Displaced by Big Data: Evidence from Active Fund Managers

Maxime Bonelli and Thierry Foucault*

November 21, 2023

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 possibility, 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 that are more likely to rely on traditional sources of expertise (e.g., specialized industry knowledge) to generate their signals, leading them to divest from covered stocks. These results suggest that big data has the potential to displace high-skill workers in finance.

Keywords: Big data, active mutual funds, stock-picking skill, quantitative investment.

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 Augustin landier and seminar participants at Toulouse Business School 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intelligence (AI).[1] As in other industries, this revolution poses a risk to workers whose skills may become outdated. A case in point (see our opening quote) are active asset managers.[2] 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.[3] 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, such data should enable asset managers to obtain more precise signals about stock returns, provided that they have the required skills for exploiting

[1] Acemoglu et al. (2022) (Figure 2) shows that the finance/insurance sector is the most exposed to artificial intelligence.

[2] 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*.

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

quantitative data. As these managers trade on their information, stock prices become more informative, reducing the return on private information for asset managers relying on more traditional skills (e.g., expert knowledge of a particular industry; see Section II). If so, we should observe a decline in the stock picking ability of traditional asset managers (a measure of their skill) when new alternative data becomes available for the stocks they have developed specific expertise in ("the displacement hypothesis").

To test this hypothesis, we study the evolution of active mutual funds' stock picking ability in retailers' stocks after a new source of information becomes available for these stocks, namely daily store-level car counts based on satellite imagery. Satellite imagery is one important type of alternative data, used by asset managers.[4] It has predictive power for firms' future cash-flows or stock returns (see Zhu, 2019; Kang et al., 2021; Katona et al., 2019) and it enhances price informativeness (Zhu, 2019). Thus, using stock coverage initiation by a satellite data provider (in our case RS Metrics) is a way to study whether the availability of alternative data affects asset managers' stock picking skills.

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. We observe the portfolio holdings of about 4,000 different active mutual funds between 2009 and 2021. With this data, in each quarter, we measure a fund 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 in the weight of this stock in the fund's portfolio from its weight in a benchmark portfolio, which we define as the market portfolio.[5] Intuitively, this measure captures a fund's ability to tilt its stock

[4] There is anecdotal evidence that some funds buy data obtained from satellite images. 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.

[5] A similar measure, but a the fund level, is used by Kacperczyk et al. (2014). See also Grinblatt and Titman (1993a), Daniel et al. (1997) and Jiang and Zheng (2018) for related measures of a fund's skill. All our findings result are unchanged when we use other measures of funds' stock picking ability, namely using the S&P 500 index (rather than the market portfolio) as the benchmark portfolio or the change in the weight

holding in the direction of future return adjusted for systematic risk. We test whether funds' stock picking ability in treated stocks declines after these stocks become covered by satellite imagery using a difference-in-differences specification controlling for 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 the 1-year horizon, *"Picking"* has a mean and median of zero, with a standard deviation of 0.36 (only 5% of the realizations for picking at the 1-year horizon are larger than 0.40). At this horizon, we find that *"Picking"* drops by 0.11 for covered stocks in our sample, 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). 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.

Our displacement hypothesis implies that this effect should be stronger for funds that have developed expertise in covered stocks using traditional methods for obtaining private signals. We test this implication by considering three different measures of expertise: (i) a fund's stock picking ability in a given stock prior coverage of this stock (higher ability reveals 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 stores 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

of a stock in the fund's portfolio rather than the deviation in the weight of this stock in the fund's portfolio from its weight in the market portfolio.

drop is more than twice larger for funds that have a high stock picking ability in covered stocks before these stocks start being covered. This is consistent with our hypothesis that asset managers who have developed specific expertise are the most at risk of being displaced by new data sources.

In contrast, we find significantly weaker effects for funds that are more diversified (hold more stocks) and have a lower volatility. This is important because Abis (2022) shows that such funds are more likely to be quantitative funds.[6] These funds have the required expertise to exploit alternative data and should therefore be less negatively affected (or even positively affected), as we find for funds that have the characteristics of quant funds.[7]

We also find that the number of funds holding covered stocks and the rank of these stocks in funds' portfolios (we rank stocks according to their weight in funds' portfolios) drop after they start being covered. Moreover, these effects are stronger for funds that we classify as experts.[8] These findings suggest that funds relying on traditional sources of expertise to generate their signals choose to exit (a form of displacement) the stocks in which the availability of alternative data reduces the value of their expertise.

According to our hypothesis, these results are due to the fact that the availability of alternative data makes the stock price of covered stocks more informative. To test whether this is the case, we use two (inverse) measures of stock price informativeness: (i) the absolute

[6]Quant funds are those who rely on big data and systematic statistical analysis, for instance using machine learning techniques, to find their trading signals (see Abis, 2022; Harvey et al., 2017; Narang, 2013). 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 and *"Igor Tulchinsky's Worldquant"*, Wall Street Journal, April 13, 2017 available at https://www.wsj.com/articles/with-125-ph-d-s-in-15-countries-a-quant-alpha-factory-hunts-for-investing-edge-1491471008.

[7]Our data do not give us the possibility to observe directly whether a fund is a quantitative fund or not.

[8]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.

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 contained less information. With either measure, we find that, as predicted, price informativeness significantly increases for the covered stocks after coverage initiation.

Finally, we show that the former findings are unlikely to be due to an improvement in public information after a stock becomes covered. Indeed, there is no significant change in sell-side equity analysts' earnings forecasts errors for stocks after coverage initiation by RS Metric.[9] Thus, the improvement in price informativeness (and the corresponding decline in funds' stock picking abilities in covered stocks) is more likely to stem from the fact that coverage enables some funds (those buying the data from RS Metric) to use a new source of private information.

Our paper contributes to the quickly growing literature on the effects of alternative data on 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.[10] Dessaint et al. (2022) find that the availability of new alternative data reduces (improves) the informativeness of equity analysts' long-term (short-term) forecasts. Our paper is more specifically related to those that focus on the information contained in satellite images and the effects of this information (e.g., Zhu, 2019; Katona et al., 2019; Kang et al., 2021; Gerken and Painter, 2022).[11] To

[9]We use the quality of sell-side analysts' forecasts as a proxy for the quality of public signals available to investors (see Lee and So (2017) and Chen et al. (2020) for evidence that that analysts provide valuable public information).

[10]Dessaint et al. (2022) survey more than 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.

[11]Katona et al. (2019) show how store-level car counts, from satellite imagery, for US retailers can be used to design profitable trading strategies ahead of earnings announcements. They also 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

our knowledge, our paper is the first to show that alternative data reduces funds' stock picking ability.[12] 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. This finding also suggests that the availability of alternative data reduces the value of traditional sources of 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; Kacperczyk et al., 2014). 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).[13] 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;

their geographical location. Kang et al. (2021) use a similar approach to show that institutional investors rely on local information to form their portfolios.

[12]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) 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).

[13]There is also a broader literature, less related to our paper, on the effect of fund size or industry size on asset managers' performance. See Berk and van Binsbergen (2015), Pastor et al. (2015), Barras et al. (2022) and Berk et al. (2020) for a survey.

Dell'Acqua et al., 2023).[14] Using online job vacancies on Burning Glass from 2010 to 2022 and various measures of AI exposure at firms' establishment levels, Acemoglu et al. (2022) finds that AI exposure is associated with a decline in previously sought skills and an increase in new skills. 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.

The rest of the paper is organized as follows. The next section presents the hypotheses tested in our paper. In Section III, we describe our data and our measures of fund managers' stock picking ability. Sections IV to VI present our main empirical findings while Section VII concludes.

II Hypotheses Development

Our main hypothesis is that the availability of new data for a stock (such as information obtained from satellite images data) reduces traditional asset managers' stock picking ability. This hypothesis follows from Dugast and Foucault (2022), who consider a rational expectations model with two types of asset managers, "experts" and "data miners", who differ in the way they produce information about the payoff of a risky asset. Experts receive signals about the asset payoff whose precision, τ, increases with their level of expertise, which is fixed (e.g., pre-determined by education or past industry experience).[15]. In contrast, "data miners" obtain their signals by exploring various datasets with statistical tools (e.g., machine learning techniques). This exploration yields a signal of precision τ distributed in $[\tau^*, \tau_{dm}^{max}]$, where τ^* is endogenous (determined by data miners' exploration intensity) and τ_{dm}^{max} (the precision of the best available signals for data miners) is determined by the set of available

[14] 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

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

data. They interpret experts as discretionary asset managers and data miners as quantitative asset managers ("quants").

