)					-0.133	-0.134
					(0.269)	(0.270)
$\frac{CAR[-1,2]}{CAR[-21,2]}$ (Year -1)						0.000
						(0.000)
$ACAR[-1, 2]$ (Year -1)						0.176
						(0.230)
Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	62,820	62,820	62,820	62,820	62,820	62,820
R^2	0.01	0.01	0.01	0.01	0.01	0.01

Table B.3: Controlling for Reversal in Picking Abilities

The table presents the results of our study controlling for potential reversal in picking abilities. Regressions are estimated at the fund-stock-quarter level. The dependent variable is *Picking* calculated at different horizons ranging from one quarter to one year, and is defined in equation (3). *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. $Covered \times Post$ is a dummy equal to one after RS Metrics initiates coverage of the stock. The table presents estimation results of specifications that include interactions with a dummy variable, "High Picking Pre", which equals one if the stock is covered by RS Metrics and the fund has a picking ability above the median for that stock before the release of satellite data imagery. All specifications include a control variable "Past Picking" corresponding to the lagged value of the dependent variable. The lag is defined such that the stock return used to computed "Past Picking" does not intersect with the current quarter. Specifically, in columns (1) and (2), "Past Picking" corresponds to Picking 1-Q in the previous quarter. In columns (3) and (4), it corresponds to Picking 2-Q lagged by two quarters. In columns (5) and (6), it corresponds to Picking 3-Q lagged by three quarters. In columns (7) and (8), it corresponds to Picking 4-Q lagged by four quarters. Standard errors are double-clustered at the fund and stock levels.

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.026***	-0.002	-0.055**	-0.000	-0.084*	0.008	-0.124*	0.009
	(0.010)	(0.005)	(0.026)	(0.010)	(0.045)	(0.015)	(0.073)	(0.022)
Covered × Post × High Picking Pre		-0.057***		-0.123***		-0.200**		-0.285**
		(0.016)		(0.042)		(0.079)		(0.129)
Past Picking	-6.543***	-6.525***	-11.817***	-11.731***	-17.194***	-17.116***	-20.745***	-20.661***
	(1.480)	(1.495)	(2.077)	(2.093)	(1.193)	(1.201)	(1.408)	(1.424)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fund × Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1.13e+07	1.12e+07	9732394	9628513	8366435	8279047	7175094	7101982
R^2	0.20	0.21	0.29	0.29	0.36	0.36	0.41	0.41

Table B.4: Alternative Data and Analysts' Forecast Errors

The table presents the results of our study on the impact of alternative data on analysts' earnings forecast errors. Regressions are estimated at the stock-quarter level. The dependent variable is standardized analysts' forecasting error calculated at different horizons: for the subsequent quarter (column 1), for the second (column 2), third (column 3), and fourth quarter (column 4) ahead, as well as for the forthcoming year (column 5). We compute the standardized analysts' earnings forecast errors as the absolute value of the corresponding actual earnings minus the average of the most recent analyst forecasts, divided by the standard deviation of those forecasts. *Covered* is a dummy equal to one if the stock is eventually covered by RS Metrics. *Covered* × *Post* is a dummy equal to one after RS Metrics initiates coverage of the stock. Control variables include the one-quarter lagged log of market capitalization, log of the book-to-market ratio, leverage (debt over assets), and log of assets, as well as the log of the daily volatility of the stock during the month preceding the earnings announcement period. All regressions focuses solely on stocks within the covered industries. Covered industries are NAICS sectors in which RS Metrics covers at least one company (cf., Appendix Table B.1). Standard errors are clustered at the stock level.

	Std. FE 1-Q	Std. FE 2-Q	Std. FE 3-Q	Std. FE 4-Q	Std. FE 1-Y
	(1)	(2)	(3)	(4)	(5)
Covered × Post	-0.059	-0.199	-0.065	0.192	-0.154
	(0.191)	(0.206)	(0.255)	(0.266)	(0.266)
Control Variables	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Observations	44,197	44,541	42,355	39,352	11,021
R^2	0.14	0.12	0.15	0.16	0.23