The realized gross excess return, $R^g(s_\tau)$, of an asset manager who obtains a signal, s_τ of precision τ is:

$$R^g(s_\tau) = w^*(s_\tau) \times R^e, \tag{1}$$

where w^* is the fraction of the asset manager's capital in the risky asset, and R^e is the excess return on the risky asset in equilibrium. For a fixed asset manager's ability τ, Dugast and Foucault (2022) obtains (see their (see eq.(26)) that the average excess gross return of an asset manager is:

$$E(R^g(s_\tau)) = \frac{\tau}{\rho^r \sigma_\omega^2 \mathcal{I}(\tau^*; \tau_{dm}^{max})}, \tag{2}$$

where ρ^r is relative risk aversion, σ_ω^2 is the volatility of the asset, and $\mathcal{I}(\tau^*; \tau_{dm}^{max})$ is the informativeness of the asset price in equilibrium, which depends on τ^* (the lowest precision of the signals discovered by quants) and τ_{dm}^{max} (the highest precision of the signals discovered by quants).

The availability of new data raises the precision of the best available signal for quants, τ_{dm}^{max}. Dugast and Foucault (2022) show that, in equilibrium, such a shock always triggers an increase in price informativeness (that is, $\mathcal{I}$ increases with τ_{dm}^{max}). The reason is that quants who find the best signals (those with a precision equal to τ_{dm}^{max}) take larger bets for a given realization of their signal as the precision of this signal (τ_{dm}^{max}) increases. Thus, the equilibrium price reflects their signal more and is therefore more informative about the asset payoff since this signal is more precise. As a result, experts' average performance drops (see eq.(2)). Indeed, as the price is more informative, their ability to generate profits using their private information is reduced. This is also true for quant funds, holding the precision of their signal fixed. However, this precision is not fixed since some can take advantage of the increase in the precision of the best signals. The performance of some of these funds can therefore increase even though the average performance of all funds decreases.

Our data does not allow us to study separately discretionary and quant funds (as we have no direct information on the fund's type). However, we expect discretionary funds to dominate since according to Abis (2022) (Figure 6), as of 2018, discretionary funds account for more than 80% of all active U.S. equity funds (and controls more than 90% of assets managed by active funds). Hence, given the previous prediction, we expect the availability of new alternative data (in our tests, satellite coverage of U.S retailers) to reduce active funds' stock picking ability. Moreover, we expect this effect to be stronger for funds that are more likely to rely on expertise (human judgement rather than quantitative analysis of massive amount of data) to obtain their signals. In contrast, it should be weaker for funds that are more likely to be quantitative funds. Last, we expect the drop in active funds' stock picking ability to be driven by an increase in price informativeness for the stocks affected by the availability of new alternative data. In the rest of the paper, we test these predictions.

III Data and Measurements

We test these predictions studying mutual funds' stock picking ability in retailers' stocks, after an alternative source of data becomes available for those stocks. Specifically, we consider the introduction of store-level car counts based on satellite imagery. In this section, we describe our data sets and measures of mutual fund skills.

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.[16] 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.[17] To address the possibility of incubation bias, we further exclude observations

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

[17]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',

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 the address where the fund is located.

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.[18] 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 3,960 funds holding 9,788 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.[19]

Our last data set comes from RS Metrics.[20]. RS Metrics is the first major data vendor that has released parking lot traffic data for U.S. retail firms (e.g., Walmart, Home Depot, Best Buy, Starbucks, Tractor Supply Co., etc.) based on satellite imagery starting from the second quarter of the year 2009. RS Metrics obtains satellite images of parking lots of U.S. retailers's stores (see Figure A.1 in the Appendix for examples) from diverse providers. Each store is

'StreetTRACKS', '100', '400', '500', '600', '1000', '1500', '2000', '3000', '5000', 'money market', 'money mkt'.

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

[19]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.

[20]RSMetrics sells data from satellite imagery. See https://rsmetrics.com/ for details information about its products

monitored multiple times in a given month, which enables RS Metrics to provide store-level information (using machine learning techniques and human judgement to analyze images) about parking lot capacity and utilization on a specific day, across major U.S retailers.[21] 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 the exact date 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. Figure I shows the number of new coverage initiations in each quarter during our sample periods.[22]

[Insert Figure I about here]

B Measuring Funds' Skill

To capture the precision of fund managers' signals, we use a measure of stock picking skill similar to Kacperczyk et al. (2014) and related to eq.(1) (see Section II).[23] Intuitively, a fund with stock picking skill should tilt its holdings in a given stock in the direction of its future

[21] 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. (2019) for additional details on the data and RS Metrics' technology.

[22] 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 I.1).

[23] The main difference with Kacperczyk et al. (2014) 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 skill measure across stocks held by each fund). A related approach to measure funds' performance is used in Jiang and Zheng (2018).

excess return. Thus, in quarter t, we measure the stock picking skill of fund f in stock i at horizon h by:

$$Picking_{f,i,t}^{h} = 100 \times (w_{i,t}^{f} - w_{i,t}^{m})(R_{t+h}^{i} - \beta_{i,t} R_{t+h}^{m}), \tag{3}$$

where $w_{i,t}^{f}$ is the fraction of fund f's assets held in stock i at the end of quarter t, $w_{i,t}^{m}$ is the fraction of total market capitalization in asset i (its weight in the "market portfolio") at the end of quarter t, R_{t+h}^{i} is the return of stock i over the following h quarters, R_{t+h}^{m} 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 (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 a given stock is high if it overweights (underweights) this stock relative to its weight in the market portfolio when the latter has subsequently high (low) idiosyncratic returns. $Picking_{f,i,t}^{h}$ is a proxy for $w^{*} \times R^{e}$ in eq.(1). Indeed, $(w_{i,t}^{f} - w_{i,t}^{m})$ in eq.(3) measures the extent to which fund f's position in stock i deviates from a passive benchmark (that taken by a mean-variance investor who can invest in the market portfolio and the riskless asset). It plays the role of w^{*} in eq.(1). Moreover $(R_{t+h}^{i} - \beta_{i,t} R_{t+h}^{m})$ measures the excess return of stock i, similar to R^{e} in eq.(1).

For robustness, we consider two alternative measures of skills. In the first one, we replace $w_{i,t}^{m}$ in (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. In our second measure, called $Trading_{f,i,t}^{h}$, we replace $w_{i,t}^{m}$ in (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 $t-4$.[24] This measure captures the ability of fund f to change its holdings in a given stock in the direction of subsequent

[24] 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).

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 $Picking^h_{f,i,t}$ (reported in Section IV). Thus, we only report them in the online appendix and, for brevity, do not recall this point when discussing the results regarding $Picking^h_{f,i,t}$.

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.[25] 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) 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]

[25]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%.

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 satellite traffic data about a firm contains information useful to forecast future sales and stock returns for this firm (see Zhu, 2019; Katona et al., 2019, for evidence) and there is anecdotal evidence that some funds (most likely quant funds) use this data (see Footnote 4). Thus, it should raise the precision of signals used by these funds and therefore *reduce* (as explained in Section II) the stock-picking ability of other funds (in particular, those who rely solely on human expertise to obtain private information).

To test this displacement hypothesis, 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 skill 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 is a stock fixed effect and $\gamma_{f\times t}$ is a fund $\times$quarter fixed effect. Standard errors are double-clustered at the fund and stock levels. Stock fixed effects capture time-invariant determinants of fund picking skills 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 or management team.

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 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 all stocks 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 fund's portfolio not covered by RS Metrics.

[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, in our sample, funds holding covered stocks have significantly higher stock picking ability in these stocks than all funds' stock picking ability in non covered stocks. In terms of magnitude, 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, the ability of funds to correctly scale up or down their positions in covered stocks in anticipation of future idiosyncratic returns for these stocks 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 $\times$ 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. Results are unchanged with this approach. This result rules out the possibility that change in funds' picking skills in covered stocks is due to a composition effects (change in the funds holding covered stocks before and after coverage initiation).

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

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

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 but becomes 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 the view that coverage initiation is the cause of this drop rather than the reverse. 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.

Finally, as a robustness test, we re-estimate our main specifications presented in Table II with a narrower group of control stocks. Specifically, one year prior to the initiation of coverage, we match based on industry (NAICS 2-digit sector) and market capitalization each covered stock with five non-covered stocks. The results with this matching approach are qualitatively similar to those reported in Table II (see Table I.2 in the online appendix).

B Selection of Covered Stocks

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 (like for instance market capitalization). 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 (since we include time fixed effects in our specifications). This possibility cannot fully be discarded since there is no limit to the list of potential time-varying confounding variables. However, we can study in more details the coverage decision.

To this end, 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 Appendix Table B.2. 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, with the results from the previous section, mitigate the possibility of reverse causality (a situation in which RS Metrics initiates coverage in response to a strong demand by funds because of a decline in the quality of their information for a stock). Lastly, measures of stock price informativeness similarly do not appear to be predictors for coverage initiation (see Column 6).[26] 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. It is noteworthy that the R2 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 RSMetrics' clients extend beyond the financial industry. Another reason is that RSMetrics' decision to cover a stock might be in response to the request of a single client. In any case, our results in this section further alleviate concerns regarding a potential reverse causality between mutual fund skills and the initiation of stock coverage.

As a robustness test, we re-estimate our main specifications presented in Table II, however,

[26]A more thorough discussion on stock price informativeness measures and analyst forecast errors is provided in Section VI.

we narrow our analysis to include only a select group of non-covered stocks. Specifically, one year prior to the initiation of coverage, we match each covered stock with five non-covered stocks that do not experience coverage by RS-Metrics. We ensure that these control stocks belong to the same industry (NAICS 2-digit sector) and we select the five stocks with the closest market capitalization to that of the corresponding covered stock. The results of this matching analysis are presented in Internet Appendix Table I.2, and show that our conclusion remains the same.

C Cross-Sectional Heterogeneity

As explained in Section II, we expect the negative effect of the availability of new alternative data on funds' stock picking ability to be stronger (and therefore easier to detect statistically) for funds that have more ex-ante expertise. In this section, we study whether this is the case.

To do so, we measure a fund's expertise in three ways: (i) the level of Picking in covered stocks before coverage initiation by RS Metrics (funds with high stock-picking skill in a stock are more likely to be able to generate private signals about this stock), (ii) a fund's focus on industries covered by RS Metrics (funds who concentrate their holdings in an industry are more likely to possess private signals about stocks in this industry; see Kacperczyk et al., 2005) and (iii) a fund's geographical proximity to the stores of firms' covered by RS Metrics (funds that are geographically close to firms' operations are more likely to have private information about these firms; see Coval and Moskowitz, 2001). As explained below, 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 Skills

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 the release of alternative data since $HighPickingPre_{f,i}$ cannot be observed (by definition) for these funds.

We emphasize that we include fund $\times$ 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.

[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.

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' on 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.

To obtain a more granular view of the variation of the effect of coverage across funds with different picking abilities (expertise), we sort funds into ten categories of expertise (rather than just two). Specifically, we estimate:

$$Picking^h_{f,i,t} = \sum_{d=1}^{10} \beta_d \{Covered \times Post\}_{i,t} \times Decile(d)PickingPre_{f,i} \\ + \alpha_{f \times i} + \gamma_{f \times t} + \epsilon_{f,i,t}, \tag{9}$$

where $Decile(d)PickingPre_{f,i}$ is a dummy variable indicating whether fund f is in the d^{th} decile of picking skills for stock i before its coverage begins (with $d = 1$ being the decile of bottom performers and $d = 10$ being the decile of top performers). Thus, β_d estimates the effect of coverage initiation for stock i on a fund in the d^{th} decile of picking skills. Standard errors are double-clustered at the fund and stock levels.

[Insert Figure IV about here]

Figure IV reports the estimates of β_d for each decile when Picking is measured for an horizon of four quarters ($h = 4$). Horizontal lines represent 95% confidence intervals. Interestingly, coverage initiation has a significantly negative effect on the stock-picking ability of the funds in the three highest deciles of expertise but a significant positive effect of the funds in the two lowest deciles. This finding indicates that the average effect of coverage initiation on Picking (Table II) masks important heterogeneity across funds: Funds who have developed expertise in covered stocks are the most negatively affected by the availability of new data. Our interpretation is that this is due to the acquisition of this data by other funds, who, by trading on it, makes prices more informative and reduces experts' ability to profit from their information. The funds acquiring the new data should, at least in the short run, experience an improvement in their stock-picking ability. Figure IV suggests that these other funds were those with low stock picking abilities prior to the availability of the new data, maybe precisely because their "technology" to produce information is different and requires quantitative data. We explore this point in more details in Section D.

C.2 Industry Expertise

Our findings so far suggest that the negative effect of coverage initiation is driven by funds that were superior stock pickers ("experts") in the covered stocks in the pre-coverage period. This observation is consistent with our hypothesis that experts relying on traditional methods to produce information are "displaced" by the availability of alternative data (because using this data requires different skills). In this section and the next, we consider two possible ways in which experts can develop a superior ability to obtain private signals: (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).[27]

We then estimate the following specification:

$$\begin{aligned} Picking^h_{f,i,t} = \; & \beta\{Covered \times Post\}_{i,t} \\ & + \beta_{expert}\{Covered \times Post\}_{i,t} \times IndustryExpert_{f,i} \\ & + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t}, \end{aligned} \tag{10}$$

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) 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

[27]Most of the covered stocks are in the retail trade industry but there is substantial variation across funds in industry specialization.

experts' stock picking ability, above and beyond the baseline effect (measured by β) for all funds holding this stock. Again, we include fund $\times$ 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.

Table IV reports estimates of eq.(10). 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 we control for a dummy variable $Covered_{i,t} \times IndustrySpecialist_{f,i}$ (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 find 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.

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). Specifically, funds that are located in proximity to the main stores of firms covered by RS Metrics should enjoy an informational advantage prior to the coverage of these firms by RS Metrics and should therefore be, more negatively, affected by coverage initiation. To test this prediction, we utilize the store-level data provided by RS Metrics to identify the primary metropolitan statistical area (MSA) for each covered stock. This primary MSA corresponds to the geographical area where the majority of the firm's parking lots are located. We then classify a fund as "Local" for a particular stock if the fund is located in the same MSA as the stock's primary MSA. To ensure that the primary MSA provides valuable information, we only consider a fund as "Local" if at least 5% of the firm's parking lots are situated within the primary MSA.[28]

We then estimate the following specification:

$$\begin{aligned} Picking^{h}_{f,i,t} = \beta\{Covered \times Post\}_{i,t} \\ + \beta_{local}\{Covered \times Post\}_{i,t} \times Local_{f,i} \\ + \alpha_{f\times i} + \gamma_{f\times t} + \epsilon_{f,i,t}, \end{aligned} \tag{11}$$

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 the fund and stock levels.

[Insert Table V about here]

Table V reports estimates of eq.(11). In columns (1), (3), (5), and (7), we estimate this equation without fund-stock fixed effects but we include stock fixed effects and we control for a dummy 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, in line with Coval and Moskowitz (2001), mutual funds located in areas where covered companies have the majority of their stores 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 those of other funds.

[28] 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%.

D Who are the new Experts?

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). The reason is that these funds lack the expertise required to exploit alternative data. In contrast, we expect quantitative funds to possess this expertise and to be those subscribing to RS Metrics data. Thus, these funds should experience a weaker drop in their stock picking ability, or possibly even an improvement, since they get access to a new source of information.

We cannot directly test this hypothesis because we do not know which funds buy data from the provider considered in our study and which are quantitative funds in our sample. However, Abis (2022) finds empirically that quantitative funds tend to hold a larger number of stocks and have a lower return volatility compared to non-quantitative funds. Thus, we expect effects to be weaker for funds holding a large number of stocks or with a low return volatility because, according to Abis (2022), they are more likely to be quantitative funds. To examine this, in each quarter, we rank funds based on the number of stocks they hold and define a dummy variable "High Nb. Stocks", equal to 1 for funds that hold more stocks than the median fund in the majority of quarters. Similarly, we rank funds based on the volatility of their returns over the previous 24 months and define a dummy variable "Low Volatility" equal to 1 for funds with a return volatility lower than the median fund. We then estimate a specification similar to eq.(10), interacting "Covered × Post" with either "High Nb. Stocks" or "Low Volatility."

We present estimates of this test in Table B.3 of the Appendix. Columns (1), (3), (5), and (7) show that the coefficient on the triple interaction term "Covered × Post × High Nb. Stocks" is positive and statistically significant. The magnitude of this coefficient is approximately 50% of the coefficient on the interaction term "Covered × Post". Thus, the negative effect of alternative data release on the stock picking abilities of funds is significantly lower for more diversified funds. In Columns (2), (4), (6), and (8), the coefficient on the triple interaction term "Covered × Post × Low Volatility" is also positive, although not statistically significant in most specifications. Nonetheless, the magnitude of this coefficient is approximately 50% of the coefficient on the interaction term "Covered × Post". Overall, these results provide suggestive evidence that funds with characteristics indicative of being quantitative experience a weaker decline in their stock picking ability after coverage initiation.

V Fund Divestment as Evidence of Displacement

As found in the previous section, the availability of new alternative data for a stock has a negative effect on experts' ability to exploit their private information in this stock. This effect should lead experts to reduce the amount they invest in covered stocks or to exit the market for these stocks. In either case, experts are "displaced". To study whether this is the case, we first analyze the rank of covered stocks in mutual fund portfolios, which captures the within-fund size rank of a given investment. A stock's rank is determined according to the percentage of total net assets of the fund's portfolio invested into it. 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 investments, 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) rather than Picking.[29] Results are presented in columns (1) and (2) of Table VI. After coverage initiation, covered stocks experience a decrease in their relative investment rank of 9%, 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

[29]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" (thus, stocks in which their Picking ability should be high).

coverage, relative to the fund's variation in holdings of uncovered stocks.

[Insert Table VI about Here]

To test whether divestment is stronger for funds that have more expertise, we estimate the same specifications as in equations (7), (10) and (11), using the negative of the logarithm of the stock rank as 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. Considering that, on average, a stock is held by 57 distinct funds, this indicates an exit of approximately 11 funds. These results suggest that funds tend to divest from covered stocks subsequent to the release of alternative data, in line with the finding that this release reduces the value of their expertise (measured by their stock picking ability) in these stocks.

[Insert Figure V about Here]

Figure V 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 six 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. Importantly, this negative trend 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.

Last, we also find a notable decline in both the average age of funds and the proportion of funds older than 5 years or 10 years subsequent to the coverage initiation (see Table I.7 in the online appendix). Thus, after coverage initiation, funds holding covered stocks are younger. This is consistent with the idea that the funds with the ability to exploit alternative data have different characteristics. In particular, they are likely to be younger as the rise of quantitative funds is a relatively new phenomenon, which has accelerated in recent years (see Abis, 2022).

One question arises regarding how funds adjust their portfolio allocation when divesting from stocks post-coverage initiation. Intuitively, they should redirect their investments towards stocks where they can leverage their expertise without facing competition from funds using alternative data (that is, uncovered stocks in which they are experts). 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 × 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 × Post" is a dummy equal to one when RS-Metrics starts covering a stock whose majority of parking lots are in the same Metropolitan Statistical Area (MSA) as the focal stock's headquarters.[30]

[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 a non-covered stocks that are peer 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 I.8 in the online appendix).

VI Mechanism: The Role of Price Informativeness

In the previous sections, we have shown that the availability of new alternative data for a stock reduces funds' stock picking ability and investments in that stock. Moreover, these effects are stronger for funds that appear to have greater expertise, either due to industry specialization or geographical location. According to the mechanism described in Section II, they are due to the fact that the availability of new data makes the price of covered stocks more informative. In this section, we test whether this is the case.

[30] Company headquarters data is sourced from Compustat. 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.

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. The ACAR assesses the extent to which prices reflect contemporaneous cash flows, and is computed as post-earnings announcement returns adjusted for predicted returns from a factor model. Lower announcement price reactions are indicative of information being incorporated in stock prices before the announcement and thus greater price informativeness.

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 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. We exclude observations with estimation windows comprising fewer than 63 (preceding) trading days, i.e., one calendar quarter.

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 high jump ratio indicates that stock prices are less informative prior to the announcement since their variation post announcement account for a larger 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 mth day prior to earnings announcements to the nth 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:

$$Informativeness_{i,t} = \beta\{Covered \times Post\}_{i,t} + \delta X_{i,t} + \alpha_i + \gamma_t + \epsilon_{f,i,t}, \tag{12}$$

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. Covered industries encompass 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.(12). 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. (2019) find that coverage of a stock by satellite data providers has no significant effect on stock price informativeness, maybe because their tests lack power to detect this effect. 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 Analysts' Forecast Accuracy

A expected, we find an improvement in price informativeness for covered stocks. Our interpretation is that this improvement reflects trades by quant funds buying RS Metrics data. Another (non exclusive) possibility is that it stems from more accurate forecasts by security analysts following covered stocks and buying RS Metrics data. In either case, 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, the exact channel is not exactly identical: In the first case, alternative data gives rise to a new source of private information while in the second it enhances the precision of public signals (sell-side equity analysts' forecasts). To study whether the second channel operates, we study whether analysts' earnings forecasts become more accurate after a stock becomes covered, relative to their forecasts for non covered stocks.

To this end, 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. (12) 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. (12). The results, presented in Table X, 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.[31] This suggests that the improvement in price informativeness for covered stocks is unlikely to be due to an improvement in public signals available to investors.

[Insert Table X about here]

VII Conclusion

This paper examines the impact of big data on the skills of active fund managers. We test whether the availability of satellite imagery data tracking retailer firms' parking lots affects the stock picking abilities of active mutual fund managers in stocks covered by this data. We provide empirical evidence that the availability of such alternative data reduces the comparative advantage of traditional asset managers who rely on specialized industry knowledge and human judgment. Specifically, we find that active mutual funds' stock picking ability declines in covered stocks after the introduction of satellite imagery data for these stocks. This decline is particularly pronounced for funds that heavily rely on traditional sources of expertise, indicating that these managers are at a higher risk of being displaced by new data sources.

These results highlight that the emergence of alternative data, such as satellite imagery, can reduce the value-added by managers who rely on traditional methods to generate their signals. They suggest that, as in other industries, the big data revolution has the potential to displace high-skilled workers in the finance industry, who lack the skills to leverage information contained in big data.

[31]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.

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.

Berk, J. and van Binsbergen, J. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 110(1):1–20.

Berk, J. B., Van Binsbergen, J. H., and Miller, M. (2020). Mutual funds: Skill and performance. *The Journal of Portfolio Management*, 46(5):17–31.

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.

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.

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.

Harvey, C. R., Rattray, S., Sinclair, A., and Van Hemert, O. (2017). Man vs. machine: Comparing discretionary andsystematic hedge fund performance. *The Journal of Portfolio Management*, 43(4):55–69.

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. (2019). On the capital market consequences of alternative data: Evidence from outer space. Working Paper.

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.

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 3,960 funds and 9,788 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)	108,381	1,475.755	5,956.635	13.300	76.900	281.400	982.800	5,655.000
Nb. Stocks	108,381	123.230	211.098	29.000	47.000	74.000	115.000	361.000
Age (Years)	108,381	16.954	12.539	2.333	8.833	15.250	21.583	36.583
Return	108,381	0.006	0.050	-0.086	-0.015	0.010	0.032	0.076
Flow	107,686	0.280	30.799	-0.053	-0.015	-0.006	0.004	0.057

Panel B: Stock-Quarter level

	Obs	Mean	Sd	5%	25%	50%	75%	95%
Beta	229,409	1.007	0.558	0.141	0.674	0.995	1.323	1.916
Return	229,409	0.012	0.153	-0.191	-0.049	0.008	0.064	0.211
Nb. Funds Holding	229,409	57.408	78.053	1.000	6.000	30.000	79.000	202.000
Covered	229,409	0.009	0.092	0.000	0.000	0.000	0.000	0.000
Covered × Post	229,409	0.006	0.078	0.000	0.000	0.000	0.000	0.000
Market Cap. (bn)	58,197	7.598	35.301	0.037	0.222	0.849	3.225	30.000
Book to Market	58,197	0.580	0.606	0.082	0.247	0.438	0.736	1.489
Assets (bn)	58,197	5.610	21.551	0.037	0.189	0.705	2.869	21.920
Leverage	58,197	0.030	0.062	0.000	0.000	0.007	0.031	0.141
$ACAR[0,2]$	58,197	0.069	0.076	0.004	0.021	0.048	0.092	0.204
$ACAR[-1,2]$	58,197	0.072	0.078	0.004	0.022	0.049	0.095	0.209
$CAR[0,2]/CAR[-21,2]$	19,164	0.480	0.461	-0.234	0.185	0.478	0.768	1.206
$CAR[-1,2]/CAR[-21,2]$	19,164	0.504	0.460	-0.214	0.212	0.507	0.792	1.222
Std. FE 1-Q	44,197	2.487	2.687	0.000	0.680	1.600	3.250	8.500
Std. FE 2-Q	44,544	2.722	2.940	0.059	0.750	1.789	3.500	9.000
Std. FE 3-Q	42,359	3.190	3.660	0.111	0.900	2.000	4.000	11.000
Std. FE 4-Q	39,354	3.500	4.100	0.143	1.000	2.000	4.400	12.364
Std. FE 1-Y	11,007	2.442	2.873	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
Covered	1.28e+07	0.026	0.159	0.000	0.000	0.000	0.000	0.000
Covered × Post	1.28e+07	0.020	0.139	0.000	0.000	0.000	0.000	0.000

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 skills. 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. 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 × Post	-0.017***	-0.021***	-0.021**	-0.032**	-0.042***	-0.038**	-0.052**	-0.067***	-0.058*	-0.081***	-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 × 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 × 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 skill 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)	(0.012)	(0.012)	(0.018)	(0.019)	(0.026)	(0.027)
Covered × Post × Industry Specialist	-0.139**	-0.161***	-0.346**	-0.419***	-0.575**	-0.711***	-0.863**	-1.086***
	(0.066)	(0.054)	(0.165)	(0.142)	(0.276)	(0.235)	(0.407)	(0.358)
Covered × Industry Specialist	0.105***		0.249***		0.397***		0.565***	
	(0.035)		(0.085)		(0.133)		(0.180)	
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 skills 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. $Covered \times Post$ 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, "Local", which equals one if the fund is located in the Metropolitan Statistical Area (MSA) where the firm covered by RS Metrics has the majority of its parking lots, as identified through satellite imagery data. 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)	(0.015)	(0.016)	(0.023)	(0.026)	(0.032)	(0.038)
Covered × Post × Local	-0.039**	-0.058***	-0.080**	-0.119**	-0.122*	-0.185**	-0.178*	-0.280**
	(0.017)	(0.020)	(0.040)	(0.058)	(0.064)	(0.092)	(0.092)	(0.133)
Covered × Local	0.026**		0.053**		0.079*		0.113**	
	(0.012)		(0.026)		(0.040)		(0.057)	
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	9409266	9409266	9130009	9130009	8844590	8844590	8557824	8557824
R^2	0.10	0.20	0.12	0.28	0.14	0.34	0.14	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* $\times$ *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**	-0.173***	-0.204***
	(0.037)	(0.038)	(0.062)	(0.064)
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	229,260	101,785
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 skill 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 Metropolitan Statistical Area (MSA) where the firm covered by RS Metrics has the majority of its parking lots, as identified through satellite imagery data. Standard errors are double-clustered at the fund and stock levels.

	Stock Rank			
	(1)	(2)	(3)	(4)
Covered × Post	0.004	-0.072**	-0.071**	-0.068*
	(0.050)	(0.035)	(0.035)	(0.035)
Covered × Post × High Picking Pre-Period	-0.197***			
	(0.054)			
Covered × Post × Industry Specialist		-0.378**		
		(0.176)		
Covered × Post × Sector Fund			-0.331**	
			(0.154)	
Covered × Post × Local				-0.148
				(0.136)
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	9410363
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 majority 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***		0.110***
	(0.030)		(0.030)
Local Peer Covered × Post		0.085***	0.083***
		(0.027)	(0.027)
Year-Quarter FE	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
Observations	227,325	227,325	227,325
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* × *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	-0.087**	-0.088**	-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	19,164	19,164	58,197	58,197
R^2	0.22	0.23	0.36	0.37

Table X: 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. $Covered \times Post$ 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.

	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	-0.041	-0.228	-0.061	0.200	-0.149
	(0.188)	(0.211)	(0.254)	(0.264)	(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	44,544	42,359	39,354	11,007
R^2	0.14	0.12	0.15	0.16	0.24

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.

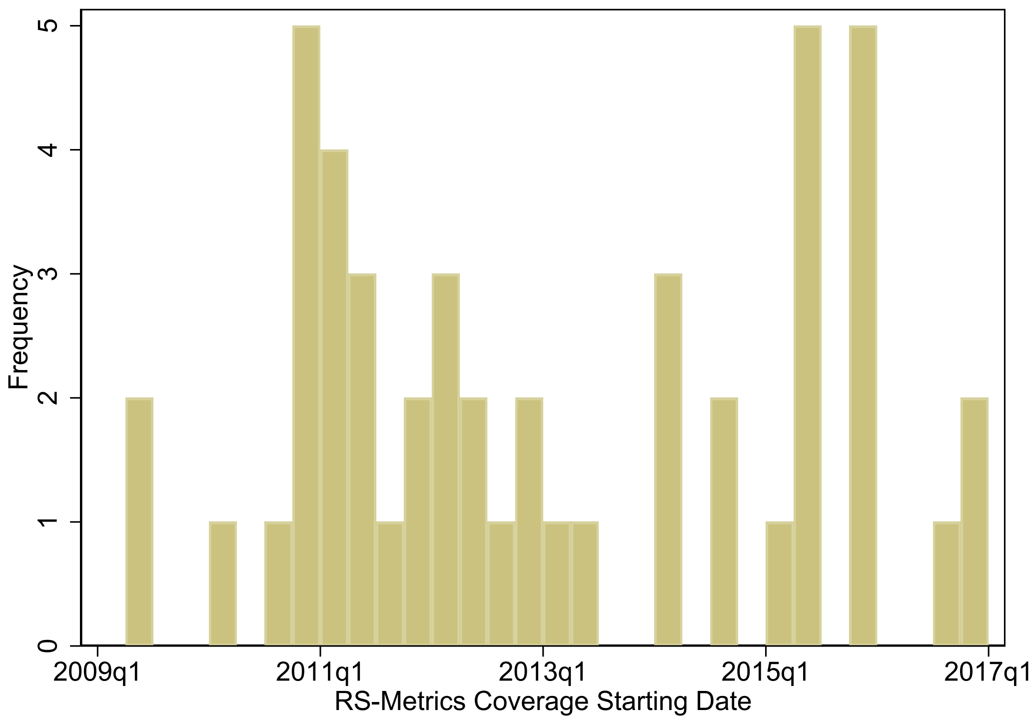


Figure II: Alternative Data and Fund Picking Skills

This figure reports the dynamic effect of the release of alternative data on fund picking skills. 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.

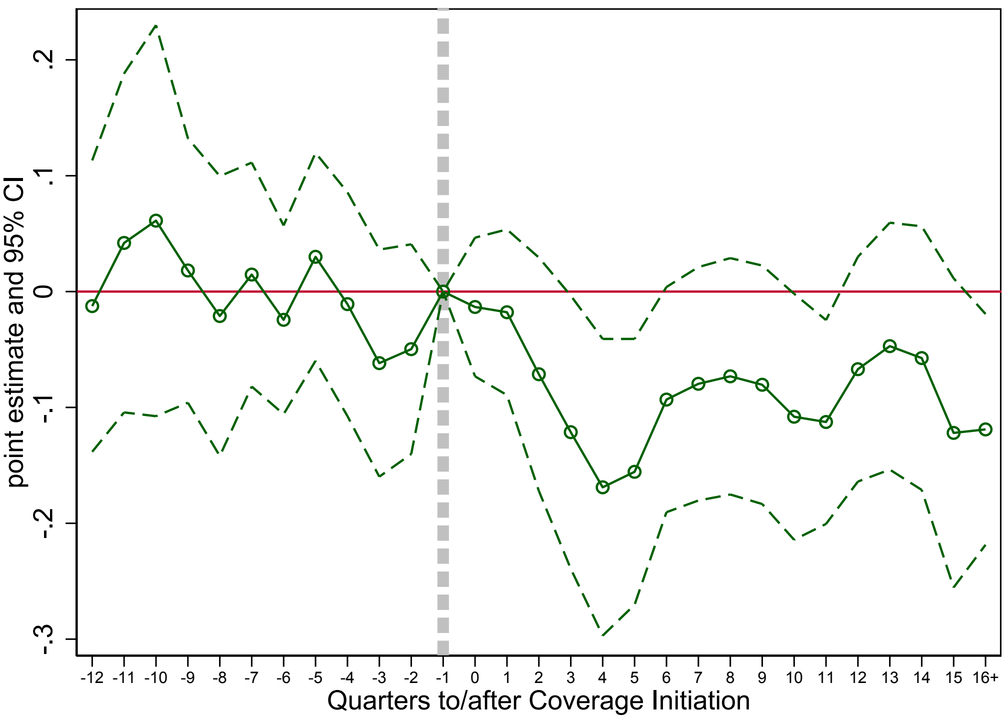
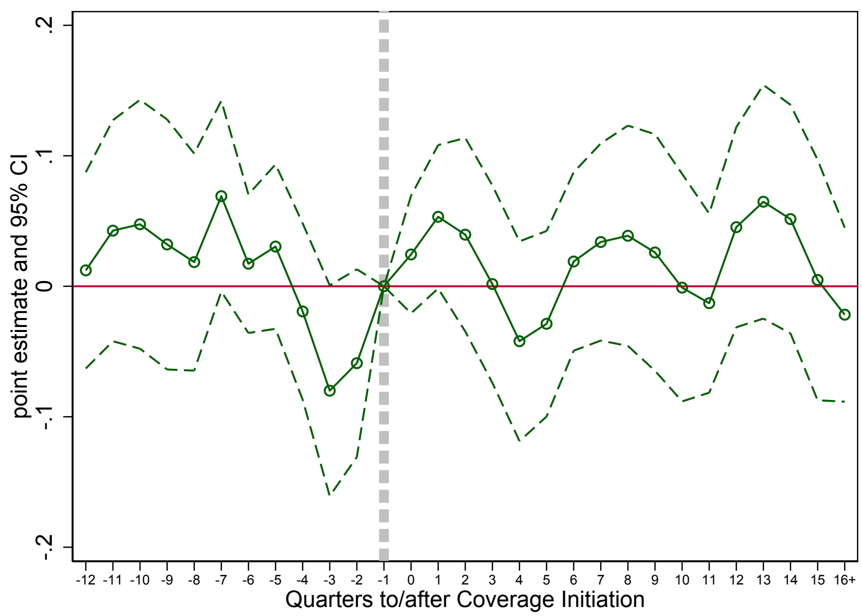


Figure III: Fund Picking Skills depending on Skills before Coverage Initiation

This figure reports the dynamic effect of the release of alternative data on fund picking skills. 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

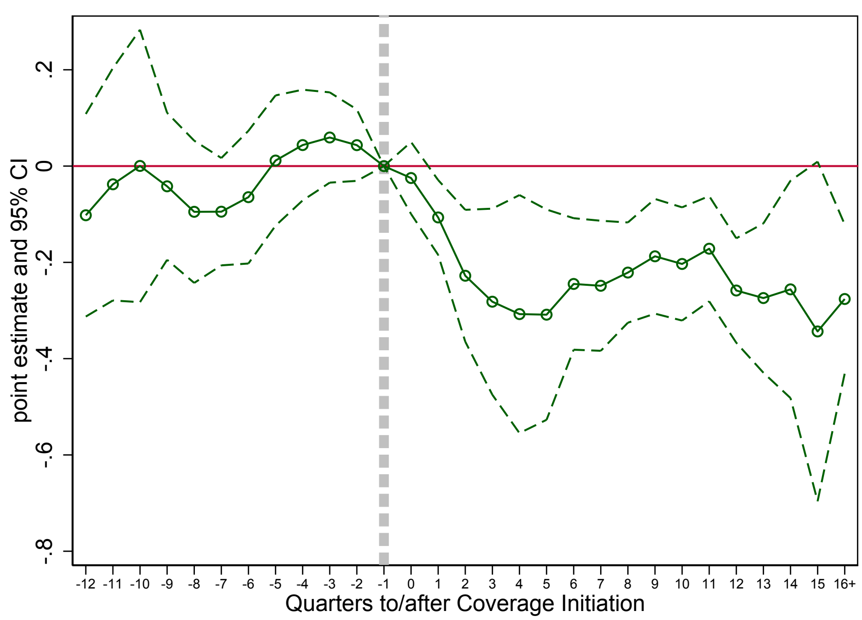


Figure IV: Heterogeneous Effect based on Skills before Coverage Initiation

This figure reports the heterogeneous effect of the release of alternative data depending on funds' ability to pick stocks before coverage starts. Each diamond corresponds to the coefficient on the interaction between Treated × Post and a specific decile dummy corresponding to the fund's picking ability before coverage initiation. Horizontal lines are 95% confidence intervals. For instance decile 1 is equal to one if the fund's picking skill in the covered stock was on average in the first decile of the distribution of picking skills before the release of alternative data. The dependent variable is Picking 4-Q, defined in equation (3). Standard errors are clustered at the fund and stock levels.

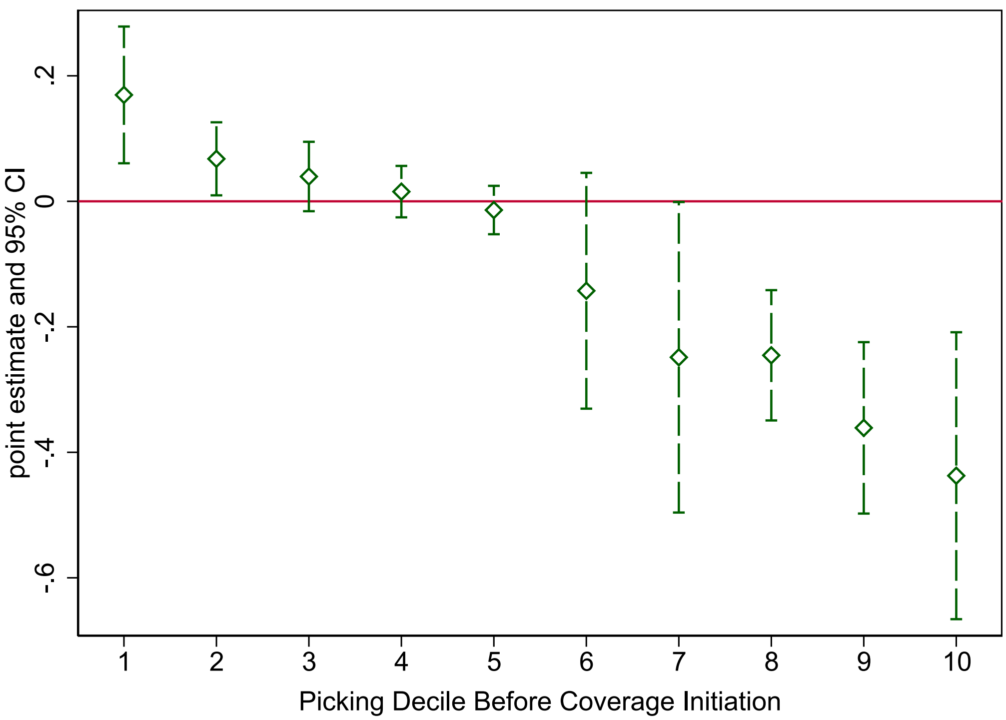
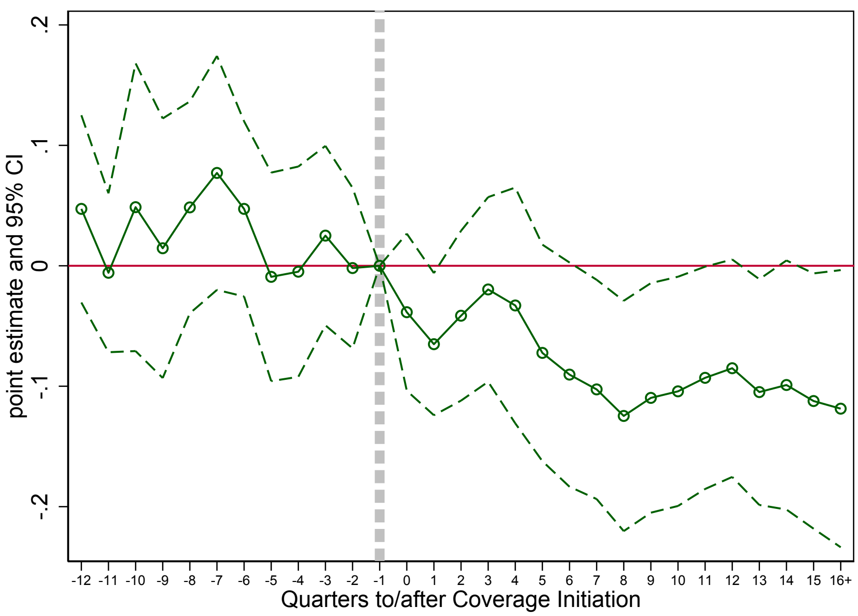


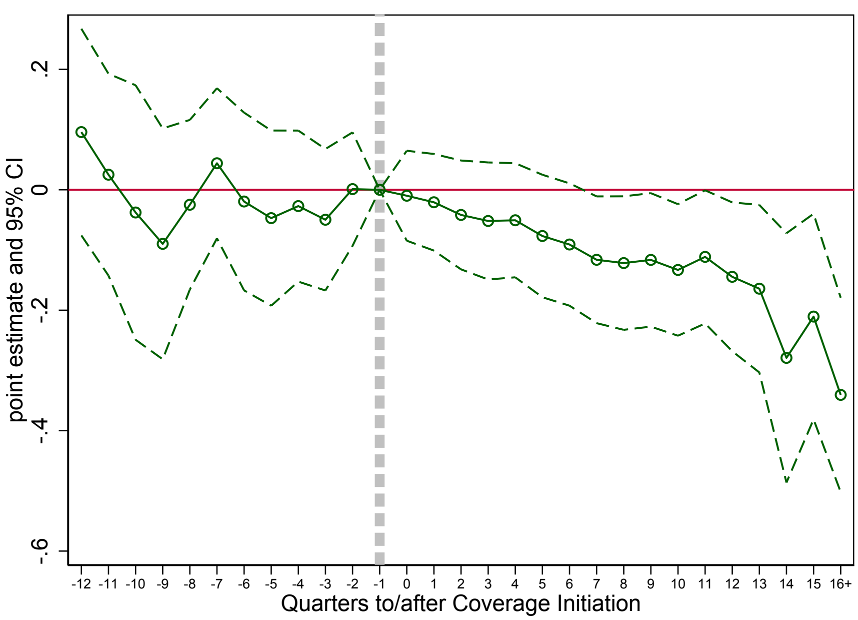
Figure V: Alternative Data and Fund Divestment

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

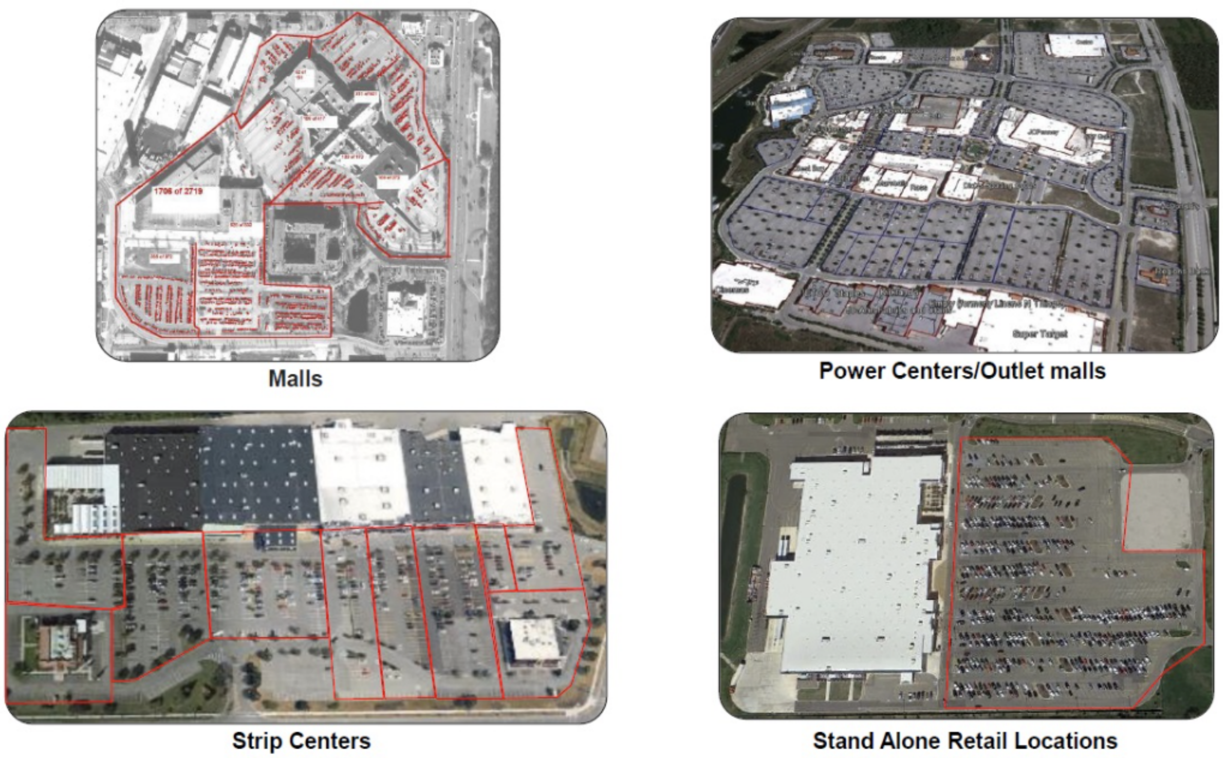


Appendix

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.)	0.043***	0.046***	0.046***	0.044***	0.045***	0.045***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.016)
Log(Assets)	-0.017	-0.018	-0.018	-0.018	-0.018	-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)		0.030	0.030	0.030	0.032	0.032
		(0.032)	(0.031)	(0.032)	(0.033)	(0.033)
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.)			-0.002	-0.001	-0.001	-0.011
			(0.022)	(0.022)	(0.022)	(0.020)
Analyst FE 1-Q (Year -1)			0.002	0.002	0.002	0.002
			(0.003)	(0.003)	(0.003)	(0.003)
Log(Nb. Funds Holding)				0.004	0.004	0.003
				(0.015)	(0.015)	(0.015)
Picking 1-Q (Year -1)					-0.037	-0.038
					(0.285)	(0.285)
$\frac{CAR[-1,2]}{CAR[-21,2]}$ (Year -1)						0.000
						(0.000)
$ACAR[-1, 2]$ (Year -1)						0.165
						(0.238)
Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	62,779	62,779	62,779	62,779	62,779	62,779
R^2	0.01	0.01	0.01	0.01	0.01	0.01

Table B.3: Proxies for Quantitative Funds

The table presents the results of our study on the differential impact of alternative data on likely quantitative funds. 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, "High Nb. Stocks", which equals one if the fund holds most of the time a larger number of stocks than the median fund in that month, as well as interactions with a dummy variable, "Low Volatility", which equals one if the fund's return volatility is most of the time lower than that of the median fund in that month. Volatility is computed as the rolling 24-month standard deviation of fund returns. It is not computed for funds with less than 24 months of return over our sample period. 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	-0.038***	-0.029**	-0.075**	-0.062**	-0.121**	-0.101**	-0.189**	-0.154**
	(0.014)	(0.011)	(0.034)	(0.026)	(0.056)	(0.043)	(0.080)	(0.062)
Covered × Post × High Nb. Stocks	0.023**		0.047*		0.075*		0.116*	
	(0.011)		(0.028)		(0.045)		(0.064)	
Covered × High Nb. Stocks	-0.033***		-0.068***		-0.106***		-0.151***	
	(0.009)		(0.021)		(0.033)		(0.045)	
Covered × Post × Low Volatility		0.013		0.036		0.059*		0.084
		(0.009)		(0.022)		(0.036)		(0.052)
Covered × Low Volatility		-0.015**		-0.038**		-0.059**		-0.079**
		(0.008)		(0.017)		(0.028)		(0.038)
Fund × Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1.28e+07	1.27e+07	1.24e+07	1.23e+07	1.20e+07	1.19e+07	1.16e+07	1.15e+07
R^2	0.10	0.10	0.12	0.12	0.14	0.14	0.15	0.15

Online Appendix for

Displaced by Big Data: Evidence from Active Fund Managers

(not intended for publication)

Maxime Bonelli and Thierry Foucault

November 21, 2023

Contents

1 Stock Picking before 2014

Table I.1: Alternative Data and Stock Picking Skills before 2014

This table presents our main results on the effect of the release of alternative data on fund picking skills, restricting our sample to the pre-2014 period, i.e., before the introduction of the retail traffic product by Orbital Insight (the main competitor of 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* × *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 × Post	-0.019***	-0.018*	-0.021**	-0.035***	-0.028	-0.022	-0.054***	-0.049	-0.036	-0.075***	-0.077*	-0.065*
	(0.007)	(0.010)	(0.009)	(0.013)	(0.019)	(0.017)	(0.020)	(0.032)	(0.030)	(0.028)	(0.043)	(0.039)
Covered	0.025***			0.048***			0.075***			0.109***		
	(0.005)			(0.011)			(0.017)			(0.023)		
Fund × 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 × Stock FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	4721171	4721171	4721171	4687649	4687649	4687649	4652005	4652005	4652005	4615190	4615190	4615190
R^2	0.06	0.09	0.22	0.06	0.11	0.31	0.06	0.14	0.39	0.06	0.16	0.44

2 Stock Picking using a Matched Sample of Control Stocks

Table I.2: Stock Picking Skills using a Matched Sample of Control Stocks

This table reproduces our main results on the effect of the release of alternative data on fund picking skills. Regressions are estimated at the fund-stock-quarter level, but include only the observations corresponding to covered stocks and a matched sample of non-covered stocks. Specifically, one year prior to the initiation of coverage, we match each covered stock with five non-covered stocks that do not experience coverage by RS-Metrics. We ensure that these control stocks belong to the same industry (NAICS 2-digit sector) and we select the five stocks with the closest market capitalization to that of the corresponding covered stock. 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 × Post	-0.014***	-0.018***	-0.021***	-0.026**	-0.036***	-0.037**	-0.038**	-0.053***	-0.053**	-0.056**	-0.079***	-0.084**
	(0.005)	(0.006)	(0.006)	(0.011)	(0.012)	(0.014)	(0.017)	(0.019)	(0.023)	(0.024)	(0.028)	(0.033)
Covered	0.011**			0.021*			0.029*			0.041*		
	(0.005)			(0.011)			(0.017)			(0.023)		
Fund × 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 × Stock FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	740,174	740,174	740,174	722,971	722,971	722,971	705,649	705,649	705,649	688,232	688,232	688,232
R^2	0.31	0.33	0.40	0.32	0.34	0.45	0.32	0.36	0.51	0.32	0.37	0.54

3 Alternative Measures of Stock Picking Ability

In this section, we show that our results (reported in Tables II, III, IV and V in the text) regarding the evolution of fund managers' stock picking ability for covered stocks still hold when we use alternative measures of fund managers' skills, namely (i) one in which we replace $\omega_{i,t}^{m}$ (the weight of stock i in the market portfolio) by $\omega_{i,t}^{SP500}$ (the weight of stock i in the SP500 index) and (ii) one (called *Trading*) where we replace $\omega_{i,t}^{m}$ by $\omega_{i,t-4}^{f}$ (the weight of stock i in fund f four quarters ago). See the text for additional details.

Table I.3: Alternative Measures of Skills and High-skill Funds pre-Coverage

This table reproduces our results 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 in Table III, but uses alternative measures of fund skills as dependent variables: fund picking skills measured using the composition of the S&P 500 index (Panel A) and stock trading skills measured using the change in the stock's weight in the fund's portfolio over four quarters (Panel B).

Panel A: S&P500-based Picking skills

	Picking 1-Q	Picking 2-Q	Picking 3-Q	Picking 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	0.002	0.013	0.025*	0.019
	(0.005)	(0.009)	(0.015)	(0.018)
Covered × Post × High Picking Pre	-0.054***	-0.113***	-0.184***	-0.252***
	(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	0.33	0.37

Panel B: Trading skills

	Trading 1-Q	Trading 2-Q	Trading 3-Q	Trading 4-Q
	(1)	(2)	(3)	(4)
Covered × Post	0.004	0.017**	0.022*	0.012
	(0.004)	(0.008)	(0.012)	(0.014)
Covered × Post × High Trading Pre	-0.043***	-0.090***	-0.130***	-0.165***
	(0.006)	(0.012)	(0.018)	(0.023)
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	0.33	0.36

Table I.4: Alternative Measures of Skills and Industry Specialists

This table reproduces our results on the differential impact of alternative data on funds with industry expertise in Panel A of Table IV, but uses alternative measures of fund skills as dependent variables: fund picking skills measured using the composition of the S&P 500 index (Panel A) and stock trading skills measured using the change in the stock's weight in the fund's portfolio over four quarters (Panel B).

Panel A: S&P500-based Picking Skills

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.017***	-0.014**	-0.032***	-0.021*	-0.052***	-0.031*	-0.083***	-0.055**
	(0.005)	(0.006)	(0.011)	(0.011)	(0.018)	(0.018)	(0.025)	(0.026)
Covered × Post × Industry Specialist	-0.137**	-0.159***	-0.342**	-0.414***	-0.568**	-0.703***	-0.853**	-1.074***
	(0.065)	(0.054)	(0.164)	(0.140)	(0.273)	(0.233)	(0.403)	(0.354)
Covered × Industry Specialist	0.103***		0.246***		0.391***		0.557***	
	(0.035)		(0.084)		(0.131)		(0.178)	
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	0.13	0.33	0.14	0.37

Panel B: Trading Skills

	Trading 1-Q		Trading 2-Q		Trading 3-Q		Trading 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.015***	-0.015***	-0.027***	-0.022**	-0.045***	-0.034**	-0.072***	-0.058***
	(0.005)	(0.005)	(0.010)	(0.009)	(0.015)	(0.014)	(0.021)	(0.020)
Covered × Post × Industry Specialist	0.006	0.028***	-0.030*	0.019	-0.088**	-0.018	-0.162***	-0.077
	(0.007)	(0.006)	(0.016)	(0.015)	(0.036)	(0.029)	(0.058)	(0.049)
Covered × Industry Specialist	0.032		0.098*		0.167**		0.232**	
	(0.021)		(0.056)		(0.084)		(0.102)	
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	0.07	0.20	0.08	0.28	0.09	0.33	0.09	0.36

Table I.5: Alternative Measures of Skills and Sector Funds

This table reproduces our results on the differential impact of alternative data on funds with industry expertise in Panel B of Table IV, but uses alternative measures of fund skills as dependent variables: fund picking skills measured using the composition of the S&P 500 index (Panel A) and stock trading skills measured using the change in the stock's weight in the fund's portfolio over four quarters (Panel B).

Panel A: S&P500-based Picking Skills

	Picking 1-Q		Picking 2-Q		Picking 3-Q		Picking 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.018***	-0.014**	-0.033***	-0.021*	-0.053***	-0.031*	-0.084***	-0.055**
	(0.005)	(0.006)	(0.011)	(0.011)	(0.018)	(0.018)	(0.025)	(0.025)
Covered × Post × Sector Fund	-0.103	-0.131**	-0.253	-0.338**	-0.424	-0.577**	-0.628*	-0.879**
	(0.063)	(0.053)	(0.157)	(0.140)	(0.258)	(0.233)	(0.380)	(0.356)
Covered × Sector Fund	0.088**		0.211**		0.338**		0.478***	
	(0.035)		(0.084)		(0.131)		(0.180)	
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	0.13	0.33	0.14	0.37

Panel B: Trading Skills

	Trading 1-Q		Trading 2-Q		Trading 3-Q		Trading 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.015***	-0.015***	-0.027***	-0.022**	-0.045***	-0.034**	-0.072***	-0.058***
	(0.005)	(0.005)	(0.010)	(0.009)	(0.015)	(0.014)	(0.021)	(0.019)
Covered × Post × Sector Fund	0.004	0.024***	-0.030*	0.015	-0.085**	-0.022	-0.138**	-0.072*
	(0.006)	(0.005)	(0.018)	(0.013)	(0.038)	(0.025)	(0.060)	(0.043)
Covered × Sector Fund	0.022		0.076		0.135*		0.186**	
	(0.019)		(0.050)		(0.076)		(0.094)	
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	0.07	0.20	0.08	0.28	0.09	0.33	0.09	0.36

Table I.6: Alternative Measures of Skills and Geographical Location

This table reproduces our results on the differential impact of alternative data on fund picking skills depending on fund location in Table V, but uses alternative measures of fund skills as dependent variables: fund picking skills measured using the composition of the S&P 500 index (Panel A) and stock trading skills measured using the change in the stock's weight in the fund's portfolio over four quarters (Panel B).

Panel A: S&P500-based Picking Skills

	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**	-0.037***	-0.030*	-0.061***	-0.048*	-0.095***	-0.079**
	(0.006)	(0.007)	(0.014)	(0.016)	(0.022)	(0.025)	(0.031)	(0.037)
Covered × Post × Local	-0.037**	-0.057***	-0.078**	-0.117**	-0.119*	-0.182**	-0.173*	-0.276**
	(0.016)	(0.020)	(0.039)	(0.057)	(0.062)	(0.090)	(0.089)	(0.131)
Covered × Local	0.026**		0.052**		0.076*		0.110**	
	(0.012)		(0.025)		(0.039)		(0.056)	
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	9409390	9409390	9130133	9130133	8844714	8844714	8557945	8557945
R^2	0.10	0.20	0.12	0.28	0.13	0.34	0.14	0.38

Panel B: Trading Skills

	Trading 1-Q		Trading 2-Q		Trading 3-Q		Trading 4-Q	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Covered × Post	-0.013**	-0.012**	-0.027**	-0.019*	-0.046***	-0.032**	-0.072***	-0.055***
	(0.005)	(0.005)	(0.011)	(0.010)	(0.016)	(0.014)	(0.022)	(0.020)
Covered × Post × Local	-0.009	-0.016**	-0.018	-0.027*	-0.027	-0.040	-0.039	-0.063*
	(0.006)	(0.007)	(0.012)	(0.014)	(0.020)	(0.025)	(0.029)	(0.036)
Covered × Local	0.010*		0.019*		0.023		0.027	
	(0.005)		(0.011)		(0.017)		(0.025)	
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	9409390	9409390	9130133	9130133	8844714	8844714	8557945	8557945
R^2	0.07	0.21	0.08	0.29	0.09	0.34	0.09	0.37

4 Age of Funds

Table I.7: Age of Funds Holding Covered Stocks

The table presents the results of our study of the age of fund holding covered stocks. All regressions are estimated at the stock-quarter level and the dependent variable. In column (1) and (2), the dependent variable is the logarithm of the average age of funds holding the stock in a given quarter. In column (3) and (4), the dependent variable is the fraction of funds holding the stocks that are older than 5 years. In column (5) and (6), the dependent variable is the fraction of funds holding the stocks that are older than 10 years. Columns (1), (3) and (5) encompasses all stocks in our sample, while column (2), (4) and (6) 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 stock level.

			Proportion of Funds Older than			
	Log(Age of Funds Holding the Stock)		5-Year		10-Year	
	(1)	(2)	(3)	(4)	(5)	(6)
Covered × Post	-0.056***	-0.056***	-0.015**	-0.013**	-0.029***	-0.029***
	(0.020)	(0.020)	(0.006)	(0.006)	(0.011)	(0.011)
Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Only Stocks in Covered Industries	No	Yes	No	Yes	No	Yes
Observations	229,260	101,785	229,260	101,785	229,260	101,785
R^2	0.65	0.61	0.48	0.44	0.57	0.51

5 Stock Picking Excluding Industry and Geographical Peers

Table I.8: Stock Picking Skills Excluding Industry and Geographical Peers

This table reproduces our main results on the effect of the release of alternative data on fund picking skills. Regressions are estimated at the fund-stock-quarter level, but exclude the observations corresponding to non-covered stocks that are in the industries covered by RS-metrics or that are headquartered in a Metropolitan Statistical Area (MSA) where one of the covered stock has the majority of its parking lots. 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. 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 × Post	-0.018***	-0.020***	-0.017***	-0.035***	-0.038***	-0.028**	-0.056***	-0.062***	-0.043**	-0.085***	-0.100***	-0.076***
	(0.005)	(0.006)	(0.006)	(0.011)	(0.012)	(0.012)	(0.017)	(0.019)	(0.019)	(0.024)	(0.027)	(0.027)
Covered	0.024***			0.049***			0.075***			0.106***		
	(0.004)			(0.009)			(0.014)			(0.020)		
Fund × 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 × Stock FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	4195784	4195784	4195784	4060491	4060491	4060491	3922965	3922965	3922965	3783816	3783816	3783816
R^2	0.14	0.17	0.27	0.15	0.20	0.37	0.15	0.22	0.43	0.15	0.23	0.49